

The History of Physics  
Volume I  
*One Hundred of Selected Famous Papers*

---



The History of Physics  
Volume I  
*One Hundred of Selected Famous Papers*

---

**Javier Saramian**

Javier Saramian  
2014

**The History of Physics: One Hundred of Selected Famous Papers**

2014 by Javier Saramian

First Printing: 2014

1<sup>st</sup> Edition

Javier Saramian  
Gonzalo de Berceo 23-25  
Logroño, La Rioja 26005

To my lovely niece. I will try to plant the seed of the curiosity on you.

Also I would like to thank all the scientists, specially to the physicists, that tries to uncover the mysteries of the reality and the universe and to the divulgars that attempt to communicate all the beauty of this subject.

*Veritas In Hoc Scientia*



# Contents

---

## Volume I

<b>Book of Lemmas</b>	212	Archimedes BC	1
<b>On the Equilibrium of Planes. On Floating Bodies</b> <i>First mathematical physicist on record</i>			
<b>Discourse on the Method</b> <i>Framework for scientific method's guiding principles</i>	1637	Descartes, René	45
<b>Dialogues Concerning Two New Sciences</b> <i>Mechanics, kinematics, theory of inertia</i>	1638	Galilei, Galileo	59
<b>Philosophiae Naturalis Principia Mathematica</b> <i>Newton's laws of motion</i>	1687	Newton, Isaac	89
<b>Rules of Reasoning in Natural Philosophy</b> <i>Scientific method bases</i>	1729	Newton, Isaac	111
<b>On a General Method in Dynamics</b> <i>Development of the principle of "Varying Action"</i>	1834	Hamilton, William R.	113
<b>A Dynamical Theory of the Electromagnetic Field</b> <i>Theory of electromagnetism</i>	1865	Maxwell, James C.	179
<b>On the Relative Motion of the Earth and the Luminiferous Ether</b> <i>Evidence for an upper limit velocity of light</i>	1887	Michelson, Albert A. Morley, Edward W.	233
<b>On the Invisible Radiations Emitted by the Salts of Uranium</b> <i>Confirmation of the effect of spontaneous radioactivity</i>	1896	Becquerel, Henri	247
<b>Cathode Rays</b> <i>Discovery of the electron, the first elementary particle</i>	1897	Thomson, J. Joseph	253
<b>On an Improvement of Wien's Equation for the Spectrum</b> <i>Discovery of the Plank's radiation law for the energy spectrum. Beginnings of the quantum era in physics</i>	1900	Planck, Max	177
<b>On the Theory of the Energy Distribution Law of the Normal Spectrum</b> <i>Quantum hypothesis and explanation of the black body radiation spectrum</i>	1900	Planck, Max	281
<b>On the Law of Distribution of Energy in the Normal Spectrum</b> <i>Quantum hypothesis and final version of Planck's formula</i>	1901	Planck, Max	289
<b>Concerning an Heuristic Point of View Toward the Emission and Transformation of Light</b> <i>Explanation of the photoelectric effect using the light as corpuscular objects</i>	1905	Einstein, Albert	301
<b>On the Electrodynamics of Moving Bodies</b> <i>Invention of the theory of special relativity. Beginnings of the relativistic era in physics</i>	1905	Einstein, Albert	317
<b>Does the Inertia of a Body Depend on its Energy Content?</b> <i>Invention of the theory of special relativity, <math>E = mc^2</math></i>	1905	Einstein, Albert	343
<b>On the Relativity Principle and the Conclusions drawn from it</b> <i>Beginning of the long development of general relativity. Equivalence principle, gravitational redshift, and bending of light</i>	1907	Einstein, Albert	347
<b>On the Influence of Gravitation on the Propagation of Light</b> <i>General relativity replaces both special relativity and Newton's theory of gravitation. Equivalence principle only holds locally.</i>	1911	Einstein, Albert	361
<b>On an Expansion Apparatus for Making Visible the Tracks of Ionising Particles in Gases</b> <i>Invention of the cloud chamber to visualize tracks of ionizing particles</i>	1912	Wilson, Charles	371
<b>Penetrating Radiation in Seven Free Ballon Flights</b> <i>Conclusive evidence for the cosmic rays</i>	1912	Hess, Victor	393

<b>On the Constitution of Atoms and Molecules</b>	1913	Bohr, Niels	399
<i>Invention of the quantum theory of atomic spectra based on the Rutherford model of atomic structure - Bohr's atom</i>			
<b>On the Constitution of Atoms and Molecules 2</b>	1913	Bohr, Niels	423
<i>Bohr's quantum theory of atomic spectra. Evidence that radioactivity is a nuclear property</i>			
<b>The Structure of the Atom</b>	1914	Rutherford, Ernest	449
<i>Confirmation of the existence of atomic nuclei. First indication of the existence of the proton</i>			
<b>On the gravitational field of a mass point according to Einstein's theory</b>	1916	Schwarzschild, Karl	461
<i>Schwarzschild metric</i>			
<b>The Foundation of the General Theory of Relativity</b>	1916	Einstein, Albert	469
<i>Theory of general relativity</i>			
<b>Invariant Variant Problems</b>	1918	Noether, Emmy	525
<i>Noether theorem connecting symmetries and conserved quantities</i>			
<b>On the Quantum Theory of Line Spectra</b>	1918	Bohr, Niels	539
<i>Bohr's invention of correspondence principle</i>			
<b>Collision of a Particle with Light Atoms. IV. Anomalous Effect in Nitrogen</b>	1919	Rutherford, Ernest	579
<i>Discovery of the proton. Evidence for it as a nucleus constituent</i>			
<b>Nuclear Constitution of Atoms</b>	1920	Rutherford, Ernest	587
<i>Rutherford neutron hypothesis</i>			
<b>On the Curvature of Space</b>	1922	Friedmann, Aleksandr	617
<i>Friedmann Cosmology</i>			
<b>Waves and Quanta</b>	1923	De Broglie, Louis	623
<i>Suggestion of the corpuscular-wave dualism for electrons</i>			
<b>The Spectrum of Scattered X-Rays</b>	1923	Compton, Arthur	627
<i>Direct experimental confirmation that the photon is an elementary particle, the Compton effect</i>			
<b>Investigations on X Rays and Beta Rays by the Cloud Method. Part 1. - X Rays</b>	1923	Wilson, Charles	633
<i>Experimental confirmation of the ionization process predicted by Compton for a corpuscular photon</i>			
<b>Planck's Law and Light Quantum Hypothesis</b>	1924	Bose, Satyendra	657
<i>Discovery of Bose-Einstein quantum statistics for particles of integer spins. New derivation of Planck's radiation law</i>			
<b>On the Connexion between the Completion of Electron Groups in an Atom with the Complex Structure of Spectra</b>	1925	Pauli, Wolfgang	661
<i>Discovery of the exclusion principle - the Pauli principle</i>			
<b>Quantum-Theoretical Re-Interpretation of Kinematic and Mechanical Relations</b>	1925	Heisenberg, Werner	675
<i>Foundation of quantum mechanics, Heisenberg approach</i>			
<b>On Quantum Mechanics</b>	1925	Bohr, Niels	691
<i>Invention of matrix formalism for the Heisenberg quantum mechanics. Systems with one degree of freedom</i>			
<b>On Quantum Mechanics II</b>	1925	Bohr, Niels	721
<i>Matrix formalism with arbitrary many degrees of freedom</i>			

## Volume II

<b>On Quantizing an Ideal Monatomic Gas</b>	1926	Fermi, Enrico	787
<i>Invention of statistics for ensembles of particles obeying Pauli principle - Fermi-Dirac quantum statistics</i>			
<b>On the Quantum Mechanics of Collisions</b>	1926	Bohr, Niels	797
<i>Statistical interpretation, probability density of quantum mechanics. Quantum theory of scattering. Born approximation</i>			
<b>Quantisation as a Problem of Proper Values - Part 1</b>	1926	Schrödinger, Erwin	801
<i>Creation of wave mechanics. Invention of the Schrödinger wave equation</i>			
<b>The Physical Content of Quantum Kinematics and Mechanics</b>	1927	Heisenberg, Werner	813
<i>Heisenberg discovery of the uncertainty principle</i>			
<b>The Quantum Theory of Dispersion</b>	1927	Dirac, Paul	837
<i>Foundations of quantum electrodynamics - QED</i>			
<b>The Quantum Theory of the Electron</b>	1928	Dirac, Paul	857
<i>Discovery of the relativistic wave equation for the electron. Prediction of the magnetic moment of the electron</i>			
<b>A Relation between Distance and Radial Velocity</b>	1929	Hubble, Edwin	873
<i>Evidence for the expansion of the universe</i>			
<b>Gravitation and the Electron</b>	1929	Weyl, Hermann	879
<i>Combination of electrodynamics and gravitation</i>			
<b>The Production of High Speed Protons Without the use of High Voltages</b>	1931	Lawrence, Ernest	893
<i>Tests of the first cyclotron</i>			
<b>Quantised Singularities in the Electromagnetic Field</b>	1931	Dirac, Paul	895
<i>Prediction of the anti-electron (<math>e^+</math>), anti-proton (anti-<math>p</math>), and an indication of the possible existence of magnetic monopoles</i>			
<b>A 1,500,000 Volt Electrostatic Generator</b>	1931	Van de Graaff, Robert	909
<i>Invention of the Van de Graaff electrostatic accelerator</i>			
<b>The Existence of a Neutron</b>	1932	Chadwick, James	911
<i>Discovery of the neutron</i>			
<b>The Electrostatic Production of High Voltage for Nuclear Investigations</b>	1933	Van de Graaff, Robert Compton, Arthur	927
<i>Invention of electrostatic accelerators. Further development</i>			
<b>The Positive Electron</b>	1933	Anderson, Carl	943
<i>Discovery of the positron, the first antiparticle, predicted by Dirac</i>			
<b>The Highly Collapsed Configurations of a Stellar Mass</b>	1935	Chandrasekhar, Subrahmanyan	951
<i>Chandrasekhar Limit</i>			
<b>Can Quantum Mechanical Description of Physical Reality be considered complete?</b>	1935	Einstein, Albert Podolsky, Boris Rosen, Nathan	971
<i>EPR paradox</i>			
<b>On the Interaction of Elementary Particles</b>	1935	Yukawa, Hideki	975
<i>Yukawa field theory of nuclear forces. Prediction of heavy quanta, the pion particles, as mediators of strong interactions</i>			
<b>Visible Radiation Produced by Electrons Moving in a Medium with Velocities Exceeding that of Light</b>	1937	Cherenkov, Pavel	985
<i>Confirmation of the Frank-Tamm theory of the Vavilov- Cerenkov effect</i>			
<b>New Evidence for the Existence of a Particle of Mass</b>	1937	Street, Jabez Stevenson, E	989
<i>Muon existence confirmation</i>			
<b>On Massive neutron cores</b>	1939	Oppenheimer, Robert	995
<i>Prediction of neutron stars</i>			
<b>Forces in Molecules</b>	1939	Feynman, Richard	1003
<i>Feynman-Hellmann theorem, as an efficient approach to the calculation of forces in molecules</i>			

<b>On Continued Gravitational Contraction</b>	1939	Oppenheimer, Robert	1007
<i>First notion of black hole</i>			
<b>Acceleration of Electrons by Magnetic Induction</b>	1940	Kerst, Donald	1013
<i>Kerst proposal for betatron accelerator</i>			
<b>The Connection Between Spin and Statistics</b>	1940	Pauli, Wolfgang	1017
<i>Theorem on the connection between spin and statistics</i>			
<b>Expanding Universe and the Origin of Elements</b>	1946	Gamov, George	1031
<i>Explanation to the observed chemical elements abundance-curve by unequilibrium process of elements formation. Birth of the Big Bang model</i>			
<b>Space-Time Approach to Non-Relativistic Quantum Mechanism</b>	1948	Feynman, Richard	1035
<i>Path integral formalism</i>			
<b>A Generalized Theory of Gravitation</b>	1948	Einstein, Albert	1071
<i>Summary of the General Relativity</i>			
<b>The Origin of Chemical Elements</b>	1948	Gamov, George	1077
<i>Theory of big-bang nucleosynthesis</i>			
<b>A Relativistic Cut-Off for Quantum Electrodynamics</b>	1948	Feynman, Richard	1079
<i>Proposal to modify classical electrodynamics to a form suitable for quantization</i>			
<b>Space-Time Approach to Quantum Electrodynamics</b>	1949	Feynman, Richard	1089
<i>Development of the covariant quantum electrodynamic theory. Feynman method</i>			
<b>The Theory of Positrons</b>	1949	Feynman, Richard	1143
<i>Creation of the covariant quantum electrodynamic theory. Feynman method</i>			
<b>Mathematical Formulation of the Quantum Theory of Electromagnetic Interaction</b>	1950	Feynman, Richard	1155
<i>Mathematical proof of the validity of the Feynman rules for calculations of amplitudes in QED</i>			
<b>Equation of State Calculations by Fast Computing Machines</b>	1953	Teller, Edward	1173
<i>The birth of computational physics</i>			
<b>Conservation of Isotopic Spin and Isotopic Gauge Invariance</b>	1954	Yang, Chen Ning Mills, Robert	1179
<i>Introduction of local gauge isotopic spin invariance in quantum field theory: Yang-Mills theory</i>			
<b>Relative State Formulation of Quantum Mechanics</b>	1957	Everett, Hugh	1185
<i>"Many worlds" interpretation</i>			
<b>Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. I</b>	1961	Nambu, Yoichiro	1207
<i>Nambu-Jona-Lasinio nonlinear model of hadrons</i>			
<b>Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. II</b>	1961	Nambu, Yoichiro	1221
<i>Nambu-Jona-Lasinio nonlinear model of hadrons</i>			
<b>A Schematic Model of Baryons and Mesons</b>	1964	Gell-Man Murray	1231
<i>Introduction of quarks as fundamental building blocks for hadrons.</i>			
<b>Broken Symmetries and the Masses of Gauge Bosons</b>	1964	Higgs, Peter	1233
<i>Higgs mechanism of mass generation for vector gauge fields</i>			
<b>Inhomogeneous Electron</b>	1964	Hohenberg, Pierre Walter, Kohn	1235
<i>Density Functional Theory</i>			
<b>On the Einstein Podolsky Rosen Paradox</b>	1964	Bell, John	1243
<i>Bell theorem</i>			
<b>Three Triplet Model with Double SU(3) Symmetry</b>	1965	Nambu, Yoichiro	1249
<i>Suggestion of the existence of three triplets of quarks</i>			
<b>Self-Consistent Equations including Exchange and Correlation Effects</b>	1965	Kohn, Walter Sham, Lu Jeu	1255
<i>Density Functional Theory</i>			
<b>Quantum Theory of Gravity. I. The Canonical Theory</b>	1967	deWitt, Bryce	1261
<i>Wheeler-DeWitt equation of quantum gravity</i>			

<b>A Model for Leptons</b>	1967 Weinberg, Steven	1297
<i>Lagrangian for the electroweak synthesis, estimations of the W and Z masses</i>		
<b>Construction of a Crossing-symmetric, Regge-behaved Amplitude for Linearly Rising Trajectories</b>	1968 Veneziano, Gabriele	1301
<i>First formulation of string theory</i>		
<b>On the interpretation of measurement in quantum theory</b>	1969 Zeh, Heinz-Dieter	1311
<i>Quantum decoherence</i>		
<b>Renormalization of Massless Yang-Mills Fields</b>	1971 Hooft, Gerard 't	1319
<i>Rigorous proofs of renormalizability of the massless Yang-Mills quantum gauge fields theory</i>		
<b>Black Holes and Entropy</b>	1972 Bekenstein, Jacob	1347
<i>Black hole thermodynamics</i>		
<b>Ultraviolet Behavior of Non Abelian Gauge Theory</b>	1973 Gross, David	1361
<i>Asymptotic freedom</i>		
<b>Unity of All Elementary-Particle Forces</b>	1974 Georgi, Howard	1365
<i>Grand unified theory of all forces except gravity</i>	Glashow, Sheldon	
<b>Particle Creation by Black Holes</b>	1975 Hawking, Stephen	1369
<i>Hawking radiation and black holes temperature</i>		
<b>Inflationary universe. A possible solution to the horizon and flatness problems</b>	1980 Guth, Alan	1391
<i>Theory of inflation</i>		
<b>Simulating Physics with Computers</b>	1981 Feynman, Richard	1401
<i>Argument for quantum computation</i>		
<b>A new inflationary universe scenario, A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems</b>	1981 Linde, Andrei	1423
<i>Extended theory of inflation</i>		
<b>Quantum theory, the Church-Turing principle and the universal quantum computer</b>	1985 Deutsch, David	1429
<i>Foundation of the universal quantum Turing model of computation</i>		
<b>New variables for classical and quantum gravity</b>	1986 Ashtekar, Abhay	1449
<i>Ashtekar variables</i>		
<b>Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels</b>	1992 Bennett, Charles	1453
<i>Brassard, Gilles</i>		
<i>Superdense coding</i>		
<b>Dimensional Reduction in Quantum Gravity</b>	1993 Hooft, Gerard 't	1459
<i>Holographic principle</i>		
<b>String Theory Dynamics in Various Dimensions</b>	1995 Witten, Edward	1473
<i>M-theory</i>		
<b>The Large N Limit of Superconformal field</b>	1997 Maldacena, Juan	1529
<i>AdS/CFT Correspondence</i>		
<b>The Hierarchy Problem and New Dimensions at a Millimeter</b>	1998 Arkani-Hammed, Nima	1551
<i>Proposed an alternative scenario to explain the weakness of gravity relative to the other forces</i>		



**The Works of Archimedes**  
**On the Equilibrium of Planes - On Floating Bodies**

Archimedes  
212 BC

ON THE EQUILIBRIUM OF PLANES  
OR  
THE CENTRES OF GRAVITY OF PLANES.

BOOK I.

---

"I POSTULATE the following:

1. Equal weights at equal distances are in equilibrium, and equal weights at unequal distances are not in equilibrium but incline towards the weight which is at the greater distance.
2. If, when weights at certain distances are in equilibrium, something be added to one of the weights, they are not in equilibrium but incline towards that weight to which the addition was made.
3. Similarly, if anything be taken away from one of the weights, they are not in equilibrium but incline towards the weight from which nothing was taken.
4. When equal and similar plane figures coincide if applied to one another, their centres of gravity similarly coincide.
5. In figures which are unequal but similar the centres of gravity will be similarly situated. By points similarly situated in relation to similar figures I mean points such that, if straight lines be drawn from them to the equal angles, they make equal angles with the corresponding sides.

6. If magnitudes at certain distances be in equilibrium, (other) magnitudes equal to them will also be in equilibrium at the same distances.

7. In any figure whose perimeter is concave in (one and) the same direction the centre of gravity must be within the figure."

### **Proposition 1.**

*Weights which balance at equal distances are equal.*

For, if they are unequal, take away from the greater the difference between the two. The remainders will then not balance [Post. 3]; which is absurd.

Therefore the weights cannot be unequal.

### **Proposition 2.**

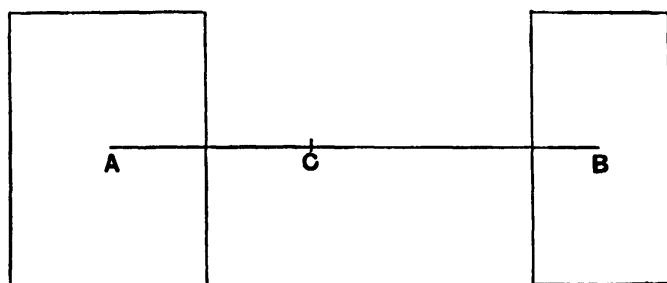
*Unequal weights at equal distances will not balance but will incline towards the greater weight.*

For take away from the greater the difference between the two. The equal remainders will therefore balance [Post. 1]. Hence, if we add the difference again, the weights will not balance but incline towards the greater [Post. 2].

### **Proposition 3.**

*Unequal weights will balance at unequal distances, the greater weight being at the lesser distance.*

Let  $A, B$  be two unequal weights (of which  $A$  is the greater) balancing about  $C$  at distances  $AC, BC$  respectively.



Then shall  $AC$  be less than  $BC$ . For, if not, take away from  $A$  the weight  $(A - B)$ . The remainders will then incline

towards  $B$  [Post. 3]. But this is impossible, for (1) if  $AC = CB$ , the equal remainders will balance, or (2) if  $AC > CB$ , they will incline towards  $A$  at the greater distance [Post. 1].

Hence  $AC < CB$ .

*Conversely*, if the weights balance, and  $AC < CB$ , then  $A > B$ .

#### Proposition 4.

*If two equal weights have not the same centre of gravity, the centre of gravity of both taken together is at the middle point of the line joining their centres of gravity.*

[Proved from Prop. 3 by *reductio ad absurdum*. Archimedes assumes that the centre of gravity of both together is on the straight line joining the centres of gravity of each, saying that this had been proved before ( $\pi\tau\delta\epsilon\delta\epsilon\kappa\tau\alpha$ ). The allusion is no doubt to the lost treatise *On levers* ( $\pi\tau\pi\zeta\gamma\omega\nu$ ).]

#### Proposition 5.

*If three equal magnitudes have their centres of gravity on a straight line at equal distances, the centre of gravity of the system will coincide with that of the middle magnitude.*

[This follows immediately from Prop. 4.]

COR 1. *The same is true of any odd number of magnitudes if those which are at equal distances from the middle one are equal, while the distances between their centres of gravity are equal.*

COR. 2. *If there be an even number of magnitudes with their centres of gravity situated at equal distances on one straight line, and if the two middle ones be equal, while those which are equidistant from them (on each side) are equal respectively, the centre of gravity of the system is the middle point of the line joining the centres of gravity of the two middle ones.*

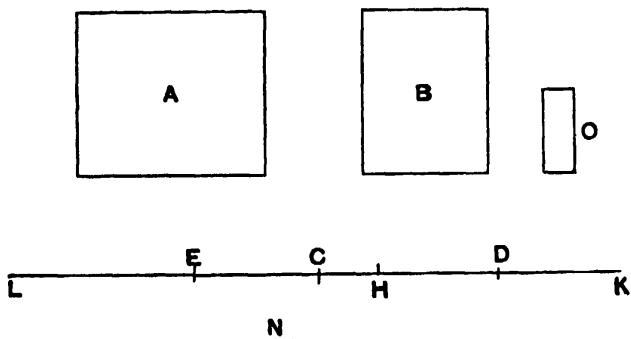
### Propositions 6, 7.

*Two magnitudes, whether commensurable [Prop. 6] or incommensurable [Prop. 7], balance at distances reciprocally proportional to the magnitudes.*

I. Suppose the magnitudes  $A, B$  to be commensurable, and the points  $A, B$  to be their centres of gravity. Let  $DE$  be a straight line so divided at  $C$  that

$$A : B = DC : CE.$$

We have then to prove that, if  $A$  be placed at  $E$  and  $B$  at  $D$ ,  $C$  is the centre of gravity of the two taken together.



Since  $A, B$  are commensurable, so are  $DC, CE$ . Let  $N$  be a common measure of  $DC, CE$ . Make  $DH, DK$  each equal to  $CE$ , and  $EL$  (on  $CE$  produced) equal to  $CD$ . Then  $EH = CD$ , since  $DH = CE$ . Therefore  $LH$  is bisected at  $E$ , as  $HK$  is bisected at  $D$ .

Thus  $LH, HK$  must each contain  $N$  an even number of times.

Take a magnitude  $O$  such that  $O$  is contained as many times in  $A$  as  $N$  is contained in  $LH$ , whence

$$A : O = LH : N.$$

$$\begin{aligned} \text{But } B : A &= CE : DC \\ &= HK : LH. \end{aligned}$$

Hence, *ex aequali*,  $B : O = HK : N$ , or  $O$  is contained in  $B$  as many times as  $N$  is contained in  $HK$ .

Thus  $O$  is a common measure of  $A, B$ .

Divide  $LH$ ,  $HK$  into parts each equal to  $N$ , and  $A$ ,  $B$  into parts each equal to  $O$ . The parts of  $A$  will therefore be equal in number to those of  $LH$ , and the parts of  $B$  equal in number to those of  $HK$ . Place one of the parts of  $A$  at the middle point of each of the parts  $N$  of  $LH$ , and one of the parts of  $B$  at the middle point of each of the parts  $N$  of  $HK$ .

Then the centre of gravity of the parts of  $A$  placed at equal distances on  $LH$  will be at  $E$ , the middle point of  $LH$  [Prop. 5, Cor. 2], and the centre of gravity of the parts of  $B$  placed at equal distances along  $HK$  will be at  $D$ , the middle point of  $HK$ .

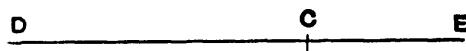
Thus we may suppose  $A$  itself applied at  $E$ , and  $B$  itself applied at  $D$ .

But the system formed by the parts  $O$  of  $A$  and  $B$  together is a system of equal magnitudes even in number and placed at equal distances along  $LK$ . And, since  $LE = CD$ , and  $EC = DK$ ,  $LC = CK$ , so that  $C$  is the middle point of  $LK$ . Therefore  $C$  is the centre of gravity of the system ranged along  $LK$ .

Therefore  $A$  acting at  $E$  and  $B$  acting at  $D$  balance about the point  $C$ .

II. Suppose the magnitudes to be incommensurable, and let them be  $(A + a)$  and  $B$  respectively. Let  $DE$  be a line divided at  $C$  so that

$$(A + a) : B = DC : CE.$$



Then, if  $(A + a)$  placed at  $E$  and  $B$  placed at  $D$  do not balance about  $C$ ,  $(A + a)$  is either too great to balance  $B$ , or not great enough.

Suppose, if possible, that  $(A + a)$  is too great to balance  $B$ . Take from  $(A + a)$  a magnitude  $a$  smaller than the deduction which would make the remainder balance  $B$ , but such that the remainder  $A$  and the magnitude  $B$  are commensurable.

Then, since  $A, B$  are commensurable, and

$$A : B < DC : CE,$$

$A$  and  $B$  will not balance [Prop. 6], but  $D$  will be depressed.

But this is impossible, since the deduction  $a$  was an insufficient deduction from  $(A+a)$  to produce equilibrium, so that  $E$  was still depressed.

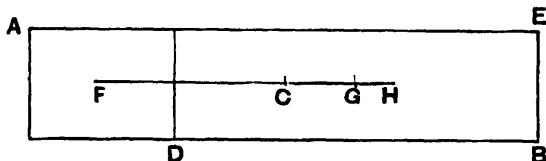
Therefore  $(A+a)$  is not too great to balance  $B$ ; and similarly it may be proved that  $B$  is not too great to balance  $(A+a)$ .

Hence  $(A+a), B$  taken together have their centre of gravity at  $C$ .

### Proposition 8.

*If  $AB$  be a magnitude whose centre of gravity is  $C$ , and  $AD$  a part of it whose centre of gravity is  $F$ , then the centre of gravity of the remaining part will be a point  $G$  on  $FC$  produced such that*

$$GC : CF = (AD) : (DE).$$



For, if the centre of gravity of the remainder  $(DE)$  be not  $G$ , let it be a point  $H$ . Then an absurdity follows at once from Props. 6, 7.

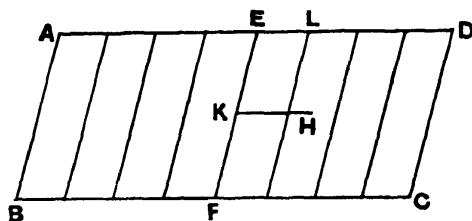
### Proposition 9.

*The centre of gravity of any parallelogram lies on the straight line joining the middle points of opposite sides.*

Let  $ABCD$  be a parallelogram, and let  $EF$  join the middle points of the opposite sides  $AD, BC$ .

If the centre of gravity does not lie on  $EF$ , suppose it to be  $H$ , and draw  $HK$  parallel to  $AD$  or  $BC$  meeting  $EF$  in  $K$ .

Then it is possible, by bisecting  $ED$ , then bisecting the halves, and so on continually, to arrive at a length  $EL$  less



than  $KH$ . Divide both  $AE$  and  $ED$  into parts each equal to  $EL$ , and through the points of division draw parallels to  $AB$  or  $CD$ .

We have then a number of equal and similar parallelograms, and, if any one be applied to any other, their centres of gravity coincide [Post. 4]. Thus we have an even number of equal magnitudes whose centres of gravity lie at equal distances along a straight line. Hence the centre of gravity of the whole parallelogram will lie on the line joining the centres of gravity of the two middle parallelograms [Prop. 5, Cor. 2].

But this is impossible, for  $H$  is outside the middle parallelograms.

Therefore the centre of gravity cannot but lie on  $EF$ .

### Proposition 10.

*The centre of gravity of a parallelogram is the point of intersection of its diagonals.*

For, by the last proposition, the centre of gravity lies on each of the lines which bisect opposite sides. Therefore it is at the point of their intersection; and this is also the point of intersection of the diagonals.

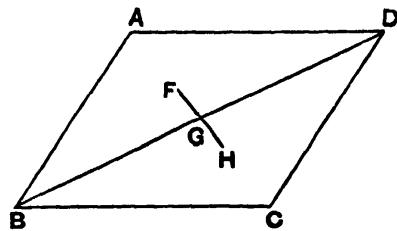
*Alternative proof.*

Let  $ABCD$  be the given parallelogram, and  $BD$  a diagonal. Then the triangles  $ABD$ ,  $CDB$  are equal and similar, so that [Post. 4], if one be applied to the other, their centres of gravity will fall one upon the other.

Suppose  $F$  to be the centre of gravity of the triangle  $ABD$ . Let  $G$  be the middle point of  $BD$ .

Join  $FG$  and produce it to  $H$ , so that  $FG = GH$ .

If we then apply the triangle  $ABD$  to the triangle  $CDB$  so that  $AD$  falls on  $CB$  and  $AB$  on  $CD$ , the point  $F$  will fall on  $H$ .



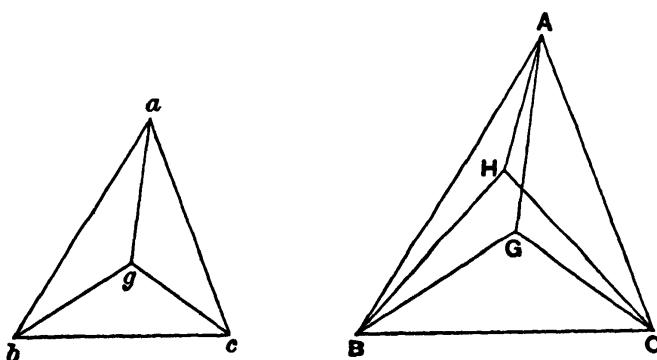
But [by Post. 4]  $F$  will fall on the centre of gravity of  $CDB$ . Therefore  $H$  is the centre of gravity of  $CDB$ .

Hence, since  $F, H$  are the centres of gravity of the two equal triangles, the centre of gravity of the whole parallelogram is at the middle point of  $FH$ , i.e. at the middle point of  $BD$ , which is the intersection of the two diagonals.

### Proposition 11.

*If  $abc, ABC$  be two similar triangles, and  $g, G$  two points in them similarly situated with respect to them respectively, then, if  $g$  be the centre of gravity of the triangle  $abc$ ,  $G$  must be the centre of gravity of the triangle  $ABC$ .*

Suppose  $ab : bc : ca = AB : BC : CA$ .



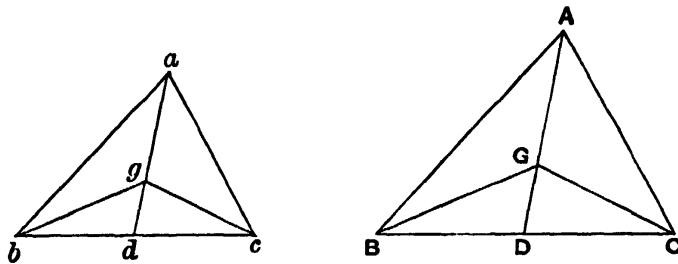
The proposition is proved by an obvious *reductio ad absurdum*. For, if  $G$  be not the centre of gravity of the triangle  $ABC$ , suppose  $H$  to be its centre of gravity.

Post. 5 requires that  $g, H$  shall be similarly situated with respect to the triangles respectively; and this leads at once to the absurdity that the angles  $HAB, GAB$  are equal.

**Proposition 12.**

*Given two similar triangles abc, ABC, and d, D the middle points of bc, BC respectively, then, if the centre of gravity of abc lie on ad, that of ABC will lie on AD.*

Let  $g$  be the point on  $ad$  which is the centre of gravity of  $abc$ .



Take  $G$  on  $AD$  such that

$$ad : ag = AD : AG,$$

and join  $gb, gc, GB, GC$ .

Then, since the triangles are similar, and  $bd, BD$  are the halves of  $bc, BC$  respectively,

$$ab : bd = AB : BD,$$

and the angles  $abd, ABD$  are equal.

Therefore the triangles  $abd, ABD$  are similar, and

$$\angle bad = \angle BAD.$$

Also

$$ba : ad = BA : AD,$$

while, from above,       $ad : ag = AD : AG$ .

Therefore  $ba : ag = BA : AG$ , while the angles  $bag, BAG$  are equal.

Hence the triangles  $bag, BAG$  are similar, and

$$\angle abg = \angle ABG.$$

And, since the angles  $abd, ABD$  are equal, it follows that

$$\angle gbd = \angle GBD.$$

In exactly the same manner we prove that

$$\angle gac = \angle GAC,$$

$$\angle acg = \angle ACG,$$

$$\angle gcd = \angle GCD.$$

Therefore  $g$ ,  $G$  are similarly situated with respect to the triangles respectively; whence [Prop. 11]  $G$  is the centre of gravity of  $ABC$ .

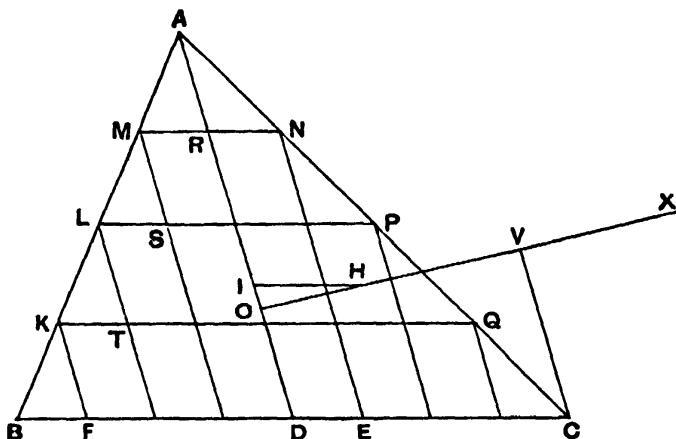
### Proposition 13.

*In any triangle the centre of gravity lies on the straight line joining any angle to the middle point of the opposite side.*

Let  $ABC$  be a triangle and  $D$  the middle point of  $BC$ . Join  $AD$ . Then shall the centre of gravity lie on  $AD$ .

For, if possible, let this not be the case, and let  $H$  be the centre of gravity. Draw  $HI$  parallel to  $CB$  meeting  $AD$  in  $I$ .

Then, if we bisect  $DC$ , then bisect the halves, and so on, we shall at length arrive at a length, as  $DE$ , less than  $HI$ .



Divide both  $BD$  and  $DC$  into lengths each equal to  $DE$ , and through the points of division draw lines each parallel to  $DA$  meeting  $BA$  and  $AC$  in points as  $K$ ,  $L$ ,  $M$  and  $N$ ,  $P$ ,  $Q$  respectively.

Join  $MN$ ,  $LP$ ,  $KQ$ , which lines will then be each parallel to  $BC$ .

We have now a series of parallelograms as  $FQ$ ,  $TP$ ,  $SN$ , and  $AD$  bisects opposite sides in each. Thus the centre of gravity of each parallelogram lies on  $AD$  [Prop. 9], and therefore the centre of gravity of the figure made up of them all lies on  $AD$ .

Let the centre of gravity of all the parallelograms taken together be  $O$ . Join  $OH$  and produce it; also draw  $CV$  parallel to  $DA$  meeting  $OH$  produced in  $V$ .

Now, if  $n$  be the number of parts into which  $AC$  is divided,

$$\triangle ADC : (\text{sum of triangles on } AN, NP, \dots)$$

$$= AC^2 : (AN^2 + NP^2 + \dots)$$

$$= n^2 : n$$

$$= n : 1$$

$$= AC : AN.$$

Similarly

$$\triangle ABD : (\text{sum of triangles on } AM, ML, \dots) = AB : AM.$$

And

$$AC : AN = AB : AM.$$

It follows that

$$\triangle ABC : (\text{sum of all the small } \Delta\text{s}) = CA : AN$$

$$> VO : OH, \text{ by parallels.}$$

Suppose  $OV$  produced to  $X$  so that

$$\triangle ABC : (\text{sum of small } \Delta\text{s}) = XO : OH,$$

whence, *dividendo*,

$$(\text{sum of parallelograms}) : (\text{sum of small } \Delta\text{s}) = XH : HO.$$

Since then the centre of gravity of the triangle  $ABC$  is at  $H$ , and the centre of gravity of the part of it made up of the parallelograms is at  $O$ , it follows from Prop. 8 that the centre of gravity of the remaining portion consisting of all the small triangles taken together is at  $X$ .

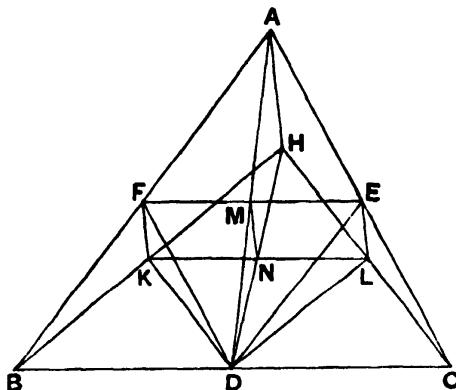
But this is impossible, since all the triangles are on one side of the line through  $X$  parallel to  $AD$ .

Therefore the centre of gravity of the triangle cannot but lie on  $AD$ .

*Alternative proof.*

Suppose, if possible, that  $H$ , not lying on  $AD$ , is the centre of gravity of the triangle  $ABC$ . Join  $AH, BH, CH$ . Let  $E, F$  be the middle points of  $CA, AB$  respectively, and join  $DE, EF, FD$ . Let  $EF$  meet  $AD$  in  $M$ .

Draw  $FK, EL$  parallel to  $AH$  meeting  $BH, CH$  in  $K, L$  respectively. Join  $KD, HD, LD, KL$ . Let  $KL$  meet  $DH$  in  $N$ , and join  $MN$ .



Since  $DE$  is parallel to  $AB$ , the triangles  $ABC, EDC$  are similar.

And, since  $CE = EA$ , and  $EL$  is parallel to  $AH$ , it follows that  $CL = LH$ . And  $CD = DB$ . Therefore  $BH$  is parallel to  $DL$ .

Thus in the similar and similarly situated triangles  $ABC, EDC$  the straight lines  $AH, BH$  are respectively parallel to  $EL, DL$ ; and it follows that  $H, L$  are similarly situated with respect to the triangles respectively.

But  $H$  is, by hypothesis, the centre of gravity of  $ABC$ . Therefore  $L$  is the centre of gravity of  $EDC$ . [Prop. 11]

Similarly the point  $K$  is the centre of gravity of the triangle  $FBD$ .

And the triangles  $FBD, EDC$  are equal, so that the centre of gravity of both together is at the middle point of  $KL$ , i.e. at the point  $N$ .

The remainder of the triangle  $ABC$ , after the triangles  $FBD, EDC$  are deducted, is the parallelogram  $AFDE$ , and the centre of gravity of this parallelogram is at  $M$ , the intersection of its diagonals.

It follows that the centre of gravity of the whole triangle  $ABC$  must lie on  $MN$ ; that is,  $MN$  must pass through  $H$ , which is impossible (since  $MN$  is parallel to  $AH$ ).

Therefore the centre of gravity of the triangle  $ABC$  cannot but lie on  $AD$ .

**Proposition 14.**

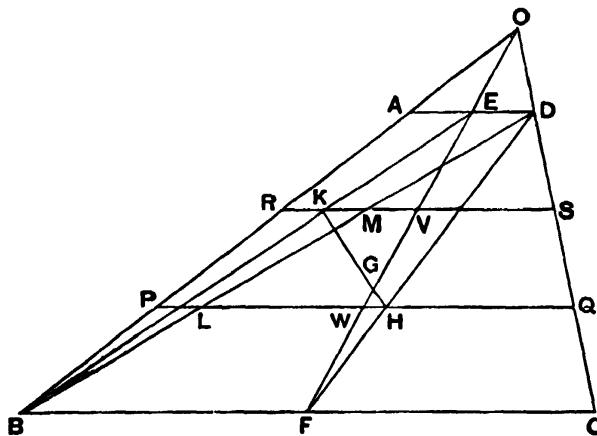
It follows at once from the last proposition that *the centre of gravity of any triangle is at the intersection of the lines drawn from any two angles to the middle points of the opposite sides respectively.*

**Proposition 15.**

*If  $AD, BC$  be the two parallel sides of a trapezium  $ABCD$ ,  $AD$  being the smaller, and if  $AD, BC$  be bisected at  $E, F$  respectively, then the centre of gravity of the trapezium is at a point  $G$  on  $EF$  such that*

$$GE : GF = (2BC + AD) : (2AD + BC).$$

Produce  $BA, CD$  to meet at  $O$ . Then  $FE$  produced will also pass through  $O$ , since  $AE = ED$ , and  $BF = FC$ .



Now the centre of gravity of the triangle  $OAD$  will lie on  $OE$ , and that of the triangle  $OCB$  will lie on  $OF$ . [Prop. 13]

It follows that the centre of gravity of the remainder, the trapezium  $ABCD$ , will also lie on  $OF$ . [Prop. 8]

Join  $BD$ , and divide it at  $L, M$  into three equal parts. Through  $L, M$  draw  $PQ, RS$  parallel to  $BC$  meeting  $BA$  in  $P, R$ ,  $FE$  in  $W, V$ , and  $CD$  in  $Q, S$  respectively.

Join  $DF, BE$  meeting  $PQ$  in  $H$  and  $RS$  in  $K$  respectively.

Now, since

$$BL = \frac{1}{3} BD,$$

$$FH = \frac{1}{3} FD.$$

Therefore  $H$  is the centre of gravity of the triangle  $DBC^*$ .

Similarly, since  $EK = \frac{1}{3}BE$ , it follows that  $K$  is the centre of gravity of the triangle  $ADB$ .

Therefore the centre of gravity of the triangles  $DBC$ ,  $ADB$  together, i.e. of the trapezium, lies on the line  $HK$ .

But it also lies on  $OF$ .

Therefore, if  $OF$ ,  $HK$  meet in  $G$ ,  $G$  is the centre of gravity of the trapezium.

Hence [Props. 6, 7]

$$\begin{aligned}\triangle DBC : \triangle ABD &= KG : GH \\ &= VG : GW.\end{aligned}$$

But  $\triangle DBC : \triangle ABD = BC : AD$ .

Therefore  $BC : AD = VG : GW$ .

It follows that

$$\begin{aligned}(2BC + AD) : (2AD + BC) &= (2VG + GW) : (2GW + VG) \\ &= EG : GF.\end{aligned}$$

Q. E. D.

\* This easy deduction from Prop. 14 is assumed by Archimedes without proof.

# ON THE EQUILIBRIUM OF PLANES.

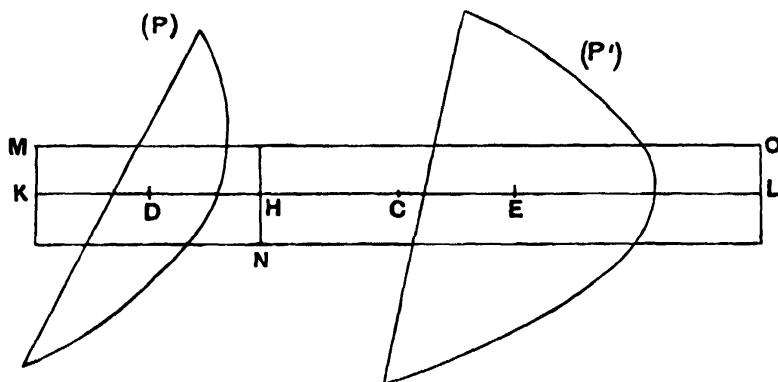
## BOOK II.

### Proposition 1.

If  $P, P'$  be two parabolic segments and  $D, E$  their centres of gravity respectively, the centre of gravity of the two segments taken together will be at a point  $C$  on  $DE$  determined by the relation

$$P : P' = CE : CD^*.$$

In the same straight line with  $DE$  measure  $EH, EL$  each equal to  $DC$ , and  $DK$  equal to  $DH$ ; whence it follows at once that  $DK = CE$ , and also that  $KC = CL$ .



\* This proposition is really a particular case of Props. 6, 7 of Book I. and is therefore hardly necessary. As, however, Book II. relates exclusively to parabolic segments, Archimedes' object was perhaps to emphasize the fact that the magnitudes in I. 6, 7 might be parabolic segments as well as rectilinear figures. His procedure is to substitute for the segments rectangles of equal area, a substitution which is rendered possible by the results obtained in his separate treatise on the *Quadrature of the Parabola*.

Apply a rectangle  $MN$  equal in area to the parabolic segment  $P$  to a base equal to  $KH$ , and place the rectangle so that  $KH$  bisects it, and is parallel to its base.

Then  $D$  is the centre of gravity of  $MN$ , since  $KD = DH$ .

Produce the sides of the rectangle which are parallel to  $KH$ , and complete the rectangle  $NO$  whose base is equal to  $HL$ . Then  $E$  is the centre of gravity of the rectangle  $NO$ .

$$\begin{aligned} \text{Now } (MN) : (NO) &= KH : HL \\ &= DH : EH \\ &= CE : CD \\ &= P : P'. \end{aligned}$$

$$\text{But } (MN) = P.$$

$$\text{Therefore } (NO) = P'.$$

Also, since  $C$  is the middle point of  $KL$ ,  $C$  is the centre of gravity of the whole parallelogram made up of the two parallelograms  $(MN)$ ,  $(NO)$ , which are equal to, and have the same centres of gravity as,  $P$ ,  $P'$  respectively.

Hence  $C$  is the centre of gravity of  $P$ ,  $P'$  taken together.

### Definition and lemmas preliminary to Proposition 2.

"If in a segment bounded by a straight line and a section of a right-angled cone [a parabola] a triangle be inscribed having the same base as the segment and equal height, if again triangles be inscribed in the remaining segments having the same bases as the segments and equal height, and if in the remaining segments triangles be inscribed in the same manner, let the resulting figure be said to be **inscribed in the recognised manner** (*γνωρίμως ἐγγράφεσθαι*) in the segment.

And it is plain

(1) that the lines joining the two angles of the figure so inscribed which are nearest to the vertex of the segment, and the next

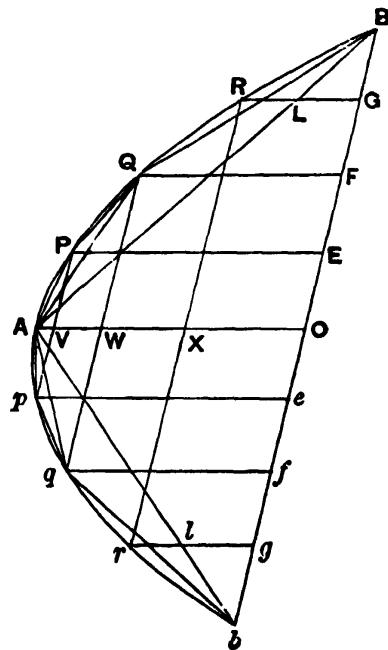
pairs of angles in order, will be parallel to the base of the segment,

- (2) that the said lines will be bisected by the diameter of the segment, and
- (3) that they will cut the diameter in the proportions of the successive odd numbers, the number one having reference to [the length adjacent to] the vertex of the segment.

And these properties will have to be proved in their proper places (*ἐν ταῖς τάξεσιν*).

[The last words indicate an intention to give these propositions in their proper connexion with systematic proofs; but the intention does not appear to have been carried out, or at least we know of no lost work of Archimedes in which they could have appeared. The results can however be easily derived from propositions given in the *Quadrature of the Parabola* as follows.

(1) Let *BRQPApqr&b* be a figure inscribed 'in the recognised manner' in the parabolic segment *BAb* of which *Bb* is the base, *A* the vertex and *AO* the diameter.



Bisect each of the lines *BQ*, *BA*, *QA*, *Aq*, *Ab*, *qb*, and through the middle points draw lines parallel to *AO* meeting *Bb* in *G*, *F*, *E*, *e*, *f*, *g* respectively.

These lines will then pass through the vertices  $R, Q, P, p, q, r$  of the respective parabolic segments [*Quadrature of the Parabola*, Prop. 18], i.e. through the angular points of the inscribed figure (since the triangles and segments are of equal height).

Also  $BG = GF = FE = EO$ , and  $Oe = ef = fg = gb$ . But  $BO = Ob$ , and therefore all the parts into which  $Bb$  is divided are equal.

If now  $AB, RG$  meet in  $L$ , and  $Ab, rg$  in  $l$ , we have

$$\begin{aligned} BG : GL &= BO : OA, \text{ by parallels,} \\ &= bO : OA \\ &= bg : gl, \end{aligned}$$

whence  $GL = gl$ .

Again [*ibid.*, Prop. 4]

$$\begin{aligned} GL : LR &= BO : OG \\ &= bO : Og \\ &= gl : lr; \end{aligned}$$

and, since  $GL = gl, LR = lr$ .

Therefore  $GR, gr$  are equal as well as parallel.

Hence  $GRrg$  is a parallelogram, and  $Rr$  is parallel to  $Bb$ .

Similarly it may be shown that  $Pp, Qq$  are each parallel to  $Bb$ .

(2) Since  $RGgr$  is a parallelogram, and  $RG, rg$  are parallel to  $AO$ , while  $GO = Og$ , it follows that  $Rr$  is bisected by  $AO$ .

And similarly for  $Pp, Qq$ .

(3) Lastly, if  $V, W, X$  be the points of bisection of  $Pp, Qq, Rr$ ,

$$\begin{aligned} AV : AW : AX : AO &= PV^2 : QW^2 : RX^2 : BO^2 \\ &= 1 : 4 : 9 : 16, \end{aligned}$$

whence  $AV : VW : WX : XO = 1 : 3 : 5 : 7.$ ]

**Proposition 2.**

*If a figure be ‘inscribed in the recognised manner’ in a parabolic segment, the centre of gravity of the figure so inscribed will lie on the diameter of the segment.*

For, in the figure of the foregoing lemmas, the centre of gravity of the trapezium  $BRrb$  must lie on  $XO$ , that of the trapezium  $RQqr$  on  $WX$ , and so on, while the centre of gravity of the triangle  $PAp$  lies on  $AV$ .

Hence the centre of gravity of the whole figure lies on  $AO$ .

**Proposition 3.**

*If  $BAB'$ ,  $bab'$  be two similar parabolic segments whose diameters are  $AO$ ,  $ao$  respectively, and if a figure be inscribed in each segment ‘in the recognised manner,’ the number of sides in each figure being equal, the centres of gravity of the inscribed figures will divide  $AO$ ,  $ao$  in the same ratio.*

[Archimedes enunciates this proposition as true of *similar* segments, but it is equally true of segments which are not similar, as the course of the proof will show.]

Suppose  $BRQPAP'Q'R'B'$ ,  $brqpap'q'r'b'$  to be the two figures inscribed ‘in the recognised manner.’ Join  $PP'$ ,  $QQ'$ ,  $RR'$  meeting  $AO$  in  $L$ ,  $M$ ,  $N$ , and  $pp'$ ,  $qq'$ ,  $rr'$  meeting  $ao$  in  $l$ ,  $m$ ,  $n$ .

Then [Lemma (3)]

$$\begin{aligned} AL : LM : MN : NO \\ = 1 : 3 : 5 : 7 \\ = al : lm : mn : no, \end{aligned}$$

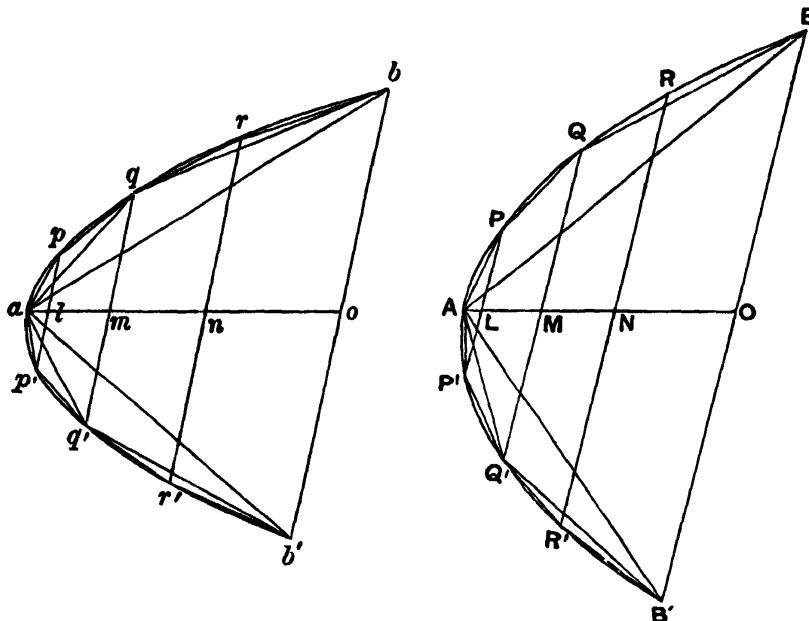
so that  $AO$ ,  $ao$  are divided in the same proportion.

Also, by reversing the proof of Lemma (3), we see that

$$PP' : pp' = QQ' : qq' = RR' : rr' = BB' : bb'.$$

Since then  $RR' : BB' = rr' : bb'$ , and these ratios respectively determine the proportion in which  $NO$ ,  $no$  are divided

by the centres of gravity of the trapezia  $BRR'B'$ ,  $brr'b'$  [I. 15], it follows that the centres of gravity of the trapezia divide  $NO$ ,  $no$  in the same ratio.



Similarly the centres of gravity of the trapezia  $RQQ'R'$ ,  $rqq'r'$  divide  $MN$ ,  $mn$  in the same ratio respectively, and so on.

Lastly, the centres of gravity of the triangles  $PAP'$ ,  $pap'$  divide  $AL$ ,  $al$  respectively in the same ratio.

Moreover the corresponding trapezia and triangles are, each to each, in the same proportion (since their sides and heights are respectively proportional), while  $AO$ ,  $ao$  are divided in the same proportion.

Therefore the centres of gravity of the complete inscribed figures divide  $AO$ ,  $ao$  in the same proportion.

#### Proposition 4.

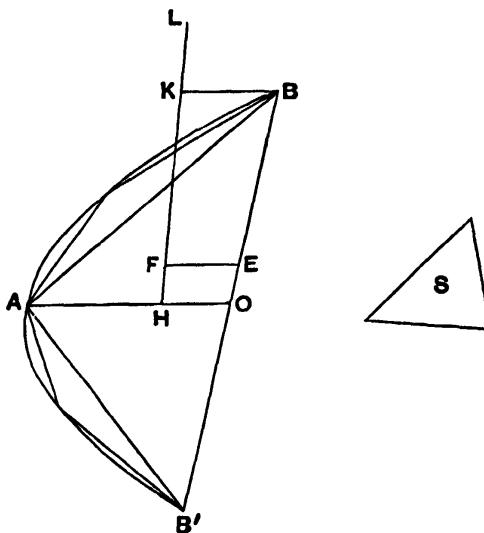
*The centre of gravity of any parabolic segment cut off by a straight line lies on the diameter of the segment.*

Let  $BAB'$  be a parabolic segment,  $A$  its vertex and  $AO$  its diameter.

Then, if the centre of gravity of the segment does not lie on  $AO$ , suppose it to be, if possible, the point  $F$ . Draw  $FE$  parallel to  $AO$  meeting  $BB'$  in  $E$ .

Inscribe in the segment the triangle  $ABB'$  having the same vertex and height as the segment, and take an area  $S$  such that

$$\Delta ABB' : S = BE : EO.$$



We can then inscribe in the segment 'in the recognised manner' a figure such that the segments of the parabola left over are together less than  $S$ . [For Prop. 20 of the *Quadrature of the Parabola* proves that, if in any segment the triangle with the same base and height be inscribed, the triangle is greater than half the segment; whence it appears that, each time that we increase the number of the sides of the figure inscribed 'in the recognised manner,' we take away more than half of the remaining segments.]

Let the inscribed figure be drawn accordingly; its centre of gravity then lies on  $AO$  [Prop. 2]. Let it be the point  $H$ .

Join  $HF$  and produce it to meet in  $K$  the line through  $B$  parallel to  $AO$ .

Then we have

$$\begin{aligned} (\text{inscribed figure}) : (\text{remainder of segmt.}) &> \Delta ABB' : S \\ &> BE : EO \\ &> KF : FH. \end{aligned}$$

Suppose  $L$  taken on  $HK$  produced so that the former ratio is equal to the ratio  $LF : FH$ .

Then, since  $H$  is the centre of gravity of the inscribed figure, and  $F$  that of the segment,  $L$  must be the centre of gravity of all the segments taken together which form the remainder of the original segment. [I. 8]

But this is impossible, since all these segments lie on one side of the line drawn through  $L$  parallel to  $AO$  [Cf. *Post.* 7].

Hence the centre of gravity of the segment cannot but lie on  $AO$ .

### Proposition 5.

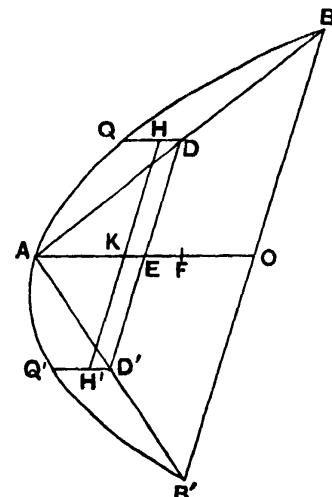
*If in a parabolic segment a figure be inscribed 'in the recognised manner,' the centre of gravity of the segment is nearer to the vertex of the segment than the centre of gravity of the inscribed figure is.*

Let  $BAB'$  be the given segment, and  $AO$  its diameter. First, let  $ABB'$  be the triangle inscribed 'in the recognised manner.'

Divide  $AO$  in  $F$  so that  $AF = 2FO$ ;  $F$  is then the centre of gravity of the triangle  $ABB'$ .

Bisect  $AB, AB'$  in  $D, D'$  respectively, and join  $DD'$  meeting  $AO$  in  $E$ . Draw  $DQ, D'Q'$  parallel to  $OA$  to meet the curve.  $QD, Q'D'$  will then be the diameters of the segments whose bases are  $AB, AB'$ , and the centres of gravity of those segments will lie respectively on  $QD, Q'D'$  [Prop. 4]. Let them be  $H, H'$ , and join  $HH'$  meeting  $AO$  in  $K$ .

Now  $QD, Q'D'$  are equal\*, and therefore the segments of which they are the diameters are equal [*On Conoids and Spheroids*, Prop. 3].



\* This may either be inferred from Lemma (1) above (since  $QQ', DD'$  are both parallel to  $BB'$ ), or from Prop. 19 of the *Quadrature of the Parabola*, which applies equally to  $Q$  or  $Q'$ .

Also, since  $QD, Q'D'$  are parallel\*, and  $DE = ED'$ ,  $K$  is the middle point of  $HH'$ .

Hence the centre of gravity of the equal segments  $AQB, A'Q'B'$  taken together is  $K$ , where  $K$  lies between  $E$  and  $A$ . And the centre of gravity of the triangle  $ABB'$  is  $F$ .

It follows that the centre of gravity of the whole segment  $BAB'$  lies between  $K$  and  $F$ , and is therefore nearer to the vertex  $A$  than  $F$  is.

Secondly, take the five-sided figure  $BQAQ'B'$  inscribed 'in the recognised manner,'  $QD, Q'D'$  being, as before, the diameters of the segments  $AQB, A'Q'B'$ .

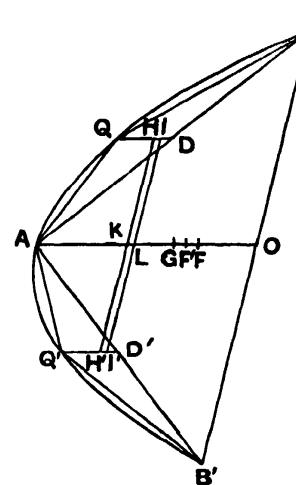
Then, by the first part of this proposition, the centre of gravity of the segment  $AQB$  (lying of course on  $QD$ ) is nearer to  $Q$  than the centre of gravity of the triangle  $AQB$  is. Let the centre of gravity of the segment be  $H$ , and that of the triangle  $I$ .

Similarly let  $H'$  be the centre of gravity of the segment  $A'Q'B'$ , and  $I'$  that of the triangle  $A'Q'B'$ .

It follows that the centre of gravity of the two segments  $AQB, A'Q'B'$  taken together is  $K$ , the middle point of  $HH'$ , and that of the two triangles  $AQB, A'Q'B'$  is  $L$ , the middle point of  $II'$ .

If now the centre of gravity of the triangle  $ABB'$  be  $F$ , the centre of gravity of the whole segment  $BAB'$  (i.e. that of the triangle  $ABB'$  and the two segments  $AQB, A'Q'B'$  taken together) is a point  $G$  on  $KF$  determined by the proportion

$$(\text{sum of segments } AQB, A'Q'B') : \triangle ABB' = FG : GK. \quad [\text{I. 6, 7}]$$



\* There is clearly some interpolation in the text here, which has the words *καὶ ἐπει παραλληλογραμμὸν ἔστι τὸ ΘΖΗΙ*. It is not yet proved that  $H'D'DH$  is a parallelogram; this can only be inferred from the fact that  $H, H'$  divide  $QD, Q'D'$  respectively in the same ratio. But this latter property does not appear till Prop. 7, and is then only enunciated of similar segments. The interpolation must have been made before Eutocius' time, because he has a note on the phrase, and explains it by gravely assuming that  $H, H'$  divide  $QD, Q'D'$  respectively in the same ratio.

And the centre of gravity of the inscribed figure  $BQAQ'B'$  is a point  $F'$  on  $LF$  determined by the proportion

$$(\Delta AQB + \Delta AQ'B') : \Delta ABB' = FF' : F'L. \quad [\text{I. 6, 7}]$$

[Hence  $FG : GK > FF' : F'L$ ,

or  $GK : FG < F'L : FF'$ ,

and, *componendo*,  $FK : FG < FL : FF'$ , while  $FK > FL$ .]

Therefore  $FG > FF'$ , or  $G$  lies nearer than  $F'$  to the vertex  $A$ .

Using this last result, and proceeding in the same way, we can prove the proposition for *any* figure inscribed ‘in the recognised manner.’

### Proposition 6.

*Given a segment of a parabola cut off by a straight line, it is possible to inscribe in it ‘in the recognised manner’ a figure such that the distance between the centres of gravity of the segment and of the inscribed figure is less than any assigned length.*

Let  $BAB'$  be the segment,  $AO$  its diameter,  $G$  its centre of gravity, and  $ABB'$  the triangle inscribed ‘in the recognised manner.’

Let  $D$  be the assigned length and  $S$  an area such that

$$AG : D = \Delta ABB' : S.$$

In the segment inscribe ‘in the recognised manner’ a figure such that the sum of the segments left over is less than  $S$ . Let  $F$  be the centre of gravity of the inscribed figure.

We shall prove that  $FG < D$ .

For, if not,  $FG$  must be either equal to, or greater than,  $D$ .

And clearly

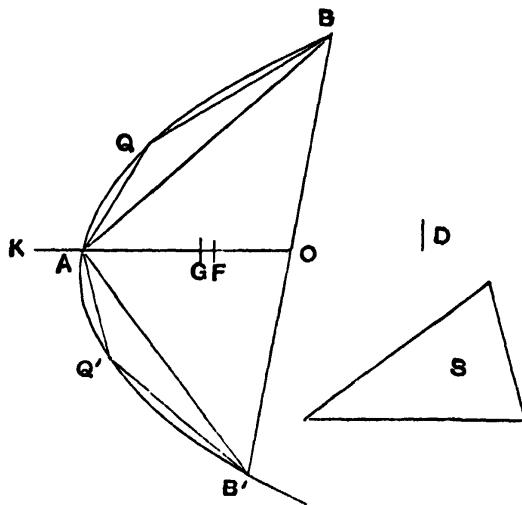
$$(\text{inscribed fig.}) : (\text{sum of remaining segmts.})$$

$$> \Delta ABB' : S$$

$$> AG : D$$

$$> AG : FG, \text{ by hypothesis (since } FG < D\text{).}$$

Let the first ratio be equal to the ratio  $KG : FG$  (where  $K$  lies on  $GA$  produced); and it follows that  $K$  is the centre of gravity of the small segments taken together. [I. 8]



But this is impossible, since the segments are all on the same side of a line drawn through  $K$  parallel to  $BB'$ .

Hence  $FG$  cannot but be less than  $D$ .

### Proposition 7.

*If there be two similar parabolic segments, their centres of gravity divide their diameters in the same ratio.*

[This proposition, though enunciated of *similar* segments only, like Prop. 3 on which it depends, is equally true of *any* segments. This fact did not escape Archimedes, who uses the proposition in its more general form for the proof of Prop. 8 immediately following.]

Let  $BAB'$ ,  $bab'$  be the two similar segments,  $AO$ ,  $ao$  their diameters, and  $G$ ,  $g$  their centres of gravity respectively.

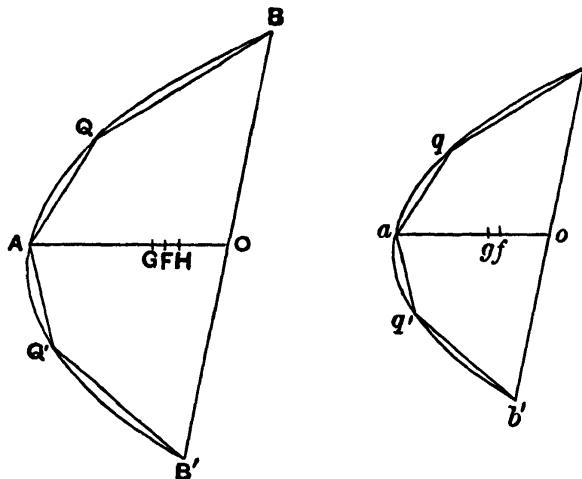
Then, if  $G$ ,  $g$  do not divide  $AO$ ,  $ao$  respectively in the same ratio, suppose  $H$  to be such a point on  $AO$  that

$$AH : HO = ag : go;$$

and inscribe in the segment  $BAB'$  'in the recognised manner' a figure such that, if  $F$  be its centre of gravity,

$$GF < GH.$$

[Prop. 6]



Inscribe in the segment  $bab'$  'in the recognised manner' a similar figure; then, if  $f$  be the centre of gravity of this figure,

$$ag < af. \quad [\text{Prop. 5}]$$

And, by Prop. 3,  $af : fo = AF : FO$ .

But  $AF : FO < AH : HO$

$$< ag : go, \text{ by hypothesis.}$$

Therefore  $af : fo < ag : go$ ; which is impossible.

It follows that  $G, g$  cannot but divide  $AO, ao$  in the same ratio.

### Proposition 8.

If  $AO$  be the diameter of a parabolic segment, and  $G$  its centre of gravity, then

$$AG = \frac{3}{2} GO.$$

Let the segment be  $BAB'$ . Inscribe the triangle  $ABB'$  'in the recognised manner,' and let  $F$  be its centre of gravity.

Bisect  $AB, AB'$  in  $D, D'$ , and draw  $DQ, D'Q'$  parallel to  $OA$  to meet the curve, so that  $QD, Q'D'$  are the diameters of the segments  $AQB, AQ'B'$  respectively.

Let  $H, H'$  be the centres of gravity of the segments  $AQB, AQ'B'$  respectively. Join  $QQ', HH'$  meeting  $AO$  in  $V, K$  respectively.

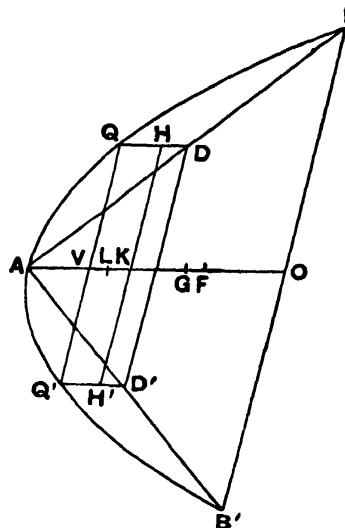
*K* is then the centre of gravity of the two segments  $AQB$ ,  $AQ'B'$  taken together.

Now  $AG : GO = QH : HD$ ,  
[Prop. 7]

$$\text{whence } AO : OG = QD : HD.$$

But  $AO = 4QD$  [as is easily proved by means of Lemma (3), p. 206].

$$\begin{aligned} \text{Therefore } & OG = 4HD; \\ \text{and, by subtraction, } & AG = 4QH. \end{aligned}$$



Also, by Lemma (2),  $QQ'$  is parallel to  $BB'$  and therefore to  $DD'$ . It follows from Prop. 7 that  $HH'$  is also parallel to  $QQ'$  or  $DD'$ , and hence

$$\text{Therefore } AG = 4VK, \text{ and } AV + KG = 3VK.$$

Measuring  $VL$  along  $VK$  so that  $VL = \frac{1}{3} AV$ , we have

$$\begin{aligned} \text{Again } A O &= 4AV \quad [\text{Lemma (3)}] \\ &= 3AL, \text{ since } AV = 3VL, \end{aligned}$$

Now, by I. 6, 7,

$$\triangle ABB' : (\text{sum of segmts. } AQB, AQ'B') = KG : GF,$$

and  $\Delta ABB' = 3 \text{ (sum of segments } AQB, AQB')\}$

[since the segment  $ABB'$  is equal to  $\frac{4}{3} \Delta ABB'$  (*Quadrature of the Parabola*, Props. 17, 24)].

$$\text{Hence } KG = 3GF.$$

But  $KG = 3LK$ , from (1) above.

$$\text{Therefore } LF = LK + KG + GF \\ = 5GF$$

And, from (2),

$$LF = (AO - AL - OF) = \frac{1}{3} AO = OF.$$

Therefore  
and

$$\text{But } AO = 3OF = 15GF.$$

Therefore, by subtraction,

$$AG = 9GF \\ = \frac{3}{2} GO.$$

**Proposition 9 (Lemma).**

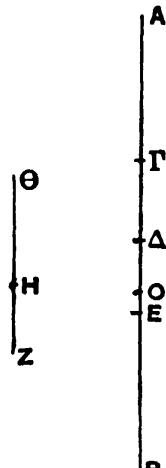
*If  $a, b, c, d$  be four lines in continued proportion and in descending order of magnitude, and if*

$$d : (a - d) = x : \frac{3}{5}(a - c),$$

and  $(2a + 4b + 6c + 3d) : (5a + 10b + 10c + 5d) = y : (a - c)$ ,  
it is required to prove that

$$x + y = \frac{2}{5}a.$$

[The following is the proof given by Archimedes, with the only difference that it is set out in algebraical instead of geometrical notation. This is done in the particular case simply in order to make the proof easier to follow. Archimedes exhibits his lines in the figure reproduced in the margin, but, now that it is possible to use algebraical notation, there is no advantage in using the figure and the more cumbrous notation which only obscures the course of the proof. The relation between Archimedes' figure and the letters used below is as follows;



$\Delta B = a$ ,  $\Gamma B = b$ ,  $\Delta B = c$ ,  $E B = d$ ,  $Z H = x$ ,  $H \Theta = y$ ,  $\Delta O = z$ .]

We have

substance

$$\frac{a-b}{b} = \frac{b-c}{c} = \frac{c-d}{d},$$

and therefore

N

$$\frac{2(a+b)}{2c} = \frac{a+b}{c} = \frac{a+b}{b} \cdot \frac{b}{c} = \frac{a-c}{b-c} \cdot \frac{b-c}{c-d} = \frac{a-c}{c-d}.$$

And, in like manner,

$$\frac{b+c}{d} = \frac{b+c}{c} \cdot \frac{c}{d} = \frac{a-c}{c-d}.$$

It follows from the last two relations that

$$\frac{a-c}{c-d} = \frac{2a+3b+c}{2c+d} \dots \dots \dots \quad (3).$$

Suppose  $z$  to be so taken that

$$\frac{2a+4b+4c+2d}{2c+d} = \frac{a-c}{z} \dots \dots \dots \quad (4),$$

so that  $z < (c-d)$ .

$$\text{Therefore } \frac{a-c+z}{a-c} = \frac{2a+4b+6c+3d}{2(a+d)+4(b+c)}.$$

And, by hypothesis,

$$\frac{a-c}{y} = \frac{5(a+d)+10(b+c)}{2a+4b+6c+3d},$$

$$\text{so that } \frac{a-c+z}{y} = \frac{5(a+d)+10(b+c)}{2(a+d)+4(b+c)} = \frac{5}{2} \dots \dots \dots \quad (5).$$

Again, dividing (3) by (4) crosswise, we obtain

$$\frac{z}{c-d} = \frac{2a+3b+c}{2(a+d)+4(b+c)},$$

$$\text{whence } \frac{c-d-z}{c-d} = \frac{b+3c+2d}{2(a+d)+4(b+c)} \dots \dots \dots \quad (6).$$

But, by (2),

$$\frac{c-d}{d} = \frac{a-b}{b} = \frac{3(b-c)}{3c} = \frac{2(c-d)}{2d},$$

$$\text{so that } \frac{c-d}{d} = \frac{(a-b)+3(b-c)+2(c-d)}{b+3c+2d} \dots \dots \dots \quad (7).$$

Combining (6) and (7), we have

$$\frac{c-d-z}{d} = \frac{(a-b)+3(b-c)+2(c-d)}{2(a+d)+4(b+c)},$$

$$\text{whence } \frac{c-z}{d} = \frac{3a+6b+3c}{2(a+d)+4(b+c)} \dots \dots \dots \quad (8).$$

And, since [by (1)]

$$\frac{c-d}{c+d} = \frac{b-c}{b+c} = \frac{a-b}{a+b},$$

we have

$$\frac{c-d}{a-c} = \frac{c+d}{b+c+a+b},$$

$$\text{whence } \frac{a-d}{a-c} = \frac{a+2b+2c+d}{a+2b+c} = \frac{2(a+d) + 4(b+c)}{2(a+c) + 4b} \dots\dots(9).$$

$$\text{Thus } \frac{a-d}{\frac{3}{5}(a-c)} = \frac{2(a+d) + 4(b+c)}{\frac{3}{5}\{2(a+c) + 4b\}},$$

and therefore, by hypothesis,

$$\frac{d}{x} = \frac{2(a+d) + 4(b+c)}{\frac{3}{5}\{2(a+c) + 4b\}}.$$

$$\text{But, by (8), } \frac{c-z}{d} = \frac{3a+6b+3c}{2(a+d)+4(b+c)};$$

and it follows, *ex aequali*, that

$$\frac{c-z}{x} = \frac{3(a+c) + 6b}{\{2(a+c) + 4b\}} = \frac{5}{3} \cdot \frac{3}{2} = \frac{5}{2}.$$

And, by (5),

$$-\frac{a - c + z}{y} = \frac{5}{2}.$$

Therefore

$$\frac{5}{2} = \frac{a}{x+y},$$

or

$$x + y = \frac{2}{5}a.$$

### **Proposition 10.**

If  $PP'B'B$  be the portion of a parabola intercepted between two parallel chords  $PP'$ ,  $BB'$  bisected respectively in  $N$ ,  $O$  by the diameter  $ANO$  ( $N$  being nearer than  $O$  to  $A$ , the vertex of the segments), and if  $NO$  be divided into five equal parts of which  $LM$  is the middle one ( $L$  being nearer than  $M$  to  $N$ ), then, if  $G$  be a point on  $LM$  such that

$$LG : GM = BO^2 \cdot (2PN + BO) : PN^2 \cdot (2BO + PN),$$

*G will be the centre of gravity of the area  $PP'B'B$ .*

Take a line  $ao$  equal to  $AO$ , and  $an$  on it equal to  $AN$ . Let  $p, q$  be points on the line  $ao$  such that

[whence  $ao : aq = aq : an = an : ap$ , or  $ao, aq, an, ap$  are lines in continued proportion and in descending order of magnitude].

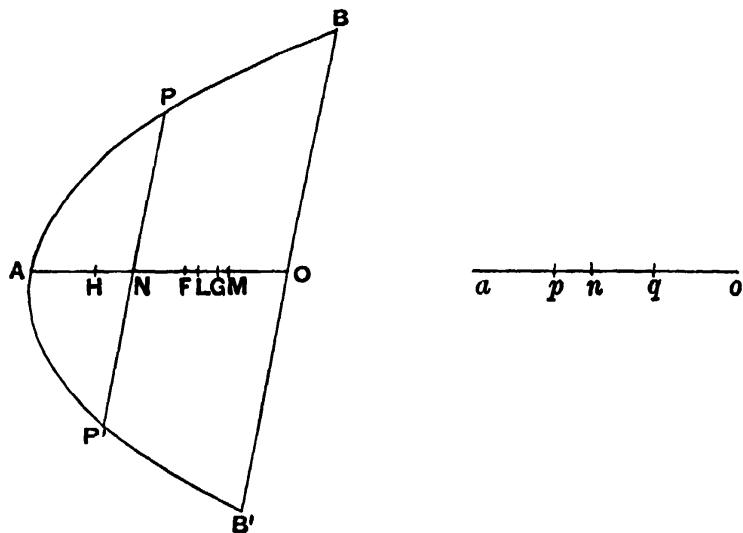
Measure along  $GA$  a length  $GF$  such that

Then, since  $PN$ ,  $BO$  are ordinates to  $ANO$ ,

$$\begin{aligned}BO^2 : PN^2 &= AO : AN \\&= ao : an \\&= ao^2 : ag^2, \text{ by (1),}\end{aligned}$$

so that

and



Thus (segment  $BAB'$ ) : (segment  $PAP'$ )

$$= \Delta BAB' : \Delta PAP' \\ = BO^3 : PN^3 \\ = ao : ap,$$

whence

$$\begin{aligned}
 (\text{area } PP'B'B) : (\text{segment } PAP') &= op : ap \\
 &= OL : GF, \text{ by (3),} \\
 &= \frac{3}{5}ON : GF \dots\dots\dots (6).
 \end{aligned}$$

$$\text{Now } BO^2 \cdot (2PN + BO) : BO^3 = (2PN + BO) : BO \\ = (2aq + ao) : ao, \text{ by (4),}$$

$$\text{and } PN^3 : PN^2 \cdot (2BO + PN) = PN : (2BO + PN)$$

$$= aq : (2ao + aq), \text{ by (4),}$$

$$= ap : (2an + ap), \text{ by (2).}$$

Hence, *ex aequali*,

$$BO^2 \cdot (2PN + BO) : PN^2 \cdot (2BO + PN) = (2aq + ao) : (2an + ap),$$

so that, by hypothesis,

$$LG : GM = (2aq + ao) : (2an + ap).$$

*Componendo*, and multiplying the antecedents by 5,

$$ON : GM = \{5(ao + ap) + 10(aq + an)\} : (2an + ap).$$

$$\text{But } ON : OM = 5 : 2$$

$$= \{5(ao + ap) + 10(aq + an)\} : \{2(ao + ap) + 4(aq + an)\}.$$

It follows that

$$ON : OG = \{5(ao + ap) + 10(aq + an)\} : (2ao + 4aq + 6an + 3ap).$$

Therefore

$$(2ao + 4aq + 6an + 3ap) : \{5(ao + ap) + 10(aq + an)\} = OG : ON$$

$$= OG : on.$$

And

$$ap : (ao - ap) = ap : op$$

$$= GF : OL, \text{ by hypothesis,}$$

$$= GF : \frac{3}{5}on,$$

while  $ao, aq, an, ap$  are in continued proportion.

Therefore, by Prop. 9,

$$GF + OG = OF = \frac{2}{5}ao = \frac{2}{5}OA.$$

Thus  $F$  is the centre of gravity of the segment  $BAB'$ . [Prop. 8]

Let  $H$  be the centre of gravity of the segment  $PAP'$ , so that  $AH = \frac{3}{5}AN$ .

And, since  $AF = \frac{3}{5}AO$ ,

we have, by subtraction,  $HF = \frac{3}{5}ON$ .

But, by (6) above,

$$\begin{aligned} (\text{area } PP'B'B) : (\text{segment } PAP') &= \frac{3}{5}ON : GF \\ &= HF : FG. \end{aligned}$$

Thus, since  $F, H$  are the centres of gravity of the segments  $BAB'$ ,  $PAP'$  respectively, it follows [by I. 6, 7] that  $G$  is the centre of gravity of the area  $PP'B'B$ .

# ON FLOATING BODIES.

## BOOK I.

---

### *Postulate 1.*

“Let it be supposed that a fluid is of such a character that, its parts lying evenly and being continuous, that part which is thrust the less is driven along by that which is thrust the more; and that each of its parts is thrust by the fluid which is above it in a perpendicular direction if the fluid be sunk in anything and compressed by anything else.”

### **Proposition 1.**

*If a surface be cut by a plane always passing through a certain point, and if the section be always a circumference [of a circle] whose centre is the aforesaid point, the surface is that of a sphere.*

For, if not, there will be some two lines drawn from the point to the surface which are not equal.

Suppose  $O$  to be the fixed point, and  $A, B$  to be two points on the surface such that  $OA, OB$  are unequal. Let the surface be cut by a plane passing through  $OA, OB$ . Then the section is, by hypothesis, a circle whose centre is  $O$ .

Thus  $OA = OB$ ; which is contrary to the assumption. Therefore the surface cannot but be a sphere.

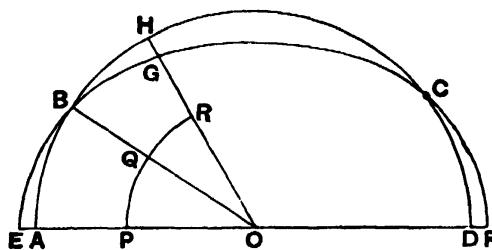
**Proposition 2.**

*The surface of any fluid at rest is the surface of a sphere whose centre is the same as that of the earth.*

Suppose the surface of the fluid cut by a plane through  $O$ , the centre of the earth, in the curve  $ABCD$ .

$ABCD$  shall be the circumference of a circle.

For, if not, some of the lines drawn from  $O$  to the curve will be unequal. Take one of them,  $OB$ , such that  $OB$  is greater than some of the lines from  $O$  to the curve and less than others. Draw a circle with  $OB$  as radius. Let it be  $EBF$ , which will therefore fall partly within and partly without the surface of the fluid.



Draw  $OGH$  making with  $OB$  an angle equal to the angle  $EOB$ , and meeting the surface in  $H$  and the circle in  $G$ . Draw also in the plane an arc of a circle  $PQR$  with centre  $O$  and within the fluid.

Then the parts of the fluid along  $PQR$  are uniform and continuous, and the part  $PQ$  is compressed by the part between it and  $AB$ , while the part  $QR$  is compressed by the part between  $QR$  and  $BH$ . Therefore the parts along  $PQ$ ,  $QR$  will be unequally compressed, and the part which is compressed the less will be set in motion by that which is compressed the more.

Therefore there will not be rest; which is contrary to the hypothesis.

Hence the section of the surface will be the circumference of a circle whose centre is  $O$ ; and so will all other sections by planes through  $O$ .

Therefore the surface is that of a sphere with centre  $O$ .

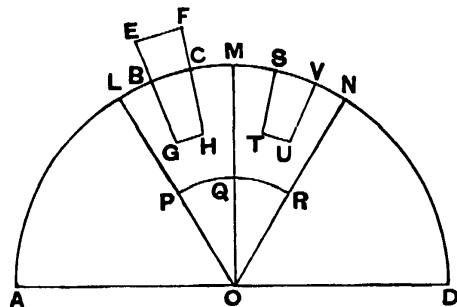
### Proposition 3.

*Of solids those which, size for size, are of equal weight with a fluid will, if let down into the fluid, be immersed so that they do not project above the surface but do not sink lower.*

If possible, let a certain solid  $EFHG$  of equal weight, volume for volume, with the fluid remain immersed in it so that part of it,  $EBCF$ , projects above the surface.

Draw through  $O$ , the centre of the earth, and through the solid a plane cutting the surface of the fluid in the circle  $ABCD$ .

Conceive a pyramid with vertex  $O$  and base a parallelogram at the surface of the fluid, such that it includes the immersed portion of the solid. Let this pyramid be cut by the plane of  $ABCD$  in  $OL, OM$ . Also let a sphere within the fluid and below  $GH$  be described with centre  $O$ , and let the plane of  $ABCD$  cut this sphere in  $PQR$ .



Conceive also another pyramid in the fluid with vertex  $O$ , continuous with the former pyramid and equal and similar to it. Let the pyramid so described be cut in  $OM, ON$  by the plane of  $ABCD$ .

Lastly, let  $STUV$  be a part of the fluid within the second pyramid equal and similar to the part  $BGHC$  of the solid, and let  $SV$  be at the surface of the fluid.

Then the pressures on  $PQ, QR$  are unequal, that on  $PQ$  being the greater. Hence the part at  $QR$  will be set in motion

by that at  $PQ$ , and the fluid will not be at rest; which is contrary to the hypothesis.

Therefore the solid will not stand out above the surface.

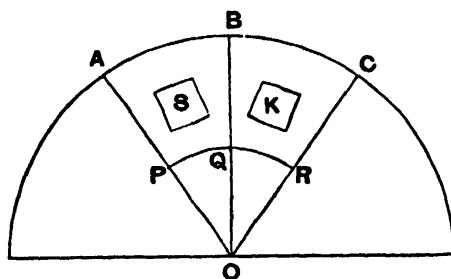
Nor will it sink further, because all the parts of the fluid will be under the same pressure.

#### **Proposition 4.**

*A solid lighter than a fluid will, if immersed in it, not be completely submerged, but part of it will project above the surface.*

In this case, after the manner of the previous proposition, we assume the solid, if possible, to be completely submerged and the fluid to be at rest in that position, and we conceive (1) a pyramid with its vertex at  $O$ , the centre of the earth, including the solid, (2) another pyramid continuous with the former and equal and similar to it, with the same vertex  $O$ , (3) a portion of the fluid within this latter pyramid equal to the immersed solid in the other pyramid, (4) a sphere with centre  $O$  whose surface is below the immersed solid and the part of the fluid in the second pyramid corresponding thereto. We suppose a plane to be drawn through the centre  $O$  cutting the surface of the fluid in the circle  $ABC$ , the solid in  $S$ , the first pyramid in  $OA$ ,  $OB$ , the second pyramid in  $OB$ ,  $OC$ , the portion of the fluid in the second pyramid in  $K$ , and the inner sphere in  $PQR$ .

Then the pressures on the parts of the fluid at  $PQ$ ,  $QR$  are unequal, since  $S$  is lighter than  $K$ . Hence there will not be rest; which is contrary to the hypothesis.

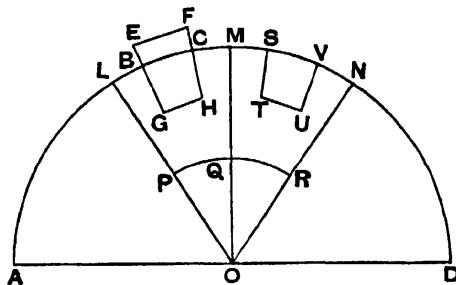


Therefore the solid  $S$  cannot, in a condition of rest, be completely submerged.

**Proposition 5.**

*Any solid lighter than a fluid will, if placed in the fluid, be so far immersed that the weight of the solid will be equal to the weight of the fluid displaced.*

For let the solid be  $EGHF$ , and let  $BGH$  be the portion of it immersed when the fluid is at rest. As in Prop. 3, conceive a pyramid with vertex  $O$  including the solid, and another pyramid with the same vertex continuous with the former and equal and similar to it. Suppose a portion of the fluid  $STUV$  at the base of the second pyramid to be equal and similar to the immersed portion of the solid; and let the construction be the same as in Prop. 3.



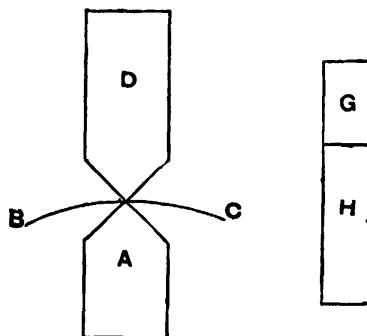
Then, since the pressure on the parts of the fluid at  $PQ$ ,  $QR$  must be equal in order that the fluid may be at rest, it follows that the weight of the portion  $STUV$  of the fluid must be equal to the weight of the solid  $EGHF$ . And the former is equal to the weight of the fluid displaced by the immersed portion of the solid  $BGH$ .

**Proposition 6.**

*If a solid lighter than a fluid be forcibly immersed in it, the solid will be driven upwards by a force equal to the difference between its weight and the weight of the fluid displaced.*

For let  $A$  be completely immersed in the fluid, and let  $G$  represent the weight of  $A$ , and  $(G + H)$  the weight of an equal volume of the fluid. Take a solid  $D$ , whose weight is  $H$

and add it to  $A$ . Then the weight of  $(A + D)$  is less than that of an equal volume of the fluid; and, if  $(A + D)$  is immersed in the fluid, it will project so that its weight will be equal to the weight of the fluid displaced. But its weight is  $(G + H)$ .



Therefore the weight of the fluid displaced is  $(G + H)$ , and hence the volume of the fluid displaced is the volume of the solid  $A$ . There will accordingly be rest with  $A$  immersed and  $D$  projecting.

Thus the weight of  $D$  balances the upward force exerted by the fluid on  $A$ , and therefore the latter force is equal to  $H$ , which is the difference between the weight of  $A$  and the weight of the fluid which  $A$  displaces.

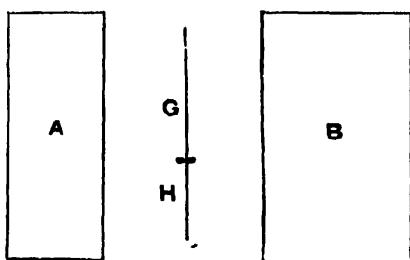
### Proposition 7.

*A solid heavier than a fluid will, if placed in it, descend to the bottom of the fluid, and the solid will, when weighed in the fluid, be lighter than its true weight by the weight of the fluid displaced.*

(1) The first part of the proposition is obvious, since the part of the fluid under the solid will be under greater pressure, and therefore the other parts will give way until the solid reaches the bottom.

(2) Let  $A$  be a solid heavier than the same volume of the fluid, and let  $(G + H)$  represent its weight, while  $G$  represents the weight of the same volume of the fluid.

Take a solid  $B$  lighter than the same volume of the fluid, and such that the weight of  $B$  is  $G$ , while the weight of the same volume of the fluid is  $(G + H)$ .



Let  $A$  and  $B$  be now combined into one solid and immersed. Then, since  $(A + B)$  will be of the same weight as the same volume of fluid, both weights being equal to  $(G + H) + G$ , it follows that  $(A + B)$  will remain stationary in the fluid.

Therefore the force which causes  $A$  by itself to sink must be equal to the upward force exerted by the fluid on  $B$  by itself. This latter is equal to the difference between  $(G + H)$  and  $G$  [Prop. 6]. Hence  $A$  is depressed by a force equal to  $H$ , i.e. its weight in the fluid is  $H$ , or the difference between  $(G + H)$  and  $G$ .

---

[This proposition may, I think, safely be regarded as decisive of the question how Archimedes determined the proportions of gold and silver contained in the famous crown (cf. Introduction, Chapter I.). The proposition suggests in fact the following method.

Let  $W$  represent the weight of the crown,  $w_1$  and  $w_2$  the weights of the gold and silver in it respectively, so that  $W = w_1 + w_2$ .

(1) Take a weight  $W$  of pure gold and weigh it in a fluid. The apparent loss of weight is then equal to the weight of the fluid displaced. If  $F_1$  denote this weight,  $F_1$  is thus known as the result of the operation of weighing.

It follows that the weight of fluid displaced by a weight  $w_1$  of gold is  $\frac{w_1}{W} \cdot F_1$ .

(2) Take a weight  $W$  of pure silver and perform the same operation. If  $F_2$  be the loss of weight when the silver is weighed in the fluid, we find in like manner that the weight of fluid displaced by  $w_2$  is  $\frac{w_2}{W} \cdot F_2$ .

(3) Lastly, weigh the crown itself in the fluid, and let  $F$  be the loss of weight. Therefore the weight of fluid displaced by the crown is  $F$ .

$$\text{It follows that } \frac{w_1}{W} \cdot F_1 + \frac{w_2}{W} \cdot F_2 = F,$$

or

$$w_1 F_1 + w_2 F_2 = (w_1 + w_2) F,$$

whence

$$\frac{w_1}{w_2} = \frac{F_2 - F}{F - F_1}.$$

This procedure corresponds pretty closely to that described in the poem *de ponderibus et mensuris* (written probably about 500 A.D.)\* purporting to explain Archimedes' method. According to the author of this poem, we first take two equal weights of pure gold and pure silver respectively and weigh them against each other when both immersed in water; this gives the relation between their weights in water and therefore between their loss of weight in water. Next we take the mixture of gold and silver and an equal weight of pure silver and weigh them against each other in water in the same manner.

The other version of the method used by Archimedes is that given by Vitruvius†, according to which he measured successively the *volumes* of fluid displaced by three equal weights, (1) the crown, (2) the same weight of gold, (3) the same weight of silver, respectively. Thus, if as before the weight of the crown is  $W$ , and it contains weights  $w_1$  and  $w_2$  of gold and silver respectively,

- (1) the crown displaces a certain quantity of fluid,  $V$  say.
- (2) the weight  $W$  of gold displaces a certain volume of

\* Torelli's *Archimedes*, p. 364; Hultsch, *Metrol. Script.* II. 95 sq., and *Prolegomena* § 118.

† *De architect.* ix. 3.

fluid,  $V_1$  say; therefore a weight  $w_1$  of gold displaces a volume  $\frac{w_1}{W} \cdot V_1$  of fluid.

(3) the weight  $W$  of silver displaces a certain volume of fluid, say  $V_2$ ; therefore a weight  $w_2$  of silver displaces a volume  $\frac{w_2}{W} \cdot V_2$  of fluid.

$$\text{It follows that } V = \frac{w_1}{W} \cdot V_1 + \frac{w_2}{W} \cdot V_2,$$

$$\text{whence, since } W = w_1 + w_2,$$

$$\frac{w_1}{w_2} = \frac{V_2 - V}{V - V_1};$$

and this ratio is obviously equal to that before obtained, viz.  
 $\frac{F_2 - F}{F - F_1}$ .]

### *Postulate 2.*

“Let it be granted that bodies which are forced upwards in a fluid are forced upwards along the perpendicular [to the surface] which passes through their centre of gravity.”

### **Proposition 8.**

*If a solid in the form of a segment of a sphere, and of a substance lighter than a fluid, be immersed in it so that its base does not touch the surface, the solid will rest in such a position that its axis is perpendicular to the surface; and, if the solid be forced into such a position that its base touches the fluid on one side and be then set free, it will not remain in that position but will return to the symmetrical position.*

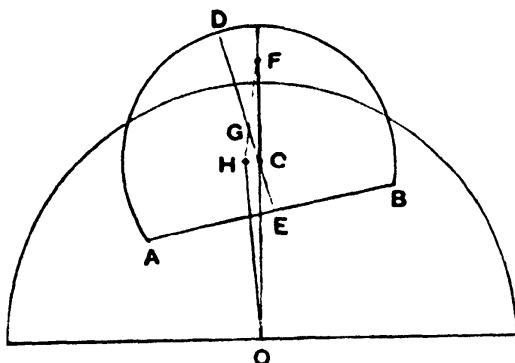
[The proof of this proposition is wanting in the Latin version of Tartaglia. Commandinus supplied a proof of his own in his edition.]

### **Proposition 9.**

*If a solid in the form of a segment of a sphere, and of a substance lighter than a fluid, be immersed in it so that its base is completely below the surface, the solid will rest in such a position that its axis is perpendicular to the surface.*

[The proof of this proposition has only survived in a mutilated form. It deals moreover with only one case out of three which are distinguished at the beginning, viz. that in which the segment is greater than a hemisphere, while figures only are given for the cases where the segment is equal to, or less than, a hemisphere.]

Suppose, first, that the segment is greater than a hemisphere. Let it be cut by a plane through its axis and the centre of the earth; and, if possible, let it be at rest in the position shown in the figure, where  $AB$  is the intersection of the plane with the base of the segment,  $DE$  its axis,  $C$  the centre of the sphere of which the segment is a part,  $O$  the centre of the earth.



The centre of gravity of the portion of the segment outside the fluid, as  $F$ , lies on  $OC$  produced, its axis passing through  $C$ .

Let  $G$  be the centre of gravity of the segment. Join  $FG$ , and produce it to  $H$  so that

$FG : GH = (\text{volume of immersed portion}) : (\text{rest of solid}).$   
Join  $OH$ .

Then the weight of the portion of the solid outside the fluid acts along  $FO$ , and the pressure of the fluid on the immersed portion along  $OH$ , while the weight of the immersed portion acts along  $HO$  and is by hypothesis less than the pressure of the fluid acting along  $OH$ .

Hence there will not be equilibrium, but the part of the segment towards  $A$  will ascend and the part towards  $B$  descend, until  $DE$  assumes a position perpendicular to the surface of the fluid.



# **Discourse on the Method**

Descartes, René  
1637

Good sense is, of all things among men, the most equally distributed; for every one thinks himself so abundantly provided with it, that those even who are the most difficult to satisfy in everything else, do not usually desire a larger measure of this quality than they already possess. And in this it is not likely that all are mistaken the conviction is rather to be held as testifying that the power of judging aright and of distinguishing truth from error, which is properly what is called good sense or reason, is by nature equal in all men; and that the diversity of our opinions, consequently, does not arise from some being endowed with a larger share of reason than others, but solely from this, that we conduct our thoughts along different ways, and do not fix our attention on the same objects. For to be possessed of a vigorous mind is not enough; the prime requisite is rightly to apply it. The greatest minds, as they are capable of the highest excellences, are open likewise to the greatest aberrations; and those who travel very slowly may yet make far greater progress, provided they keep always to the straight road, than those who, while they run, forsake it.

For myself, I have never fancied my mind to be in any respect more perfect than those of the generality; on the contrary, I have often wished that I were equal to some others in promptitude of thought, or in clearness and distinctness of imagination, or in fullness and readiness of memory. And besides these, I know of no other qualities that contribute to the perfection of the mind; for as to the reason or sense, inasmuch as it is that alone which constitutes us men, and distinguishes us from the brutes, I am disposed to believe that it is to be found complete in each individual; and on this point to adopt the common opinion of philosophers, who say that the difference of greater and less holds only among the accidents, and not among the forms or natures of individuals of the same species.

I will not hesitate, however, to avow my belief that it has been my singular good fortune to have very early in life fallen in with certain tracks which have conducted me to considerations and maxims, of which I have formed a method that gives me the means, as I think, of gradually augmenting my knowledge, and of raising it by little and little to the highest point which the mediocrity of my talents and the brief duration of my life will permit me to reach. For I have already reaped from it such fruits that, although I have been accustomed to think lowly enough of myself, and although when I look with the eye of a philosopher at the varied courses and pursuits of mankind at large, I find scarcely one which does not appear in vain and useless, I nevertheless derive the

highest satisfaction from the progress I conceive myself to have already made in the search after truth, and cannot help entertaining such expectations of the future as to believe that if, among the occupations of men as men, there is any one really excellent and important, it is that which I have chosen.

After all, it is possible I may be mistaken; and it is but a little copper and glass, perhaps, that I take for gold and diamonds. I know how very liable we are to delusion in what relates to ourselves, and also how much the judgments of our friends are to be suspected when given in our favor. But I shall endeavor in this discourse to describe the paths I have followed, and to delineate my life as in a picture, in order that each one may also be able to judge of them for himself, and that in the general opinion entertained of them, as gathered from current report, I myself may have a new help towards instruction to be added to those I have been in the habit of employing.

My present design, then, is not to teach the method which each ought to follow for the right conduct of his reason, but solely to describe the way in which I have endeavored to conduct my own. They who set themselves to give precepts must of course regard themselves as possessed of greater skill than those to whom they prescribe; and if they err in the slightest particular, they subject themselves to censure. But as this tract is put forth merely as a history, or, if you will, as a tale, in which, amid some examples worthy of imitation, there will be found, perhaps, as many more which it were advisable not to follow, I hope it will prove useful to some without being hurtful to any, and that my openness will find some favor with all.

From my childhood, I have been familiar with letters; and as I was given to believe that by their help a clear and certain knowledge of all that is useful in life might be acquired, I was ardently desirous of instruction. But as soon as I had finished the entire course of study, at the close of which it is customary to be admitted into the order of the learned, I completely changed my opinion. For I found myself involved in so many doubts and errors, that I was convinced I had advanced no farther in all my attempts at learning, than the discovery at every turn of my own ignorance. And yet I was studying in one of the most celebrated schools in Europe, in which I thought there must be learned men, if such were anywhere to be found. I had been taught all that others learned there; and not contented with the sciences actually taught us, I had, in addition, read all the books that had fallen into my hands, treating of such branches as are esteemed the most curious and rare. I knew the

judgment which others had formed of me; and I did not find that I was considered inferior to my fellows, although there were among them some who were already marked out to fill the places of our instructors. And, in fine, our age appeared to me as flourishing, and as fertile in powerful minds as any preceding one. I was thus led to take the liberty of judging of all other men by myself, and of concluding that there was no science in existence that was of such a nature as I had previously been given to believe.

I still continued, however, to hold in esteem the studies of the schools. I was aware that the languages taught in them are necessary to the understanding of the writings of the ancients; that the grace of fable stirs the mind; that the memorable deeds of history elevate it; and, if read with discretion, aid in forming the judgment; that the perusal of all excellent books is, as it were, to interview with the noblest men of past ages, who have written them, and even a studied interview, in which are discovered to us only their choicest thoughts; that eloquence has incomparable force and beauty; that poesy has its ravishing graces and delights; that in the mathematics there are many refined discoveries eminently suited to gratify the inquisitive, as well as further all the arts and lessen the labour of man; that numerous highly useful precepts and exhortations to virtue are contained in treatises on morals; that theology points out the path to heaven; that philosophy affords the means of discoursing with an appearance of truth on all matters, and commands the admiration of the more simple; that jurisprudence, medicine, and the other sciences, secure for their cultivators honors and riches; and, in fine, that it is useful to bestow some attention upon all, even upon those abounding the most in superstition and error, that we may be in a position to determine their real value, and guard against being deceived.

But I believed that I had already given sufficient time to languages, and likewise to the reading of the writings of the ancients, to their histories and fables. For to hold converse with those of other ages and to travel, are almost the same thing. It is useful to know something of the manners of different nations, that we may be enabled to form a more correct judgment regarding our own, and be prevented from thinking that everything contrary to our customs is ridiculous and irrational, a conclusion usually come to by those whose experience has been limited to their own country. On the other hand, when too much time is occupied in traveling, we become strangers to our native country; and the over curious in the customs of the past are generally ignorant of those of the present. Besides, fictitious narratives lead us to imagine the possibility of

many events that are impossible; and even the most faithful histories, if they do not wholly misrepresent matters, or exaggerate their importance to render the account of them more worthy of perusal, omit, at least, almost always the meanest and least striking of the attendant circumstances; hence it happens that the remainder does not represent the truth, and that such as regulate their conduct by examples drawn from this source, are apt to fall into the extravagances of the knight-errants of romance, and to entertain projects that exceed their powers.

I esteemed eloquence highly, and was in raptures with poesy; but I thought that both were gifts of nature rather than fruits of study. Those in whom the faculty of reason is predominant, and who most skillfully dispose their thoughts with a view to render them clear and intelligible, are always the best able to persuade others of the truth of what they lay down, though they should speak only in the language of Lower Brittany, and be wholly ignorant of the rules of rhetoric; and those whose minds are stored with the most agreeable fancies, and who can give expression to them with the greatest embellishment and harmony, are still the best poets, though unacquainted with the art of poetry.

I was especially delighted with the mathematics, on account of the certitude and evidence of their reasonings; but I had not as yet a precise knowledge of their true use; and thinking that they but contributed to the advancement of the mechanical arts, I was astonished that foundations, so strong and solid, should have had no loftier superstructure reared on them. On the other hand, I compared the disquisitions of the ancient moralists to very towering and magnificent palaces with no better foundation than sand and mud: they laud the virtues very highly, and exhibit them as estimable far above anything on earth; but they give us no adequate criterion of virtue, and frequently that which they designate with so fine a name is but apathy, or pride, or despair, or parricide.

I revered our theology, and aspired as much as any one to reach heaven: but being given assuredly to understand that the way is not less open to the most ignorant than to the most learned, and that the revealed truths which lead to heaven are above our comprehension, I did not presume to subject them to the impotency of my reason; and I thought that in order competently to undertake their examination, there was need of some special help from heaven, and of being more than man.

Of philosophy I will say nothing, except that when I saw that it had been cultivated for many ages by the most distinguished men, and that yet there is not a single matter within its sphere which is not still in dispute, and nothing, therefore, which is above doubt, I did not presume to

anticipate that my success would be greater in it than that of others; and further, when I considered the number of conflicting opinions touching a single matter that may be upheld by learned men, while there can be but one true, I reckoned as well-nigh false all that was only probable.

As to the other sciences, inasmuch as these borrow their principles from philosophy, I judged that no solid superstructures could be reared on foundations so infirm; and neither the honor nor the gain held out by them was sufficient to determine me to their cultivation: for I was not, thank Heaven, in a condition which compelled me to make merchandise of science for the bettering of my fortune; and though I might not profess to scorn glory as a cynic, I yet made very slight account of that honor which I hoped to acquire only through fictitious titles. And, in fine, of false sciences I thought I knew the worth sufficiently to escape being deceived by the professions of an alchemist, the predictions of an astrologer, the impostures of a magician, or by the artifices and boasting of any of those who profess to know things of which they are ignorant.

For these reasons, as soon as my age permitted me to pass from under the control of my instructors, I entirely abandoned the study of letters, and resolved no longer to seek any other science than the knowledge of myself, or of the great book of the world. I spent the remainder of my youth in traveling, in visiting courts and armies, in holding intercourse with men of different dispositions and ranks, in collecting varied experience, in proving myself in the different situations into which fortune threw me, and, above all, in making such reflection on the matter of my experience as to secure my improvement. For it occurred to me that I should find much more truth in the reasonings of each individual with reference to the affairs in which he is personally interested, and the issue of which must presently punish him if he has judged amiss, than in those conducted by a man of letters in his study, regarding speculative matters that are of no practical moment, and followed by no consequences to himself, farther, perhaps, than that they foster his vanity the better the more remote they are from common sense; requiring, as they must in this case, the exercise of greater ingenuity and art to render them probable. In addition, I had always a most earnest desire to know how to distinguish the true from the false, in order that I might be able clearly to discriminate the right path in life, and proceed in it with confidence.

It is true that, while busied only in considering the manners of other men, I found here, too, scarce any ground for settled conviction, and remarked hardly less contradiction among them than in the opinions of the philosophers. So that the greatest advantage I derived from the study

consisted in this, that, observing many things which, however extravagant and ridiculous to our apprehension, are yet by common consent received and approved by other great nations, I learned to entertain too decided a belief in regard to nothing of the truth of which I had been persuaded merely by example and custom; and thus I gradually extricated myself from many errors powerful enough to darken our natural intelligence, and incapacitate us in great measure from listening to reason. But after I had been occupied several years in thus studying the book of the world, and in essaying to gather some experience, I at length resolved to make myself an object of study, and to employ all the powers of my mind in choosing the paths I ought to follow, an undertaking which was accompanied with greater success than it would have been had I never quitted my country or my books.

I was then in Germany, attracted thither by the wars in that country, which have not yet been brought to a termination; and as I was returning to the army from the coronation of the emperor, the setting in of winter arrested me in a locality where, as I found no society to interest me, and was besides fortunately undisturbed by any cares or passions, I remained the whole day in seclusion, with full opportunity to occupy my attention with my own thoughts. Of these one of the very first that occurred to me was, that there is seldom so much perfection in works composed of many separate parts, upon which different hands had been employed, as in those completed by a single master. Thus it is observable that the buildings which a single architect has planned and executed, are generally more elegant and commodious than those which several have attempted to improve, by making old walls serve for purposes for which they were not originally built. Thus also, those ancient cities which, from being at first only villages, have become, in course of time, large towns, are usually but ill laid out compared with the regularity constructed towns which a professional architect has freely planned on an open plain; so that although the several buildings of the former may often equal or surpass in beauty those of the latter, yet when one observes their indiscriminate juxtaposition, there a large one and here a small, and the consequent crookedness and irregularity of the streets, one is disposed to allege that chance rather than any human will guided by reason must have led to such an arrangement. And if we consider that nevertheless there have been at all times certain officers whose duty it was to see that private buildings contributed to public ornament, the difficulty of reaching high perfection with but the materials of others to operate on, will be readily acknowledged. In the same way I fancied that those nations which, starting from a semi-barbarous state and advancing to civilization by slow degrees, have had their laws successively determined, and, as it were, forced upon them simply by experience of the hurtfulness of particular crimes and disputes, would by this process come to be possessed of less perfect institutions than those which, from the commencement of their association as communities, have followed the appointments of some wise legislator. It is thus quite certain that the constitution of the true religion, the ordinances of which are derived from God, must be incomparably superior to that of every other. And, to speak of human affairs, I believe that the pre-eminence of Sparta was due not to the goodness of each of its laws in particular, for many of these were very strange, and even opposed to good morals, but to the circumstance that, originated by a single individual, they all tended to a

single end. In the same way I thought that the sciences contained in books (such of them at least as are made up of probable reasonings, without demonstrations), composed as they are of the opinions of many different individuals massed together, are farther removed from truth than the simple inferences which a man of good sense using his natural and unprejudiced judgment draws respecting the matters of his experience. And because we have all to pass through a state of infancy to manhood, and have been of necessity, for a length of time, governed by our desires and preceptors (whose dictates were frequently conflicting, while neither perhaps always counseled us for the best), I farther concluded that it is almost impossible that our judgments can be so correct or solid as they would have been, had our reason been mature from the moment of our birth, and had we always been guided by it alone.

It is true, however, that it is not customary to pull down all the houses of a town with the single design of rebuilding them differently, and thereby rendering the streets more handsome; but it often happens that a private individual takes down his own with the view of erecting it anew, and that people are even sometimes constrained to this when their houses are in danger of falling from age, or when the foundations are insecure. With this before me by way of example, I was persuaded that it would indeed be preposterous for a private individual to think of reforming a state by fundamentally changing it throughout, and overturning it in order to set it up amended; and the same I thought was true of any similar project for reforming the body of the sciences, or the order of teaching them established in the schools: but as for the opinions which up to that time I had embraced, I thought that I could not do better than resolve at once to sweep them wholly away, that I might afterwards be in a position to admit either others more correct, or even perhaps the same when they had undergone the scrutiny of reason. I firmly believed that in this way I should much better succeed in the conduct of my life, than if I built only upon old foundations, and leaned upon principles which, in my youth, I had taken upon trust. For although I recognized various difficulties in this undertaking, these were not, however, without remedy, nor once to be compared with such as attend the slightest reformation in public affairs. Large bodies, if once overthrown, are with great difficulty set up again, or even kept erect when once seriously shaken, and the fall of such is always disastrous. Then if there are any imperfections in the constitutions of states (and that many such exist the diversity of constitutions is alone sufficient to assure us), custom has without doubt materially smoothed their inconveniences, and has even

managed to steer altogether clear of, or insensibly corrected a number which sagacity could not have provided against with equal effect; and, in fine, the defects are almost always more tolerable than the change necessary for their removal; in the same manner that highways which wind among mountains, by being much frequented, become gradually so smooth and commodious, that it is much better to follow them than to seek a straighter path by climbing over the tops of rocks and descending to the bottoms of precipices.

Hence it is that I cannot in any degree approve of those restless and busy meddlers who, called neither by birth nor fortune to take part in the management of public affairs, are yet always projecting reforms; and if I thought that this tract contained aught which might justify the suspicion that I was a victim of such folly, I would by no means permit its publication. I have never contemplated anything higher than the reformation of my own opinions, and basing them on a foundation wholly my own. And although my own satisfaction with my work has led me to present here a draft of it, I do not by any means therefore recommend to every one else to make a similar attempt. Those whom God has endowed with a larger measure of genius will entertain, perhaps, designs still more exalted; but for the many I am much afraid lest even the present undertaking be more than they can safely venture to imitate. The single design to strip one's self of all past beliefs is one that ought not to be taken by every one. The majority of men is composed of two classes, for neither of which would this be at all a befitting resolution: in the first place, of those who with more than a due confidence in their own powers, are precipitate in their judgments and want the patience requisite for orderly and circumspect thinking; whence it happens, that if men of this class once take the liberty to doubt of their accustomed opinions, and quit the beaten highway, they will never be able to thread the byway that would lead them by a shorter course, and will lose themselves and continue to wander for life; in the second place, of those who, possessed of sufficient sense or modesty to determine that there are others who excel them in the power of discriminating between truth and error, and by whom they may be instructed, ought rather to content themselves with the opinions of such than trust for more correct to their own reason.

For my own part, I should doubtless have belonged to the latter class, had I received instruction from but one master, or had I never known the diversities of opinion that from time immemorial have prevailed among men of the greatest learning. But I had become aware, even so early as during my college life, that no opinion, however absurd and incredible,

can be imagined, which has not been maintained by some one of the philosophers; and afterwards in the course of my travels I remarked that all those whose opinions are decidedly repugnant to ours are not in that account barbarians and savages, but on the contrary that many of these nations make an equally good, if not better, use of their reason than we do. I took into account also the very different character which a person brought up from infancy in France or Germany exhibits, from that which, with the same mind originally, this individual would have possessed had he lived always among the Chinese or with savages, and the circumstance that in dress itself the fashion which pleased us ten years ago, and which may again, perhaps, be received into favor before ten years have gone, appears to us at this moment extravagant and ridiculous. I was thus led to infer that the ground of our opinions is far more custom and example than any certain knowledge. And, finally, although such be the ground of our opinions, I remarked that a plurality of suffrages is no guarantee of truth where it is at all of difficult discovery, as in such cases it is much more likely that it will be found by one than by many. I could, however, select from the crowd no one whose opinions seemed worthy of preference, and thus I found myself constrained, as it were, to use my own reason in the conduct of my life.

But like one walking alone and in the dark, I resolved to proceed so slowly and with such circumspection, that if I did not advance far, I would at least guard against falling. I did not even choose to dismiss summarily any of the opinions that had crept into my belief without having been introduced by reason, but first of all took sufficient time carefully to satisfy myself of the general nature of the task I was setting myself, and ascertain the true method by which to arrive at the knowledge of whatever lay within the compass of my powers.

Among the branches of philosophy, I had, at an earlier period, given some attention to logic, and among those of the mathematics to geometrical analysis and algebra, — three arts or sciences which ought, as I conceived, to contribute something to my design. But, on examination, I found that, as for logic, its syllogisms and the majority of its other precepts are of avail rather in the communication of what we already know, or even as the art of Lully, in speaking without judgment of things of which we are ignorant, than in the investigation of the unknown; and although this science contains indeed a number of correct and very excellent precepts, there are, nevertheless, so many others, and these either injurious or superfluous, mingled with the former, that it is almost quite as difficult to effect a severance of the true from the false as it is to extract a

Diana or a Minerva from a rough block of marble. Then as to the analysis of the ancients and the algebra of the moderns, besides that they embrace only matters highly abstract, and, to appearance, of no use, the former is so exclusively restricted to the consideration of figures, that it can exercise the understanding only on condition of greatly fatiguing the imagination; and, in the latter, there is so complete a subjection to certain rules and formulas, that there results an art full of confusion and obscurity calculated to embarrass, instead of a science fitted to cultivate the mind. By these considerations I was induced to seek some other method which would comprise the advantages of the three and be exempt from their defects. And as a multitude of laws often only hampers justice, so that a state is best governed when, with few laws, these are rigidly administered; in like manner, instead of the great number of precepts of which logic is composed, I believed that the four following would prove perfectly sufficient for me, provided I took the firm and unwavering resolution never in a single instance to fail in observing them.

The first was never to accept anything for true which I did not clearly know to be such; that is to say, carefully to avoid precipitancy and prejudice, and to comprise nothing more in my judgement than what was presented to my mind so clearly and distinctly as to exclude all ground of doubt.

The second, to divide each of the difficulties under examination into as many parts as possible, and as might be necessary for its adequate solution.

The third, to conduct my thoughts in such order that, by commencing with objects the simplest and easiest to know, I might ascend by little and little, and, as it were, step by step, to the knowledge of the more complex; assigning in thought a certain order even to those objects which in their own nature do not stand in a relation of antecedence and sequence.

And the last, in every case to make enumerations so complete, and reviews so general, that I might be assured that nothing was omitted.

The long chains of simple and easy reasonings by means of which geometers are accustomed to reach the conclusions of their most difficult demonstrations, had led me to imagine that all things, to the knowledge of which man is competent, are mutually connected in the same way, and that there is nothing so far removed from us as to be beyond our reach, or so hidden that we cannot discover it, provided only we abstain from accepting the false for the true, and always preserve in our thoughts the order necessary for the deduction of one truth from

another. And I had little difficulty in determining the objects with which it was necessary to commence, for I was already persuaded that it must be with the simplest and easiest to know, and, considering that of all those who have hitherto sought truth in the sciences, the mathematicians alone have been able to find any demonstrations, that is, any certain and evident reasons, I did not doubt but that such must have been the rule of their investigations. I resolved to commence, therefore, with the examination of the simplest objects, not anticipating, however, from this any other advantage than that to be found in accustoming my mind to the love and nourishment of truth, and to a distaste for all such reasonings as were unsound. But I had no intention on that account of attempting to master all the particular sciences commonly denominated mathematics: but observing that, however different their objects, they all agree in considering only the various relations or proportions subsisting among those objects, I thought it best for my purpose to consider these proportions in the most general form possible, without referring them to any objects in particular, except such as would most facilitate the knowledge of them, and without by any means restricting them to these, that afterwards I might thus be the better able to apply them to every other class of objects to which they are legitimately applicable. Perceiving further, that in order to understand these relations I should sometimes have to consider them one by one and sometimes only to bear them in mind, or embrace them in the aggregate, I thought that, in order the better to consider them individually, I should view them as subsisting between straight lines, than which I could find no objects more simple, or capable of being more distinctly represented to my imagination and senses; and on the other hand, that in order to retain them in the memory or embrace an aggregate of many, I should express them by certain characters the briefest possible. In this way I believed that I could borrow all that was best both in geometrical analysis and in algebra, and correct all the defects of the one by help of the other.

And, in point of fact, the accurate observance of these few precepts gave me, I take the liberty of saying, such ease in unraveling all the questions embraced in these two sciences, that in the two or three months I devoted to their examination, not only did I reach solutions of questions I had formerly deemed exceedingly difficult but even as regards questions of the solution of which I continued ignorant, I was enabled, as it appeared to me, to determine the means whereby, and the extent to which a solution was possible; results attributable to the circumstance that I commenced with the simplest and most general truths, and that



## **Dialogues Concerning Two New Sciences**

Galilei, Galileo  
1638

## THIRD DAY

### CHANGE OF POSITION. [*De Motu Locali*][190]

This discussion is divided into three parts; the first part deals with motion which is steady or uniform; the second treats of motion as we find it accelerated in nature; the third deals with the so-called violent motions and with projectiles.

## UNIFORM MOTION[191]

In dealing with steady or uniform motion, we need a single definition which I give as follows:

### Definition

By steady or uniform motion, I mean one in which the distances traversed by the moving particle during any equal intervals of time, are themselves equal.

### Caution

We must add to the old definition (which defined steady motion simply as one in which equal distances are traversed in equal times) the word "any," meaning by this, all equal intervals of time; for it may happen that the moving body will traverse equal distances during some equal intervals of time and yet the distances traversed during some small portion of these time-intervals may not be equal, even though the time-intervals be equal.

From the above definition, four axioms follow, namely:

### Axiom I

In the case of one and the same uniform motion, the distance traversed during a longer interval of time is greater than the distance traversed during a shorter interval of time.

### Axiom II

In the case of one and the same uniform motion, the time required to traverse a greater distance is longer than the time required for a less distance.

### Axiom III

In one and the same interval of time, the distance traversed at a greater speed is larger than the distance traversed at a less speed.

### Axiom IV[192]

The speed required to traverse a longer distance is greater than that required to traverse a shorter distance during the same time-interval.

## Theorem I, Proposition I

If a moving particle, carried uniformly at a constant speed, traverses two distances the time-intervals required are to each other in the ratio of these distances.

Let a particle move uniformly with constant speed through two distances AB, BC, and let the time required to traverse AB be represented by DE; the time required to traverse BC, by EF; then I say that the distance AB is to the distance BC as the time DE is to the time EF.

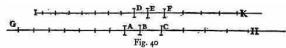


Fig. 40

Let the distances and times be extended on both sides towards G, H and I, K; let AG be divided into any number whatever of spaces each equal to AB, and in like manner lay off in DI exactly the same number of time-intervals each equal to DE. Again lay off in CH any number whatever of distances each equal to BC; and in FK exactly the same number of time-intervals each equal to EF; then will the distance BG and the time EI be equal and arbitrary multiples of the distance BA and the time ED; and likewise the distance HB and the time KE are equal and arbitrary multiples of the distance CB and the time FE.

And since DE is the time required to traverse AB, the whole time EI will be required for the whole distance BG, and when the motion is uniform there will be in EI as many time-intervals each equal to DE as there are distances in BG each equal to BA; and likewise it follows that KE represents the time required to traverse HB.

Since, however, the motion is uniform, it follows that if the distance GB is equal to the distance BH, then must also the time IE be equal to the time EK; and if GB is greater than BH, then also IE will be greater than EK; and if less, less.\* There[193] are then four quantities, the first AB, the second BC, the third DE, and the fourth EF; the time IE and the distance GB are arbitrary multiples of the first and the third, namely of the distance AB and the time DE.

But it has been proved that *both* of these latter quantities are either equal to, greater than, or less than the time EK and the space BH, which are arbitrary multiples of the second and the fourth. Therefore the first is to the second, namely the distance AB is to the distance BC, as the third is to the fourth, namely the time DE is to the time EF.

q. e. d.

## Theorem II, Proposition II,

If a moving particle traverses two distances in equal intervals of time, these distances will bear to each other the same ratio as the speeds. And conversely if the distances are as the speeds then the times are equal.

Referring to Fig. 40, let AB and BC represent the two distances traversed in equal time-intervals, the distance AB for instance with the velocity DE, and the distance BC with the velocity EF. Then, I say, the distance AB is to the distance BC as the velocity DE is to the velocity EF. For if equal multiples of both distances and speeds be taken, as above, namely, GB and IE of AB and DE respectively, and in like manner HB and KE of BC and EF, then one may infer, in the same manner as above, that the multiples GB and IE are either less than, equal to, or greater than equal multiples of BH and EK. Hence the theorem is established.

### Theorem III, Proposition III

In the case of unequal speeds, the time-intervals required to traverse a given space are to each other inversely as the speeds.

Let the larger of the two unequal speeds be indicated by A; the smaller, by B; and let the motion corresponding to both traverse the given space CD. Then I say the time required to traverse the distance CD at speed A is to the time required to traverse the same distance at speed B, as the speed B is to the speed A. For let CD be to CE as A is to B; then, from the preceding, it follows that the time required to complete the distance CD at speed A is the same as [194] the time necessary to complete CE at speed B; but the time needed to traverse the distance CE at speed B is to the time required to traverse the distance CD at the same speed as CE is to CD; therefore the time in which CD is covered at speed A is to the time in which CD is covered at speed B as CE is to CD, that is, as speed B is to speed A.

q. e. d.

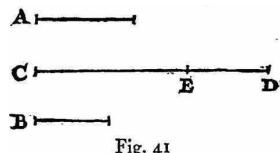


Fig. 41

Fig. 41

### Theorem IV, Proposition IV

If two particles are carried with uniform motion, but each with a different speed, the distances covered by them during unequal intervals of time bear to each other the compound ratio of the speeds and time intervals.

Let the two particles which are carried with uniform motion be E and F and let the ratio of the speed of the body E be to that of the body F as A is to B; but let the ratio of the time consumed by the motion of E be to the time consumed by the motion of F as C is to D. Then, I say, that the distance covered by E, with speed A in time C, bears to the space traversed by F with speed B in time D a ratio which is the product of the ratio of the speed A to the speed B by the ratio of the time C to the time D. For if G is the distance traversed by E at speed A during the time-interval C, and if G is to I as

the speed A is to the speed B; and if also the time-interval C is to the time-interval D as I is to L, then it follows that I is the distance traversed by F in the same time that G is traversed by E since G is to I in the same ratio as the speed A to the speed B. And since I is to L in the same ratio as the time-intervals C and D, if I is the distance traversed by F during the interval C, then L will be the distance traversed by F during the interval D at the speed B.

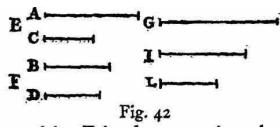


Fig. 42

But the ratio of G to L is the product of the ratios G to I and I to L, that is, of the ratios of the speed A to the speed B and of the time-interval C to the time-interval D.

q. e. d.

### Theorem V, Proposition V[195]

If two particles are moved at a uniform rate, but with unequal speeds, through unequal distances, then the ratio of the time-intervals occupied will be the product of the ratio of the distances by the inverse ratio of the speeds.

Let the two moving particles be denoted by A and B, and let the speed of A be to the speed of B in the ratio of V to T; in like manner let the distances traversed be in the ratio of S to R; then I say that the ratio of the time-interval during which the motion of A occurs to the time-interval occupied by the motion of B is the product of the ratio of the speed T to the speed V by the ratio of the distance S to the distance R.

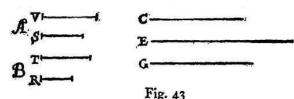


Fig. 43

Let C be the time-interval occupied by the motion of A, and let the time-interval C bear to a time-interval E the same ratio as the speed T to the speed V.

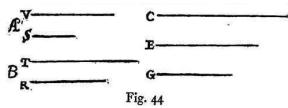
And since C is the time-interval during which A, with speed V, traverses the distance S and since T, the speed of B, is to the speed V, as the time-interval C is to the time-interval E, then E will be the time required by the particle B to traverse the distance S. If now we let the time-interval E be to the time-interval G as the distance S is to the distance R, then it follows that G is the time required by B to traverse the space R.

Since the ratio of C to G is the product of the ratios C to E and E to G (while also the ratio of C to E is the inverse ratio of the speeds of A and B respectively, i. e., the ratio of T to V); and since the ratio of E to G is the same as that of the distances S and R respectively, the proposition is proved.

## Theorem VI, Proposition VI[196]

If two particles are carried at a uniform rate, the ratio of their speeds will be the product of the ratio of the distances traversed by the inverse ratio of the time-intervals occupied.

Let A and B be the two particles which move at a uniform rate; and let the respective distances traversed by them have the ratio of V to T, but let the time-intervals be as S to R. Then I say the speed of A will bear to the speed of B a ratio which is the product of the ratio of the distance V to the distance T and the time-interval R to the time-interval S.



Let C be the speed at which A traverses the distance V during the time-interval S; and let the speed C bear the same ratio to another speed E as V bears to T; then E will be the speed at which B traverses the distance T during the time-interval S. If now the speed E is to another speed G as the time-interval R is to the time-interval S, then G will be the speed at which the particle B traverses the distance T during the time-interval R. Thus we have the speed C at which the particle A covers the distance V during the time S and also the speed G at which the particle B traverses the distance T during the time R. The ratio of C to G is the product of the ratio C to E and E to G; the ratio of C to E is by definition the same as the ratio of the distance V to distance T; and the ratio of E to G is the same as the ratio of R to S. Hence follows the proposition.

SALV.

The preceding is what our Author has written concerning uniform motion. We pass now to a new and more discriminating consideration of naturally accelerated motion, such as that generally experienced by heavy falling bodies; following is the title and introduction.

## NATURALLY ACCELERATED MOTION[197]

The properties belonging to uniform motion have been discussed in the preceding section; but accelerated motion remains to be considered.

And first of all it seems desirable to find and explain a definition best fitting natural phenomena. For anyone may invent an arbitrary type of motion and discuss its properties; thus, for instance, some have imagined helices and conchoids as described by certain motions which are not met with in nature, and have very commendably established the properties which these curves possess in virtue of their definitions; but we have decided to consider the phenomena of bodies falling with an acceleration such as actually occurs in nature and to make this definition of accelerated motion exhibit the essential features of observed accelerated motions. And this, at last, after repeated efforts we trust we have succeeded in doing. In this belief we are confirmed mainly by the consideration that experimental results are seen to agree with and exactly correspond with those properties which have been, one after another, demonstrated by us. Finally, in the investigation of naturally accelerated motion we were led, by hand as it were, in following the habit and custom of nature herself, in all her various other processes, to employ only those means which are most common, simple and easy.

For I think no one believes that swimming or flying can be accomplished in a manner simpler or easier than that instinctively employed by fishes and birds.

When, therefore, I observe a stone initially at rest falling from an elevated position and continually acquiring new increments of speed, why should I not believe that such increases take place in a manner which is exceedingly simple and rather obvious to everybody? If now we examine the matter carefully we find no addition or increment more simple than that which repeats itself always in the same manner. This we readily understand when we consider the intimate relationship between time and motion; for just as uniformity of motion is defined by and conceived through equal times and equal spaces (thus we call a motion uniform when equal distances are traversed during equal time-intervals), so also we may, in a similar manner, through equal time-intervals, conceive additions of speed as taking place without complication; thus we may picture to our[198] mind a motion as uniformly and continuously accelerated when, during any equal intervals of time whatever, equal increments of speed are given to it. Thus if any equal intervals of time whatever have elapsed, counting from the time at which the moving body left its position of rest and began to descend, the amount of speed acquired during the first two time-intervals will be double that acquired during the first time-interval alone; so the amount added during three of these time-intervals will be treble; and that in four, quadruple that of the first time-interval. To put the matter more clearly, if a body were to continue its motion with the same speed which it had acquired during the first time-interval and were to retain this same uniform speed, then its motion would be twice as slow as that which it would have if its velocity had been acquired during *two* time-intervals.

And thus, it seems, we shall not be far wrong if we put the increment of speed as proportional to the increment of time; hence the definition of motion which we are about to discuss may be stated as follows: A motion is said to be uniformly accelerated, when starting from rest, it acquires, during equal time-intervals, equal increments of speed.

SAGR.

Although I can offer no rational objection to this or indeed to any other definition, devised by any author whomsoever, since all definitions are arbitrary, I may nevertheless without offense be allowed to doubt whether such a definition as the above, established in an abstract manner, corresponds to and describes that kind of accelerated motion which we meet in nature in the case of freely falling bodies. And since the Author apparently maintains that the motion described in his definition is that of freely falling bodies, I would like to clear my mind of certain difficulties in order that I may later apply myself more earnestly to the propositions and their demonstrations.

SALV.

It is well that you and Simplicio raise these difficulties. They are, I imagine, the same which occurred to me when I first saw this treatise, and which were removed either by discussion with the Author himself, or by turning the matter over in my own mind.

SAGR.

When I think of a heavy body falling from rest, that is, starting with zero speed and gaining speed in proportion to the[199] time from the beginning of the motion; such a motion as would, for instance, in eight beats of the pulse acquire eight degrees of speed; having at the end of the fourth beat acquired four degrees; at the end of the second, two; at the end of the first, one: and since time is divisible without limit, it follows from all these considerations that if the earlier speed of a body is less than its present speed in a constant ratio, then there is no degree of speed however small (or, one may say, no degree of slowness however great) with which we may not find this body travelling after starting from infinite slowness, i. e., from rest. So that if that speed which it had at the end of the fourth beat was such that, if kept uniform, the body would traverse two miles in an hour, and if keeping the speed which it had at the end of the second beat, it would traverse one mile an hour, we must infer that, as the instant of starting is more and more nearly approached, the body moves so slowly that, if it kept on moving at this rate, it would not traverse a mile in an hour, or in a day, or in a year or in a thousand years; indeed, it would not traverse a span in an even greater time; a phenomenon which baffles the imagination, while our senses show us that a heavy falling body suddenly acquires great speed.

SALV.

This is one of the difficulties which I also at the beginning, experienced, but which I shortly afterwards removed; and the removal was effected by the very experiment

which creates the difficulty for you. You say the experiment appears to show that immediately after a heavy body starts from rest it acquires a very considerable speed: and I say that the same experiment makes clear the fact that the initial motions of a falling body, no matter how heavy, are very slow and gentle. Place a heavy body upon a yielding material, and leave it there without any pressure except that owing to its own weight; it is clear that if one lifts this body a cubit or two and allows it to fall upon the same material, it will, with this impulse, exert a new and greater pressure than that caused by its mere weight; and this effect is brought about by the [weight of the] falling body together with the velocity acquired during the fall, an effect which will be greater and greater according to the height of the fall, that is according as the velocity of the falling body becomes greater. From the quality and intensity of the blow we are thus enabled to accurately estimate the speed of a falling body. But tell me, gentlemen, is it not true that if a block be allowed to fall upon a stake from a height of four cubits and drives it into the earth,[200] say, four finger-breadths, that coming from a height of two cubits it will drive the stake a much less distance, and from the height of one cubit a still less distance; and finally if the block be lifted only one finger-breadth how much more will it accomplish than if merely laid on top of the stake without percussion? Certainly very little. If it be lifted only the thickness of a leaf, the effect will be altogether imperceptible. And since the effect of the blow depends upon the velocity of this striking body, can any one doubt the motion is very slow and the speed more than small whenever the effect [of the blow] is imperceptible? See now the power of truth; the same experiment which at first glance seemed to show one thing, when more carefully examined, assures us of the contrary.

But without depending upon the above experiment, which is doubtless very conclusive, it seems to me that it ought not to be difficult to establish such a fact by reasoning alone. Imagine a heavy stone held in the air at rest; the support is removed and the stone set free; then since it is heavier than the air it begins to fall, and not with uniform motion but slowly at the beginning and with a continuously accelerated motion. Now since velocity can be increased and diminished without limit, what reason is there to believe that such a moving body starting with infinite slowness, that is, from rest, immediately acquires a speed of ten degrees rather than one of four, or of two, or of one, or of a half, or of a hundredth; or, indeed, of any of the infinite number of small values [of speed]? Pray listen. I hardly think you will refuse to grant that the gain of speed of the stone falling from rest follows the same sequence as the diminution and loss of this same speed when, by some impelling force, the stone is thrown to its former elevation: but even if you do not grant this, I do not see how you can doubt that the ascending stone, diminishing in speed, must before coming to rest pass through every possible degree of slowness.

SIMP.

But if the number of degrees of greater and greater slowness is limitless, they will never be all exhausted, therefore such an ascending heavy body will never reach rest, but will continue to move without limit always at a slower rate; but this is not the observed fact.

SALV.

This would happen, Simplicio, if the moving body were to maintain its speed for any length of time at each degree of velocity; but it merely passes each point without delaying more than an instant: and since each time-interval however[201] small may be divided into an infinite number of instants, these will always be sufficient [in number] to correspond to the infinite degrees of diminished velocity.

That such a heavy rising body does not remain for any length of time at any given degree of velocity is evident from the following: because if, some time-interval having been assigned, the body moves with the same speed in the last as in the first instant of that time-interval, it could from this second degree of elevation be in like manner raised through an equal height, just as it was transferred from the first elevation to the second, and by the same reasoning would pass from the second to the third and would finally continue in uniform motion forever.

SAGR.

From these considerations it appears to me that we may obtain a proper solution of the problem discussed by philosophers, namely, what causes the acceleration in the natural motion of heavy bodies? Since, as it seems to me, the force [*virtù*] impressed by the agent projecting the body upwards diminishes continuously, this force, so long as it was greater than the contrary force of gravitation, impelled the body upwards; when the two are in equilibrium the body ceases to rise and passes through the state of rest in which the impressed impetus [*impeto*] is not destroyed, but only its excess over the weight of the body has been consumed—the excess which caused the body to rise. Then as the diminution of the outside impetus [*impeto*] continues, and gravitation gains the upper hand, the fall begins, but slowly at first on account of the opposing impetus [*virtù impressa*], a large portion of which still remains in the body; but as this continues to diminish it also continues to be more and more overcome by gravity, hence the continuous acceleration of motion.

SIMP.

The idea is clever, yet more subtle than sound; for even if the argument were conclusive, it would explain only the case in which a natural motion is preceded by a violent motion, in which there still remains active a portion of the external force [*virtù esterna*]; but where there is no such remaining portion and the body starts from an antecedent state of rest, the cogency of the whole argument fails.

SAGR.

I believe that you are mistaken and that this distinction between cases which you make is superfluous or rather nonexistent. But, tell me, cannot a projectile receive from the projector either a large or a small force [*virtù*] such as will throw it to a height of a hundred cubits, and even twenty or four or one?

Undoubtedly, yes.

SAGR.

So therefore this impressed force [*virtù impressa*] may exceed the resistance of gravity so slightly as to raise it only a finger-breadth; and finally the force [*virtù*] of the projector may be just large enough to exactly balance the resistance of gravity so that the body is not lifted at all but merely sustained. When one holds a stone in his hand does he do anything but give it a force impelling [*virtù impellente*] it upwards equal to the power [*facoltà*] of gravity drawing it downwards? And do you not continuously impress this force [*virtù*] upon the stone as long as you hold it in the hand? Does it perhaps diminish with the time during which one holds the stone?

And what does it matter whether this support which prevents the stone from falling is furnished by one's hand or by a table or by a rope from which it hangs? Certainly nothing at all. You must conclude, therefore, Simplicio, that it makes no difference whatever whether the fall of the stone is preceded by a period of rest which is long, short, or instantaneous provided only the fall does not take place so long as the stone is acted upon by a force [*virtù*] opposed to its weight and sufficient to hold it at rest.

SALV.

The present does not seem to be the proper time to investigate the cause of the acceleration of natural motion concerning which various opinions have been expressed by various philosophers, some explaining it by attraction to the center, others to repulsion between the very small parts of the body, while still others attribute it to a certain stress in the surrounding medium which closes in behind the falling body and drives it from one of its positions to another. Now, all these fantasies, and others too, ought to be examined; but it is not really worth while. At present it is the purpose of our Author merely to investigate and to demonstrate some of the properties of accelerated motion (whatever the cause of this acceleration may be)—meaning thereby a motion, such that the momentum of its velocity [*i momenti della sua velocità*] goes on increasing after departure from rest, in simple proportionality to the time, which is the same as saying that in equal time-intervals the body receives equal increments of velocity; and if we find the properties [of accelerated motion] which will be demonstrated later are realized in freely falling and accelerated bodies, we may conclude that the assumed definition includes such a motion of falling bodies and that their speed [*accelerazione*] goes on increasing as the time and the duration of the motion.

So far as I see at present, the definition might have been put a little more clearly perhaps without changing the fundamental idea, namely, uniformly accelerated motion is such that its speed increases in proportion to the space traversed; so that, for example, the speed acquired by a body in falling four cubits would be double that acquired in falling two cubits and this latter speed would be double that acquired in the first cubit. Because there is no doubt but that a heavy body falling from the height of six cubits has, and strikes with, a momentum [*impeto*] double that it had at the end of three cubits, triple that which it had at the end of one.

SALV.

It is very comforting to me to have had such a companion in error; and moreover let me tell you that your proposition seems so highly probable that our Author himself admitted, when I advanced this opinion to him, that he had for some time shared the same fallacy. But what most surprised me was to see two propositions so inherently probable that they commanded the assent of everyone to whom they were presented, proven in a few simple words to be not only false, but impossible.

SIMP.

I am one of those who accept the proposition, and believe that a falling body acquires force [*vires*] in its descent, its velocity increasing in proportion to the space, and that the momentum [*momento*] of the falling body is doubled when it falls from a doubled height; these propositions, it appears to me, ought to be conceded without hesitation or controversy.

SALV.

And yet they are as false and impossible as that motion should be completed instantaneously; and here is a very clear demonstration of it. If the velocities are in proportion to the spaces traversed, or to be traversed, then these spaces are traversed in equal intervals of time; if, therefore, the velocity with which the falling body traverses a space of eight feet were double that with which it covered the first four feet (just as the one distance is double the other) then the time-intervals required for these passages would be equal. But for one and the same body to fall eight feet and four feet in the same time is possible only in the case of instantaneous [discontinuous] motion;[204] but observation shows us that the motion of a falling body occupies time, and less of it in covering a distance of four feet than of eight feet; therefore it is not true that its velocity increases in proportion to the space.

The falsity of the other proposition may be shown with equal clearness. For if we consider a single striking body the difference of momentum in its blows can depend only upon difference of velocity; for if the striking body falling from a double height were to deliver a blow of double momentum, it would be necessary for this body to strike with a doubled velocity; but with this doubled speed it would traverse a doubled space in the same time-interval; observation however shows that the time required for fall from the greater height is longer.

SAGR.

You present these recondite matters with too much evidence and ease; this great facility makes them less appreciated than they would be had they been presented in a more abstruse manner. For, in my opinion, people esteem more lightly that knowledge which they acquire with so little labor than that acquired through long and obscure discussion.

SALV.

If those who demonstrate with brevity and clearness the fallacy of many popular beliefs were treated with contempt instead of gratitude the injury would be quite bearable; but on the other hand it is very unpleasant and annoying to see men, who claim to be peers of anyone in a certain field of study, take for granted certain conclusions which later are quickly and easily shown by another to be false. I do not describe such a feeling as one of envy, which usually degenerates into hatred and anger against those who discover such fallacies; I would call it a strong desire to maintain old errors, rather than accept newly discovered truths. This desire at times induces them to unite against these truths, although at heart believing in them, merely for the purpose of lowering the esteem in which certain others are held by the unthinking crowd. Indeed, I have heard from our Academician many such fallacies held as true but easily refutable; some of these I have in mind.

SAGR.

You must not withhold them from us, but, at the proper time, tell us about them even though an extra session be necessary. But now, continuing the thread of our talk, it would[205] seem that up to the present we have established the definition of uniformly accelerated motion which is expressed as follows:

A motion is said to be equally or uniformly accelerated when, starting from rest, its momentum (*celeritatis momenta*) receives equal increments in equal times.

SALV.

This definition established, the Author makes a single assumption, namely,

The speeds acquired by one and the same body moving down planes of different inclinations are equal when the heights of these planes are equal.

By the height of an inclined plane we mean the perpendicular let fall from the upper end of the plane upon the horizontal line drawn through the lower end of the same plane. Thus, to illustrate, let the line AB be horizontal, and let the planes CA and CD be inclined to it; then the Author calls the perpendicular CB the “height” of the planes CA and CD; he supposes that the speeds acquired by one and the same body, descending along the planes CA and CD to the terminal points A and D are equal since the heights of these planes are the same, CB; and also it must be understood that this speed is that which would be acquired by the same body falling from C to B.

SAGR.

Your assumption appears to me so reasonable that it ought to be conceded without question, provided of course there are no chance or outside resistances, and that the planes are hard and smooth, and that the figure of the moving body is perfectly round, so that neither plane nor moving body is rough. All resistance and opposition having been removed, my reason tells me at once that a heavy and perfectly round ball

descending along the lines CA, CD, CB would reach the terminal points A, D, B, with equal momenta [*impeti eguali*].

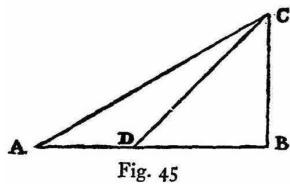


Fig. 45

SALV.

Your words are very plausible; but I hope by experiment to increase the probability to an extent which shall be little short of a rigid demonstration.

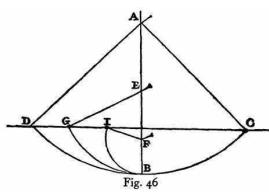


Fig. 46

This experiment leaves no room for doubt as to the truth of our supposition; for since the two arcs CB and DB are equal and similarly placed, the momentum [*momento*] acquired by the fall through the arc CB is the same as that gained by fall through the arc DB; but the momentum [*momento*] acquired at B, owing to fall through CB, is able to lift the same body [*mobile*] through the arc BD; therefore, the momentum acquired in the fall BD is equal to that which lifts the same body through the same arc from B to D; so, in general, every momentum acquired by fall through an arc is equal to that which can lift the same body through the same arc. But all these momenta [*momenti*] which cause a rise through the arcs BD, BG, and BI are equal, since they are produced by the same momentum, gained by fall through CB, as experiment shows. Therefore all the momenta gained by fall through the arcs DB, GB, IB are equal.

SAGR.

The argument seems to me so conclusive and the experiment so well adapted to establish the hypothesis that we may, indeed, consider it as demonstrated.

SALV.

I do not wish, Sagredo, that we trouble ourselves too much about this matter, since we are going to apply this principle mainly in motions which occur on plane surfaces, and

not upon curved, along which acceleration varies in a manner greatly different from that which we have assumed for planes.

So that, although the above experiment shows us that the descent of the moving body through the arc CB confers upon it momentum [*momento*] just sufficient to carry it to the same height through any of the arcs BD, BG, BI, we are not able, by similar means, to show that the event would be identical in the case of a perfectly round ball descending along planes whose inclinations are respectively the same as the chords of these arcs. It seems likely, on the other hand, that, since these planes form angles at the point B, they will present an obstacle to the ball which has descended along the chord CB, and starts to rise along the chord BD, BG, BI.

In striking these planes some of its momentum [*impeto*] will be lost and it will not be able to rise to the height of the line CD; but this obstacle, which interferes with the experiment, once removed, it is clear that the momentum [*impeto*] (which gains [208] in strength with descent) will be able to carry the body to the same height. Let us then, for the present, take this as a postulate, the absolute truth of which will be established when we find that the inferences from it correspond to and agree perfectly with experiment. The author having assumed this single principle passes next to the propositions which he clearly demonstrates; the first of these is as follows:

### Theorem I, Proposition I

The time in which any space is traversed by a body starting from rest and uniformly accelerated is equal to the time in which that same space would be traversed by the same body moving at a uniform speed whose value is the mean of the highest speed and the speed just before acceleration began.

Let us represent by the line AB the time in which the space CD is traversed by a body which starts from rest at C and is uniformly accelerated; let the final and highest value of the speed gained during the interval AB be represented by the line EB drawn at right angles to AB; draw the line AE, then all lines drawn from equidistant points on AB and parallel to BE will represent the increasing values of the speed, beginning with the instant A. Let the point F bisect the line EB; draw FG parallel to BA, and GA parallel to FB, thus forming a parallelogram AGFB which will be equal in area to the triangle AEB, since the side GF bisects the side AE at the point I; for if the parallel lines in the triangle AEB are extended to GI, then the sum of all the parallels contained in the quadrilateral is equal to the sum of those contained in the triangle AEB; for those in the triangle IEF are equal to those contained in the triangle GIA, while those included in the trapezium AIFB are common. Since each and every instant of time in the time-interval AB has its corresponding point on the line AB, from which points parallels drawn in and limited by the triangle AEB represent the increasing values of the growing velocity, and since parallels contained within the rectangle represent the values of a speed which is not increasing, but constant, it appears, in like manner, that the momenta [*momenta*] assumed by the moving body may also be represented, in the case of the accelerated motion, by the increasing parallels of the triangle [209] AEB, and, in the case of the uniform motion, by the parallels of the rectangle GB. For, what the momenta may lack in the first part of the

accelerated motion (the deficiency of the momenta being represented by the parallels of the triangle AGI) is made up by the momenta represented by the parallels of the triangle IEF.

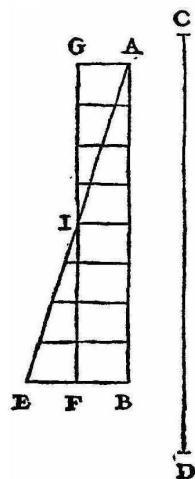


Fig. 47

Fig. 47

Hence it is clear that equal spaces will be traversed in equal times by two bodies, one of which, starting from rest, moves with a uniform acceleration, while the momentum of the other, moving with uniform speed, is one-half its maximum momentum under accelerated motion.

q. e. d.

### Theorem II, Proposition II

The spaces described by a body falling from rest with a uniformly accelerated motion are to each other as the squares of the time-intervals employed in traversing these distances.

Let the time beginning with any instant A be represented by the straight line AB in which are taken any two time-intervals AD and AE. Let HI represent the distance through which the body, starting from rest at H, falls with uniform acceleration. If HL represents the space traversed during the time-interval AD, and HM that covered during the interval AE, then the space MH stands to the space LH in a ratio which is the square of the ratio of the time AE to the time AD; or we may say simply that the distances HM and HL are related as the squares of AE and AD.

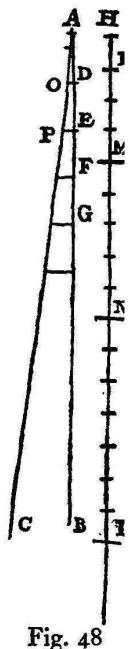


Fig. 48

Fig. 48

Draw the line AC making any angle whatever with the line AB; and from the points D and E, draw the parallel lines DO and EP; of these two lines, DO represents the greatest velocity attained during the interval AD, while EP represents the maximum velocity acquired during the interval AE. But it has just been proved that so far as distances traversed are concerned it is precisely the same whether a body falls from rest with a uniform acceleration or whether it falls during an equal time-interval with a constant speed which is one-half the maximum speed attained during the accelerated motion. It follows therefore that the distances HM and HL are the same as would be traversed, during the time-intervals AE and AD, by uniform velocities equal to one-half those represented by DO and EP respectively. If, therefore, one can show that the distances HM and HL are in the same ratio as the squares of the time-intervals AE and AD, our proposition will be proven.

q. e. d.

Evidently then the ratio of the distances is the square of the ratio of the final velocities, that is, of the lines EP and DO, since these are to each other as AE to AD.

## COROLLARY I

Hence it is clear that if we take any equal intervals of time whatever, counting from the beginning of the motion, such as AD, DE, EF, FG, in which the spaces HL, LM, MN, NI are traversed, these spaces will bear to one another the same ratio as the series of odd numbers, 1, 3, 5, 7; for this is the ratio of the differences of the squares of the lines [which represent time], differences which exceed one another by equal amounts, this excess being equal to the smallest line [viz. the one representing a single

time-interval]: or we may say [that this is the ratio] of the differences of the squares of the natural numbers beginning with unity.

While, therefore, during equal intervals of time the velocities increase as the natural numbers, the increments in the distances traversed during these equal time-intervals are to one another as the odd numbers beginning with unity.

SAGR.

Please suspend the discussion for a moment since there just occurs to me an idea which I want to illustrate by means of a diagram in order that it may be clearer both to you and to me.

Let the line AI represent the lapse of time measured from the initial instant A; through A draw the straight line AF making any angle whatever; join the terminal points I and F; divide the time AI in half at C; draw CB parallel to IF. Let us consider CB as the maximum value of the velocity which increases from zero at the beginning, in simple proportionality to the intercepts on the triangle ABC of lines drawn parallel to BC; or what is the same thing, let us suppose the velocity to increase in proportion to the time; then I admit without question, in view of the preceding argument, that the space described by a body falling in the aforesaid manner will be equal to the space traversed by the same body during the same length of time travelling with a uniform speed equal to EC, the half of BC. Further let us imagine that the[211] body has fallen with accelerated motion so that, at the instant C, it has the velocity BC. It is clear that if the body continued to descend with the same speed BC, without acceleration, it would in the next time-interval CI traverse double the distance covered during the interval AC, with the uniform speed EC which is half of BC; but since the falling body acquires equal increments of speed during equal increments of time, it follows that the velocity BC, during the next time-interval CI will be increased by an amount represented by the parallels of the triangle BFG which is equal to the triangle ABC. If, then, one adds to the velocity GI half of the velocity FG, the highest speed acquired by the accelerated motion and determined by the parallels of the triangle BFG, he will have the uniform velocity with which the same space would have been described in the time CI; and since this speed IN is three times as great as EC it follows that the space described during the interval CI is three times as great as that described during the interval AC. Let us imagine the motion extended over another equal time-interval IO, and the triangle extended to APO; it is then evident that if the motion continues during the interval IO, at the constant rate IF acquired by acceleration during the time AI, the space traversed during the interval IO will be four times that traversed during the first interval AC, because the speed IF is four times the speed EC. But if we enlarge our triangle so as to include FPQ which is equal to ABC, still assuming the acceleration to be constant, we shall add to the uniform speed an increment RQ, equal to EC; then the value of the equivalent uniform speed during the time-interval IO will be five times that during the first time-interval AC; therefore the space traversed will be quintuple that during the first interval AC. It is thus evident by simple computation that a moving body starting from rest and acquiring velocity at a rate proportional to the time, will, during equal intervals of time, traverse distances which are related to each other as the odd numbers beginning with unity, 1, 3, 5;\* or considering the total

space traversed, that covered[212] in double time will be quadruple that covered during unit time; in triple time, the space is nine times as great as in unit time. And in general the spaces traversed are in the duplicate ratio of the times, i. e., in the ratio of the squares of the times.

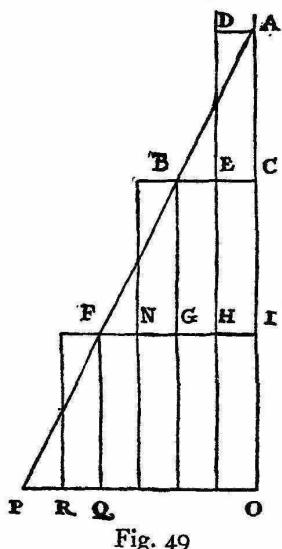


Fig. 49

Fig. 49

SIMP.

In truth, I find more pleasure in this simple and clear argument of Sagredo than in the Author's demonstration which to me appears rather obscure; so that I am convinced that matters are as described, once having accepted the definition of uniformly accelerated motion. But as to whether this acceleration is that which one meets in nature in the case of falling bodies, I am still doubtful; and it seems to me, not only for my own sake but also for all those who think as I do, that this would be the proper moment to introduce one of those experiments—and there are many of them, I understand—which illustrate in several ways the conclusions reached.

SALV.

The request which you, as a man of science, make, is a very reasonable one; for this is the custom—and properly so—in those sciences where mathematical demonstrations are applied to natural phenomena, as is seen in the case of perspective, astronomy, mechanics, music, and others where the principles, once established by well-chosen experiments, become the foundations of the entire superstructure. I hope therefore it will not appear to be a waste of time if we discuss at considerable length this first and most fundamental question upon which hinge numerous consequences of which we have in this book only a small number, placed there by the Author, who has done so much to open a pathway hitherto closed to minds of speculative turn. So far as experiments go they have not been neglected by the Author; and often, in his

company, I have attempted in the following manner to assure myself that the acceleration actually experienced by falling bodies is that above described.

A piece of wooden moulding or scantling, about 12 cubits long, half a cubit wide, and three finger-breadths thick, was taken; on its edge was cut a channel a little more than one finger in breadth; having made this groove very straight, smooth, and polished, and having lined it with parchment, also as smooth and polished as possible, we rolled along it a hard, smooth, and very round bronze ball. Having placed this[213] board in a sloping position, by lifting one end some one or two cubits above the other, we rolled the ball, as I was just saying, along the channel, noting, in a manner presently to be described, the time required to make the descent. We repeated this experiment more than once in order to measure the time with an accuracy such that the deviation between two observations never exceeded one-tenth of a pulse-beat. Having performed this operation and having assured ourselves of its reliability, we now rolled the ball only one-quarter the length of the channel; and having measured the time of its descent, we found it precisely one-half of the former. Next we tried other distances, comparing the time for the whole length with that for the half, or with that for two-thirds, or three-fourths, or indeed for any fraction; in such experiments, repeated a full hundred times, we always found that the spaces traversed were to each other as the squares of the times, and this was true for all inclinations of the plane, i. e., of the channel, along which we rolled the ball. We also observed that the times of descent, for various inclinations of the plane, bore to one another precisely that ratio which, as we shall see later, the Author had predicted and demonstrated for them.

For the measurement of time, we employed a large vessel of water placed in an elevated position; to the bottom of this vessel was soldered a pipe of small diameter giving a thin jet of water, which we collected in a small glass during the time of each descent, whether for the whole length of the channel or for a part of its length; the water thus collected was weighed, after each descent, on a very accurate balance; the differences and ratios of these weights gave us the differences and ratios of the times, and this with such accuracy that although the operation was repeated many, many times, there was no appreciable discrepancy in the results.

SIMP.

I would like to have been present at these experiments; but feeling confidence in the care with which you performed them, and in the fidelity with which you relate them, I am satisfied and accept them as true and valid

SALV.

Then we can proceed without discussion.

Secondly, it follows that, starting from any initial point, if we take any two distances, traversed in any time-intervals whatsoever, these time-intervals bear to one another the same ratio as one of the distances to the mean proportional of the two distances.

For if we take two distances ST and SY measured from the initial point S, the mean proportional of which is SX, the time of fall through ST is to the time of fall through SY as ST is to SX; or one may say the time of fall through SY is to the time of fall through ST as SY is to SX. Now since it has been shown that the spaces traversed are in the same ratio as the squares of the times; and since, moreover, the ratio of the space SY to the space ST is the square of the ratio SY to SX, it follows that the ratio of the times of fall through SY and ST is the ratio of the respective distances SY and SX.

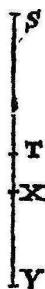


Fig. 50

## SCHOLIUM

The above corollary has been proven for the case of vertical fall; but it holds also for planes inclined at any angle; for it is to be assumed that along these planes the velocity increases in the same ratio, that is, in proportion to the time, or, if you prefer, as the series of natural numbers.\*

SALV.

Here, Sagredo, I should like, if it be not too tedious to Simplicio, to interrupt for a moment the present discussion in order to make some additions on the basis of what has already been proved and of what mechanical principles we have already learned from our Academician. This addition I make for the better establishment on logical and experimental grounds, of the principle which we have above considered; and what is more important, for the purpose of deriving it geometrically, after first demonstrating a single lemma which is fundamental in the science of motion [*impeti*].

SAGR.

If the advance which you propose to make is such as will confirm and fully establish these sciences of motion, I will gladly devote to it any length of time. Indeed, I shall not only[215] be glad to have you proceed, but I beg of you at once to satisfy the curiosity which you have awakened in me concerning your proposition; and I think that Simplicio is of the same mind.

SIMP.

Quite right.

SALV.

Since then I have your permission, let us first of all consider this notable fact, that the momenta or speeds [*i momenti o le velocità*] of one and the same moving body vary with the inclination of the plane.

The speed reaches a maximum along a vertical direction, and for other directions diminishes as the plane diverges from the vertical. Therefore the impetus, ability, energy, [*l'impeto, il talento, l'energia*] or, one might say, the momentum [*il momento*] of descent of the moving body is diminished by the plane upon which it is supported and along which it rolls.

For the sake of greater clearness erect the line AB perpendicular to the horizontal AC; next draw AD, AE, AF, etc., at different inclinations to the horizontal. Then I say that all the momentum of the falling body is along the vertical and is a maximum when it falls in that direction; the momentum is less along DA and still less along EA, and even less yet along the more inclined plane FA. Finally on the horizontal plane the momentum vanishes altogether; the body finds itself in a condition of indifference as to motion or rest; has no inherent tendency to move in any direction, and offers no resistance to being set in motion. For just as a heavy body or system of bodies cannot of itself move upwards, or recede from the common center [*comun centro*] toward which all heavy things tend, so it is impossible for any body of its own accord to assume any motion other than one which carries it nearer to the aforesaid common center. Hence, along the horizontal, by which we understand a surface, every point of which is equidistant from this same common center, the body will have no momentum whatever.

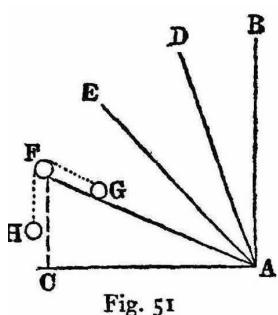


Fig. 51

It is clear that the impelling force [*impeto*] acting on a body in descent is equal to the resistance or least force [*resistenza o forza minima*] sufficient to hold it at rest. In order to measure this force and resistance [*forza e resistenza*] I propose to use the weight of another body. Let us place upon the plane FA a body G connected to the weight H by means of a cord passing over the point F; then the body H will ascend or

descend, along the perpendicular, the same distance which the body G ascends or descends along the inclined plane FA; but this distance will not be equal to the rise or fall of G along the vertical in which direction alone G, as other bodies, exerts its force [*resistenza*]. This is clear. For if we consider the motion of the body G, from A to F, in the triangle AFC to be made up of a horizontal component AC and a vertical component CF, and remember that this body experiences no resistance to motion along the horizontal (because by such a[217] motion the body neither gains nor loses distance from the common center of heavy things) it follows that resistance is met only in consequence of the body rising through the vertical distance CF. Since then the body G in moving from A to F offers resistance only in so far as it rises through the vertical distance CF, while the other body H must fall vertically through the entire distance FA, and since this ratio is maintained whether the motion be large or small, the two bodies being inextensibly connected, we are able to assert positively that, in case of equilibrium (bodies at rest) the momenta, the velocities, or their tendency to motion [*propensioni al moto*], i. e., the spaces which would be traversed by them in equal times, must be in the inverse ratio to their weights. This is what has been demonstrated in every case of mechanical motion.\* So that, in order to hold the weight G at rest, one must give H a weight smaller in the same ratio as the distance CF is smaller than FA. If we do this,  $FA:FC = \text{weight } G:\text{weight } H$ ; then equilibrium will occur, that is, the weights H and G will have the same impelling forces [*momenti eguali*], and the two bodies will come to rest.

And since we are agreed that the impetus, energy, momentum or tendency to motion of a moving body is as great as the force or least resistance [*forza o resistenza minima*] sufficient to stop it, and since we have found that the weight H is capable of preventing motion in the weight G, it follows that the less weight H whose entire force [*momento totale*] is along the perpendicular, FC, will be an exact measure of the component of force [*momento parziale*] which the larger weight G exerts along the plane FA. But the measure of the total force [*total momento*] on the body G is its own weight, since to prevent its fall it is only necessary to balance it with an equal weight, provided this second weight be free to move vertically; therefore the component of the force [*momento parziale*] on G along the inclined plane FA will bear to the maximum and total force on this same body G along the perpendicular FC the same ratio as the weight H to the weight G. This ratio is, by construction, the same which the height, FC, of the inclined plane bears to the length FA. We have here the lemma which I proposed to demonstrate and which, as you will see, has been assumed by our Author in the second part of the sixth proposition of the present treatise.

SAGR.

From what you have shown thus far, it appears to me that one might infer, arguing *ex aequali con la proportione perturbata*, that the tendencies [*momenti*] of one and the same body to move along planes differently inclined, but having the same vertical height, as FA and FI, are to each other inversely as the lengths of the planes.

Perfectly right. This point established, I pass to the demonstration of the following theorem:

If a body falls freely along smooth planes inclined at any angle whatsoever, but of the same height, the speeds with which it reaches the bottom are the same.

First we must recall the fact that on a plane of any inclination whatever a body starting from rest gains speed or momentum [*la quantita dell'impeto*] in direct proportion to the time, in agreement with the definition of naturally accelerated motion given by the Author. Hence, as he has shown in the preceding proposition, the distances traversed are proportional to the squares of the times and therefore to the squares of the speeds. The speed relations are here the same as in the motion first studied [i. e., *vertical motion*], since in each case the gain of speed is proportional to the time.

Let AB be an inclined plane whose height above the level BC is AC. As we have seen above the force impelling [*l'impeto*] a body to fall along the vertical AC is to the force which drives the same body along the inclined plane AB as AB is to AC. On the incline AB, lay off AD a third proportional to AB and AC; then the force producing motion along AC is to that along AB (i. e., along AD) as the length AC is to the length AD. And therefore the body will traverse the space AD, along the incline AB, in the same time which it would occupy in falling the vertical distance AC, (since the forces [*momenti*] are in the same ratio as these distances); also the speed at C is to the speed at D as the distance AC is to the distance AD. But, according to the definition of accelerated motion, the speed at B is to the speed of the same body at D as the time required to traverse AB is to the time required for AD; and, according to the last corollary of the second proposition, the time of passing through the distance AB bears to the time of passing through AD the same ratio as the distance AC (a mean proportional between AB and AD) to AD. Accordingly the two speeds at B and C each bear to the speed at D the same ratio, namely, that of the distances AC and AD; hence they are equal. This is the theorem which I set out to prove.

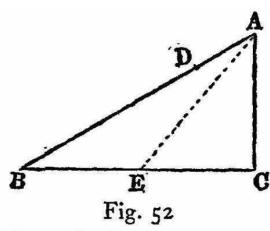


Fig. 52

From the above we are better able to demonstrate the following third proposition of the Author in which he employs the following principle, namely, the time required to traverse an inclined plane is to that required to fall through the vertical height of the plane in the same ratio as the length of the plane to its height.

In like manner it can be shown that the time required to fall through AC is to the time required for any other incline AE as the length AC is to the length AE; therefore, *ex aequali*, the time of fall along the incline AB is to that along AE as the distance AB is to the distance AE, etc.\*

One might by application of this same theorem, as Sagredo will readily see, immediately demonstrate the sixth proposition of the Author; but let us here end this digression which Sagredo has perhaps found rather tedious, though I consider it quite important for the theory of motion.

SAGR.

On the contrary it has given me great satisfaction, and indeed I find it necessary for a complete grasp of this principle.

SALV.

I will now resume the reading of the text.

If one and the same body, starting from rest, falls along an inclined plane and also along a vertical, each having the same height, the times of descent will be to each other as the lengths of the inclined plane and the vertical.

Let AC be the inclined plane and AB the perpendicular, each having the same vertical height above the horizontal, namely, BA; then I say, the time of descent of one and the same body [216] along the plane AC bears a ratio to the time of fall along the perpendicular AB, which is the same as the ratio of the length AC to the length AB. Let DG, EI and LF be any lines parallel to the horizontal CB; then it follows from what has preceded that a body starting from A will acquire the same speed at the point G as at D, since in each case the vertical fall is the same; in like manner the speeds at I and E will be the same; so also those at L and F. And in general the speeds at the two extremities of any parallel drawn from any point on AB to the corresponding point on AC will be equal.

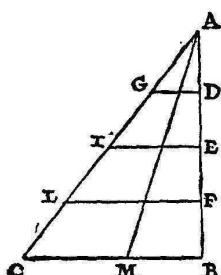


Fig. 53

Fig. 53

Thus the two distances AC and AB are traversed at the same speed. But it has already been proved[217] that if two distances are traversed by a body moving with equal speeds, then the ratio of the times of descent will be the ratio of the distances themselves; therefore, the time of descent along AC is to that along AB as the length of the plane AC is to the vertical distance AB.

q. e. d.

It seems to me that the above could have been proved clearly and briefly on the basis of a proposition already demonstrated, namely, that the distance traversed in the case of accelerated motion along AC or AB is the same as that covered[219] by a uniform speed whose value is one-half the maximum speed, CB; the two distances AC and AB having been traversed at the same uniform speed it is evident, from Proposition I, that the times of descent will be to each other as the distances.

## COROLLARY

Hence we may infer that the times of descent along planes having different inclinations, but the same vertical height stand to one another in the same ratio as the lengths of the planes. For consider any plane AM extending from A to the horizontal CB; then it may be demonstrated in the same manner that the time of descent along AM is to the time along AB as the distance AM is to AB; but since the time along AB is to that along AC as the length AB is to the length AC, it follows, *ex aequali*, that as AM is to AC so is the time along AM to the time along AC.

## Theorem IV, Proposition IV

The times of descent along planes of the same length but of different inclinations are to each other in the inverse ratio of the square roots of their heights

From a single point B draw the planes BA and BC, having the same length but different inclinations; let AE and CD be horizontal lines drawn to meet the perpendicular BD; and[220] let BE represent the height of the plane AB, and BD the height of BC; also let BI be a mean proportional to BD and BE; then the ratio of BD to BI is equal to the square root of the ratio of BD to BE. Now, I say, the ratio of the times of descent along BA and BC is the ratio of BD to BI; so that the time of descent along BA is related to the height of the other plane BC, namely BD as the time along BC is related to the height BI. Now it must be proved that the time of descent along BA is to that along BC as the length BD is to the length BI.

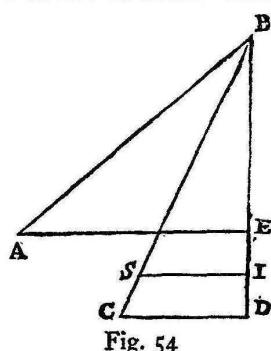


Fig. 54

Draw IS parallel to DC; and since it has been shown that the time of fall along BA is to that along the vertical BE as BA is to BE; and also that the time along BE is to that

along BD as BE is to BI; and likewise that the time along BD is to that along BC as BD is to BC, or as BI to BS; it follows, *ex æquali*, that the time along BA is to that along BC as BA to BS, or BC to BS. However, BC is to BS as BD is to BI; hence follows our proposition.

### Theorem V, Proposition V

The times of descent along planes of different length, slope and height bear to one another a ratio which is equal to the product of the ratio of the lengths by the square root of the inverse ratio of their heights.

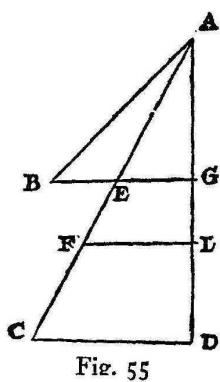


Fig. 55

### Theorem VI, Proposition VI

If from the highest or lowest point in a vertical circle there be drawn any inclined planes meeting the circumference the times of descent along these chords are each equal to the other.

By use of the principles of mechanics [*ex mechanicis*] one may obtain the same result, namely, that a falling body will require equal times to traverse the distances CA and DA, indicated in the following figure. Lay off BA equal to DA, and let fall the [222] perpendiculars BE and DF; it follows from the principles of mechanics that the component of the momentum [*momentum ponderis*] acting along the inclined plane ABC is to the total momentum [i. e., the momentum of the body falling freely] as BE is to BA; in like manner the momentum along the plane AD is to its total momentum [i. e., the momentum of the body falling freely] as DF is to DA, or to BA. Therefore the momentum of this same weight along the plane DA is to that along the plane ABC as the length DF is to the length BE; for this reason, this same weight will in equal times according to the second proposition of the first book, traverse spaces along the planes CA and DA which are to each other as the lengths BE and DF. But it can be shown that CA is to DA as BE is to DF. Hence the falling body will traverse the two paths CA and DA in equal times.

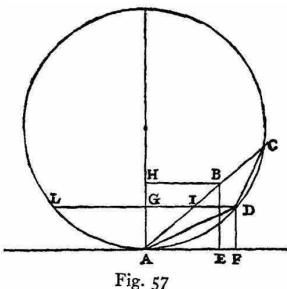


Fig. 57

Fig. 57

Moreover the fact that CA is to DA as BE is to DF may be demonstrated as follows: Join C and D; through D, draw the line DGL parallel to AF and cutting the line AC in I; through B draw the line BH, also parallel to AF. Then the angle ADI will be equal to the angle DCA, since they subtend equal arcs LA and DA, and since the angle DAC is common, the sides of the triangles, CAD and DAI, about the common angle will be proportional to each other; accordingly as CA is to DA so is DA to IA, that is as BA is to IA, or as HA is to GA, that is as BE is to DF.

e. d.

The same proposition may be more easily demonstrated as follows: On the horizontal line AB draw a circle whose diameter DC is vertical. From the upper end of this diameter draw any inclined plane, DF, extending to meet the circumference; then, I say, a body will occupy the same time in falling along the plane DF as along the diameter DC. For draw FG parallel to AB and perpendicular to DC; join FC; and since the time of fall along DC is to that along DG as the mean proportional[223] between CD and GD is to GD itself; and since also DF is a mean proportional between DC and DG, the angle DFC inscribed in a semicircle being a right-angle, and FG being perpendicular to DC, it follows that the time of fall along DC is to that along DG as the length FD is to GD. But it has already been demonstrated that the time of descent along DF is to that along DG as the length DF is to DG; hence the times of descent along DF and DC each bear to the time of fall along DG the same ratio; consequently they are equal.

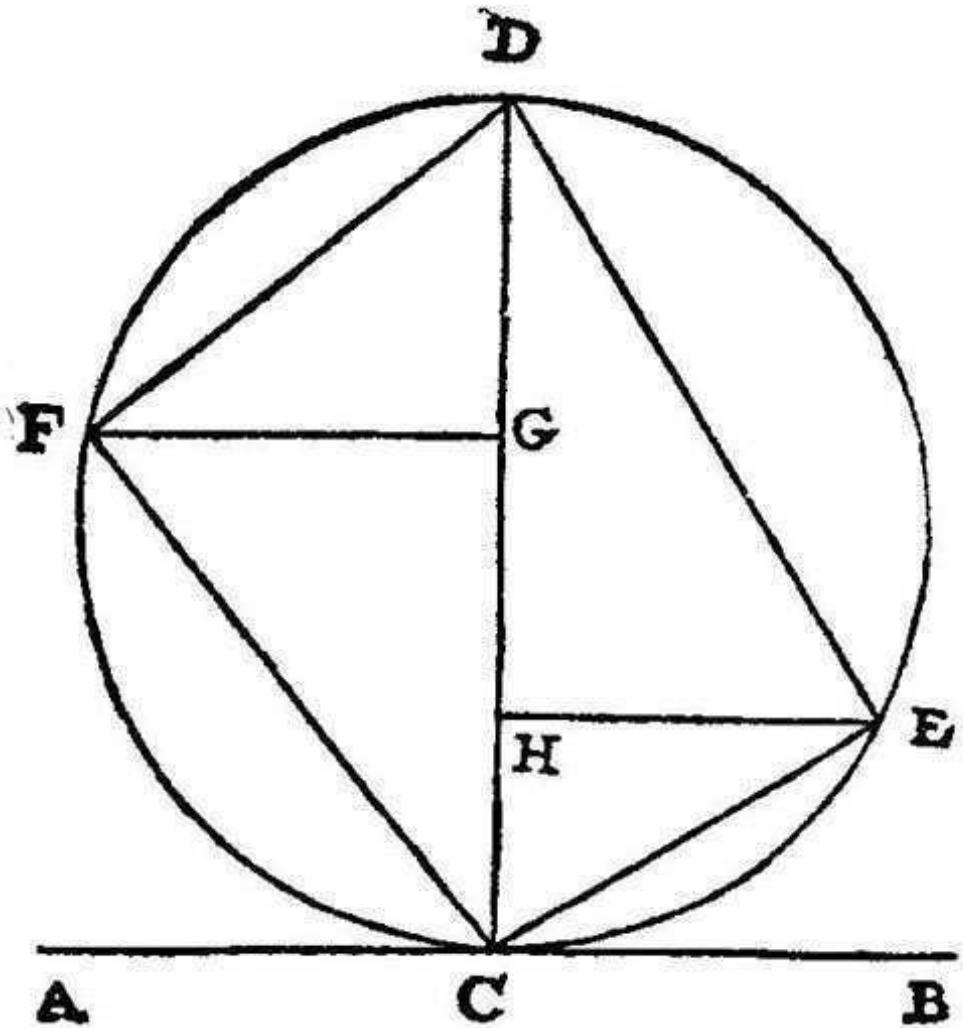


Fig. 58

Fig. 58

In like manner it may be shown that if one draws the chord CE from the lower end of the diameter, also the line EH parallel to the horizon, and joins the points E and D, the time of descent along EC, will be the same as that along the diameter, DC.

**Philosophiae Naturalis Principia Mathematica**

Newton, Isaac  
1687

---

## DEFINITIONS.

---

### DEFINITION I

*The quantity of matter is a measure of the same arising jointly from the density and magnitude [volume].*

**AIR** with the density doubled, in a volume also doubled, shall be quadrupled; in triple the volume, six times as great. You understand the same about snow and powders with the condensing from melting or by compression. And the account of all bodies is the same which are condensed in different ways by whatever causes. Here meanwhile I have no account of a medium, if which there were, freely pervading the interstitia of the parts [of the body]. But I understand this quantity everywhere in what follows under the name of the body or of the mass. That becomes known through the weight of any body: for the proportion to the weight is to be found through experiments with the most accurate of pendulums set up, as will be shown later.

### DEFINITION II.

*The quantity of motion is a measure of the same arising from the velocity and quantity of matter jointly.*

The whole motion is the sum of the motions within the single parts ; and therefore in a body twice as great, with equal velocity, it is doubled, & with the velocity doubled four times as great.

### DEFINITION III.

*The innate force of matter is the resisting force, by which each individual body, however great it is in itself, persists in its state either of rest or of moving uniformly straight forwards.*

This innate [or *vis insita*] force is proportional always to its body, and nor does it differ at all from the inertia of the mass, unless in the required manner of being considered. From the inertia of matter it arises, that each body may be disturbed with difficulty either from its state of rest or from its state of motion. From which the *vis insita* will be possible also to be called by a most significant name the *vis inertiae* [force of inertia]. Truly the body exercises this force only in the change made of its state by some other force impressed on itself ; and the exercise is under that difference with respect to

resistance and impetus: Resistance: in as much as the body is resisting a change in its state by the force acting; Impetus: in as much as the same body, with the force of resistance requiring to concede to the obstacle, tries to change the state of this obstacle. One commonly attributes resistance to states of rest and impetus to states of movement : but motion and rest, as they are considered commonly, are distinguished only in turn from each other; nor are [bodies] truly at rest which may be regarded commonly as being in a state of rest.

[The use of the word *directum*, direct or straight forward rather than straight line, as is given in texts on mechanics removes a circular argument from the definition, as the body can only so move in the absence of forces, and cannot be part of the definition as well as a consequence; and there are of course no lines drawn in space, although we could in principle detect deviations of motion along a given direction. Clearly, this was Newton's original meaning, where he uses the word *directum*, and his thoughts on the predominance of Mechanics over Geometry are set out in the Preface to the first edition.]

#### **DEFINITION IV.**

*The impressed force is the action exercised on the body, to changing the state either of rest or of motion uniform in direction.*

This force is in position only during the action, nor remains in the body after the action. For the body may persevere in any new state by the force of inertia only. Moreover the impressed force is of diverse origins, as from a blow, from pressure, or from the centripetal force.

#### **DEFINITION V.**

*It is the centripetal force, by which bodies are drawn, impelled, or tend in some manner from all sides towards some point, as towards a centre.*

Gravity is [a force] of this kind, by which bodies tend towards the centre of the earth ; the magnetic force, by which iron seeks a loadstone; and that force, whatever it may be, by which the planets are drawn perpetually from rectilinear motion, and are forced to revolve along curved lines. A stone rotating in a sling is trying to depart from the turning hand; and in its attempt has stretched the sling, and with that the more the faster it revolves, and it flies off as soon as it is released. I call the force contrary to that endeavour the centripetal force, by which the sling continually pulls the stone back to the hand and keeps it in orbit, as it is directed to the hand or the centre of the orbit. And the account is the same of all bodies, which are driven in a circle. All these are trying to recede from the centre of the orbit; and unless some other force shall be present trying the opposite to this, by which they may be confined and retained in the orbits, and each thus I call centripetal, they will depart with a uniform motion in straight lines. A projectile, if it were abandoned by the force of gravity, would not be deflected towards the earth, but would go in a straight line to the heavens; and that with a uniform motion, but only if the

air resistance may be removed. By its gravity it is drawn from the rectilinear course and always is deflected to the earth, and that more or less for its gravity and with the velocity of the motion. So that the smaller were the gravity for a quantity of matter or the greater the velocity with which it was projected, by that the less will it deviate from a rectilinear course and the further it will go on. If a leaden sphere is projected from the peak of some mountain with a given velocity along a horizontal line by the force of gunpowder, it may go on in a curved line for a distance of two miles, before it falls to earth : since here with the velocity doubled it may go on twice as far as it were, and with ten times the velocity ten times as far as it were: but only if the resistance of the air is removed. And by increasing the velocity it may be possible to increase the distance to any desired distance in which it is projected, and the curvature of the line that it may describe be lessened , thus so that it may fall only according to a distance of ten or thirty or ninety degrees ; or also so that it may encircle the whole earth or finally depart into the heavens, and from the departing speed to go on indefinitely. And by the same account, by which the projectile may be turned by the force of gravity in orbit and may be able to encircle the whole earth, also the moon is able, either by the force of gravity, but only it shall be of gravity, or some other force , by which it may be acted on, always to be drawn back from a rectilinear course towards the earth, and to be turning in its orbit : and without such a force the moon would not be able to be retained in its orbit. This force, if it should be just a little less, would not be enough to turn the moon from a rectilinear course : if just a little greater, would turn the moon more and it would be led from its orbit towards the earth. Certainly it is required, that it shall be of a just magnitude : and it is required of mathematicians to find the force, by which a body will be able to be retained carefully in some given orbit ; and in turn to find the curved path, in which a body departing from some given place with a given velocity may be deflected by a given force. Moreover the magnitude of this centripetal force is of three kinds, absolute, accelerative, and motive.

[ Newton uses some of his later dynamical ideas to refine the centripetal force acting on a body under the influence of a large mass into three parts: the *absolute* force, which depends primarily on the magnitudes of the large mass and small mass, e.g. if the centripetal force were produced on a body near the sun, or near the earth, all else being equal; the *accelerative* force is simply the acceleration due to gravity on a small mass at some location: the force of gravity on a unit mass (i.e.  $g$ ) ; by *motive* force Newton means the rate of change  $\frac{d(mv)}{dt}$  of the quantity of motion  $mv$  or momentum, which in turn he calls simply *motion*.]

## DEFINITION VI.

*The absolute magnitude of a centripetal force is a measure of the same, greater or less, for the effectiveness of the cause of propagating that from the centre into the orbital regions.*

So that the magnetic force for the size of the loadstone either extends more in one loadstone of greater strength, or lesser in another.

### DEFINITION VII.

*The accelerative magnitude of the centripetal force is the measure of this proportional to velocity, that it generates in a given time.*

As the strength of the same loadstone is greater in a smaller distance, smaller in greater : or the force of gravity is greater in valleys, less at the tops of high mountains, and small still (as it will become apparent afterwards) at greater distances from the globe of the earth ; but at equal distances it is the same on all sides, because therefore all falling bodies (heavy or light, large or small) with the air resistance removed, accelerate equally.

### DEFINITION VIII.

*The motive magnitude of the centripetal force is the measure of this, proportional to the motion, which it generates in a given time.*

As the weight is greater in a greater body, less in a smaller ; and in the same body greater near the earth, less in the heavens. This magnitude is the centripetency or the propensity of the whole body to move towards the centre, and (as thus I have said) the weight; and it becomes known always by that force equal and opposite to itself, by which the descent of the body can be impeded.

And the magnitudes of these forces for the sake of brevity can be called the motive, accelerative, and absolute forces, and for the sake of being distinct refer to bodies attracted towards the centre, to the locations of these [moving] bodies, and to the centre of the forces: there is no doubt that the motive force for a body, as the attempt of the whole towards the centre [of the attracting body] is composed from the attempts of all the parts ; and the accelerative force at the position of the body, as a certain effectiveness, spread out in the orbit from the centre through the individual locations to the bodies towards moving the bodies which are in these places ; but the absolute force towards the centre, as being provided by some cause, without which the motive forces may not be propagated through the regions in the revolution; or for that cause there shall be some central body (such is the loadstone at the centre of the magnetic force, or the earth from the centre of the force of gravity) or some other [cause] which may not be apparent. The concept here is only mathematical : For I do not consider the causes and physical seats of the forces.

Therefore the accelerative force is to the motive force as the velocity is to the motion. For the quantity of the motion arises from the velocity and also from the quantity of matter; the motive force arises from the accelerative force taken jointly with the same quantity of matter. For the sum of the actions of the accelerative force on the individual particles of the body is the whole motive force [*i.e.* the weight]. From which next to the

surface of the earth, where the accelerative gravity or the gravitating force is the same in all bodies, the motive gravity or weight is as the body : but if it may ascend to regions were the accelerative weight shall be less, the weight equally may be diminished, and it will be always as the body and accelerative gravity jointly. Thus in regions were the accelerative gravity is twice as small, the weight of the body small by two or three times will be four or six times as small. Again I name attraction and impulses, in the same sense, accelerative and motive forces. But for these words attraction, impulse, or of any propensity towards the centre, I use indifferently and interchangeably among themselves; these forces are required only to be considered from the mathematical point of view and not physically. From which the reader may be warned, lest by words of this kind he may think me to define somewhere either a kind or manner of action or a physical account, or to attribute truly real forces to the centres (which are mathematical points); if perhaps I have said to draw from the centre or to be forces of the centres.

### *Scholium.*

Up to this stage it has been considered to explain a few notable words, and in the following in what sense they shall be required to be understood. Time, space, position and motion, are on the whole the most notable. Yet it is required to note that ordinary people may not conceive these quantities otherwise than from the relation they bear to perception. And thence certain prejudices may arise, with which removed it will be agreed to distinguish between the absolute and the relative, the true and the apparent, the mathematical and the common usage.

I. Absolute time, true and mathematical, flows equably in itself and by its nature without a relation to anything external, and by another name is called duration. Relative, apparent, and common time is some sensible external measure of duration you please (whether with accurate or with unequal intervals) which commonly is used in place of true time; as in the hour, day, month, year.

II. Absolute space, by its own nature without relation to anything external, always remains similar and immovable: relative [space] is some mobile measure or dimension of this [absolute] space, which is defined by its position to bodies according to our senses, and by ordinary people is taken for an unmoving space: as in the dimension of a space either underground, in the air, or in the heavens, defined by its situation relative to the earth. Absolute and relative spaces are likewise in kind and magnitude; but they do not always endure in the same position. For if the earth may move, for example, the space of our air, because relative to and with respect to the earth it always remains the same, now there will be one part of absolute space into which the air moves, now another part of this; and thus always it will be moving absolutely.

III. The position [or place] is a part of space which a body occupies, and for that reason it is either an absolute or relative space. A part of space, I say, not the situation of [places within] the body, nor of the surrounding surface. For there are always equal

positions[within] equal solid shapes; but not so surfaces as most are unequal on account of dissimilarities of the figures;

[for a surface is liable to change, due to air resistance, etc.]

Truly positions may not have a magnitude on speaking properly, nor are they [to be considered] as places rather than as affectations of places [*i.e.* the position is not a physical property of the body, but rather an indication of where the body is situated at some time in space]. The whole motion is the same as the sum of the motions of the parts, that is, the translation of all from its place is the same as the sum of the translations of the parts from their places ; and thus the place of the whole is the same as the sum of all the parts of the places and therefore both inside and with the whole body.

IV. An absolute motion is the translation of a body from one absolute place into another absolute place, a relative [motion] from a relative [place] into a relative [place]. Thus in a ship which is carried along in full sail, the relative position of the body is that region of the ship in which the body moves about, or the part of the whole cavity of the ship [hold] which the body fills up, and which thus is moving together with the ship : and relative quiet is the state of being of the body in that same ship or in the part of the hold. But the persistence of the body is true rest in the same part of space in which the ship is not moving, in which the ship itself together with the hold and all the contents may be moving. From which if the earth truly is at rest, the body which relatively at rest in the ship, truly will be moving and absolutely with that velocity by which the ship is moving on the earth. If the earth also is moving; there is the true and absolute motion of the body, partially from the motion of the ship truly in an unmoving space, partially from the motion of the ship relative to the earth: and if the body is moving relatively in the ship, the true motion of this arises, partially from the true motion of the earth in motionless space, partially from the relative motion both of the ship on the earth as well as of the body in the ship ; and from these relative motions the motion of the body relative to the earth arises. So that if that part of the earth, where the ship is moving, truly is moving to the east with a speed of 10010 parts; and by the wind in the sails the ship is carried to the west with a velocity of ten parts ; moreover a sailor may be walking on the ship towards the east with a velocity of one part : truly the sailor will be moving and absolutely in the immobile space with a velocity of 10001 to the east, and relative to the earth towards the east with a speed of nine parts.

Absolute time is distinguished from relative time in astronomy by the common equation of time. For the natural days are unequal, which commonly may be taken as equal for the measure of time. Astronomers correct this inequality, so that they measure the motion of the heavens from the truer time. It is possible, that there shall be no uniform motion, by which the time may be measured accurately. All motions are able to be accelerated and retarded, but the flow of absolute time is unable to change. The duration or the perseverance of the existence of things is the same, either the movement shall be fast or slow or none at all: hence this is distinguished by merit from the sensibilities of their measurement, and from the same [the passage of time] is deduced through an astronomical equation. But a need prevails for phenomena in the determination of this equation, at some stage through an experiment with pendulum clocks, then also by the eclipses of a satellite of Jupiter.

As the order of the parts of time is unchangeable, thus too the parts of space. These could be moved from their own places (as thus I may say), and they will be moved away from each other [*i.e.* out of sequence]. For the times and the spaces are themselves of this [kind] and as if the places of all things: in time according to an order of successions, and in space according to an order of positions, to be put in place everywhere. Concerning the essence of these things, it is that they shall be regarded as places : and it is absurd to move the first places. These therefore are absolute places ; and only the translations from these places are absolute motions.

In truth since these parts of space are unable to be seen, and to be distinguished from each other by our senses; we use in turn perceptible measures of these. For we define all places from the positions and distances of things from some body, which we regard as fixed : and then also we may consider all motion with respect to the aforementioned place, as far as we may conceive bodies to be transferred from the same. Thus in exchange of absolute places and motions we make use of relative ones ; not to be an inconvenience in human affairs : but required to be abstracted from the senses in [natural] philosophical matters. And indeed it can happen, that actually no body may be at rest, to which the position and motion may be referred to.

But rest and motion, both absolute and relative, are distinguished in turn from each other by their properties, causes and effects. The property of rest is, that bodies truly at rest are at rest among themselves. And thus since it shall be possible, that some body in the regions of fixed [stars], or far beyond, may remain absolutely at rest ; moreover it is impossible to know in turn from the situation of bodies in our regions, whether or not any of these given at a remote position may serve [to determine true rest in the absolute space for local bodies]; true rest cannot be defined from the situation of these bodies between themselves.

A property of motion is, that the parts which maintain given positions to the whole, share in the motion of the whole. For all the rotating parts are trying to recede from the axis of the motion, and the impetus of the forwards motion arises from the impetus of the individual parts taken together. Therefore with the motion for circulating bodies [*e.g.* planets], they do move in circles [*i.e.* orbits] in which they are relatively at rest. And therefore true and absolute motion cannot be defined by a translation from the vicinity of such bodies, which [otherwise] may be regarded as being in a state of rest. For external bodies [introduced by way of example] must not only seem as being in a state of rest, but also truly to be at rest. Otherwise everything included also participates in the true orbiting motion, besides a translation from the vicinity of the orbiting body ; and with that translation taken away they are not truly at rest, but they may be seen only at rest in this manner. For the orbiting bodies are to the included, as the total exterior part to the interior part, or as a shell to the kernel. But with the shell moving also the kernel is moving, or a part of the whole, without a translation from the vicinity of the shell.

The relation to the preceding property is this, because in the place moved a single location is moved : and thus a body, which is moved with the place moved, also shares the motion of its place. Therefore all [relative] motions, which are made from moved places, are only parts of both the whole and absolute motions, and every whole motion is composed from the motion of the body from its first place, and from the motion of this place from its own place in turn, and thus henceforth ; until at last it may arrive at a

stationary place, as in the example of the sailor mentioned above. From which whole and absolute motions can be defined only from unmoved places: and therefore above I have referred to these as immovable places, and relative places to be moveable places. But they are not immovable places, unless all the given positions may serve in turn from infinity to infinity ; and so always they remain immovable, and I call the space which they constitute immovable.

The causes, by which true and relative motions can be distinguished from each other in turn, are the forces impressed on bodies according to the motion required to be generated. True motion neither can be generated nor changed, other than by forces impressed on the motion of the body itself: but relative motion can be generated and changed without forces being impressed on this body. For it suffices that they be impressed on other bodies alone to which the motion shall be relative, so that with these yielding, that relation may be changed from which it consisted, of rest or relative motion. On the contrary true motion always is changed from the forces impressed on a moving body ; but the relative motion from these forces is not changed by necessity. For if the same forces thus may be impressed on other bodies also, for which a relation is made, thus so that the relative situation will be conserved on which the relative motion is founded. Therefore all relative motion can be changed where the true may be conserved, and to be conserved where the true may be changed ; and therefore true motion in relations of this kind are considered the least.

The effects, by which absolute and relative motions are to be distinguished from each other, are the forces of receding from the axis of circular motion. For none of these forces in circular motion are in mere relative motion, but are in a true [circular] motion greater from true absolute motion for a quantity of motion. If a vessel may hang from a long thread, and always is turned in a circle, while the thread becomes very stiff, then it may be filled with water, and together with the water remains at rest; then by another force it is set in motion suddenly in the opposite sense and with the thread loosening itself, it may persevere a long time in this motion; the surface of the water from the beginning was flat, just before the motion of the vessel: But after the vessel, with the force impressed a little on the water, has the effect that this too begins to rotate sensibly; itself to recede a little from the middle, and to ascend the sides of the vessel, adopting a concave figure, (as I have itself tested) and by moving faster from the motion it will rise always more and more, while the revolutions by being required to be completed in the same times with the vessel, it may come to relative rest with the same vessel. Here the ascent indicates an attempt to recede from the axis of the motion, and by such an attempt it becomes known, and the true and absolute circular motion of the water is measured, and this generally is contrary to the relative motion. In the beginning, when the motion of the water was a maximum relative to the vessel, that motion did not incite any attempt to recede from the axis: the water did not seek circumference by requiring to ascend the sides of the vessel, but remained flat, and therefore the true circular motion had not yet begun. Truly later, when the relative motion had decreased, the ascent of this to the sides of the vessel indicated an attempt of receding from the axis; and this trial showed this true circular motion always increasing, and finally made a maximum when the water remained at rest relative to the vessel. Whereby this trial does not depend on the translation of the water

with respect to orbiting bodies, and therefore true circular motion cannot be defined by such translations. Truly the circular motion of each revolving body is unique, corresponding to a singular and adequate effort to be performed : but relative motions are for innumerable and varied external relations ; and corresponding to a relation, generally they are lacking in true effects, unless in as much as they share in that true and single motion. And by those who wish, within a system of these [rotational motions], our heavens [*i.e.* local space] to revolve in a circle below the heavens of the fixed stars [*i.e.* distant interstellar space], and the planets to defer with it ; the individual parts of the heavens, and the planets which truly are moving, which indeed within their nearby heavens [*i.e.* the local part of space relative to themselves] are relatively at rest, truly are moving. For they change their positions in turn (as otherwise the system truly passes into rest) and together with the deferred heavens they participate in the motion of these, and so that the parts of the revolving total are trying to recede from the axes of these.

Relative quantities are not therefore these quantities themselves, the names of which they bear, but those perceptible measures (true or mistaken) of them which are used by ordinary people in place of the measured quantities. [Thus, a length is related to a standard length, etc.] But if the significances of words are required to be defined from the use; these measures perceptible [to the senses] are to be particularly understood by these names : Time, Space, Location and Motion ; and the discourse will be contrary to custom and purely mathematical, if measured quantities here are understood. [Here Newton is expressing the fact that he uses such quantities in our sense as abstract variables, rather than as mere units for measuring the amounts of physical quantities, as one might use in arithmetic.] Hence they carry the strength of holy scriptures, which may be interpreted there by these names regarding measured quantities. Nor do they corrupt mathematics or [natural] philosophy any less, who combine true quantities with the relations of these and with common measures.

Indeed it is most difficult to know the true motion of bodies and actually to discriminate from apparent motion ; therefore because the parts of that immobile space, in which bodies truly are moving, do not meet the senses. Yet the cause is not yet quite desperate. For arguments are able to be chosen, partially from apparent motions which are the differences of true motions, partially from forces which are the causes and effects of true motions. So that if two globes, to be connected in turn at a given distance from the intervening thread, may be revolving about the common centre of gravity; the exertion of the globes to recede from the axis of the motion might become known from the tension in the thread, and thence the quantity of the circular motion can be computed. Then if any forces acting equally likewise may be impressed mutually on the faces of the globes to increase or diminish the circular motion, the increase or decrease in the circular motion may become known from the increase or decrease in the tension of the thread ; and thence finally the faces of the globes on which the impressed forces must be impressed, so that the motion may be increased maximally; that is, the faces to the rear, or which are following in the circular motion. But with the faces which are following known, and with the opposite faces which precede, the determination of the motion may be known. In this manner both the quantity and the determination of the motion of this circle may be found in a vacuum however great, where nothing may stand out externally and perceptibly by which the globes may be able to be brought together [in comparison]. If now bodies may

be put in place in that space with a long distance maintained between themselves, such as the fixed stars are in regions of the heavens : indeed it may not be possible to know from the relative translation of the globes among the bodies, whether from these or those a motion may be required to be given. But if attention is turned to the string, and the tension of that itself is taken to be as the required motion of the globes ; it is possible to conclude that the motion is that of the globes, and [the distant] bodies to be at rest ; & then finally from the translation of the globes among the bodies, the determination of this motion can be deduced. But the true motion from these causes, are to be deduced from the effects and from the apparent differences, or on the contrary from the motions or forces, either true or apparent, the causes and effects of these to be found, will be taught in greater detail in the following. For towards this end I have composed the following treatise.

# AXIOMS, OR THE LAWS OF MOTION.

---

## LAW I.

*Every body perseveres either in its state of resting or of moving uniformly in a direction, unless that is compelled to change its state by impressed forces.*

**P**rojectiles persevere in their motion, unless in as much as they may be retarded by the resistance of the air, and they are impelled downwards by the force of gravity. A child's spinning top, the parts of which by requiring to stick together always, withdraw themselves from circular motion, does not stop rotating, unless perhaps it may be slowed down by the air. But the greater bodies of the planets and comets preserve both their progressive and circular motions for a long time made in spaces with less resistance .

## LAW II.

*The change of motion is proportional to the [magnitude of the] impressed motive force, and to be made along the right line by which that force is impressed.*

If a force may generate some motion ; twice the force will double it, three times triples, if it were impressed either once at the same time, or successively and gradually. And this motion (because it is determined always in the same direction generated by the same force) if the body were moving before, either is added to the motion of that in the same direction, or in the contrary direction is taken away, or the oblique is added to the oblique, and where from that each successive determination is composed.

## LAW III.

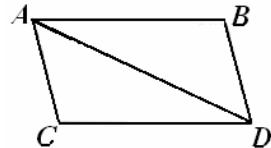
*To an action there is always an equal and contrary reaction : or the actions of two bodies between themselves are always mutually equal and directed in opposite directions.*

Anything pressing or pulling another, by that is pressed or pulled just as much. If anyone presses a stone with a finger, the finger of this person is pressed by the stone. If a horse pulls a stone tied to a rope, and also the horse (as thus I may say) is drawn back equally by the stone: for the rope stretched the same on both sides requiring itself to be loosened will draw upon the horse towards the stone, and the stone towards the horse; and yet it may impede the progress of the one as much as it advances the progress of the other. If some body striking another body will have changed the latter's motion in some manner by the former's force, the same change too will be undergone in turn on its own motion in the contrary direction by the force of the other (on account of the mutual pressing together). By these actions equal changes are made, not of the velocities but of the motions ; obviously in bodies not impeded otherwise. For changes of the velocity, are made likewise in the contrary parts, because the motions are changed equally, they are inversely proportional with the bodies [*i.e.* with their masses]. This law is obtained with attractions also, as will be approved in a nearby scholium.

## COROLLARY I.

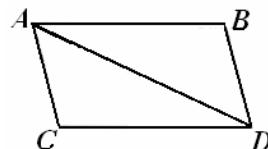
*A body with forces added together describes the diagonal of the parallelogram in the same time, in which the separate sides are described.*

If the body in a given time, by the force alone  $M$  impressed at the place  $A$ , may be carried with a uniform motion from  $A$  to  $B$ ; and by the single force  $N$  impressed at the same place, may be carried from  $A$  to  $C$ : the parallelogram  $ABDC$  may be completed, and by each force that body may be carried in the same time on the diagonal from  $A$  to  $D$ . For because the force  $N$  acts along the line  $AC$  parallel to  $BD$  itself, this force by law II will not at all change the velocity required to approach that line  $BD$  generated by the other force. Therefore the body approaches the line  $BD$  in the same time, whether the force  $N$  may be impressed or not ; and thus at the end of the time it may be found somewhere on that line  $BD$ . By the same argument at the end of the same time the body will be found somewhere on the line  $CD$ , and on that account it is necessary to be found at the concurrence  $D$  of each of the lines. Moreover it will go in rectilinear motion from  $A$  to  $D$  by law I.

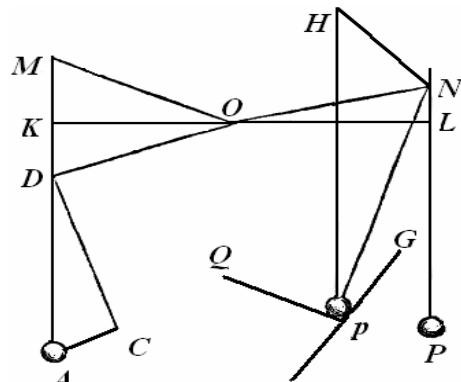


## COROLLARY II.

*And hence the composition of the force directed along AD is apparent from any oblique forces AB and BD, and in turn the resolution of any force directed along AD into any oblique forces AB and BD. Which composition and resolution indeed is confirmed abundantly from mechanics.*



So that if from the centre of some wheel  $O$  the unequal radii  $OM, ON$  emerge, the weights  $A$  and  $P$  may be sustained by the threads  $MA, NP$ , and the forces of the weights are required towards moving the wheel: Through the centre  $O$  a right line  $KOL$  is drawn, meeting the threads perpendicularly at  $K$  and  $L$ , and with the centre  $O$ , and  $OL$  the greater of the intervals  $OK, OL$ , a circle is described meeting the thread  $MA$  in  $D$ : and  $AC$  shall be parallel to the right line made  $OD$ , and the perpendicular  $DC$ . Because it is of no importance, whether the points of the threads  $K, L, D$  shall be fastened or not to the plane of the wheel; the weights will prevail the same, and if they may be suspended from the points  $K$  and  $L$  or  $D$  and  $L$ . But the total force of the weight  $A$  is set out by the line  $AD$ , and this is resolved into the forces  $AC, CD$ , of which  $AC$  by pulling [drawing in the original text] the radius  $OD$  directly from the centre provides no force to the required wheel movement ; but the other force  $DC$ , by pulling on the radius  $DO$  perpendicularly, accomplishes the same, as if it pulls the radius  $OL$  the equal of  $OD$  itself; and that is, the same weight  $P$ , but only if that weight shall be to the weight  $A$  as the force  $DC$  to the force  $DA$ , that is (on account of the similar triangles  $ADC, DOK$ ,) as  $OK$  to  $OD$  or  $OL$ . Therefore the weights  $A$  and  $P$ , which are inversely as the radii  $OK$  and  $OL$  placed in line, will exert the same influence, and thus remain in equilibrium : which is the most noticeable property of scales, levers, and of a wheel and axle. If either weight shall be greater than in this ratio, the force of this will be so much greater requiring the wheel to rotate.



But if a weight  $p$ , equal to the weight  $P$ , is suspended partly by the thread  $Np$ , and partly by resting on the oblique plane  $pG$  : [the forces]  $pH$  and  $NH$  are acting, the former perpendicular to the horizontal, the latter perpendicular to the plane  $pG$ ; and if the force of the weight  $p$  acting downwards, is set out by the line  $pH$ , this can be resolved into the forces  $pN, HN$ . If some plane  $pQ$  shall be perpendicular to  $pN$ , cutting the other plane  $pG$  in a line parallel to the horizontal [here we have to think in 3 dimensions]; and the weight  $p$  lies only on the planes  $pQ, pG$  ; that it may press on these planes by these forces  $pN, HN$ , without doubt the plane  $pQ$  perpendicularly by the force  $pN$ , and the plane  $pG$  by the force  $HN$ . And thus if the plane  $pQ$  may be removed, so that the weight may stretch the thread ; because now the thread in turn by being required to sustain the weight, performs the function of the plane removed, may be extended by that same force  $pN$ , which before

acted on the plane. From which, the tension of this oblique thread will be to the tension of the thread of the other perpendicular  $PN$ , as  $pN$  to  $pH$ . And thus if the weight  $p$  shall be to the weight  $A$  in a ratio, which is composed from the reciprocal ratio of the minimum distances of their threads  $pN, AM$  from the centre of the wheel, and in the direct ratio  $pH$  to  $pN$ ; the same weights likewise will prevail for the wheel being moved, and thus mutually will sustain each other, as any test can prove.

But the weight  $p$ , pressing on these two oblique planes, can be compared to a wedge making a split between the internal surfaces: and thence the forces of the wedge and of the hammer become known: seeing that since the force by which the weight  $p$  presses hard on the plane  $pQ$  to the force, by which the same either by its weight or impelled by the blow of the hammer acting along the line  $pH$  in the plane, shall be as  $pN$  to  $pH$ ; and to the force, by which it presses hard on the other plane  $pG$ , as  $pN$  to  $NH$ . But also the force of a screw can be deduced by a similar division of the forces; obviously which wedge is pushed against by a lever. Therefore the uses of this corollary appear to be the widest, and by extending widely the truth of this prevails; since from now with what has been said, all mechanical devices may depend on explanations shown in different ways by authors. And from these indeed we may derive easily the forces of machines, which from wheels, revolving cylinders, pulleys, levers, stretched cords and weights directly or obliquely ascending, and with the rest from the powers of mechanics are accustomed to be assembled, and as also the forces of tendons requiring the bones of animals to be moved.

[The diagram has been altered a little, as in the original the oblique angle appears to be  $90^0$ . Note that Newton pays a little attention here to statics and simple machines.]

### COROLLARY III.

*The quantity of motion which is deduced by taking the sum of the motions of the contributing factors in the same direction and the difference of the contributing factors in the opposite direction , may not change from the action of bodies among themselves.*

And indeed the action and the contrary reaction to that are equal by law III, and thus by law II bring about equal changes in the motions towards the contrary directions. Therefore if the motions are made in the same direction; whatever is added to the motion of the departing body, is taken from the motion of the following body thus, so that the sum may remain the same as before. If the bodies may get in each other's way, an equal among will be taken from the motion of each, and thus the difference of the contributing factors of the motions in the opposite directions will remain the same.

So that if a spherical body  $A$  shall be three times greater than a spherical body  $B$ , and it may have two parts of velocity ; and  $B$  may follow on the same right line with a velocity of ten parts, and thus the motion of  $A$  shall be to the motion of  $B$  itself, as 6 to 10 : the motion from these may be put to be of 6 parts and of 10 parts, and the sum will be of 16 parts. Therefore in the meeting of the bodies, if the body  $A$  may gain a motion of 3, 4, or 5 parts, the body  $B$  will lose just as many parts, and thus the body  $A$  will go on after the reflection with 9, 10 or 11 parts, and  $B$  with 7, 6, or 5 parts, with the same sum of the parts present as before. If the body  $A$  may gain 9, 10, 11, or 12, and thus may move past

the meeting with 15, 16, 17, or 18 parts ; the body *B*, by losing as many parts as *A* may gain, may be progressing with one part, with 9 parts lost, or it may be at rest with the 10 parts of its motion missing, or with one part it may be moving backwards with one part more (as thus I may say) missing from its motion, or it may be moving backwards with two parts on account of subtracting 12 parts of the motion forwards. And thus the sum of the motions in the same direction  $15+1$  or  $16+0$ , and the differences in the opposite directions  $17-1$  and  $18-2$  will be always 16 parts, as before the meeting and the reflection. But with the motions known with which bodies go on after the reflection, the velocities of which may be found, on putting that to be to the velocity before the reflection, as the motion after is to the motion before. So that in the final case, where the motion of the body *A* were of six parts before reflection and of eight parts afterwards, and the velocity of two parts before the reflection; the velocity of six parts after the reflection may be found, on being required to say, that the motion of 6 parts before the reflection to the 18 parts afterwards, thus of 2 parts of velocity before the reflection to 6 parts of velocity afterwards.

But if bodies are incident between themselves mutually obliquely either non spherical or with differing rectilinear motions, and the motions of these may be required after reflection ; the position of the plane in which the meeting bodies are touching at the point of concurrence is required to be known: then the motion of each body (by Corol.II) is required to be separated into two parts, one perpendicular to this plane, the other parallel to the same : but the parallel motions, therefore because the bodies act in turn between themselves along a line perpendicular to this plane, are required to have retained the same motion before and after the reflection, and the changes in the perpendicular motions thus required are to be attributed equally in opposite directions, so that both the sum of the acting together and the difference of the contrary may remain the same as before. From reflections of this kind also circular motions are accustomed to arise about their own centres. But I will not consider these cases in the following, and it would be exceedingly long to consider showing all this here.

#### COROLLARY IV.

*The common centre of gravity of two or more bodies, from the actions of the bodies between themselves, does not change its state either of motion or of rest ; and therefore the common centre of gravity of all bodies in the mutual actions between themselves (with external actions and impediments excluded) either is at rest or is moving uniformly in direction.*

For if two points may be progressing with a uniform motion in right lines, and the distances of these is divided in a give ratio, the dividing point either is at rest or it is progressing uniformly in the right line. This is shown later in a corollary to lemma XXIII of this work, if the motion of the points is made in the same plane ; and by the same account can be demonstrated, if these motions are not made in the same plane. Hence if some bodies are moving uniformly in right lines, the common centre of gravity of any two either is at rest or progresses uniformly in a right line; because the line joining the centres of these bodies therefore is required to be progressing uniformly in right lines, it

is divided by this common centre in a given ratio. And similarly the common centre of these two and of any third either is at rest or progressing uniformly in a right line ; because the distances between the common centre of the two bodies and of the third body from that therefore is divided in a given ratio . In the same manner also the common centre of these three and of some fourth body either is at rest or is moving uniformly in a straight line; because from that the distance between the common centre of the three and the centre of the fourth body therefore is divided in a given ratio, and thus *ad infinitum*. Therefore in a system of bodies, which in the interactions among themselves in turn and in general are free from all extrinsic forces, and therefore singly are moving uniformly in individual right lines, the common centre of gravity of all either is at rest or is moving in a direction uniformly.

Again in a system of two bodies acting on each other in turn, since the distances of the centres of each from the common centre of gravity shall be reciprocally as the bodies [*i.e.* the masses of the bodies]; the relative motions of the bodies from the same shall be equally between themselves either approaching to that centre or receding from the same. Hence that centre, from the equal changes of the motions made in opposite directions, and thus from the actions of these bodies between themselves, neither will move forwards nor be retarded nor suffers a change in its state as far as motion or rest is concerned. But in a system of several bodies, because the common centre of gravity of any two bodies acting mutually between themselves on account of that action allows no change at all in its state ; and of the remaining , the action from which does not hinder these, the common centre of gravity thence suffers nothing; but the distance of these two centres of gravity is divided by the common centre of all the bodies into parts to which the total [masses] of the bodies are reciprocally proportional ; and thus with the state of their moving or resting maintained from these two centres, the common centre of all also maintains its own state: because it is evident that common centre of all on account of the actions of two bodies between themselves at no time changes its own state as far as motion and rest are concerned. But in such a system all the actions of the bodies between each other, are either between two bodies, or between the actions of two composite bodies ; and therefore at no time do they adopt a change of everything from the common centre in the state of this of motion or rest. Whereby since that centre where the bodies do not act among themselves in turn, either is at rest, or is progressing uniformly in some right motion, will go on the same, without the opposition of bodies, without actions between themselves, either to be at rest or always to be progressing uniformly in a direction ; unless it may be disturbed from that state by some external forces. Therefore it is the same law of a system of many bodies, which of a solitary body, as far as persevering in a state of motion or of rest. Indeed the progressive motion either of a solitary body or of a system of bodies must always be considered from the motion of the centre of gravity.

#### COROLLARY V.

*The motions of bodies between themselves included in a given space are the same, whether that space be at rest, or the same may be moving in a direction without circular motion.*

For the differences of the motions given in the same direction , and the sums to be given in the opposite directions, are the same initially in each case (by hypothesis), and with these sums or differences the collisions and impulses arise with which the bodies strike each other. Therefore by law II the effects of the collisions are equal in each case ; and therefore the motion between themselves in one case will remain equal to the motions between themselves in another case. The same may be proven clearly by an experiment. All motions between bodies occur in the same way on a ship, whither that is at rest or is moving uniformly in a direction.

## COROLLARY VI.

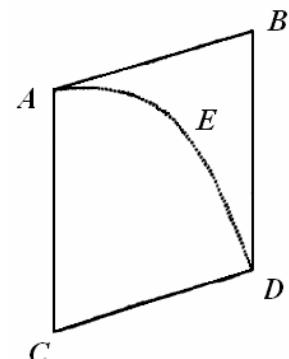
*If bodies may be moving in some manner among themselves, and from the forces with equal accelerations may be impressed to move along parallel lines ; all will go on to move in the same manner between themselves, and if by these forces they are not to be disturbed.*

For these forces equally (for the quantities of the bodies required to be moved) and by acting along parallel lines, all the bodies will be moved equally (as far as velocity is concerned) by law II, and thus at no time will the positions and motions of these between each other be changed.

### *Scholium.*

Thus far I have treated the fundamentals with the usual mathematics and confirmed many times by experiment. By the first two laws and from the first two corollaries Galileo found the fall of weights to be in the duplicate ration of the time, and the motion of projectiles to be in a parabola; by agreeing with experiment, unless as far as the those motions are retarded a little by the resistance of the air. With a body falling uniform gravity, by acting equally in equal small intervals of time, will impress equal forces on that body, and generates equal velocities: and in the total time the total force impressed and the total velocity it generates is proportional to the time. And the distances described in the proportional times, are as the velocities and times taken together ; that is in the duplicate ratio of the times. And with the body projected up uniform gravity impresses forces and velocities taken proportional to the times ; and the times required to rise the greatest heights are as the velocities required to be taken away, and these heights are as the velocities and the times taken together or, in the duplicate ratio of the velocities.

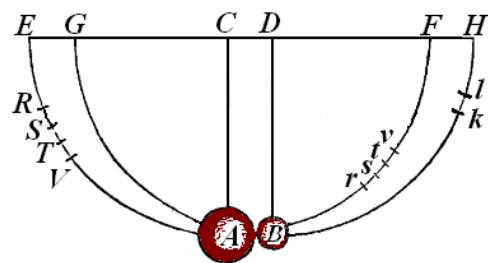
[*Duplicate ratio* means of course, *as the square.*] And the motion arising of a body projected along some right line is composed from the motion arising from gravity. So that if the body *A* by its motion of projection only in a given time can describe the right line *AB* and from its motion only of requiring to fall, can describe the height *AC* in the same time: the parallelogram *ABDC* may be completed, and that body will be found at the end of the time at the place *D* from the composed motion ; and the curved line *AED*, which that body



describes, will be a parabola which the right line  $AB$  touches at  $A$ , and the ordinate  $BD$  of which is as  $AB$  squared. The demonstrations of the times of oscillations of pendulums will depend on the same laws and corollaries, from the daily experience with clocks. From the very same, and with the third law Sir Christopher Wren, Dr. John Wallis and Christian Huygens, easily the principle outstanding geometers of the age, have discovered separately the rules of hard bodies colliding and rebounding, and almost at the same time they communicated the same with the Royal Society, among themselves (as regards these laws) everything is in agreement: and indeed first Wallis, then Wren and Huygens produced the discovery. But the truth has been established by Wren in person before the Royal Society through an experiment with pendulums : which also the most illustrious Mariotte soon deemed worthy to explain in a whole book. Truly, so that this experiment may agree to the precision with theories, and account is required to be had, both of the resistance of the air, as well as of the elasticity of the colliding bodies.

The spherical bodies  $A, B$  may hang from parallel threads and with  $AC, BD$  equal from the centres  $C, D$ . From these centres and intervals the bisected semicircles  $EAF, GBH$  are described, with radii  $CA, DB$ . The body  $A$  may be drawn to some point  $R$  of the arc  $EAF$ , and thence may be released (with the body  $B$  taken aside), and after one oscillation it may return to the point  $V$ .  $RV$  is the retardation from the air resistance.  $ST$  is made the fourth part of  $RV$  placed in the middle, thus evidently so that  $RS$  and  $TV$  are equal, and  $RS$  to  $ST$  shall be as 3 to 2. And thus  $ST$  will show the retardation in falling from  $S$  to  $A$  approximately. [A rule gained from experience perhaps for light damping.]

Body  $B$  may be restored to its place  $B$ . Body  $A$  may fall from the point  $S$ , and the velocity of that at the place of reflection  $A$  will be without so great a sensible error, and as if it had fallen from the location  $T$  in a vacuum. Therefore this velocity is set out by the chord of the arc  $TA$ . For the velocity of the pendulum at the lowest point is as the chord of the arc, that it has described on falling, the proposition is well known from geometry. After the reflection the body  $A$  may arrive at the location  $s$ , and the body  $B$  at the location  $k$ . The body  $B$  may be taken away and the position  $v$  may be found ; from which if the body  $A$  may be sent off and after one oscillation may return to the location  $r$ , let  $st$  be the fourth part of that  $rv$  placed in the middle, thus so that it may be considered that  $rs$  and  $tv$  are equal ; and the velocity may be set out by the chord of the arc  $tA$ , that the body  $A$  had approximately after the reflection at the place  $A$ . For  $t$  will be that true and correct place, to which the body  $A$ , with the resistance of the air removed, ought to be able to rise. The location  $k$  to which the body  $B$  has risen is required to be corrected by a similar method, and requiring to find the location  $l$ , to which that body ought to ascend in a vacuum. With this done it is possible to test everything, in the same way as if we were placed in a vacuum. Yet the body  $A$  will be required to adopt (as thus I may say) the chord  $TA$  of the arc, which shows the velocity of this, so that the motion may be had approximately at the place  $A$  before the reflection ; then the chord  $tA$  of the arc, so that the motion of this may be had approximately at the place  $A$  after the reflection. And thus the body  $B$  will be

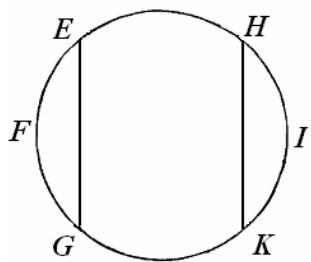


required to adopt the chord of the arc  $Bl$ , so that the approximate motion of this may be had after the reflection. And by a similar method, where the bodies are sent off at the same time from two different locations, the motions of each are required to be found both before, as well as after the reflection ; and then finally the motions among themselves are to be brought together and the effect of the reflection deduced. In this manner with the matter requiring to be tested with ten feet pendulums, and that with bodies both unequal as well as equal, and by arranging so that bodies may concur from the greatest intervals, such as 8, 12, or 16 feet ; I have found [the text has *reperi* or *you find* in the singular command mode, which has been taken as a misprint, rather than *repperi* which has been adopted for translation] always within an error of 3 inches in the measurement, where the bodies themselves were meeting each other directly, equal changes were obtained in the contrary parts of the motions for the bodies, and thus the actions and reactions always to be equal. So that if the body  $A$  were incident on body  $B$  at rest with 9 parts of motion, and with 7 parts removed it went on with 2 parts after the reflection ; body  $B$  was rebounding with these 7 parts. If the bodies were going against each other,  $A$  with 12 parts and  $B$  with 6, and  $A$  was returning with 2 parts ;  $B$  was returning with 8, each with the removal of 14 parts. From the motion of  $A$  12 parts are removed and nothing remains: 2 other parts are taken away , and there is made a motion of 2 parts in the opposite direction: and thus from the motion of the body  $B$  of 6 parts by requiring 14 parts to be taken away, 8 parts are made in the opposite direction. But if the bodies were going in the same direction,  $A$  faster with 14 parts, and  $B$  slower with 5 parts, and after the reflection  $A$  was going on with 5 parts ;  $B$  was going on with 14 parts, with the translation made of 9 parts from  $A$  to  $B$ . And thus for the rest. From a running together and collision of bodies at no time does the quantity of motion change, which is deduced from the sum of the motions acting in the same directions or from the differences in contrary directions. For the error of an inch or two I may attribute to the difficulty required in performing the individual measurements accurately enough. It was with difficulty, not only that the pendulums thus be dropped at the same time, so that the bodies could strike each other at the lowest position  $A B$ ; but also the locations  $s, k$  to be noted, to which the bodies were ascending after the collision. But the unequal density of the parts of the pendulous bodies, and the construction from other causes of irregularity, were leading to errors.

Again lest anyone may object to the rule requiring approval which this experiment has found, to presuppose the bodies either to be completely hard, or perhaps perfectly elastic, none are to be found of this kind in natural compositions ; because I add now experiments described which succeed equally with soft or hard bodies, without doubt by no means depending on the condition of hardness. For if that rule is required to be extended to bodies which are not perfectly hard, the reflection is to be diminished only in a certain proportion to the magnitude of the elastic force. In the theory of *Wren* and *Huygens* absolutely hard bodies return with the speed of the encounter. Most surely that will be proven with perfectly elastic bodies. In imperfectly elastic bodies the return speed is required to be diminished likewise with the elastic force ; therefore because that elastic force, (unless where the parts of bodies are struck by their coming together, or extended a little as if they suffer under a hammer,) and certainly shall be required to be determined (as much as I know) and may be made so that bodies may return with a relative velocity in turn, which shall be in a given ratio to the relative velocity of approach. This I have

tested thus with balls of wool closely piled together and strongly constricted. First by dropping pendulums and by measuring the reflection, I have found the magnitude of the elastic force; then with this force I have determined the reflections in other cases of concurrence, and they have answered the trials. Always the wool have returned with a relative velocity, which is to be to the relative velocity of concurrence as around 5 to 9. Balls of steel return with almost the same velocity ; others from cork with a little less : but with glass the proportion was around 15 to 16. And with this agreed upon, the third law has agreed with theory as far as impacts and reflections are concerned, which clearly agree with experiment.

I briefly show the matter for attractions thus. With any two bodies *A* and *B* mutually attracting each other, consider some obstacle placed between each, by which the meeting of these may be impeded. If either body *A* is drawn more towards the other body *B*, than that other *B* towards the first *A*, the obstacle will be urged more by the pressing of body *A* than by the pressing of body *B*; and hence will not stay in equilibrium. The pressing will prevail stronger, and it will act so that the system of the two bodies and the obstacle may move in the direction towards *B*, and in motions in free spaces always by accelerating, may depart to infinity. Which is absurd and contrary to the first law. For by the first law the system must persevere in its state of rest or of uniform motion in a direction, and hence the bodies will press equally on the obstacle, and on that account are drawn equally in turn. I have tested this with a loadstone and iron. If these placed in their own vessels touching separately they may float next to each other in still water ; neither propels the other, but from the equality of the attraction they sustain mutual attempts between themselves, and finally they remain in an established equilibrium.



Thus also is the gravity between the earth and the mutual parts of this. The earth *FI* is cut by some place *EG* into two parts *EGF* and *EGI*: and the mutual weights of these shall be equal mutually between themselves. For if by another plane *HK* which shall be parallel to the first part *EG*, the greater part *EGI* shall be cut into the two parts *EGKH* and *HKI*, of which *HKI* shall be equal to the first part cut *EGF*: it is evident that

the middle part *EGKH* by its own weight will not be inclined to either of the extreme parts, but between each in equilibrium, thus so that I may say, it may be suspended and it is at rest. But the extreme part *HKI* by its own weight presses on the middle part, and will urge that into the other extreme part *EGF*; and thus the force by which the sum of the parts *HKI* & *EGKH*, *EGI* tends towards the third part *EGF*, is equal to the weight of the third part *HKI*, that is to the weight of the third part *EGF*. And therefore the weights of the two parts *EGI*, *EGF* are mutually in equilibrium, as I had wished to show. And unless these weights shall be equal, the whole earth floating on the free aether may go towards the greater weight, and from that required flight would go off to infinity.

Just as bodies in coming together and reflecting may exert the same influence [on each other], the velocities of which are reciprocally as their innate forces: thus the agents exert the same influence in the movements of mechanical devices, and by contrary exertions mutually sustain each other, the velocities of which, following the determination of the forces considered, are reciprocally as the forces. Thus the weights exert the same

influence towards moving the arms of scales, which with the scales oscillating are reciprocally as the velocities of these up and down: that is, the weights, if they ascend up and down rightly [*i.e.* vertically], exert the same influences, which are reciprocally as the distances of the points from which they are suspended from the axis of the scales; if, by an oblique plane or from other obstacles to the motion, the ascents or descents are oblique, they exert the same influence, which are reciprocally as the ascent and descent, just as made along the perpendicular: and that on account of the determination of the weight acting downwards.

Similarly with a pulley or a pulley system, the force of the hand directly drawing on the rope which shall be to the weight , ascending either directly or obliquely, as the perpendicular speed of ascent to the speed of the hand pulling directly on the rope will sustain the weight. In clocks and similar devices, which have been constructed from little wheels joined together, the forces required contrary to promoting and retarding the motions of the wheels if they are reciprocally as the speeds of the wheels on which they are impressed, will mutually sustain each other. The force of a screw required to press upon a body is to the force of hand turning the handle, as the rotational speed of the handle in that part where it is pressed on by the hand, to the speed of progress of the screw towards the body pressed. The forces by which a wedge urges the two parts of wood to be split are to the force of the hammer to the wedge, as the progress of the wedge following a determined force impressed by the hammer on itself, to the speed by which the parts of the wood, following lines perpendicular to the faces of the wedge. And an account of all machines is the same.

The effectiveness and use of these consists in this only, that by diminishing the speed we augment the force, and vice versa: From which the general problem is solved in all kinds of suitable mechanical devices : *a given weight is to be moved by a given force*, or some given resistance is to be overcome by a given force. For if machines may be formed thus, so that the speeds of the driving force and of the resistance shall be reciprocally as the forces ; the driving force will sustain the resistance : and it may overcome the same with a greater difference of the speeds. Certainly if the disparity of the speeds shall be so great, so that all resistance may also overcome, which is accustomed to arise both from the slipping and friction of nearby bodies between each other, as well as from the cohesion and in turn of the separation and continued elevation of bodies ; with all that resistance overcome, the excess force will produce an acceleration motion proportional to itself, partially within the parts of the machine, and partially within the resisting body. It is not the intention of this work to treat everything mechanical. I have wished only to show, both how wide and sure the third law of motion shall be. For if the action of the driving force may be estimated from the speed and this force taken jointly ; and similarly the reaction of the resistance may be estimated conjointly from the velocities of the individual parts of this, and from the friction of these, from the cohesion, and from the weight, and from the acceleration arising ; the action and the reaction will always be equal to each other in turn, in every use of instruments. And as far as the action is propagated by the instrument and finally may be impressed on any resisting body, the final determination of this will always be contrary to the determined reaction.

# Newton's *Principia*: Rules of Reasoning in Natural Philosophy

Trans. A. Motte, 1729

## RULE I

*We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances.*

To this purpose the philosophers say that Nature does nothing in vain, and more is in vain when less will serve; for Nature is pleased with simplicity, and affects not the pomp of superfluous causes.

## RULE II

*Therefore to the same natural effects we must, as far as possible, assign the same causes.*

As to respiration in a man and in a beast; the descent of stones in Europe and in America; the light of our culinary fire and of the sun; the reflection of light in the earth, and in the planets.

## RULE III

*The qualities of bodies, which admit neither intension nor remission of degrees, and which are found to belong to all bodies within the reach of our experiments, are to be esteemed the universal qualities of all bodies whatsoever.*

For since the qualities of bodies are only known to us by experiments, we are to hold for universal all such as universally agree with experiments; and such as are not liable to diminution can never be quite taken away. We are certainly not to relinquish the evidence of experiments for the sake of dreams and vain fictions of our own devising; nor are we to recede from the analogy of Nature, which uses to be simple, and always consonant to itself. We no other way know the extension of bodies than by our senses, nor do these reach it in all bodies; but because we perceive extension in all that are sensible, therefore we ascribe it universally to all others also. That abundance of bodies are hard, we learn by experience; and because the hardness of the whole arises from the hardness of the parts, we therefore justly infer the hardness of the undivided particles not only of the bodies we feel but of all others. That all bodies are impenetrable, we gather not from reason, but from sensation. The bodies which we handle we find impenetrable, and thence conclude impenetrability to be an universal property of all bodies whatsoever. That all bodies are moveable, and endowed with certain powers (which we call the *vires inertiae*) of persevering in their motion, or in their rest we only infer from the like properties observed in the bodies which we have

seen. The extension, hardness, impenetrability, mobility, and *vis inertiae* of the whole, result from the extension hardness, impenetrability, mobility, and *vires inertiae* of the parts; and thence we conclude the least particles of all bodies to be also all extended, and hard and impenetrable, and moveable, and endowed with their proper *vires inertiae*. And this is the foundation of all philosophy. Moreover, that the divided but contiguous particles of bodies may be separated from one another, is matter of observation; and, in the particles that remain undivided, our minds are able to distinguish yet lesser parts, as is mathematically demonstrated. But whether the parts so distinguished, and not yet divided, may, by the powers of Nature, be actually divided and separated from one another, we cannot certainly determine. Yet, had we the proof of but one experiment that any undivided particle, in breaking a hard and solid body, offered a division, we might by virtue of this rule conclude that the undivided as well as the divided particles may be divided and actually separated to infinity.

Lastly, if it universally appears, by experiments and astronomical observations, that all bodies about the earth gravitate towards the earth, and that in proportion to the quantity of matter which they severally contain, that the moon likewise, according to the quantity of its matter, gravitates towards the earth; that, on the other hand, our sea gravitates towards the moon; and all the planets mutually one towards another; and the comets in like manner towards the sun; we must, in consequence of this rule, universally allow that all bodies whatsoever are endowed with a principle of mutual gravitation. For the argument from the appearances concludes with more force for the universal gravitation of all bodies than that for their impenetrability; of which, among those in the celestial regions, we have no experiments, nor any manner of observation. Not that I affirm gravity to be essential to bodies: by their *vis insita* I mean nothing but their *vis inertiae*. This is immutable. Their gravity is diminished as they recede from the earth.

#### RULE IV

*In experimental philosophy we are to look upon propositions collected by general induction from phænomena as accurately or very nearly true, notwithstanding any contrary hypotheses that may be imagined, till such time as other phænomena occur, by which they may either be made more accurate, or liable to exceptions.*

This rule we must follow, that the argument of induction may not be evaded by hypotheses.

# **ON A GENERAL METHOD IN DYNAMICS**

**By**

**William Rowan Hamilton**

(Philosophical Transactions of the Royal Society, part II for 1834, pp. 247–308.)

Edited by David R. Wilkins

2000

## NOTE ON THE TEXT

This edition is based on the original publication in the *Philosophical Transactions of the Royal Society*, part II for 1834.

The following errors in the original published text have been corrected:

a term  $w^{(n)}$  in the last summand on the right hand side of equation (S<sup>5</sup>.) has been corrected to  $w^{(n-1)}$ ;

a minus sign (–) missing from equation (K<sup>6</sup>.) has been inserted.

The paper *On a General Method in Dynamics* has also been republished in *The Mathematical Papers of Sir William Rowan Hamilton, Volume II: Dynamics*, edited for the Royal Irish Academy by A. W. Conway and A. J. McConnell, and published by Cambridge University Press in 1940.

David R. Wilkins  
Dublin, February 2000

*On a General Method in Dynamics; by which the Study of the Motions of all free Systems of attracting or repelling Points is reduced to the Search and Differentiation of one central Relation, or characteristic Function.* By WILLIAM ROWAN HAMILTON, Member of several scientific Societies in the British Dominions, and of the American Academy of Arts and Sciences, Andrews' Professor of Astronomy in the University of Dublin, and Royal Astronomer of Ireland. Communicated by Captain BEAUFORT, R.N. F.R.S.

Received April 1,—Read April 10, 1834.

[*Philosophical Transactions of the Royal Society*, part II for 1834, pp. 247–308.]

*Introductory Remarks.*

The theoretical development of the laws of motion of bodies is a problem of such interest and importance, that it has engaged the attention of all the most eminent mathematicians, since the invention of dynamics as a mathematical science by GALILEO, and especially since the wonderful extension which was given to that science by NEWTON. Among the successors of those illustrious men, LAGRANGE has perhaps done more than any other analyst, to give extent and harmony to such deductive researches, by showing that the most varied consequences respecting the motions of systems of bodies may be derived from one radical formula; the beauty of the method so suiting the dignity of the results, as to make of his great work a kind of scientific poem. But the science of force, or of power acting by law in space and time, has undergone already another revolution, and has become already more dynamic, by having almost dismissed the conceptions of solidity and cohesion, and those other material ties, or geometrically imaginably conditions, which LAGRANGE so happily reasoned on, and by tending more and more to resolve all connexions and actions of bodies into attractions and repulsions of points: and while the science is advancing thus in one direction by the improvement of physical views, it may advance in another direction also by the invention of mathematical methods. And the method proposed in the present essay, for the deductive study of the motions of attracting or repelling systems, will perhaps be received with indulgence, as an attempt to assist in carrying forward so high an inquiry.

In the methods commonly employed, the determination of the motion of a free point in space, under the influence of accelerating forces, depends on the integration of three equations in ordinary differentials of the second order; and the determination of the motions of a system of free points, attracting or repelling one another, depends on the integration of a system of such equations, in number threefold the number of the attracting or repelling points, unless we previously diminish by unity this latter number, by considering only relative motions. Thus, in the solar system, when we consider only the mutual attractions of the sun and the ten known planets, the determination of the motions of the latter about the former is reduced, by the usual methods, to the integration of a system of thirty ordinary differential equations of the

second order, between the coordinates and the time; or, by a transformation of LAGRANGE, to the integration of a system of sixty ordinary differential equations of the first order, between the time and the elliptic elements: by which integrations, the thirty varying coordinates, or the sixty varying elements, are to be found as functions of the time. In the method of the present essay, this problem is reduced to the search and differentiation of a single function, which satisfies two partial differential equations of the first order and of the second degree: and every other dynamical problem, respecting the motions of any system, however numerous, of attracting or repelling points, (even if we suppose those points restricted by any conditions of connexion consistent with the law of living force,) is reduced, in like manner, to the study of one central function, of which the form marks out and characterizes the properties of the moving system, and is to be determined by a pair of partial differential equations of the first order, combined with some simple considerations. The difficulty is therefore at least transferred from the integration of many equations of one class to the integration of two of another: and even if it should be thought that no practical facility is gained, yet an intellectual pleasure may result from the reduction of the most complex and, probably, of all researches respecting the forces and motions of body, to the study of one characteristic function,\* the unfolding of one central relation.

The present essay does not pretend to treat fully of this extensive subject,—a task which may require the labours of many years and many minds; but only to suggest the thought and propose the path to others. Although, therefore, the method may be used in the most varied dynamical researches, it is at present only applied to the orbits and perturbations of a system with any laws of attraction or repulsion, and with one predominant mass or centre of predominant energy; and only so far, even in this one research, as appears sufficient to make the principle itself understood. It may be mentioned here, that this dynamical principle is only another form of that idea which has already been applied to optics in the *Theory of systems of rays*, and that an intention of applying it to the motion of systems of bodies was announced† at the publication of that theory. And besides the idea itself, the manner of calculation also, which has been thus exemplified in the sciences of optics and dynamics, seems not confined to those two sciences, but capable of other applications; and the peculiar combination which it involves, of the principles of variations with those of partial differentials, for the determination and use of an important class of integrals, may constitute, when it shall be matured by the future labours of mathematicians, a separate branch of analysis.

WILLIAM R. HAMILTON.

*Observatory, Dublin, March 1834.*

---

\* LAGRANGE and, after him, LAPLACE and others, have employed a single function to express the different forces of a system, and so to form in an elegant manner the differential equations of its motion. By this conception, great simplicity has been given to the statement of the problem of dynamics; but the solution of that problem, or the expression of the motions themselves, and of their integrals, depends on a very different and hitherto unimagined function, as it is the purpose of this essay to show.

† Transactions of the Royal Irish Academy, Vol. xv, page 80. A notice of this dynamical principle was also lately given in an article “On a general Method of expressing the Paths of Light and of the Planets,” published in the Dublin University Review for October 1833.

*Integration of the Equations of Motion of a System, characteristic Function of such Motion, and Law of varying Action.*

1. The known differential equations of motion of a system of free points, repelling or attracting one another according to any functions of their distances, and not disturbed by any foreign force, may be comprised in the following formula:

$$\Sigma .m(x'' \delta x + y'' \delta y + z'' \delta z) = \delta U. \quad (1.)$$

In this formula the sign of summation  $\Sigma$  extends to all the points of the system;  $m$  is, for any one such point, the constant called its mass;  $x'', y'', z''$ , are its component accelerations, or the second differential coefficients of its rectangular coordinates  $x, y, z$ , taken with respect to the time;  $\delta x, \delta y, \delta z$ , are any arbitrary infinitesimal displacements which the point can be imagined to receive in the same three rectangular directions; and  $\delta U$  is the infinitesimal variation corresponding, of a function  $U$  of the masses and mutual distances of the several points of the system, of which the form depends on the laws of their mutual actions, by the equation

$$U = \Sigma .mm,f(r), \quad (2.)$$

$r$  being the distance between any two points  $m, m_1$ , and the function  $f(r)$  being such that the derivative or differential coefficient  $f'(r)$  expresses the law of their repulsion, being negative in the case of attraction. The function which has been here called  $U$  may be named the *force-function* of a system: it is of great utility in theoretical mechanics, into which it was introduced by LAGRANGE, and it furnishes the following elegant forms for the differential equations of motion, included in the formula (1.):

$$\left. \begin{aligned} m_1 x_1'' &= \frac{\delta U}{\delta x_1}; & m_2 x_2'' &= \frac{\delta U}{\delta x_2}; & \dots & m_n x_n'' = \frac{\delta U}{\delta x_n}; \\ m_1 y_1'' &= \frac{\delta U}{\delta y_1}; & m_2 y_2'' &= \frac{\delta U}{\delta y_2}; & \dots & m_n y_n'' = \frac{\delta U}{\delta y_n}; \\ m_1 z_1'' &= \frac{\delta U}{\delta z_1}; & m_2 z_2'' &= \frac{\delta U}{\delta z_2}; & \dots & m_n z_n'' = \frac{\delta U}{\delta z_n}; \end{aligned} \right\} \quad (3.)$$

the second members of these equations being the partial differential coefficients of the first order of the function  $U$ . But notwithstanding the elegance and simplicity of this known manner of stating the principal problem of dynamics, the difficulty of solving that problem, or even of expressing its solution, has hitherto appeared insuperable; so that only seven intermediate integrals, or integrals of the first order, with as many arbitrary constants, have hitherto been found for these general equations of motion of a system of  $n$  points, instead of  $3n$  intermediate and  $3n$  final integrals, involving ultimately  $6n$  constants; nor has any integral been found which does not need to be integrated again. No general solution has been obtained assigning (as a complete solution ought to do)  $3n$  relations between the  $n$  masses  $m_1, m_2, \dots, m_n$ , the  $3n$  varying coordinates  $x_1, y_1, z_1, \dots, x_n, y_n, z_n$ , the varying time  $t$ , and the  $6n$  initial data of the problem, namely, the initial coordinates  $a_1, b_1, c_1, \dots, a_n, b_n, c_n$ , and their initial rates of increase  $a'_1, b'_1, c'_1, \dots, a'_n, b'_n, c'_n$ ; the quantities called here initial being those

which correspond to the arbitrary origin of time. It is, however, possible (as we shall see) to express these long-sought relations by the partial differential coefficients of a new central or radical function, to the search and employment of which the difficulty of mathematical dynamics becomes henceforth reduced.

2. If we put for abridgement

$$T = \frac{1}{2} \Sigma .m(x'^2 + y'^2 + z'^2), \quad (4.)$$

so that  $2T$  denotes, as in the Mécanique Analytique, the whole living force of the system; ( $x'$ ,  $y'$ ,  $z'$ , being here, according to the analogy of our foregoing notation, the rectangular components of velocity of the point  $m$ , or the first differential coefficients of its coordinates taken with respect to the time;) an easy and well known combination of the differential equations of motion, obtained by changing in the formula (1.) the variations to the differentials of the coordinates, may be expressed in the following manner,

$$dT = dU, \quad (5.)$$

and gives, by integration, the celebrated law of living force, under the form

$$T = U + H. \quad (6.)$$

In this expression, which is one of the seven known integrals already mentioned, the quantity  $H$  is independent of the time, and does not alter in the passage of the points of the system from one set of positions to another. We have, for example, an initial equation of the same form, corresponding to the origin of time, which may be written thus,

$$T_0 = U_0 + H. \quad (7.)$$

The quantity  $H$  may, however, receive any arbitrary increment whatever, when we pass in thought from a system moving in one way, to the same system moving in another, with the same dynamical relations between the accelerations and positions of its points, but with different initial data; but the increment of  $H$ , thus obtained, is evidently connected with the analogous increments of the functions  $T$  and  $U$ , by the relation

$$\Delta T = \Delta U + \Delta H, \quad (8.)$$

which, for the case of infinitesimal variations, may be conveniently be written thus,

$$\delta T = \delta U + \delta H; \quad (9.)$$

and this last relation, when multiplied by  $dt$ , and integrated, conducts to an important result. For it thus becomes, by (4.) and (1.),

$$\int \Sigma .m(dx \cdot \delta x' + dy \cdot \delta y' + dz \cdot \delta z') = \int \Sigma .m(dx' \cdot \delta x + dy' \cdot \delta y + dz' \cdot \delta z) + \int \delta H \cdot dt, \quad (10.)$$

that is, by the principles of the calculus of variations,

$$\delta V = \Sigma .m(x' \delta x + y' \delta y + z' \delta z) - \Sigma .m(a' \delta a + b' \delta b + c' \delta c) + t \delta H, \quad (\text{A.})$$

if we denote by  $V$  the integral

$$V = \int \Sigma .m(x' dx + y' dy + z' dz) = \int_0^t 2T dt, \quad (\text{B.})$$

namely, the accumulated living force, called often the action of the system, from its initial to its final position.

If, then, we consider (as it is easy to see that we may) the action  $V$  as a function of the initial and final coordinates, and of the quantity  $H$ , we shall have, by (A.), the following groups of equations; first, the group,

$$\left. \begin{aligned} \frac{\delta V}{\delta x_1} &= m_1 x'_1; & \frac{\delta V}{\delta x_2} &= m_2 x'_2; & \dots & \frac{\delta V}{\delta x_n} &= m_n x'_n; \\ \frac{\delta V}{\delta y_1} &= m_1 y'_1; & \frac{\delta V}{\delta y_2} &= m_2 y'_2; & \dots & \frac{\delta V}{\delta y_n} &= m_n y'_n; \\ \frac{\delta V}{\delta z_1} &= m_1 z'_1; & \frac{\delta V}{\delta z_2} &= m_2 z'_2; & \dots & \frac{\delta V}{\delta z_n} &= m_n z'_n. \end{aligned} \right\} \quad (\text{C.})$$

Secondly, the group,

$$\left. \begin{aligned} \frac{\delta V}{\delta a_1} &= -m_1 a'_1; & \frac{\delta V}{\delta a_2} &= -m_2 a'_2; & \dots & \frac{\delta V}{\delta a_n} &= -m_n a'_n; \\ \frac{\delta V}{\delta b_1} &= -m_1 b'_1; & \frac{\delta V}{\delta b_2} &= -m_2 b'_2; & \dots & \frac{\delta V}{\delta b_n} &= -m_n b'_n; \\ \frac{\delta V}{\delta c_1} &= -m_1 c'_1; & \frac{\delta V}{\delta c_2} &= -m_2 c'_2; & \dots & \frac{\delta V}{\delta c_n} &= -m_n c'_n; \end{aligned} \right\} \quad (\text{D.})$$

and finally, the equation,

$$\frac{\delta V}{\delta H} = t. \quad (\text{E.})$$

So that if this function  $V$  were known, it would only remain to eliminate  $H$  between the  $3n+1$  equations (C.) and (E.), in order to obtain all the  $3n$  intermediate integrals, or between (D.) and (E.) to obtain all the  $3n$  final integrals of the differential equations of motion; that is, ultimately, to obtain the  $3n$  sought relations between the  $3n$  varying coordinates and the time, involving also the masses and the  $6n$  initial data above mentioned; the discovery of which relations would be (as we have said) the general solution of the general problem of dynamics. We have, therefore, at least reduced that general problem to the search and differentiation of a single function  $V$ , which we shall call on this account the CHARACTERISTIC FUNCTION of motion of a system; and the equation (A.), expressing the fundamental law of its variation, we shall call the *equation of the characteristic function*, or the LAW OF VARYING ACTION.

3. To show more clearly that the action or accumulated living force of a system, or in other words, the integral of the product of the living force by the element of the time, may be regarded as a function of the  $6n + 1$  quantities already mentioned, namely, of the initial and final coordinates, and of the quantity  $H$ , we may observe, that whatever depends on the manner and time of motion of the system may be considered as such a function; because the initial form of the law of living force, when combined with the  $3n$  known or unknown relations between the time, the initial data, and the varying coordinates, will always furnish  $3n + 1$  relations, known or unknown, to connect the time and the initial components of velocities with the initial and final coordinates, and with  $H$ . Yet from not having formed the conception of the action as a *function* of this kind, the consequences that have been here deduced from the formula (A.) for the variation of that definite integral appear to have escaped the notice of LAGRANGE, and of the other illustrious analysts who have written on theoretical mechanics; although they were in possession of a formula for the variation of this integral not greatly differing from ours. For although LAGRANGE and others, in treating of the motion of a system, have shown that the variation of this definite integral vanishes when the extreme coordinates and the constant  $H$  are given, they appear to have deduced from this result only the well known law of *least action*; namely, that if the points or bodies of a system be imagined to move from a given set of initial to a given set of final positions, not as they do nor even as they could move consistently with the general dynamical laws or differential equations of motion, but so as not to violate any supposed geometrical connexions, nor that one dynamical relation between velocities and configurations which constitutes the law of living force; and if, besides, this geometrically imaginable, but dynamically impossible motion, be made to differ infinitely *little* from the actual manner of motion of the system, between the given extreme positions; then the varied value of the definite integral called action, or the accumulated living force of the system in the motion thus imagined, will differ infinitely *less* from the actual value of that integral. But when this well known law of least, or as it might be better called, of *stationary action*, is applied to the determination of the actual motion of the system, it serves only to form, by the rules of the calculus of variations, the differential equations of motion of the second order, which can always be otherwise found. It seems, therefore, to be with reason that LAGRANGE, LAPLACE, and POISSON have spoken lightly of the utility of this principle in the present state of dynamics. A different estimate, perhaps, will be formed of that other principle which has been introduced in the present paper, under the name of the *law of varying action*, in which we pass from an actual motion to another motion dynamically possible, by varying the extreme positions of the system, and (in general) the quantity  $H$ , and which serves to express, by means of a single function, not the mere differential equations of motion, but their intermediate and their final integrals.

*Verification of the foregoing Integrals.*

4. A verification, which ought not to be neglected, and at the same time an illustration of this new principle, may be obtained by deducing the known differential equations of motion from our system of intermediate integrals, and by showing the consistence of these again with our final integral system. As preliminary to such verification, it is useful to observe that the final equation (6.) of living force, when combined with the system (C.), takes this new form,

$$\frac{1}{2} \Sigma \cdot \frac{1}{m} \left\{ \left( \frac{\delta V}{\delta x} \right)^2 + \left( \frac{\delta V}{\delta y} \right)^2 + \left( \frac{\delta V}{\delta z} \right)^2 \right\} = U + H; \quad (\text{F.})$$

and that the initial equation (7.) of living force becomes by (D.)

$$\frac{1}{2} \Sigma \cdot \frac{1}{m} \left\{ \left( \frac{\delta V}{\delta a} \right)^2 + \left( \frac{\delta V}{\delta b} \right)^2 + \left( \frac{\delta V}{\delta c} \right)^2 \right\} = U_0 + H. \quad (\text{G.})$$

These two partial differential equations, initial and final, of the first order and the second degree, must both be identically satisfied by the characteristic function  $V$ : they furnish (as we shall find) the principal means of discovering the form of that function, and are of essential importance in its theory. If the form of this function were known, we might eliminate  $3n-1$  of the  $3n$  initial coordinates between the  $3n$  equations (C.); and although we cannot yet perform the actual process of this elimination, we are entitled to assert that it would remove along with the others the remaining initial coordinate, and would conduct to the equation (6.) of final living force, which might then be transformed into the equation (F.). In like manner we may conclude that all the  $3n$  final coordinates could be eliminated together from the  $3n$  equations (D.), and that the result would be the initial equation (7.) of living force, or the transformed equation (G.). We may therefore consider the law of living force, which assisted us in discovering the properties of our characteristic function  $V$ , as included reciprocally in those properties, and as resulting by elimination, in every particular case, from the systems (C.) and (D.); and in treating of either of these systems, or in conducting any other dynamical investigation by the method of this characteristic function, we are at liberty to employ the partial differential equations (F.) and (G.) which that function must necessarily satisfy.

It will now be easy to deduce, as we proposed, the known equations of motion (3.) of the second order, by differentiation and elimination of constants, from our intermediate integral system (C.), (E.), or even from a part of that system, namely, from the group (C.), when combined with the equation (F.). For we thus obtain

$$\begin{aligned} m_1 x_1'' &= \frac{d}{dt} \frac{\delta V}{\delta x_1} = x'_1 \frac{\delta^2 V}{\delta x_1^2} + x'_2 \frac{\delta^2 V}{\delta x_1 \delta x_2} + \cdots + x'_n \frac{\delta^2 V}{\delta x_1 \delta x_n} \\ &\quad + y'_1 \frac{\delta^2 V}{\delta x_1 \delta y_1} + y'_2 \frac{\delta^2 V}{\delta x_1 \delta y_2} + \cdots + y'_n \frac{\delta^2 V}{\delta x_1 \delta y_n} \\ &\quad + z'_1 \frac{\delta^2 V}{\delta x_1 \delta z_1} + z'_2 \frac{\delta^2 V}{\delta x_1 \delta z_2} + \cdots + z'_n \frac{\delta^2 V}{\delta x_1 \delta z_n} \\ &= \frac{1}{m_1} \frac{\delta V}{\delta x_1} \frac{\delta^2 V}{\delta x_1^2} + \frac{1}{m_2} \frac{\delta V}{\delta x_2} \frac{\delta^2 V}{\delta x_1 \delta x_2} + \cdots + \frac{1}{m_n} \frac{\delta V}{\delta x_n} \frac{\delta^2 V}{\delta x_1 \delta x_n} \\ &\quad + \frac{1}{m_1} \frac{\delta V}{\delta y_1} \frac{\delta^2 V}{\delta x_1 \delta y_1} + \frac{1}{m_2} \frac{\delta V}{\delta y_2} \frac{\delta^2 V}{\delta x_1 \delta y_2} + \cdots + \frac{1}{m_n} \frac{\delta V}{\delta y_n} \frac{\delta^2 V}{\delta x_1 \delta y_n} \\ &\quad + \frac{1}{m_1} \frac{\delta V}{\delta z_1} \frac{\delta^2 V}{\delta x_1 \delta z_1} + \frac{1}{m_2} \frac{\delta V}{\delta z_2} \frac{\delta^2 V}{\delta x_1 \delta z_2} + \cdots + \frac{1}{m_n} \frac{\delta V}{\delta z_n} \frac{\delta^2 V}{\delta x_1 \delta z_n} \\ &= \frac{\delta}{\delta x_1} \Sigma \cdot \frac{1}{2m} \left\{ \left( \frac{\delta V}{\delta x} \right)^2 + \left( \frac{\delta V}{\delta y} \right)^2 + \left( \frac{\delta V}{\delta z} \right)^2 \right\} = \frac{\delta}{\delta x_1} (U + H); \end{aligned} \quad \left. \right\} \quad (11.)$$

that is, we obtain

$$m_1 x_1'' = \frac{\delta U}{\delta x_1}. \quad (12.)$$

And in like manner we might deduce, by differentiation, from the integrals (C.) and from (F.) all the other known differential equations of motion, of the second order, contained in the set marked (3.); or, more concisely, we may deduce at once the formula (1.), which contains all those known equations, by observing that the intermediate integrals (C.), when combined with the relation (F.), give

$$\left. \begin{aligned} & \Sigma .m(x'' \delta x + y'' \delta y + z'' \delta z) \\ &= \Sigma \left( \frac{d}{dt} \frac{\delta V}{\delta x} \cdot \delta x + \frac{d}{dt} \frac{\delta V}{\delta y} \cdot \delta y + \frac{d}{dt} \frac{\delta V}{\delta z} \cdot \delta z \right) \\ &= \Sigma \cdot \frac{1}{m} \left( \frac{\delta V}{\delta x} \frac{\delta}{\delta x} + \frac{\delta V}{\delta y} \frac{\delta}{\delta y} + \frac{\delta V}{\delta z} \frac{\delta}{\delta z} \right) \Sigma \left( \frac{\delta V}{\delta x} \delta x + \frac{\delta V}{\delta y} \delta y + \frac{\delta V}{\delta z} \delta z \right) \\ &= \Sigma \left( \delta x \frac{\delta}{\delta x} + \delta y \frac{\delta}{\delta y} + \delta z \frac{\delta}{\delta z} \right) \Sigma \cdot \frac{1}{2m} \left\{ \left( \frac{\delta V}{\delta x} \right)^2 + \left( \frac{\delta V}{\delta y} \right)^2 + \left( \frac{\delta V}{\delta z} \right)^2 \right\} \\ &= \Sigma \left( \delta x \frac{\delta}{\delta x} + \delta y \frac{\delta}{\delta y} + \delta z \frac{\delta}{\delta z} \right) (U + H) \\ &= \delta U. \end{aligned} \right\} \quad (13.)$$

5. Again, we were to show that our intermediate integral system, composed of the equations (C.) and (E.), with the  $3n$  arbitrary constants  $a_1, b_1, c_1, \dots, a_n, b_n, c_n$ , (and involving also the auxiliary constant  $H$ ), is consistent with our final integral system of equations (D.) and (E.), which contain  $3n$  other arbitrary constants, namely  $a'_1, b'_1, c'_1, \dots, a'_n, b'_n, c'_n$ . The immediate differentials of the equations (C.), (D.), (E.), taken with respect to the time, are, for the first group,

$$\left. \begin{aligned} \frac{d}{dt} \frac{\delta V}{\delta x_1} &= m_1 x_1''; & \frac{d}{dt} \frac{\delta V}{\delta x_2} &= m_2 x_2''; & \dots & \frac{d}{dt} \frac{\delta V}{\delta x_n} &= m_n x_n''; \\ \frac{d}{dt} \frac{\delta V}{\delta y_1} &= m_1 y_1''; & \frac{d}{dt} \frac{\delta V}{\delta y_2} &= m_2 y_2''; & \dots & \frac{d}{dt} \frac{\delta V}{\delta y_n} &= m_n y_n''; \\ \frac{d}{dt} \frac{\delta V}{\delta z_1} &= m_1 z_1''; & \frac{d}{dt} \frac{\delta V}{\delta z_2} &= m_2 z_2''; & \dots & \frac{d}{dt} \frac{\delta V}{\delta z_n} &= m_n z_n''; \end{aligned} \right\} \quad (H.)$$

for the second group,

$$\left. \begin{aligned} \frac{d}{dt} \frac{\delta V}{\delta a_1} &= 0; & \frac{d}{dt} \frac{\delta V}{\delta a_2} &= 0; & \dots & \frac{d}{dt} \frac{\delta V}{\delta a_n} &= 0; \\ \frac{d}{dt} \frac{\delta V}{\delta b_1} &= 0; & \frac{d}{dt} \frac{\delta V}{\delta b_2} &= 0; & \dots & \frac{d}{dt} \frac{\delta V}{\delta b_n} &= 0; \\ \frac{d}{dt} \frac{\delta V}{\delta c_1} &= 0; & \frac{d}{dt} \frac{\delta V}{\delta c_2} &= 0; & \dots & \frac{d}{dt} \frac{\delta V}{\delta c_n} &= 0; \end{aligned} \right\} \quad (I.)$$

and finally, for the last equation,

$$\frac{d}{dt} \frac{\delta V}{\delta H} = 1. \quad (\text{K.})$$

By combining the equations (C.) with their differentials (H.), and with the relation (F.), we deduced, in the foregoing number, the known equations of motion (3.); and we are now to show the consistence of the same intermediate integrals (C.) with the group of differentials (I.) which have been obtained from the final integrals.

The first equation of the group (I.) may be developed thus:

$$\left. \begin{aligned} 0 &= x'_1 \frac{\delta^2 V}{\delta a_1 \delta x_1} + x'_2 \frac{\delta^2 V}{\delta a_1 \delta x_2} + \cdots + x'_n \frac{\delta^2 V}{\delta a_1 \delta x_n} \\ &\quad + y'_1 \frac{\delta^2 V}{\delta a_1 \delta y_1} + y'_2 \frac{\delta^2 V}{\delta a_1 \delta y_2} + \cdots + y'_n \frac{\delta^2 V}{\delta a_1 \delta y_n} \\ &\quad + z'_1 \frac{\delta^2 V}{\delta a_1 \delta z_1} + z'_2 \frac{\delta^2 V}{\delta a_1 \delta z_2} + \cdots + z'_n \frac{\delta^2 V}{\delta a_1 \delta z_n} \end{aligned} \right\} \quad (14.)$$

and the others may be similarly developed. In order, therefore, to show that they are satisfied by the group (C.), it is sufficient to prove that the following equations are true,

$$\left. \begin{aligned} 0 &= \frac{\delta}{\delta a_i} \Sigma \cdot \frac{1}{2m} \left\{ \left( \frac{\delta V}{\delta x} \right)^2 + \left( \frac{\delta V}{\delta y} \right)^2 + \left( \frac{\delta V}{\delta z} \right)^2 \right\}, \\ 0 &= \frac{\delta}{\delta b_i} \Sigma \cdot \frac{1}{2m} \left\{ \left( \frac{\delta V}{\delta x} \right)^2 + \left( \frac{\delta V}{\delta y} \right)^2 + \left( \frac{\delta V}{\delta z} \right)^2 \right\}, \\ 0 &= \frac{\delta}{\delta c_i} \Sigma \cdot \frac{1}{2m} \left\{ \left( \frac{\delta V}{\delta x} \right)^2 + \left( \frac{\delta V}{\delta y} \right)^2 + \left( \frac{\delta V}{\delta z} \right)^2 \right\}, \end{aligned} \right\} \quad (\text{L.})$$

the integer  $i$  receiving any value from 1 to  $n$  inclusive; which may be shown at once, and the required verification thereby be obtained, if we merely take the variation of the relation (F.) with respect to the initial coordinates, as in the former verification we took its variation with respect to the final coordinates, and so obtained results which agreed with the known equations of motion, and which may be thus collected,

$$\left. \begin{aligned} \frac{\delta}{\delta x_i} \Sigma \cdot \frac{1}{2m} \left\{ \left( \frac{\delta V}{\delta x} \right)^2 + \left( \frac{\delta V}{\delta y} \right)^2 + \left( \frac{\delta V}{\delta z} \right)^2 \right\} &= \frac{\delta U}{\delta x_i}; \\ \frac{\delta}{\delta y_i} \Sigma \cdot \frac{1}{2m} \left\{ \left( \frac{\delta V}{\delta x} \right)^2 + \left( \frac{\delta V}{\delta y} \right)^2 + \left( \frac{\delta V}{\delta z} \right)^2 \right\} &= \frac{\delta U}{\delta y_i}; \\ \frac{\delta}{\delta z_i} \Sigma \cdot \frac{1}{2m} \left\{ \left( \frac{\delta V}{\delta x} \right)^2 + \left( \frac{\delta V}{\delta y} \right)^2 + \left( \frac{\delta V}{\delta z} \right)^2 \right\} &= \frac{\delta U}{\delta z_i}. \end{aligned} \right\} \quad (\text{M.})$$

The same relation (F.), by being varied with respect to the quantity  $H$ , conducts to the expression

$$\frac{\delta}{\delta H} \Sigma \cdot \frac{1}{2m} \left\{ \left( \frac{\delta V}{\delta x} \right)^2 + \left( \frac{\delta V}{\delta y} \right)^2 + \left( \frac{\delta V}{\delta z} \right)^2 \right\} = 1; \quad (\text{N.})$$

and this, when developed, agrees with the equation (K.), which is a new verification of the consistence of our foregoing results. Nor would it have been much more difficult, by the help of the foregoing principles, to have integrated directly our integrals of the first order, and so to have deduced in a different way our final integral system.

6. It may be considered as still another verification of our own general integral equations, to show that they include not only the known law of living force, or the integral expressing that law, but also the six other known integrals of the first order, which contain the law of motion of the centre of gravity, and the law of description of areas. For this purpose, it is only necessary to observe that it evidently follows from the conception of our characteristic function  $V$ , that the function depends on the initial and final positions of the attracting or repelling points of a system, not as referred to any foreign standard, but only as compared with one another; and therefore that this function will not vary, if without making any real change in either initial or final configuration, or in the relation of these to each other, we alter at once all the initial and all the final positions of the points of the system, by any common motion, whether of translation or of rotation. Now by considering these coordinate translations, we obtain the three following partial differential equations of the first order, which the function  $V$  must satisfy,

$$\left. \begin{aligned} \Sigma \frac{\delta V}{\delta x} + \Sigma \frac{\delta V}{\delta a} &= 0; \\ \Sigma \frac{\delta V}{\delta y} + \Sigma \frac{\delta V}{\delta b} &= 0; \\ \Sigma \frac{\delta V}{\delta z} + \Sigma \frac{\delta V}{\delta c} &= 0; \end{aligned} \right\} \quad (\text{O.})$$

and by considering three coordinate rotations, we obtain these three other relations between the partial differential coefficients of the same order of the same characteristic function,

$$\left. \begin{aligned} \Sigma \left( x \frac{\delta V}{\delta y} - y \frac{\delta V}{\delta x} \right) + \Sigma \left( a \frac{\delta V}{\delta b} - b \frac{\delta V}{\delta a} \right) &= 0; \\ \Sigma \left( y \frac{\delta V}{\delta z} - z \frac{\delta V}{\delta y} \right) + \Sigma \left( b \frac{\delta V}{\delta c} - c \frac{\delta V}{\delta b} \right) &= 0; \\ \Sigma \left( z \frac{\delta V}{\delta x} - x \frac{\delta V}{\delta z} \right) + \Sigma \left( c \frac{\delta V}{\delta a} - a \frac{\delta V}{\delta c} \right) &= 0; \end{aligned} \right\} \quad (\text{P.})$$

and if we change the final coefficients of  $V$  to the final components of momentum, and the initial coefficients to the initial components taken negatively, according to the dynamical properties of this function expressed by the integrals (C.) and (D.), we shall change these partial differential equations (O.) (P.), to the following,

$$\Sigma .mx' = \Sigma .ma'; \quad \Sigma .my' = \Sigma .mb'; \quad \Sigma .mz' = \Sigma .mc'; \quad (15.)$$

and

$$\left. \begin{aligned} \Sigma .m(xy' - yx') &= \Sigma .m(ab' - ba'); \\ \Sigma .m(yz' - zy') &= \Sigma .m(bc' - cb'); \\ \Sigma .m(zx' - xz') &= \Sigma .m(ca' - ac'). \end{aligned} \right\} \quad (16.)$$

In this manner, therefore, we can deduce from the properties of our characteristic function the six other known integrals above mentioned, in addition to that seventh which contains the law of living force, and which assisted in the discovery of our method.

*Introduction of relative or polar Coordinates, or other marks of position of a System.*

7. The property of our characteristic function, by which it depends only on the internal or mutual relations between the positions initial and final of the points of an attracting or repelling system, suggests an advantage in employing internal or relative coordinates; and from the analogy of other applications of algebraical methods to researches of a geometrical kind, it may be expected that polar and other marks of position will also often be found useful. Supposing, therefore, that the  $3n$  final coordinates  $x_1 y_1 z_1 \dots x_n y_n z_n$  have been expressed as functions of  $3n$  other variables  $\eta_1 \eta_2 \dots \eta_{3n}$ , and that the  $3n$  initial coordinates have in like manner been expressed as functions of  $3n$  similar quantities, which we shall call  $e_1 e_2 \dots e_{3n}$ , we shall proceed to assign a general method for introducing these new marks of position into the expressions of our fundamental relations.

For this purpose we have only to transform the law of varying action, or the fundamental formula (A.), by transforming the two sums,

$$\Sigma .m(x' \delta x + y' \delta y + z' \delta z), \quad \text{and} \quad \Sigma .m(a' \delta a + b' \delta b + c' \delta c),$$

which it involves, and which are respectively equivalent to the following more developed expressions,

$$\left. \begin{aligned} \Sigma .m(x' \delta x + y' \delta y + z' \delta z) &= m_1(x'_1 \delta x_1 + y'_1 \delta y_1 + z'_1 \delta z_1) \\ &\quad + m_2(x'_2 \delta x_2 + y'_2 \delta y_2 + z'_2 \delta z_2) \\ &\quad + \&c. + m_n(x'_n \delta x_n + y'_n \delta y_n + z'_n \delta z_n); \end{aligned} \right\} \quad (17.)$$

$$\left. \begin{aligned} \Sigma .m(a' \delta a + b' \delta b + c' \delta c) &= m_1(a'_1 \delta a_1 + b'_1 \delta b_1 + c'_1 \delta c_1) \\ &\quad + m_2(a'_2 \delta a_2 + b'_2 \delta b_2 + c'_2 \delta c_2) \\ &\quad + \&c. + m_n(a'_n \delta a_n + b'_n \delta b_n + c'_n \delta c_n). \end{aligned} \right\} \quad (18.)$$

Now  $x_i$  being by supposition a function of the  $3n$  new marks of position  $\eta_1 \dots \eta_{3n}$ , its variation  $\delta x_i$ , and its differential coefficient  $x'_i$  may be thus expressed:

$$\delta x_i = \frac{\delta x_i}{\delta \eta_1} \delta \eta_1 + \frac{\delta x_i}{\delta \eta_2} \delta \eta_2 + \dots + \frac{\delta x_i}{\delta \eta_{3n}} \delta \eta_{3n}; \quad (19.)$$

$$x'_i = \frac{\delta x_i}{\delta \eta_1} \eta'_1 + \frac{\delta x_i}{\delta \eta_2} \eta'_2 + \dots + \frac{\delta x_i}{\delta \eta_{3n}} \eta'_{3n}; \quad (20.)$$

and similarly for  $y_i$  and  $z_i$ . If, then, we consider  $x'_i$  as a function, by (20.), of  $\eta'_1 \dots \eta'_{3n}$ , involving also in general  $\eta_1 \dots \eta_{3n}$ , and if we take its partial differential coefficients of the first order with respect to  $\eta'_1 \dots \eta'_{3n}$ , we find the relations,

$$\frac{\delta x'_i}{\delta \eta'_1} = \frac{\delta x_i}{\delta \eta_1}; \quad \frac{\delta x'_i}{\delta \eta'_2} = \frac{\delta x_i}{\delta \eta_2}; \quad \dots \quad \frac{\delta x'_i}{\delta \eta'_{3n}} = \frac{\delta x_i}{\delta \eta_{3n}}; \quad (21.)$$

and therefore we obtain these new expressions for the variations  $\delta x_i$ ,  $\delta y_i$ ,  $\delta z_i$ ,

$$\left. \begin{aligned} \delta x_i &= \frac{\delta x'_i}{\delta \eta'_1} \delta \eta_1 + \frac{\delta x'_i}{\delta \eta'_2} \delta \eta_2 + \cdots + \frac{\delta x'_i}{\delta \eta'_{3n}} \delta \eta_{3n}, \\ \delta y_i &= \frac{\delta y'_i}{\delta \eta'_1} \delta \eta_1 + \frac{\delta y'_i}{\delta \eta'_2} \delta \eta_2 + \cdots + \frac{\delta y'_i}{\delta \eta'_{3n}} \delta \eta_{3n}, \\ \delta z_i &= \frac{\delta z'_i}{\delta \eta'_1} \delta \eta_1 + \frac{\delta z'_i}{\delta \eta'_2} \delta \eta_2 + \cdots + \frac{\delta z'_i}{\delta \eta'_{3n}} \delta \eta_{3n}. \end{aligned} \right\} \quad (22.)$$

Substituting these expressions (22.) for the variations in the sum (17.), we easily transform it into the following,

$$\left. \begin{aligned} \Sigma .m(x' \delta x + y' \delta y + z' \delta z) &= \Sigma .m \left( x' \frac{\delta x'}{\delta \eta'_1} + y' \frac{\delta y'}{\delta \eta'_1} + z' \frac{\delta z'}{\delta \eta'_1} \right) \cdot \delta \eta_1 \\ &\quad + \Sigma .m \left( x' \frac{\delta x'}{\delta \eta'_2} + y' \frac{\delta y'}{\delta \eta'_2} + z' \frac{\delta z'}{\delta \eta'_2} \right) \cdot \delta \eta_2 \\ &\quad + \&c. + \Sigma .m \left( x' \frac{\delta x'}{\delta \eta'_{3n}} + y' \frac{\delta y'}{\delta \eta'_{3n}} + z' \frac{\delta z'}{\delta \eta'_{3n}} \right) \cdot \delta \eta_{3n} \\ &= \frac{\delta T}{\delta \eta'_1} \delta \eta_1 + \frac{\delta T}{\delta \eta'_2} \delta \eta_2 + \cdots + \frac{\delta T}{\delta \eta'_{3n}} \delta \eta_{3n}; \end{aligned} \right\} \quad (23.)$$

$T$  being the same quantity as before, namely, the half of the final living force of system, but being now considered as a function of  $\eta'_1 \dots \eta'_{3n}$ , involving also the masses, and in general  $\eta_1 \dots \eta_{3n}$ , and obtained by substituting for the quantities  $x'$   $y'$   $z'$  their values of the form (20.) in the equation of definition

$$T = \frac{1}{2} \Sigma .m(x'^2 + y'^2 + z'^2). \quad (4.)$$

In like manner we find this transformation for the sum (18.),

$$\Sigma .m(a' \delta a + b' \delta b + c' \delta c) = \frac{\delta T_0}{\delta e'_1} \delta e_1 + \frac{\delta T_0}{\delta e'_2} \delta e_2 + \cdots + \frac{\delta T_0}{\delta e'_{3n}} \delta e_{3n}. \quad (24.)$$

The law of varying action, or the formula (A.), becomes therefore, when expressed by the present more general coordinates or marks of position,

$$\delta V = \Sigma . \frac{\delta T}{\delta \eta'} \delta \eta - \Sigma . \frac{\delta T}{\delta e'} \delta e + t \delta H; \quad (\text{Q.})$$

and instead of the groups (C.) and (D.), into which, along with the equation (E.), this law resolved itself before, it gives now these other groups,

$$\frac{\delta V}{\delta \eta_1} = \frac{\delta T}{\delta \eta'_1}; \quad \frac{\delta V}{\delta \eta_2} = \frac{\delta T}{\delta \eta'_2}; \quad \cdots \quad \frac{\delta V}{\delta \eta_{3n}} = \frac{\delta T}{\delta \eta'_{3n}}; \quad (\text{R.})$$

and

$$\frac{\delta V}{\delta e_1} = -\frac{\delta T_0}{\delta e'_1}; \quad \frac{\delta V}{\delta e_2} = -\frac{\delta T_0}{\delta e'_2}; \quad \dots \quad \frac{\delta V}{\delta e_{3n}} = -\frac{\delta T_0}{\delta e'_{3n}}. \quad (\text{S.})$$

The quantities  $e_1 e_2 \dots e_{3n}$ , and  $e'_1 e'_2 \dots e'_{3n}$ , are now the initial data respecting the manner of motion of the system; and the  $3n$  final integrals, connecting these  $6n$  initial data, and the  $n$  masses, with the time  $t$ , and with the  $3n$  final or varying quantities  $\eta_1 \eta_2 \dots \eta_{3n}$ , which mark the varying positions of the  $n$  moving points of the system, are now to be obtained by eliminating the auxiliary constant  $H$  between the  $3n+1$  equations (S.) and (E.); while the  $3n$  intermediate integrals, or integrals of the first order, which connect the same varying marks of position and their first differential coefficients with the time, the masses, and the initial marks of position, are the result of elimination of the same auxiliary constant  $H$  between the equations (R.) and (E.). Our fundamental formula, and intermediate and final integrals, can therefore be very simply expressed with any new sets of coordinates; and the partial differential equations (F.) (G.), which our characteristic function  $V$  must satisfy, and which are, as we have said, essential in the theory of that function, can also easily be expressed with any such transformed coordinates, by merely combining the final and initial expressions of the law of living force,

$$T = U + H, \quad (6.)$$

$$T_0 = U_0 + H, \quad (7.)$$

with the new groups (R.) and (S.). For this purpose we must now consider the function  $U$ , of the masses and mutual distances of the several points of the system, as depending on the new marks of position  $\eta_1 \eta_2 \dots \eta_{3n}$ ; and the analogous function  $U_0$ , as depending similarly on the initial quantities  $e_1 e_2 \dots e_{3n}$ ; we must also suppose that  $T$  is expressed (as it may) as a function of its own coefficients,  $\frac{\delta T}{\delta \eta'_1}, \frac{\delta T}{\delta \eta'_2}, \dots, \frac{\delta T}{\delta \eta'_{3n}}$ , which will always be, with respect to these, homogeneous of the second dimension, and may also involve explicitly the quantities  $\eta_1 \eta_2 \dots \eta_{3n}$ ; and that  $T_0$  is expressed as a similar function of its coefficients  $\frac{\delta T_0}{\delta e'_1}, \frac{\delta T_0}{\delta e'_2}, \dots, \frac{\delta T_0}{\delta e'_{3n}}$ ; so that

$$\left. \begin{aligned} T &= F \left( \frac{\delta T}{\delta \eta'_1}, \frac{\delta T}{\delta \eta'_2}, \dots, \frac{\delta T}{\delta \eta'_{3n}} \right), \\ T_0 &= F \left( \frac{\delta T_0}{\delta e'_1}, \frac{\delta T_0}{\delta e'_2}, \dots, \frac{\delta T_0}{\delta e'_{3n}} \right); \end{aligned} \right\} \quad (25.)$$

and that then these coefficients of  $T$  and  $T_0$  are changed to their values (R.) and (S.), so as to give, instead of (F.) and (G.), two other transformed equations, namely,

$$F \left( \frac{\delta V}{\delta \eta_1}, \frac{\delta V}{\delta \eta_2}, \dots, \frac{\delta V}{\delta \eta_{3n}} \right) = U + H, \quad (\text{T.})$$

and, on account of the homogeneity and dimension of  $T_0$ ,

$$F \left( \frac{\delta V}{\delta e_1}, \frac{\delta V}{\delta e_2}, \dots, \frac{\delta V}{\delta e_{3n}} \right) = U_0 + H. \quad (\text{U.})$$

8. Nor is there any difficulty in deducing analogous transformations for the known differential equations of motion of the second order, of any system of free points, by taking the variation of the new form (T.) of the law of living force, and by attending to the dynamical meanings of the coefficients of our characteristic function. For if we observe that the final living force  $2T$ , when considered as a function of  $\eta_1 \eta_2 \dots \eta_{3n}$ , and of  $\eta'_1 \eta'_2 \dots \eta'_{3n}$ , is necessarily homogeneous of the second dimension with respect to the latter set of variables, and must therefore satisfy the condition

$$2T = \eta'_1 \frac{\delta T}{\delta \eta'_1} + \eta'_2 \frac{\delta T}{\delta \eta'_2} + \dots + \eta'_{3n} \frac{\delta T}{\delta \eta'_{3n}}, \quad (26.)$$

we shall perceive that its total variation,

$$\left. \begin{aligned} \delta T &= \frac{\delta T}{\delta \eta_1} \delta \eta_1 + \frac{\delta T}{\delta \eta_2} \delta \eta_2 + \dots + \frac{\delta T}{\delta \eta_{3n}} \delta \eta_{3n} \\ &+ \frac{\delta T}{\delta \eta'_1} \delta \eta'_1 + \frac{\delta T}{\delta \eta'_2} \delta \eta'_2 + \dots + \frac{\delta T}{\delta \eta'_{3n}} \delta \eta'_{3n}, \end{aligned} \right\} \quad (27.)$$

may be put under the form

$$\left. \begin{aligned} \delta T &= \eta'_1 \delta \frac{\delta T}{\delta \eta'_1} + \eta'_2 \delta \frac{\delta T}{\delta \eta'_2} + \dots + \eta'_{3n} \delta \frac{\delta T}{\delta \eta'_{3n}} \\ &- \frac{\delta T}{\delta \eta_1} \delta \eta_1 - \frac{\delta T}{\delta \eta_2} \delta \eta_2 - \dots - \frac{\delta T}{\delta \eta_{3n}} \delta \eta_{3n} \\ &= \Sigma \cdot \eta' \delta \frac{\delta T}{\delta \eta'} - \Sigma \cdot \frac{\delta T}{\delta \eta} \delta \eta \\ &= \Sigma \cdot \left( \eta' \delta \frac{\delta V}{\delta \eta} - \frac{\delta T}{\delta \eta} \delta \eta \right), \end{aligned} \right\} \quad (28.)$$

and therefore that the total variation of the new partial differential equation (T.) may be thus written,

$$\Sigma \cdot \left( \eta' \delta \frac{\delta V}{\delta \eta} - \frac{\delta T}{\delta \eta} \delta \eta \right) = \Sigma \cdot \frac{\delta U}{\delta \eta} \delta \eta + \delta H : \quad (V.)$$

in which, if we observe that  $\eta' = \frac{d\eta}{dt}$ , and that the quantities of the form  $\eta$  are the only ones which vary with the time, we shall see that

$$\Sigma \cdot \eta' \delta \frac{\delta V}{\delta \eta} = \Sigma \left( \frac{d}{dt} \frac{\delta V}{\delta \eta} \cdot \delta \eta + \frac{d}{dt} \frac{\delta V}{\delta e} \cdot \delta e \right) + \frac{d}{dt} \frac{\delta V}{\delta H} \cdot \delta H, \quad (29.)$$

because the identical equation  $\delta dV = d\delta V$  gives, when developed,

$$\Sigma \left( \delta \frac{\delta V}{\delta \eta} \cdot d\eta + \delta \frac{\delta V}{\delta e} \cdot de \right) + \delta \frac{\delta V}{\delta H} \cdot dH = \Sigma \left( d \frac{\delta V}{\delta \eta} \cdot \delta \eta + d \frac{\delta V}{\delta e} \cdot \delta e \right) + d \frac{\delta V}{\delta H} \cdot \delta H. \quad (30.)$$

Decomposing, therefore, the expression (V.), for the variation of half the living force, into as many separate equations as it contains independent variations, we obtain, not only the equation

$$\frac{d}{dt} \frac{\delta V}{\delta H} = 1, \quad (\text{K.})$$

which had already presented itself, and the group

$$\frac{d}{dt} \frac{\delta V}{\delta e_1} = 0, \quad \frac{d}{dt} \frac{\delta V}{\delta e_2} = 0, \quad \dots \quad \frac{d}{dt} \frac{\delta V}{\delta e_{3n}} = 0, \quad (\text{W.})$$

which might have been at once obtained by differentiation from the final integrals (S.), but also a group of  $3n$  other equations of the form

$$\frac{d}{dt} \frac{\delta V}{\delta \eta} - \frac{\delta T}{\delta \eta} = \frac{\delta U}{\delta \eta}, \quad (\text{X.})$$

which give, by the intermediate integrals (R.),

$$\frac{d}{dt} \frac{\delta T}{\delta \eta'} - \frac{\delta T}{\delta \eta} = \frac{\delta U}{\delta \eta} : \quad (\text{Y.})$$

that is, more fully,

$$\left. \begin{aligned} \frac{d}{dt} \frac{\delta T}{\delta \eta'_1} - \frac{\delta T}{\delta \eta_1} &= \frac{\delta U}{\delta \eta_1}; \\ \frac{d}{dt} \frac{\delta T}{\delta \eta'_2} - \frac{\delta T}{\delta \eta_2} &= \frac{\delta U}{\delta \eta_2}; \\ &\dots \\ \frac{d}{dt} \frac{\delta T}{\delta \eta'_{3n}} - \frac{\delta T}{\delta \eta_{3n}} &= \frac{\delta U}{\delta \eta_{3n}}. \end{aligned} \right\} \quad (\text{Z.})$$

These last transformations of the differential equations of motion of the second order, of an attracting or repelling system, coincide in all respects (a slight difference of notation excepted,) with the elegant canonical forms in the *Mécanique Analytique* of LAGRANGE; but it seemed worth while to deduce them here anew, from the properties of our characteristic function. And if we were to suppose (as it has often been thought convenient and even necessary to do,) that the  $n$  points of a system are not entirely free, nor subject only to their own mutual attractions or repulsions, but connected by any geometrical conditions, and influenced by any foreign agencies, consistent with the law of conservation of living force; so that the number of independent marks of position would be now less numerous, and the force-function  $U$  less simple than before; it might still be proved, by a reasoning very similar to the foregoing, that on these suppositions also (which however, the dynamical spirit is tending more and more to exclude,) the accumulated living force or action  $V$  of the system is a *characteristic motion-function* of the kind already explained; having the same law and formula of variation, which are susceptible of the same transformations; obliged to satisfy in the same way a final and an initial relation between its partial differential coefficients of the

first order; conducting, by the variation of one of these two relations, to the same canonical forms assigned by LAGRANGE for the differential equations of motion; and furnishing, on the same principles as before, their intermediate and their final integrals. To those imaginable cases, indeed, in which the law of living force no longer holds, our method also would not apply; but it appears to be the growing conviction of the persons who have meditated the most profoundly on the mathematical dynamics of the universe, that these are cases suggested by insufficient views of the mutual actions of body.

9. It results from the foregoing remarks, that in order to apply our method of the characteristic function to any problem of dynamics respecting any moving system, the known law of living force is to be combined with our law of varying action; and that the general expression of this latter law is to be obtained in the following manner. We are first to express the quantity  $T$ , namely, the half of the living force of the system, as a function (which will always be homogeneous of the second dimension,) of the differential coefficients or rates of increase  $\eta'_1, \eta'_2, \&c.$ , of any rectangular coordinates, or other marks of position of the system: we are next to take the variation of this homogeneous function with respect to those rates of increase, and to change the variations of those rates  $\delta\eta'_1, \delta\eta'_2, \&c.$ , to the variations  $\delta\eta_1, \delta\eta_2, \&c.$ , of the marks of position themselves; and then to subtract the initial from the final value of the result, and to equate the remainder to  $\delta V - t\delta H$ . A slight consideration will show that this general rule or process for obtaining the variation of the characteristic function  $V$ , is applicable even when the marks of position  $\eta_1, \eta_2, \&c.$  are not all independent of each other; which will happen when they have been made, from any motive of convenience, more numerous than the rectangular coordinates of the several points of the system. For if we suppose that the  $3n$  rectangular coordinates  $x_1 y_1 z_1 \dots x_n y_n z_n$  have been expressed by any transformation as functions of  $3n + k$  other marks of position,  $\eta_1 \eta_2 \dots \eta_{3n+k}$ , which must therefore be connected by  $k$  equations of condition,

$$\left. \begin{array}{l} 0 = \phi_1(\eta_1, \eta_2, \dots \eta_{3n+k}), \\ 0 = \phi_2(\eta_1, \eta_2, \dots \eta_{3n+k}), \\ \dots \dots \\ 0 = \phi_k(\eta_1, \eta_2, \dots \eta_{3n+k}), \end{array} \right\} \quad (31.)$$

giving  $k$  of the new marks of position as functions of the remaining  $3n$ ,

$$\left. \begin{array}{l} \eta_{3n+1} = \psi_1(\eta_1, \eta_2, \dots \eta_{3n}), \\ \eta_{3n+2} = \psi_2(\eta_1, \eta_2, \dots \eta_{3n}), \\ \dots \dots \\ \eta_{3n+k} = \psi_k(\eta_1, \eta_2, \dots \eta_{3n}), \end{array} \right\} \quad (32.)$$

the expression

$$T = \frac{1}{2} \Sigma .m(x'^2 + y'^2 + z'^2), \quad (4.)$$

will become, by the introduction of these new variables, a homogeneous function of the second dimension of the  $3n + k$  rates of increase  $\eta'_1, \eta'_2, \dots \eta'_{3n+k}$ , involving also in general

$\eta_1, \eta_2, \dots, \eta_{3n+k}$ , and having a variation which may be thus expressed:

$$\left. \begin{aligned} \delta T &= \left( \frac{\delta T}{\delta \eta'_1} \right) \delta \eta'_1 + \left( \frac{\delta T}{\delta \eta'_2} \right) \delta \eta'_2 + \cdots + \left( \frac{\delta T}{\delta \eta'_{3n+k}} \right) \delta \eta'_{3n+k} \\ &\quad + \left( \frac{\delta T}{\delta \eta_1} \right) \delta \eta_1 + \left( \frac{\delta T}{\delta \eta_2} \right) \delta \eta_2 + \cdots + \left( \frac{\delta T}{\delta \eta_{3n+k}} \right) \delta \eta_{3n+k}; \end{aligned} \right\} \quad (33.)$$

or in this other way,

$$\left. \begin{aligned} \delta T &= \frac{\delta T}{\delta \eta'_1} \delta \eta'_1 + \frac{\delta T}{\delta \eta'_2} \delta \eta'_2 + \cdots + \frac{\delta T}{\delta \eta'_{3n}} \delta \eta'_{3n} \\ &\quad + \frac{\delta T}{\delta \eta_1} \delta \eta_1 + \frac{\delta T}{\delta \eta_2} \delta \eta_2 + \cdots + \frac{\delta T}{\delta \eta_{3n}} \delta \eta_{3n}, \end{aligned} \right\} \quad (34.)$$

on account of the relations (32.) which give, when differentiated with respect to the time,

$$\left. \begin{aligned} \eta'_{3n+1} &= \eta'_1 \frac{\delta \psi_1}{\delta \eta_1} + \eta'_2 \frac{\delta \psi_1}{\delta \eta_2} + \cdots + \eta'_{3n} \frac{\delta \psi_1}{\delta \eta_{3n}}, \\ \eta'_{3n+2} &= \eta'_1 \frac{\delta \psi_2}{\delta \eta_1} + \eta'_2 \frac{\delta \psi_2}{\delta \eta_2} + \cdots + \eta'_{3n} \frac{\delta \psi_2}{\delta \eta_{3n}}, \\ &\dots \\ \eta'_{3n+k} &= \eta'_1 \frac{\delta \psi_k}{\delta \eta_1} + \eta'_2 \frac{\delta \psi_k}{\delta \eta_2} + \cdots + \eta'_{3n} \frac{\delta \psi_k}{\delta \eta_{3n}}, \end{aligned} \right\} \quad (35.)$$

and therefore, attending only to the variations of quantities of the form  $\eta'$ ,

$$\left. \begin{aligned} \delta \eta'_{3n+1} &= \frac{\delta \psi_1}{\delta \eta_1} \delta \eta'_1 + \frac{\delta \psi_1}{\delta \eta_2} \delta \eta'_2 + \cdots + \frac{\delta \psi_1}{\delta \eta_{3n}} \delta \eta'_{3n}, \\ \delta \eta'_{3n+2} &= \frac{\delta \psi_2}{\delta \eta_1} \delta \eta'_1 + \frac{\delta \psi_2}{\delta \eta_2} \delta \eta'_2 + \cdots + \frac{\delta \psi_2}{\delta \eta_{3n}} \delta \eta'_{3n}, \\ &\dots \\ \delta \eta'_{3n+k} &= \frac{\delta \psi_k}{\delta \eta_1} \delta \eta'_1 + \frac{\delta \psi_k}{\delta \eta_2} \delta \eta'_2 + \cdots + \frac{\delta \psi_k}{\delta \eta_{3n}} \delta \eta'_{3n}. \end{aligned} \right\} \quad (36.)$$

Comparing the two expressions (33.) and (34.), we find by (36.) the relations

$$\left. \begin{aligned} \frac{\delta T}{\delta \eta'_1} &= \left( \frac{\delta T}{\delta \eta'_1} \right) + \left( \frac{\delta T}{\delta \eta'_{3n+1}} \right) \frac{\delta \psi_1}{\delta \eta_1} + \left( \frac{\delta T}{\delta \eta'_{3n+2}} \right) \frac{\delta \psi_2}{\delta \eta_1} + \cdots + \left( \frac{\delta T}{\delta \eta'_{3n+k}} \right) \frac{\delta \psi_k}{\delta \eta_1}, \\ \frac{\delta T}{\delta \eta'_2} &= \left( \frac{\delta T}{\delta \eta'_2} \right) + \left( \frac{\delta T}{\delta \eta'_{3n+1}} \right) \frac{\delta \psi_1}{\delta \eta_2} + \left( \frac{\delta T}{\delta \eta'_{3n+2}} \right) \frac{\delta \psi_2}{\delta \eta_2} + \cdots + \left( \frac{\delta T}{\delta \eta'_{3n+k}} \right) \frac{\delta \psi_k}{\delta \eta_2}, \\ &\dots \\ \frac{\delta T}{\delta \eta'_{3n}} &= \left( \frac{\delta T}{\delta \eta'_{3n}} \right) + \left( \frac{\delta T}{\delta \eta'_{3n+1}} \right) \frac{\delta \psi_1}{\delta \eta_{3n}} + \left( \frac{\delta T}{\delta \eta'_{3n+2}} \right) \frac{\delta \psi_2}{\delta \eta_{3n}} + \cdots + \left( \frac{\delta T}{\delta \eta'_{3n+k}} \right) \frac{\delta \psi_k}{\delta \eta_{3n}}; \end{aligned} \right\} \quad (37.)$$

which give, by (32.),

$$\frac{\delta T}{\delta \eta'_1} \delta \eta_1 + \frac{\delta T}{\delta \eta'_2} \delta \eta_2 + \cdots + \frac{\delta T}{\delta \eta'_{3n}} \delta \eta_{3n} = \left( \frac{\delta T}{\delta \eta'_1} \right) \delta \eta_1 + \left( \frac{\delta T}{\delta \eta'_2} \right) \delta \eta_2 + \cdots + \left( \frac{\delta T}{\delta \eta'_{3n+k}} \right) \delta \eta_{3n+k}; \quad (38.)$$

we may therefore put the expression (Q.) under the following more general form,

$$\delta V = \Sigma \cdot \left( \frac{\delta T}{\delta \eta'} \right) \delta \eta - \Sigma \cdot \left( \frac{\delta T_0}{\delta e'} \right) \delta e + t \delta H, \quad (A^1.)$$

the coefficients  $\left( \frac{\delta T}{\delta \eta'} \right)$  being formed by treating all the  $3n+k$  quantities  $\eta'_1, \eta'_2, \dots, \eta'_{3n+k}$ , as independent; which was the extension above announced, of the rule for forming the variation of the characteristic function  $V$ .

We cannot, however, immediately decompose this new expression ( $A^1.$ ) for  $\delta V$ , as we did the expression (Q.), by treating all the variations  $\delta \eta, \delta e$ , as independent; but we may decompose it so, if we previously combine it with the final equations of condition (31.), and with the analogous initial equations of condition, namely,

$$\left. \begin{array}{l} 0 = \Phi_1(e_1, e_2, \dots, e_{3n+k}), \\ 0 = \Phi_2(e_1, e_2, \dots, e_{3n+k}), \\ \dots \dots \\ 0 = \Phi_k(e_1, e_2, \dots, e_{3n+k}), \end{array} \right\} \quad (39.)$$

which we may do by adding the variations of the connecting functions  $\phi_1, \dots, \phi_k, \Phi_1, \dots, \Phi_k$  multiplied respectively by the factors to be determined,  $\lambda_1, \dots, \lambda_k, \Lambda_1, \dots, \Lambda_k$ . In this manner the law of varying action takes this new form,

$$\delta V = \Sigma \cdot \left( \frac{\delta T}{\delta \eta'} \right) \delta \eta - \Sigma \cdot \left( \frac{\delta T_0}{\delta e'} \right) \delta e + t \delta H + \Sigma . \lambda \delta \phi + \Sigma . \Lambda \delta \Phi; \quad (B^1.)$$

and decomposes itself into  $6n+2k+1$  separate expressions, for the partial differential coefficients of the first order of the characteristic function  $V$ , namely, into the following,

$$\left. \begin{array}{l} \frac{\delta V}{\delta \eta_1} = \left( \frac{\delta T}{\delta \eta'_1} \right) + \lambda_1 \frac{\delta \phi_1}{\delta \eta_1} + \lambda_2 \frac{\delta \phi_2}{\delta \eta_1} + \cdots + \lambda_k \frac{\delta \phi_k}{\delta \eta_1}, \\ \frac{\delta V}{\delta \eta_2} = \left( \frac{\delta T}{\delta \eta'_2} \right) + \lambda_1 \frac{\delta \phi_1}{\delta \eta_2} + \lambda_2 \frac{\delta \phi_2}{\delta \eta_2} + \cdots + \lambda_k \frac{\delta \phi_k}{\delta \eta_2}, \\ \dots \dots \\ \frac{\delta V}{\delta \eta_{3n+k}} = \left( \frac{\delta T}{\delta \eta'_{3n+k}} \right) + \lambda_1 \frac{\delta \phi_1}{\delta \eta_{3n+k}} + \cdots + \lambda_k \frac{\delta \phi_k}{\delta \eta_{3n+k}}, \end{array} \right\} \quad (C^1.)$$

and

$$\left. \begin{array}{l} \frac{\delta V}{\delta e_1} = - \left( \frac{\delta T}{\delta e'_1} \right) + \Lambda_1 \frac{\delta \Phi_1}{\delta e_1} + \Lambda_2 \frac{\delta \Phi_2}{\delta e_1} + \cdots + \Lambda_k \frac{\delta \Phi_k}{\delta e_1}, \\ \frac{\delta V}{\delta e_2} = - \left( \frac{\delta T}{\delta e'_2} \right) + \Lambda_1 \frac{\delta \Phi_1}{\delta e_2} + \Lambda_2 \frac{\delta \Phi_2}{\delta e_2} + \cdots + \Lambda_k \frac{\delta \Phi_k}{\delta e_2}, \\ \dots \dots \\ \frac{\delta V}{\delta e_{3n+k}} = - \left( \frac{\delta T}{\delta e'_{3n+k}} \right) + \Lambda_1 \frac{\delta \Phi_1}{\delta e_{3n+k}} + \cdots + \Lambda_k \frac{\delta \Phi_k}{\delta e_{3n+k}}, \end{array} \right\} \quad (D^1.)$$

besides the old equation (E.). The analogous introduction of multipliers in the canonical forms of LAGRANGE, for the differential equations of motion of the second order, by which a sum such as  $\Sigma \lambda \frac{\delta\phi}{\delta\eta}$  is added to  $\frac{\delta U}{\delta\eta}$  in the second member of the formula (Y.), is also easily justified on the principles of the present essay.

*Separation of the relative motion of a system from the motion of its centre of gravity; characteristic function for such motion, and law of its variation.*

10. As an example of the foregoing transformations, and at the same time as an important application, we shall now introduce relative coordinates,  $x, y, z$ , referred to an internal origin  $x_n, y_n, z_n$ ; that is, we shall put

$$x_i = x_{ni} + x_n, \quad y_i = y_{ni} + y_n, \quad z_i = z_{ni} + z_n, \quad (40.)$$

and in like manner

$$a_i = a_{ni} + a_n, \quad b_i = b_{ni} + b_n, \quad c_i = c_{ni} + c_n; \quad (41.)$$

together with the differentiated expressions

$$x'_i = x'_{ni} + x'_n, \quad y'_i = y'_{ni} + y'_n, \quad z'_i = z'_{ni} + z'_n, \quad (42.)$$

and

$$a'_i = a'_{ni} + a'_n, \quad b'_i = b'_{ni} + b'_n, \quad c'_i = c'_{ni} + c'_n. \quad (43.)$$

Introducing the expressions (42.) for the rectangular components of velocity, we find that the value given by (4.) for the living force  $2T$  decomposes itself into the three following parts,

$$\left. \begin{aligned} 2T &= \Sigma .m(x'^2 + y'^2 + z'^2) \\ &= \Sigma .m(x'^2 + y'^2 + z'^2) + 2(x'_n \Sigma .mx' + y'_n \Sigma .my' + z'_n \Sigma .mz') \\ &\quad + (x'^2_n + y'^2_n + z'^2_n) \Sigma m; \end{aligned} \right\} \quad (44.)$$

if then we establish, as we may, the three equations of condition,

$$\Sigma .mx, = 0, \quad \Sigma .my, = 0, \quad \Sigma .mz, = 0, \quad (45.)$$

which give by (40.),

$$x_n = \frac{\Sigma .mx}{\Sigma m}, \quad y_n = \frac{\Sigma .my}{\Sigma m}, \quad z_n = \frac{\Sigma .mz}{\Sigma m}, \quad (46.)$$

so that  $x_n, y_n, z_n$  are now the coordinates of the point which is called the centre of gravity of the system, we may reduce the function  $T$  to the form

$$T = T_r + T_n, \quad (47.)$$

in which

$$T_r = \frac{1}{2} \Sigma .m(x'^2_r + y'^2_r + z'^2_r), \quad (48.)$$

and

$$T_{\text{II}} = \frac{1}{2}(x_{\text{II}}'^2 + y_{\text{II}}'^2 + z_{\text{II}}'^2) \Sigma m. \quad (49.)$$

By this known decomposition, the whole living force  $2T$  of the system is resolved into the two parts  $2T_I$  and  $2T_{\text{II}}$ , of which the former,  $2T_I$ , may be called the *relative living force*, being that which results solely from the relative velocities of the points of the system, in their motions about their common centre of gravity  $x_{\text{II}}, y_{\text{II}}, z_{\text{II}}$ ; while the latter part,  $2T_{\text{II}}$ , results only from the absolute motion of that centre of gravity in space, and is the same as if all the masses of the system were united in that common centre. At the same time, the law of living force,  $T = U + H$ , (6.), resolves itself by the law of motion of the centre of gravity into the two following separate equations,

$$T_I = U + H_I, \quad (50.)$$

and

$$T_{\text{II}} = H_{\text{II}}; \quad (51.)$$

$H_I$  and  $H_{\text{II}}$  being two new constants independent of the time  $t$ , and such that their sum

$$H_I + H_{\text{II}} = H. \quad (52.)$$

And we may in like manner decompose the action, or accumulated living force  $V$ , which is equal to the definite integral  $\int_0^t 2T dt$ , into the two following analogous parts,

$$V = V_I + V_{\text{II}}, \quad (\text{E}^1.)$$

determined by the two equations,

$$V_I = \int_0^t 2T_I dt, \quad (\text{F}^1.)$$

and

$$V_{\text{II}} = \int_0^t 2T_{\text{II}} dt. \quad (\text{G}^1.)$$

The last equation gives by (51.),

$$V_{\text{II}} = 2H_{\text{II}}t; \quad (53.)$$

a result which, by the law of motion of the centre of gravity, may be thus expressed,

$$V_{\text{II}} = \sqrt{(x_{\text{II}} - a_{\text{II}})^2 + (y_{\text{II}} - b_{\text{II}})^2 + (z_{\text{II}} - c_{\text{II}})^2} \cdot \sqrt{2H_{\text{II}} \Sigma m}: \quad (\text{H}^1.)$$

$a_{\text{II}}, b_{\text{II}}, c_{\text{II}}$  being the initial coordinates of the centre of gravity, so that

$$a_{\text{II}} = \frac{\Sigma .m a}{\Sigma m}, \quad b_{\text{II}} = \frac{\Sigma .m b}{\Sigma m}, \quad c_{\text{II}} = \frac{\Sigma .m c}{\Sigma m}. \quad (54.)$$

And for the variation  $\delta V$  of the whole function  $V$ , the rule of the last number gives

$$\left. \begin{aligned} \delta V &= \Sigma .m(x' \delta x_I - a' \delta a_I + y' \delta y_I - b' \delta b_I + z' \delta z_I - c' \delta c_I) \\ &\quad + (x_{\text{II}}' \delta x_{\text{II}} - a_{\text{II}}' \delta a_{\text{II}} + y_{\text{II}}' \delta y_{\text{II}} - b_{\text{II}}' \delta b_{\text{II}} + z_{\text{II}}' \delta z_{\text{II}} - c_{\text{II}}' \delta c_{\text{II}}) \Sigma m \\ &\quad + t \delta H + \lambda_1 \Sigma .m \delta x_I + \lambda_2 \Sigma .m \delta y_I + \lambda_3 \Sigma .m \delta z_I \\ &\quad + \Lambda_1 \Sigma .m \delta a_I + \Lambda_2 \Sigma .m \delta b_I + \Lambda_3 \Sigma .m \delta c_I; \end{aligned} \right\} \quad (\text{I}^1.)$$

while the variation of the part  $V_{\text{II}}$ , determined by the equation (H<sup>1</sup>.), is easily shown to be equivalent to the part

$$\delta V_{\text{II}} = (x'_{\text{II}} \delta x_{\text{II}} - a'_{\text{II}} \delta a_{\text{II}} + y'_{\text{II}} \delta y_{\text{II}} - b'_{\text{II}} \delta b_{\text{II}} + z'_{\text{II}} \delta z_{\text{II}} - c'_{\text{II}} \delta c_{\text{II}}) \Sigma m + t \delta H_{\text{II}}; \quad (\text{K}^1.)$$

the variation of the other part  $V_I$  may therefore be thus expressed,

$$\left. \begin{aligned} \delta V_I &= \Sigma .m(x' \delta x, -a' \delta a, +y' \delta y, -b' \delta b, +z' \delta z, -c' \delta c) \\ &\quad + t \delta H_I + \lambda_1 \Sigma .m \delta x_I + \lambda_2 \Sigma .m \delta y_I + \lambda_3 \Sigma .m \delta z_I \\ &\quad + \Lambda_1 \Sigma .m \delta a_I + \Lambda_2 \Sigma .m \delta b_I + \Lambda_3 \Sigma .m \delta c_I : \end{aligned} \right\} \quad (\text{L}^1.)$$

and it resolves itself into the following separate expressions, in which the part  $V_I$  is considered as a function of the  $6n + 1$  quantities  $x_{ri}$   $y_{ri}$   $z_{ri}$   $a_{ri}$   $b_{ri}$   $c_{ri}$   $H_I$ , of which, however, only  $6n - 5$  are really independent:

first group,

$$\left. \begin{aligned} \frac{\delta V_I}{\delta x_{r1}} &= m_1 x'_{r1} + \lambda_1 m_1; & \dots & \frac{\delta V_I}{\delta x_{rn}} &= m_n x'_{rn} + \lambda_1 m_n; \\ \frac{\delta V_I}{\delta y_{r1}} &= m_1 y'_{r1} + \lambda_2 m_1; & \dots & \frac{\delta V_I}{\delta y_{rn}} &= m_n y'_{rn} + \lambda_2 m_n; \\ \frac{\delta V_I}{\delta z_{r1}} &= m_1 z'_{r1} + \lambda_3 m_1; & \dots & \frac{\delta V_I}{\delta z_{rn}} &= m_n z'_{rn} + \lambda_3 m_n; \end{aligned} \right\} \quad (\text{M}^1.)$$

second group,

$$\left. \begin{aligned} \frac{\delta V_I}{\delta a_{r1}} &= -m_1 a'_{r1} + \Lambda_1 m_1; & \dots & \frac{\delta V_I}{\delta a_{rn}} &= -m_n a'_{rn} + \Lambda_1 m_n; \\ \frac{\delta V_I}{\delta b_{r1}} &= -m_1 b'_{r1} + \Lambda_2 m_1; & \dots & \frac{\delta V_I}{\delta b_{rn}} &= -m_n b'_{rn} + \Lambda_2 m_n; \\ \frac{\delta V_I}{\delta c_{r1}} &= -m_1 c'_{r1} + \Lambda_3 m_1; & \dots & \frac{\delta V_I}{\delta c_{rn}} &= -m_n c'_{rn} + \Lambda_3 m_n; \end{aligned} \right\} \quad (\text{N}^1.)$$

and finally,

$$\frac{\delta V_I}{\delta H_I} = t. \quad (\text{O}^1.)$$

With respect to the six multipliers  $\lambda_1 \lambda_2 \lambda_3 \Lambda_1 \Lambda_2 \Lambda_3$  which were introduced by the 3 final equations of condition (45.), and by the 3 analogous initial equations of condition,

$$\Sigma .ma_I = 0, \quad \Sigma .mb_I = 0, \quad \Sigma .mc_I = 0; \quad (55.)$$

we have, by differentiating these conditions,

$$\Sigma .mx'_I = 0, \quad \Sigma .my'_I = 0, \quad \Sigma .mz'_I = 0, \quad (56.)$$

and

$$\Sigma .ma'_I = 0, \quad \Sigma .mb'_I = 0, \quad \Sigma .mc'_I = 0; \quad (57.)$$

and therefore

$$\lambda_1 = \frac{\Sigma \frac{\delta V_r}{\delta x_r}}{\Sigma m}, \quad \lambda_2 = \frac{\Sigma \frac{\delta V_r}{\delta y_r}}{\Sigma m}, \quad \lambda_3 = \frac{\Sigma \frac{\delta V_r}{\delta z_r}}{\Sigma m}, \quad (58.)$$

and

$$\Lambda_1 = \frac{\Sigma \frac{\delta V_r}{\delta a_r}}{\Sigma m}, \quad \Lambda_2 = \frac{\Sigma \frac{\delta V_r}{\delta b_r}}{\Sigma m}, \quad \Lambda_3 = \frac{\Sigma \frac{\delta V_r}{\delta c_r}}{\Sigma m}. \quad (59.)$$

11. As an example of the determination of these multipliers, we may suppose that the part  $V_r$ , of the whole action  $V$ , has been expressed, before differentiation, as a function of  $H_r$ , and of these other  $6n - 6$  independent quantities

$$\left. \begin{array}{l} x_{r1} - x_{rn} = \xi_1, \quad x_{r2} - x_{rn} = \xi_2, \quad \dots \quad x_{r(n-1)} - x_{rn} = \xi_{n-1}, \\ y_{r1} - y_{rn} = \eta_1, \quad y_{r2} - y_{rn} = \eta_2, \quad \dots \quad y_{r(n-1)} - y_{rn} = \eta_{n-1}, \\ z_{r1} - z_{rn} = \zeta_1, \quad z_{r2} - z_{rn} = \zeta_2, \quad \dots \quad z_{r(n-1)} - z_{rn} = \zeta_{n-1}, \end{array} \right\} \quad (60.)$$

and

$$\left. \begin{array}{l} a_{r1} - a_{rn} = \alpha_1, \quad a_{r2} - a_{rn} = \alpha_2, \quad \dots \quad a_{r(n-1)} - a_{rn} = \alpha_{n-1}, \\ b_{r1} - b_{rn} = \beta_1, \quad b_{r2} - b_{rn} = \beta_2, \quad \dots \quad b_{r(n-1)} - b_{rn} = \beta_{n-1}, \\ c_{r1} - c_{rn} = \gamma_1, \quad c_{r2} - c_{rn} = \gamma_2, \quad \dots \quad c_{r(n-1)} - c_{rn} = \gamma_{n-1}; \end{array} \right\} \quad (61.)$$

that is, of the *differences* only of the *centrobaric* coordinates; or, in other words, as a function of the coordinates (initial and final) of  $n - 1$  points of the system, referred to the  $n^{\text{th}}$  point, as an internal or moveable origin: because the centrobaric coordinates  $x_{ri}$ ,  $y_{ri}$ ,  $z_{ri}$ ,  $a_{ri}$ ,  $b_{ri}$ ,  $c_{ri}$ , may themselves, by the equations of condition, be expressed as a function of these, namely,

$$x_{ri} = \xi_i - \frac{\Sigma .m\xi}{\Sigma m}, \quad y_{ri} = \eta_i - \frac{\Sigma .m\eta}{\Sigma m}, \quad z_{ri} = \zeta_i - \frac{\Sigma .m\zeta}{\Sigma m}, \quad (62.)$$

and in like manner,

$$a_{ri} = \alpha_i - \frac{\Sigma .m\alpha}{\Sigma m}, \quad b_{ri} = \beta_i - \frac{\Sigma .m\beta}{\Sigma m}, \quad c_{ri} = \gamma_i - \frac{\Sigma .m\gamma}{\Sigma m}; \quad (63.)$$

in which we are to observe, that the six quantities  $\xi_n$   $\eta_n$   $\zeta_n$   $\alpha_n$   $\beta_n$   $\gamma_n$  must be considered as separately vanishing. When  $V_r$  has been thus expressed as a function of the centrobaric coordinates, involving their differences only, it will evidently satisfy the six partial differential equations,

$$\left. \begin{array}{l} \Sigma \frac{\delta V_r}{\delta x_r} = 0, \quad \Sigma \frac{\delta V_r}{\delta y_r} = 0, \quad \Sigma \frac{\delta V_r}{\delta z_r} = 0, \\ \Sigma \frac{\delta V_r}{\delta a_r} = 0, \quad \Sigma \frac{\delta V_r}{\delta b_r} = 0, \quad \Sigma \frac{\delta V_r}{\delta c_r} = 0; \end{array} \right\} \quad (\text{P}^1.)$$

after this preparation, therefore, of the function  $V_r$ , the six multipliers determined by (58.) and (59.) will vanish, so that we shall have

$$\lambda_1 = 0, \quad \lambda_2 = 0, \quad \lambda_3 = 0, \quad \Lambda_1 = 0, \quad \Lambda_2 = 0, \quad \Lambda_3 = 0, \quad (64.)$$

and the groups (M<sup>1</sup>.) and (N<sup>1</sup>.) will reduce themselves to the two following:

$$\left. \begin{array}{l} \frac{\delta V_r}{\delta x_{r1}} = m_1 x'_{r1}; \quad \frac{\delta V_r}{\delta x_{r2}} = m_2 x'_{r2}; \quad \cdots \quad \frac{\delta V_r}{\delta x_m} = m_n x'_{rn}; \\ \frac{\delta V_r}{\delta y_{r1}} = m_1 y'_{r1}; \quad \frac{\delta V_r}{\delta y_{r2}} = m_2 y'_{r2}; \quad \cdots \quad \frac{\delta V_r}{\delta y_m} = m_n y'_{rn}; \\ \frac{\delta V_r}{\delta z_{r1}} = m_1 z'_{r1}; \quad \cdots \quad \frac{\delta V_r}{\delta z_{r2}} = m_2 z'_{r2}; \quad \cdots \quad \frac{\delta V_r}{\delta z_m} = m_n z'_{rn}; \end{array} \right\} \quad (\text{Q}^1.)$$

and

$$\left. \begin{array}{l} \frac{\delta V_r}{\delta a_{r1}} = -m_1 a'_{r1}; \quad \frac{\delta V_r}{\delta a_{r2}} = -m_2 a'_{r2}; \quad \cdots \quad \frac{\delta V_r}{\delta a_m} = -m_n a'_{rn}; \\ \frac{\delta V_r}{\delta b_{r1}} = -m_1 b'_{r1}; \quad \frac{\delta V_r}{\delta b_{r2}} = -m_2 b'_{r2}; \quad \cdots \quad \frac{\delta V_r}{\delta b_m} = -m_n b'_{rn}; \\ \frac{\delta V_r}{\delta c_{r1}} = -m_1 c'_{r1}; \quad \cdots \quad \frac{\delta V_r}{\delta c_{r2}} = -m_2 c'_{r2}; \quad \cdots \quad \frac{\delta V_r}{\delta c_m} = -m_n c'_{rn}; \end{array} \right\} \quad (\text{R}^1.)$$

analogous in all respects to the groups (C.) and (D.). We find, therefore, for the relative motion of a system about its own centre of gravity, equations of the same form as those which we had obtained before for the absolute motion of the same system of points in space. And we see that in investigating such relative motion only, it is useful to confine ourselves to the part  $V_r$  of our whole characteristic function, that is, to the *relative action* of the system, or accumulated living force of the motion about the centre of gravity; and to consider this part as the *characteristic function* of such relative motion, in a sense analogous to that which has been already explained.

This relative action, or part  $V_r$ , may, however, be otherwise expressed, and even in an infinite variety of ways, on account of the six equations of condition which connect the  $6n$  centrobaric coordinates; and every different preparation of its form will give a different set of values for the six multipliers  $\lambda_1 \lambda_2 \lambda_3 \Lambda_1 \Lambda_2 \Lambda_3$ . For example, we might eliminate, by a previous preparation, the six centrobaric coordinates of the point  $m_n$  from the expression of  $V_r$ , so as to make this expression involve only the centrobaric coordinates of the other  $n - 1$  points of the system, and then we should have

$$\frac{\delta V_r}{\delta x_m} = 0, \quad \frac{\delta V_r}{\delta y_m} = 0, \quad \frac{\delta V_r}{\delta z_m} = 0, \quad \frac{\delta V_r}{\delta a_m} = 0, \quad \frac{\delta V_r}{\delta b_m} = 0, \quad \frac{\delta V_r}{\delta c_m} = 0, \quad (\text{S}^1.)$$

and therefore, by the six last equations of the groups (M<sup>1</sup>.) and (N<sup>1</sup>.), the multipliers would take the values

$$\lambda_1 = -x'_{rn}, \quad \lambda_2 = -y'_{rn}, \quad \lambda_3 = -z'_{rn}, \quad \Lambda_1 = a'_{rn}, \quad \Lambda_2 = b'_{rn}, \quad \Lambda_3 = c'_{rn}, \quad (65.)$$

and would reduce, by (60.) and (61.), the preceding  $6n - 6$  equations of the same groups (M<sup>1</sup>.) and (N<sup>1</sup>.), to the forms

$$\left. \begin{array}{l} \frac{\delta V_r}{\delta x_{r1}} = m_1 \xi'_1, \quad \frac{\delta V_r}{\delta x_{r2}} = m_2 \xi'_2, \quad \cdots \quad \frac{\delta V_r}{\delta x_{n-1}} = m_{n-1} \xi'_{n-1}, \\ \frac{\delta V_r}{\delta y_{r1}} = m_1 \eta'_1, \quad \frac{\delta V_r}{\delta y_{r2}} = m_2 \eta'_2, \quad \cdots \quad \frac{\delta V_r}{\delta y_{n-1}} = m_{n-1} \eta'_{n-1}, \\ \frac{\delta V_r}{\delta z_{r1}} = m_1 \zeta'_1, \quad \cdots \quad \frac{\delta V_r}{\delta z_{r2}} = m_2 \zeta'_2, \quad \cdots \quad \frac{\delta V_r}{\delta z_{n-1}} = m_{n-1} \zeta'_{n-1}, \end{array} \right\} \quad (\text{T}^1.)$$

and

$$\left. \begin{aligned} \frac{\delta V_r}{\delta a_{r1}} &= -m_1 \alpha'_1, & \frac{\delta V_r}{\delta a_{r2}} &= -m_2 \alpha'_2, & \dots & \frac{\delta V_r}{\delta a_{rn-1}} &= -m_{n-1} \alpha'_{n-1}, \\ \frac{\delta V_r}{\delta b_{r1}} &= -m_1 \beta'_1, & \frac{\delta V_r}{\delta b_{r2}} &= -m_2 \beta'_2, & \dots & \frac{\delta V_r}{\delta b_{rn-1}} &= -m_{n-1} \beta'_{n-1}, \\ \frac{\delta V_r}{\delta c_{r1}} &= -m_1 \gamma'_1, & \frac{\delta V_r}{\delta c_{r2}} &= -m_2 \gamma'_2, & \dots & \frac{\delta V_r}{\delta c_{rn-1}} &= -m_{n-1} \gamma'_{n-1}. \end{aligned} \right\} \quad (\text{U}^1.)$$

12. We might also express the relative action  $V_r$ , not as a function of the centrobaric, but of some other internal coordinates, or marks of relative position. We might, for instance, express it and its variation as functions of the  $6n - 6$  independent internal coordinates  $\xi \eta \zeta \alpha \beta \gamma$  already mentioned, and of their variations, defining these without any reference to the centre of gravity, by the equations

$$\left. \begin{aligned} \xi_i &= x_i - x_n, & \eta_i &= y_i - y_n, & \zeta_i &= z_i - z_n, \\ \alpha_i &= a_i - a_n, & \beta_i &= b_i - b_n, & \gamma_i &= c_i - c_n. \end{aligned} \right\} \quad (66.)$$

For all such transformations of  $\delta V_r$  it is easy to establish a rule or law, which may be called the *law of varying relative action* (exactly analogous to the rule (B<sup>1</sup>.)), namely, the following:

$$\delta V_r = \Sigma \cdot \left( \frac{\delta T_r}{\delta \eta'_i} \right) \delta \eta_i - \Sigma \cdot \left( \frac{\delta T_{r0}}{\delta e'_i} \right) \delta e_i + t \delta H_r + \Sigma .\lambda, \delta \phi_r + \Sigma .\Lambda, \delta \Phi_r; \quad (\text{V}^1.)$$

which implies that we are to express the half  $T_r$  of the relative living force of the system as a function of the rates of increase  $\eta'_i$  of any marks of relative position; and after taking its variation with respect to these rates, to change their variations to the variations of the marks of position themselves; then to subtract the initial from the final value of the result, and to add the variations of the final and initial functions  $\phi_r, \Phi_r$ , which enter into the equations of condition, if any, of the form  $\phi_r = 0, \Phi_r = 0$ , (connecting the final and initial marks of relative position,) multiplied respectively by undetermined factors  $\lambda, \Lambda$ ; and lastly, to equate the whole result to  $\delta V_r - t \delta H_r$ ,  $H_r$  being the quantity independent of the time in the equation (50.) of relative living force, and  $V_r$  being the relative action, of which we desired to express the variation. It is not necessary to dwell here on the demonstration of this new rule (V<sup>1</sup>.), which may easily be deduced from the principles already laid down; or by the calculus of variations from the law of relative living force, combined with the differential equations of second order of relative motion.

But to give an example of its application, let us resume the problem already mentioned, namely to express  $\delta V_r$  by means of the  $6n - 5$  independent variations  $\delta \xi_i \delta \eta_i \delta \zeta_i \delta \alpha_i \delta \beta_i \delta \gamma_i \delta H_r$ . For this purpose we shall employ a known transformation of the relative living force  $2T_r$ , multiplied by the sum of the masses of the system, namely the following:

$$2T_r \Sigma m = \Sigma .m_i m_k \{(x'_i - x'_k)^2 + (y'_i - y'_k)^2 + (z'_i - z'_k)^2\}; \quad (67.)$$

the sign of summation  $\Sigma$  extending, in the second member, to all the combinations of points two by two, which can be formed without repetition. This transformation gives, by (66.),

$$\left. \begin{aligned} 2T_r \Sigma m &= m_n \Sigma .m (\xi'^2 + \eta'^2 + \zeta'^2) \\ &+ \Sigma .m_i m_k \{(\xi'_i - \xi'_k)^2 + (\eta'_i - \eta'_k)^2 + (\zeta'_i - \zeta'_k)^2\}; \end{aligned} \right\} \quad (68.)$$

the sign of summation  $\Sigma$ , extending only to the first  $n - 1$  points of the system. Applying, therefore, our general rule or law of varying relative action, and observing that the  $6n - 6$  internal coordinates  $\xi \eta \zeta \alpha \beta \gamma$  are independent, we find the following new expression:

$$\left. \begin{aligned} \delta V_r = t \delta H_r + \frac{m_n}{\Sigma m} \cdot \Sigma_r . m (\xi' \delta \xi - \alpha' \delta \alpha + \eta' \delta \eta - \beta' \delta \beta + \zeta' \delta \zeta - \gamma' \delta \gamma) \\ + \frac{1}{\Sigma m} \cdot \Sigma_r . m_i m_k \{ (\xi'_i - \xi'_k) (\delta \xi_i - \delta \xi_k) + (\eta'_i - \eta'_k) (\delta \eta_i - \delta \eta_k) \\ + (\zeta'_i - \zeta'_k) (\delta \zeta_i - \delta \zeta_k) \} \\ - \frac{1}{\Sigma m} \cdot \Sigma_r . m_i m_k \{ (\alpha'_i - \alpha'_k) (\delta \alpha_i - \delta \alpha_k) + (\beta'_i - \beta'_k) (\delta \beta_i - \delta \beta_k) \\ + (\gamma'_i - \gamma'_k) (\delta \gamma_i - \delta \gamma_k) \} : \end{aligned} \right\} \quad (W^1.)$$

which gives, besides the equation (O<sup>1</sup>.), the following groups:

$$\left. \begin{aligned} \frac{\delta V_r}{\delta \xi_i} = \frac{m_i}{\Sigma m} \cdot \Sigma . m (\xi'_i - \xi') = m_i \left( \xi'_i - \frac{\Sigma_r m \xi'}{\Sigma m} \right), \\ \frac{\delta V_r}{\delta \eta_i} = \frac{m_i}{\Sigma m} \cdot \Sigma . m (\eta'_i - \eta') = m_i \left( \eta'_i - \frac{\Sigma_r m \eta'}{\Sigma m} \right), \\ \frac{\delta V_r}{\delta \zeta_i} = \frac{m_i}{\Sigma m} \cdot \Sigma . m (\zeta'_i - \zeta') = m_i \left( \zeta'_i - \frac{\Sigma_r m \zeta'}{\Sigma m} \right), \end{aligned} \right\} \quad (X^1.)$$

and

$$\left. \begin{aligned} \frac{\delta V_r}{\delta \alpha_i} = \frac{-m_i}{\Sigma m} \cdot \Sigma . m (\alpha'_i - \alpha') = -m_i \left( \alpha'_i - \frac{\Sigma_r m \alpha'}{\Sigma m} \right), \\ \frac{\delta V_r}{\delta \beta_i} = \frac{-m_i}{\Sigma m} \cdot \Sigma . m (\beta'_i - \beta') = -m_i \left( \beta'_i - \frac{\Sigma_r m \beta'}{\Sigma m} \right), \\ \frac{\delta V_r}{\delta \gamma_i} = \frac{-m_i}{\Sigma m} \cdot \Sigma . m (\gamma'_i - \gamma') = -m_i \left( \gamma'_i - \frac{\Sigma_r m \gamma'}{\Sigma m} \right); \end{aligned} \right\} \quad (Y^1.)$$

results which may be thus summed up:

$$\left. \begin{aligned} \delta V_r = t \delta H_r + \Sigma_r . m (\xi' \delta \xi - \alpha' \delta \alpha + \eta' \delta \eta - \beta' \delta \beta + \zeta' \delta \zeta - \gamma' \delta \gamma) \\ - \frac{1}{\Sigma m} (\Sigma_r m \xi' . \Sigma_r m \delta \xi + \Sigma_r m \eta' . \Sigma_r m \delta \eta + \Sigma_r m \zeta' . \Sigma_r m \delta \zeta) \\ + \frac{1}{\Sigma m} (\Sigma_r m \alpha' . \Sigma_r m \delta \alpha + \Sigma_r m \beta' . \Sigma_r m \delta \beta + \Sigma_r m \gamma' . \Sigma_r m \delta \gamma), \end{aligned} \right\} \quad (Z^1.)$$

and might have been otherwise deduced by our rule, from this other known transformation of  $T_r$ ,

$$T_r = \frac{1}{2} \Sigma_r . m (\xi'^2 + \eta'^2 + \zeta'^2) - \frac{(\Sigma_r m \xi')^2 + (\Sigma_r m \eta')^2 + (\Sigma_r m \zeta')^2}{2 \Sigma m}. \quad (69.)$$

And to obtain, with any set of internal or relative marks of position, the two partial differential equations which the characteristic function  $V_r$  of relative motion must satisfy, and

which offer (as we shall find) the chief means of discovering its form, namely, the equations analogous to those marked (F.) and (G.), we have only to eliminate the rates of increase of the marks of position of the system, which determine the final and initial components of the relative velocities of its points, by the law of varying relative action, from the final and initial expressions of the law of relative living force; namely, from the following equations:

$$T_t = U + H_r, \quad (50.)$$

and

$$T_{t_0} = U_0 + H_r. \quad (70.)$$

The law of areas, or the property respecting rotation which was expressed by the partial differential equations (P.), will also always admit of being expressed in relative coordinates, and will assist in discovering the form of the characteristic function  $V$ ; by showing that this function involves only such internal coordinates (in number  $6n - 9$ ) as do not alter by any common rotation of all points final and initial, round the centre of gravity, or round any other internal origin; that origin being treated as fixed, and the quantity  $H_r$  as constant, in determining the effects of this rotation. The general problem of dynamics, respecting the motions of a free system of  $n$  points attracting or repelling one another, is therefore reduced, in the last analysis, by the method of the present essay, to the research and differentiation of a function  $V$ , depending on  $6n - 9$  internal or relative coordinates, and on the quantity  $H_r$ , and satisfying a pair of partial differential equations of the first order and second degree; in integrating which equations, we are to observe, that at the assumed origin of the motion, namely at the moment when  $t = 0$ , the final or variable coordinates are equal to their initial values, and the partial differential coefficient  $\frac{\delta V}{\delta H_r}$  vanishes; and, that at a moment infinitely little distant, the differential alterations of the coordinates have ratios connected with the other partial differential coefficients of the characteristic function  $V$ , by the law of varying relative action. It may be here observed, that, although the consideration of the point, called usually the centre of gravity, is very simply suggested by the process of the tenth number, yet this internal centre is even more simply indicated by our early corollaries from the law of varying action; which show that the components of relative final velocities, in any system of attracting or repelling points, may be expressed by the differences of quantities of the form  $\frac{1}{m} \frac{\delta V}{\delta x}$ ,  $\frac{1}{m} \frac{\delta V}{\delta y}$ ,  $\frac{1}{m} \frac{\delta V}{\delta z}$ : and that therefore in calculating these relative velocities, it is advantageous to introduce the final sums  $\Sigma mx$ ,  $\Sigma my$ ,  $\Sigma mz$ , and, for an analogous reason, the initial sums  $\Sigma ma$ ,  $\Sigma mb$ ,  $\Sigma mc$ , among the marks of the extreme positions of the system, in the expression of the characteristic function  $V$ ; because, in differentiating that expression for the calculation of relative velocities, those sums may be treated as constant.

*On Systems of two Points, in general; Characteristic Function of the motion of any Binary System.*

13. To illustrate the foregoing principles, which extend to any free system of points, however numerous, attracting or repelling one another, let us now consider, in particular, a system of two such points. For such a system, the known *force-function*  $U$  becomes, by (2.)

$$U = m_1 m_2 f(r), \quad (71.)$$

$r$  being the mutual distance

$$r = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \quad (72.)$$

between the two points  $m_1$ ,  $m_2$ , and  $f(r)$  being a function of this distance such that its derivative or differential coefficient  $f'(r)$  expresses the law of their repulsion or attraction, according as it is positive or negative. The known differential equations of motion, of the second order, are now, by (1.), comprised in the following formula:

$$m_1(x_1'' \delta x_1 + y_1'' \delta y_1 + z_1'' \delta z_1) + m_2(x_2'' \delta x_2 + y_2'' \delta y_2 + z_2'' \delta z_2) = m_1 m_2 \delta f(r); \quad (73.)$$

they are therefore, separately,

$$\left. \begin{aligned} x_1'' &= m_2 \frac{\delta f(r)}{\delta x_1}, & y_1'' &= m_2 \frac{\delta f(r)}{\delta y_1}, & z_1'' &= m_2 \frac{\delta f(r)}{\delta z_1}, \\ x_2'' &= m_1 \frac{\delta f(r)}{\delta x_2}, & y_2'' &= m_1 \frac{\delta f(r)}{\delta y_2}, & z_2'' &= m_1 \frac{\delta f(r)}{\delta z_2}. \end{aligned} \right\} \quad (74.)$$

The problem of integrating these equations consists in proposing to assign, by their means, six relations between the time  $t$ , the masses  $m_1$ ,  $m_2$ , the six varying coordinates  $x_1$ ,  $y_1$ ,  $z_1$ ,  $x_2$ ,  $y_2$ ,  $z_2$ , and their initial values and initial rates of increase  $a_1$ ,  $b_1$ ,  $c_1$ ,  $a_2$ ,  $b_2$ ,  $c_2$ ,  $a'_1$ ,  $b'_1$ ,  $c'_1$ ,  $a'_2$ ,  $b'_2$ ,  $c'_2$ . If we knew these six final integrals, and combined them with the initial form of the law of living force, or of the known intermediate integral

$$\frac{1}{2}m_1(x_1'^2 + y_1'^2 + z_1'^2) + \frac{1}{2}m_2(x_2'^2 + y_2'^2 + z_2'^2) = m_1 m_2 f(r) + H; \quad (75.)$$

that is, with the following formula,

$$\frac{1}{2}m_1(a_1'^2 + b_1'^2 + c_1'^2) + \frac{1}{2}m_2(a_2'^2 + b_2'^2 + c_2'^2) = m_1 m_2 f(r_0) + H, \quad (76.)$$

in which  $r_0$  is the initial distance

$$r_0 = \sqrt{(a_1 - a_2)^2 + (b_1 - b_2)^2 + (c_1 - c_2)^2}, \quad (77.)$$

and  $H$  is a constant quantity, introduced by integration; we could, by the combination of these seven relations, determine the time  $t$ , and the six initial components of velocity  $a'_1$ ,  $b'_1$ ,  $c'_1$ ,  $a'_2$ ,  $b'_2$ ,  $c'_2$ , as functions of the twelve final and initial coordinates  $x_1$ ,  $y_1$ ,  $z_1$ ,  $x_2$ ,  $y_2$ ,  $z_2$ ,  $a_1$ ,  $b_1$ ,  $c_1$ ,  $a_2$ ,  $b_2$ ,  $c_2$ , and of the quantity  $H$ , (involving also the masses:) we could therefore determine whatever else depends on the manner and time of motion of this system of two points, as a function of the same extreme coordinates and of the same quantity  $H$ . In particular, we could determine the action, or accumulated living force of the system, namely,

$$V = m_1 \int_0^t (x_1'^2 + y_1'^2 + z_1'^2) dt + m_2 \int_0^t (x_2'^2 + y_2'^2 + z_2'^2) dt, \quad (A^2.)$$

as a function of those thirteen quantities  $x_1$   $y_1$   $z_1$   $x_2$   $y_2$   $z_2$   $a_1$   $b_1$   $c_1$   $a_2$   $b_2$   $c_2$   $H$ : and might then calculate the variation of this function,

$$\left. \begin{aligned} \delta V = & \frac{\delta V}{\delta x_1} \delta x_1 + \frac{\delta V}{\delta y_1} \delta y_1 + \frac{\delta V}{\delta z_1} \delta z_1 + \frac{\delta V}{\delta x_2} \delta x_2 + \frac{\delta V}{\delta y_2} \delta y_2 + \frac{\delta V}{\delta z_2} \delta z_2 \\ & + \frac{\delta V}{\delta a_1} \delta a_1 + \frac{\delta V}{\delta b_1} \delta b_1 + \frac{\delta V}{\delta c_1} \delta c_1 + \frac{\delta V}{\delta a_2} \delta a_2 + \frac{\delta V}{\delta b_2} \delta b_2 + \frac{\delta V}{\delta c_2} \delta c_2 \\ & + \frac{\delta V}{\delta H} \delta H. \end{aligned} \right\} \quad (\text{B}^2.)$$

But the essence of our method consists in *forming previously the expression of this variation by our law of varying action*, namely,

$$\left. \begin{aligned} \delta V = & m_1(x'_1 \delta x_1 - a'_1 \delta a_1 + y'_1 \delta y_1 - b'_1 \delta b_1 + z'_1 \delta z_1 - c'_1 \delta c_1) \\ & + m_2(x'_2 \delta x_2 - a'_2 \delta a_2 + y'_2 \delta y_2 - b'_2 \delta b_2 + z'_2 \delta z_2 - c'_2 \delta c_2) \\ & + t \delta H; \end{aligned} \right\} \quad (\text{C}^2.)$$

and in *considering  $V$  as a characteristic function of the motion*, from the form of which may be deduced all the intermediate and all the final integrals of the known differential equations, by resolving the expression (C<sup>2</sup>.<sup>)</sup> into the following separate groups, (included in (C.) and (D.),)

$$\left. \begin{aligned} \frac{\delta V}{\delta x_1} = & m_1 x'_1, \quad \frac{\delta V}{\delta y_1} = m_1 y'_1, \quad \frac{\delta V}{\delta z_1} = m_1 z'_1, \\ \frac{\delta V}{\delta x_2} = & m_2 x'_2, \quad \frac{\delta V}{\delta y_2} = m_2 y'_2, \quad \frac{\delta V}{\delta z_2} = m_2 z'_2; \end{aligned} \right\} \quad (\text{D}^2.)$$

and

$$\left. \begin{aligned} \frac{\delta V}{\delta a_1} = & -m_1 a'_1, \quad \frac{\delta V}{\delta b_1} = -m_1 b'_1, \quad \frac{\delta V}{\delta c_1} = -m_1 c'_1, \\ \frac{\delta V}{\delta a_2} = & -m_2 a'_2, \quad \frac{\delta V}{\delta b_2} = -m_2 b'_2, \quad \frac{\delta V}{\delta c_2} = -m_2 c'_2; \end{aligned} \right\} \quad (\text{E}^2.)$$

besides this other equation, which had occurred before,

$$\frac{\delta V}{\delta H} = t. \quad (\text{E.})$$

By this new method, the difficulty of integrating the six known equations of motion of the second order (74.) is reduced to the search and differentiation of a single function  $V$ ; and to find the form of this function, we are to employ the following pair of partial differential equations of the first order:

$$\left. \begin{aligned} \frac{1}{2m_1} \left\{ \left( \frac{\delta V}{\delta x_1} \right)^2 + \left( \frac{\delta V}{\delta y_1} \right)^2 + \left( \frac{\delta V}{\delta z_1} \right)^2 \right\} + \frac{1}{2m_2} \left\{ \left( \frac{\delta V}{\delta x_2} \right)^2 + \left( \frac{\delta V}{\delta y_2} \right)^2 + \left( \frac{\delta V}{\delta z_2} \right)^2 \right\} \\ = m_1 m_2 f(r) + H, \end{aligned} \right\} \quad (\text{F}^2.)$$

$$\left. \begin{aligned} & \frac{1}{2m_1} \left\{ \left( \frac{\delta V}{\delta a_1} \right)^2 + \left( \frac{\delta V}{\delta b_1} \right)^2 + \left( \frac{\delta V}{\delta c_1} \right)^2 \right\} + \frac{1}{2m_2} \left\{ \left( \frac{\delta V}{\delta a_2} \right)^2 + \left( \frac{\delta V}{\delta b_2} \right)^2 + \left( \frac{\delta V}{\delta c_2} \right)^2 \right\} \\ & = m_1 m_2 f(r_0) + H, \end{aligned} \right\} \quad (\text{G}^2.)$$

combined with some simple considerations. And it easily results from the principles already laid down, that the integral of this pair of equations, adapted to the present question, is

$$\left. \begin{aligned} V &= \sqrt{(x_{II} - a_{II})^2 + (y_{II} - b_{II})^2 + (z_{II} - c_{II})^2} \cdot \sqrt{2H_{II}(m_1 + m_2)} \\ &+ \frac{m_1 m_2}{m_1 + m_2} \left( h\vartheta + \int_{r_0}^r \rho dr \right); \end{aligned} \right\} \quad (\text{H}^2.)$$

in which  $x_{II}$ ,  $y_{II}$ ,  $z_{II}$ ,  $a_{II}$ ,  $b_{II}$ ,  $c_{II}$  denote the coordinates, final and initial, of the centre of gravity of the system,

$$\left. \begin{aligned} x_{II} &= \frac{m_1 x_1 + m_2 x_2}{m_1 + m_2}, & y_{II} &= \frac{m_1 y_1 + m_2 y_2}{m_1 + m_2}, & z_{II} &= \frac{m_1 z_1 + m_2 z_2}{m_1 + m_2}, \\ a_{II} &= \frac{m_1 a_1 + m_2 a_2}{m_1 + m_2}, & b_{II} &= \frac{m_1 b_1 + m_2 b_2}{m_1 + m_2}, & c_{II} &= \frac{m_1 c_1 + m_2 c_2}{m_1 + m_2}, \end{aligned} \right\} \quad (78.)$$

and  $\vartheta$  is the angle between the final and initial distances  $r$ ,  $r_0$ : we have also put for abridgement

$$\rho = \pm \sqrt{2(m_1 + m_2) \left( f(r) + \frac{H_r}{m_1 m_2} \right) - \frac{h^2}}, \quad (79.)$$

the upper or the lower sign to be used, according as the distance  $r$  is increasing or decreasing, and have introduced three auxiliary quantities  $h$ ,  $H_r$ ,  $H_{II}$ , to be determined by this condition,

$$0 = \vartheta + \int_{r_0}^r \frac{\delta \rho}{\delta h} dr, \quad (\text{I}^2.)$$

combined with the two following,

$$\left. \begin{aligned} \frac{m_1 m_2}{m_1 + m_2} \int_{r_0}^r \frac{\delta \rho}{\delta H_r} dr &= \sqrt{(x_{II} - a_{II})^2 + (y_{II} - b_{II})^2 + (z_{II} - c_{II})^2} \cdot \sqrt{\frac{m_1 + m_2}{2H_{II}}}, \\ H_r + H_{II} &= H; \end{aligned} \right\} \quad (\text{K}^2.)$$

which auxiliary quantities, although in one view they are functions of the twelve extreme coordinates, are yet to be treated as constant in calculating the three definite integrals, or limits of sums of numerous small elements,

$$\int_{r_0}^r \rho dr, \quad \int_{r_0}^r \frac{\delta \rho}{\delta h} dr, \quad \int_{r_0}^r \frac{\delta \rho}{\delta H_r} dr.$$

The form ( $\text{H}^2.$ ), for the *characteristic function of a binary system*, may be regarded as a central or radical relation, which includes the whole theory of the motion of such a system; so that all the details of this motion may be deduced from it by the application of our general method. But because the theory of binary systems has been brought to great perfection already, by the labours of former writers, it may suffice to give briefly here a few instances of such deduction.

14. The form (H<sup>2</sup>.), for the characteristic function of a binary system, involves explicitly, when  $\rho$  is changed to its value (79.), the twelve quantities  $x_{\text{II}}$ ,  $y_{\text{II}}$ ,  $z_{\text{II}}$ ,  $a_{\text{II}}$ ,  $b_{\text{II}}$ ,  $c_{\text{II}}$ ,  $r$ ,  $r_0$ ,  $\vartheta$ ,  $h$ ,  $H$ ,  $H_{\text{II}}$ , (besides the masses  $m_1$ ,  $m_2$  which are always considered as given;) its variation may therefore be thus expressed:

$$\left. \begin{aligned} \delta V = & \frac{\delta V}{\delta x_{\text{II}}} \delta x_{\text{II}} + \frac{\delta V}{\delta y_{\text{II}}} \delta y_{\text{II}} + \frac{\delta V}{\delta z_{\text{II}}} \delta z_{\text{II}} + \frac{\delta V}{\delta a_{\text{II}}} \delta a_{\text{II}} + \frac{\delta V}{\delta b_{\text{II}}} \delta b_{\text{II}} + \frac{\delta V}{\delta c_{\text{II}}} \delta c_{\text{II}} \\ & + \frac{\delta V}{\delta r} \delta r + \frac{\delta V}{\delta r_0} \delta r_0 + \frac{\delta V}{\delta \vartheta} \delta \vartheta + \frac{\delta V}{\delta H} \delta H + \frac{\delta V}{\delta H_{\text{II}}} \delta H_{\text{II}}. \end{aligned} \right\} \quad (\text{L}^2.)$$

In this expression, if we put for abridgement

$$\lambda = \sqrt{\frac{2H_{\text{II}}(m_1 + m_2)}{(x_{\text{II}} - a_{\text{II}})^2 + (y_{\text{II}} - b_{\text{II}})^2 + (z_{\text{II}} - c_{\text{II}})^2}}, \quad (80.)$$

we shall have

$$\left. \begin{aligned} \frac{\delta V}{\delta x_{\text{II}}} &= \lambda(x_{\text{II}} - a_{\text{II}}), & \frac{\delta V}{\delta y_{\text{II}}} &= \lambda(y_{\text{II}} - b_{\text{II}}), & \frac{\delta V}{\delta z_{\text{II}}} &= \lambda(z_{\text{II}} - c_{\text{II}}), \\ \frac{\delta V}{\delta a_{\text{II}}} &= \lambda(a_{\text{II}} - x_{\text{II}}), & \frac{\delta V}{\delta b_{\text{II}}} &= \lambda(b_{\text{II}} - y_{\text{II}}), & \frac{\delta V}{\delta c_{\text{II}}} &= \lambda(c_{\text{II}} - z_{\text{II}}); \end{aligned} \right\} \quad (\text{M}^2.)$$

and if we put

$$\rho_0 = \pm \sqrt{2(m_1 + m_2) \left( f(r_0) + \frac{H}{m_1 m_2} \right) - \frac{h_2}{r_0^2}}, \quad (81.)$$

the sign of the radical being determined by the same rule as that of  $\rho$ , we shall have

$$\frac{\delta V}{\delta r} = \frac{m_1 m_2 \rho}{m_1 + m_2}, \quad \frac{\delta V}{\delta r_0} = \frac{-m_1 m_2 \rho_0}{m_1 + m_2}, \quad \frac{\delta V}{\delta \vartheta} = \frac{m_1 m_2 h}{m_1 + m_2}; \quad (\text{N}^2.)$$

besides, by the equations of condition (I<sup>2</sup>.), (K<sup>2</sup>.), we have

$$\frac{\delta V}{\delta h} = 0, \quad (\text{O}^2.)$$

and

$$\frac{\delta V}{\delta H_{\text{II}}} = \frac{\delta V}{\delta H} = \int_{r_0}^r \frac{dr}{\rho}, \quad \delta H + \delta H_{\text{II}} = \delta H. \quad (\text{P}^2.)$$

The expression (L<sup>2</sup>. ) may therefore be thus transformed:

$$\left. \begin{aligned} \delta V = & \lambda \{ (x_{\text{II}} - a_{\text{II}})(\delta x_{\text{II}} - \delta a_{\text{II}}) + (y_{\text{II}} - b_{\text{II}})(\delta y_{\text{II}} - \delta b_{\text{II}}) + (z_{\text{II}} - c_{\text{II}})(\delta z_{\text{II}} - \delta c_{\text{II}}) \} \\ & + \frac{m_1 m_2}{m_1 + m_2} (\rho \delta r - \rho_0 \delta r_0 + h \delta \vartheta) + \int_{r_0}^r \frac{\delta r}{\rho} \cdot \delta H; \end{aligned} \right\} \quad (\text{Q}^2.)$$

and may be resolved by our general method into twelve separate expressions for the final and initial components of velocities, namely,

$$\left. \begin{aligned} x'_1 &= \frac{1}{m_1} \frac{\delta V}{\delta x_1} = \frac{\lambda}{m_1 + m_2} (x_{\prime\prime} - a_{\prime\prime}) + \frac{m_2}{m_1 + m_2} \left( \rho \frac{\delta r}{\delta x_1} + h \frac{\delta \vartheta}{\delta x_1} \right), \\ y'_1 &= \frac{1}{m_1} \frac{\delta V}{\delta y_1} = \frac{\lambda}{m_1 + m_2} (y_{\prime\prime} - b_{\prime\prime}) + \frac{m_2}{m_1 + m_2} \left( \rho \frac{\delta r}{\delta y_1} + h \frac{\delta \vartheta}{\delta y_1} \right), \\ z'_1 &= \frac{1}{m_1} \frac{\delta V}{\delta z_1} = \frac{\lambda}{m_1 + m_2} (z_{\prime\prime} - c_{\prime\prime}) + \frac{m_2}{m_1 + m_2} \left( \rho \frac{\delta r}{\delta z_1} + h \frac{\delta \vartheta}{\delta z_1} \right), \\ x'_2 &= \frac{1}{m_2} \frac{\delta V}{\delta x_2} = \frac{\lambda}{m_1 + m_2} (x_{\prime\prime} - a_{\prime\prime}) + \frac{m_1}{m_1 + m_2} \left( \rho \frac{\delta r}{\delta x_2} + h \frac{\delta \vartheta}{\delta x_2} \right), \\ y'_2 &= \frac{1}{m_2} \frac{\delta V}{\delta y_2} = \frac{\lambda}{m_1 + m_2} (y_{\prime\prime} - b_{\prime\prime}) + \frac{m_1}{m_1 + m_2} \left( \rho \frac{\delta r}{\delta y_2} + h \frac{\delta \vartheta}{\delta y_2} \right), \\ z'_2 &= \frac{1}{m_2} \frac{\delta V}{\delta z_2} = \frac{\lambda}{m_1 + m_2} (z_{\prime\prime} - c_{\prime\prime}) + \frac{m_1}{m_1 + m_2} \left( \rho \frac{\delta r}{\delta z_2} + h \frac{\delta \vartheta}{\delta z_2} \right), \end{aligned} \right\} \quad (\text{R}^2.)$$

and

$$\left. \begin{aligned} a'_1 &= \frac{-1}{m_1} \frac{\delta V}{\delta a_1} = \frac{\lambda}{m_1 + m_2} (x_{\prime\prime} - a_{\prime\prime}) + \frac{m_2}{m_1 + m_2} \left( \rho_0 \frac{\delta r_0}{\delta a_1} - h \frac{\delta \vartheta}{\delta a_1} \right), \\ b'_1 &= \frac{-1}{m_1} \frac{\delta V}{\delta b_1} = \frac{\lambda}{m_1 + m_2} (y_{\prime\prime} - b_{\prime\prime}) + \frac{m_2}{m_1 + m_2} \left( \rho_0 \frac{\delta r_0}{\delta b_1} - h \frac{\delta \vartheta}{\delta b_1} \right), \\ c'_1 &= \frac{-1}{m_1} \frac{\delta V}{\delta c_1} = \frac{\lambda}{m_1 + m_2} (z_{\prime\prime} - c_{\prime\prime}) + \frac{m_2}{m_1 + m_2} \left( \rho_0 \frac{\delta r_0}{\delta c_1} - h \frac{\delta \vartheta}{\delta c_1} \right), \\ a'_2 &= \frac{-1}{m_2} \frac{\delta V}{\delta a_2} = \frac{\lambda}{m_1 + m_2} (x_{\prime\prime} - a_{\prime\prime}) + \frac{m_1}{m_1 + m_2} \left( \rho_0 \frac{\delta r_0}{\delta a_2} - h \frac{\delta \vartheta}{\delta a_2} \right), \\ b'_2 &= \frac{-1}{m_2} \frac{\delta V}{\delta b_2} = \frac{\lambda}{m_1 + m_2} (y_{\prime\prime} - b_{\prime\prime}) + \frac{m_1}{m_1 + m_2} \left( \rho_0 \frac{\delta r_0}{\delta b_2} - h \frac{\delta \vartheta}{\delta b_2} \right), \\ c'_2 &= \frac{-1}{m_2} \frac{\delta V}{\delta c_2} = \frac{\lambda}{m_1 + m_2} (z_{\prime\prime} - c_{\prime\prime}) + \frac{m_1}{m_1 + m_2} \left( \rho_0 \frac{\delta r_0}{\delta c_2} - h \frac{\delta \vartheta}{\delta c_2} \right); \end{aligned} \right\} \quad (\text{S}^2.)$$

besides the following expression for the time of motion of the system:

$$t = \frac{\delta V}{\delta H} = \int_{r_0}^r \frac{dr}{\rho}, \quad (\text{T}^2.)$$

which gives by (K<sup>2</sup>.), and by (79.), (80.),

$$t = \frac{m_1 + m_2}{\lambda}. \quad (\text{U}^2.)$$

The six equations (R<sup>2</sup>.) give the six intermediate integrals, and the six equations (S<sup>2</sup>.) give the six final integrals of the six known differential equations of motion (74.) for any binary system, if we eliminate or determine the three auxiliary quantities  $h$ ,  $H_r$ ,  $H_{\prime\prime}$ , by the

three conditions (I<sup>2</sup>.) (T<sup>2</sup>.) (U<sup>2</sup>.). Thus, if we observe that the distances  $r$ ,  $r_0$ , and the included angle  $\vartheta$ , depend only on relative coordinates, which may be thus denoted,

$$\left. \begin{aligned} x_1 - x_2 &= \xi, & y_1 - y_2 &= \eta, & z_1 - z_2 &= \zeta, \\ a_1 - a_2 &= \alpha, & b_1 - b_2 &= \beta, & c_1 - c_2 &= \gamma, \end{aligned} \right\} \quad (82.)$$

we obtain by easy combinations the three following intermediate integrals for the centre of gravity of the system:

$$x_{\text{II}}'t = x_{\text{II}} - a_{\text{II}}, \quad y_{\text{II}}'t = y_{\text{II}} - b_{\text{II}}, \quad z_{\text{II}}'t = z_{\text{II}} - c_{\text{II}}, \quad (83.)$$

and the three following final integrals,

$$a_{\text{II}}'t = x_{\text{II}} - a_{\text{II}}, \quad b_{\text{II}}'t = y_{\text{II}} - b_{\text{II}}, \quad c_{\text{II}}'t = z_{\text{II}} - c_{\text{II}}, \quad (84.)$$

expressing the well-known law of the rectilinear and uniform motion of that centre. We obtain also the three following intermediate integrals for the relative motion of one point of the system about the other:

$$\left. \begin{aligned} \xi' &= \rho \frac{\delta r}{\delta \xi} + h \frac{\delta \vartheta}{\delta \xi}, \\ \eta' &= \rho \frac{\delta r}{\delta \eta} + h \frac{\delta \vartheta}{\delta \eta}, \\ \zeta' &= \rho \frac{\delta r}{\delta \zeta} + h \frac{\delta \vartheta}{\delta \zeta}, \end{aligned} \right\} \quad (85.)$$

and the three following final integrals,

$$\left. \begin{aligned} \alpha' &= \rho_0 \frac{\delta r_0}{\delta \alpha} - h \frac{\delta \vartheta}{\delta \alpha}, \\ \beta' &= \rho_0 \frac{\delta r_0}{\delta \beta} - h \frac{\delta \vartheta}{\delta \beta}, \\ \gamma' &= \rho_0 \frac{\delta r_0}{\delta \gamma} - h \frac{\delta \vartheta}{\delta \gamma}; \end{aligned} \right\} \quad (86.)$$

in which the auxiliary quantities  $h$ ,  $H$ , are to be determined by (I<sup>2</sup>.), (T<sup>2</sup>.), and in which the dependence of  $r$ ,  $r_0$ ,  $\vartheta$ , on  $\xi$ ,  $\eta$ ,  $\zeta$ ,  $\alpha$ ,  $\beta$ ,  $\gamma$ , is expressed by the following equations:

$$\left. \begin{aligned} r &= \sqrt{\xi^2 + \eta^2 + \zeta^2}, & r_0 &= \sqrt{\alpha^2 + \beta^2 + \gamma^2}, \\ rr_0 \cos \vartheta &= \xi \alpha + \eta \beta + \zeta \gamma. \end{aligned} \right\} \quad (87.)$$

If then we put, for abridgement,

$$A = \frac{\rho}{r} + \frac{h}{r^2 \tan \vartheta}, \quad B = \frac{h}{rr_0 \sin \vartheta}, \quad C = \frac{-\rho_0}{r_0} + \frac{h}{r_0^2 \tan \vartheta}, \quad (88.)$$

we shall have these three intermediate integrals,

$$\xi' = A\xi - B\alpha, \quad \eta' = A\eta - B\beta, \quad \zeta' = A\zeta - B\gamma, \quad (89.)$$

and these three final integrals,

$$\alpha' = B\xi - C\alpha, \quad \beta' = B\eta - C\beta, \quad \gamma' = B\zeta - C\gamma, \quad (90.)$$

of the equations of relative motion. These integrals give,

$$\left. \begin{aligned} \xi\eta' - \eta\xi' &= \alpha\beta' - \beta\alpha' = B(\alpha\eta - \beta\xi), \\ \eta\zeta' - \zeta\eta' &= \beta\gamma' - \gamma\beta' = B(\beta\zeta - \gamma\eta), \\ \zeta\xi' - \xi\zeta' &= \gamma\alpha' - \alpha\gamma' = B(\gamma\xi - \alpha\zeta), \end{aligned} \right\} \quad (91.)$$

and

$$\zeta(\alpha\beta' - \beta\alpha') + \xi(\beta\gamma' - \gamma\beta') + \eta(\gamma\alpha' - \alpha\gamma') = 0; \quad (92.)$$

they contain therefore the known law of equable description of areas, and the law of a plane relative orbit. If we take for simplicity this plane for the plane  $\xi \eta$ , the quantities  $\zeta \zeta' \gamma \gamma'$  will vanish; and we may put,

$$\left. \begin{aligned} \xi &= r \cos \theta, & \eta &= r \sin \theta, & \zeta &= 0, \\ \alpha &= r_0 \cos \theta_0, & \beta &= r_0 \sin \theta_0, & \gamma &= 0, \end{aligned} \right\} \quad (93.)$$

and

$$\left. \begin{aligned} \xi' &= r' \cos \theta - \theta' r \sin \theta, & \eta' &= r' \sin \theta + \theta' r \cos \theta, & \zeta' &= 0, \\ \alpha' &= r'_0 \cos \theta_0 - \theta'_0 r_0 \sin \theta_0, & \beta' &= r'_0 \sin \theta_0 + \theta'_0 r_0 \cos \theta_0, & \gamma' &= 0, \end{aligned} \right\} \quad (94.)$$

the angles  $\theta \theta_0$  being counted from some fixed line in the plane, and being such that their difference

$$\theta - \theta_0 = \vartheta. \quad (95.)$$

These values give

$$\xi\eta' - \eta\xi' = r^2\theta', \quad \alpha\beta' - \beta\alpha' = r_0^2\theta'_0, \quad \alpha\eta - \beta\xi = rr_0 \sin \vartheta, \quad (96.)$$

and therefore, by (88.) and (91.),

$$r^2\theta' = r_0^2\theta'_0 = h; \quad (97.)$$

the quantity  $\frac{1}{2}h$  is therefore the constant areal velocity in the relative motion of the system; a result which is easily seen to be independent of the directions of the three rectangular coordinates. The same values (93.), (94.), give

$$\left. \begin{aligned} \xi \cos \theta + \eta \sin \theta &= r, & \xi' \cos \theta + \eta' \sin \theta &= r', & \alpha \cos \theta + \beta \sin \theta &= r_0 \cos \vartheta, \\ \alpha \cos \theta_0 + \beta \sin \theta_0 &= r_0, & \alpha' \cos \theta_0 + \beta' \sin \theta_0 &= r'_0, & \xi \cos \theta_0 + \eta \sin \theta_0 &= r \cos \vartheta, \end{aligned} \right\} \quad (98.)$$

and therefore, by the intermediate and final integrals, (89.), (90.),

$$r' = \rho, \quad r'_0 = \rho_0; \quad (99.)$$

results which evidently agree with the condition (T<sup>2</sup>.), and which give by (79.) and (81.), for all directions of coordinates,

$$r'^2 + \frac{h^2}{r^2} - 2(m_1 + m_2)f(r) = r'^2_0 + \frac{h^2}{r^2_0} - 2(m_1 + m_2)f(r_0) = 2H, \left( \frac{1}{m_1} + \frac{1}{m_2} \right); \quad (100.)$$

the other auxiliary quantity  $H$ , is therefore also a constant, independent of the time, and enters as such into the constant part in the expression for  $\left( r'^2 + \frac{h^2}{r^2} \right)$  the square of the relative velocity. The equation of condition (I<sup>2</sup>.), connecting these two constants  $h$ ,  $H$ , with the extreme lengths of the radius vector  $r$ , and with the angle  $\vartheta$  described by this radius in revolving from its initial to its final direction, is the equation of the plane relative orbit; and the other equation of condition (T<sup>2</sup>.), connecting the same two constants with the same extreme distances and with the time, gives the law of the velocity of mutual approach or recess.

We may remark that the part  $V_r$  of the whole characteristic function  $V$ , which represents the relative action and determines the relative motion in the system, namely,

$$V_r = \frac{m_1 m_2}{m_1 + m_2} \left( h\vartheta + \int_{r_0}^r \rho dr \right), \quad (V^2.)$$

may be put, by (I<sup>2</sup>.), under the form

$$V_r = \frac{m_1 m_2}{m_1 + m_2} \int_{r_0}^r \left( \rho - h \frac{\delta \rho}{\delta h} \right) dr, \quad (W^2.)$$

or finally, by (79.)

$$V_r = 2 \int_{r_0}^r \frac{m_1 m_2 f(r) + H_r}{\rho} dr; \quad (X^2.)$$

the condition (I<sup>2</sup>.) may also itself be transformed, by (79.), as follows:

$$\vartheta = h \int_{r_0}^r \frac{dr}{r^2 \rho}; \quad (Y^2.)$$

results which all admit of easy verifications. The partial differential equations connected with the law of relative living force, which the characteristic function  $V_r$  of relative motion must satisfy, may be put under the following forms:

$$\left. \begin{aligned} \left( \frac{\delta V_r}{\delta r} \right)^2 + \frac{1}{r^2} \left( \frac{\delta V_r}{\delta \vartheta} \right)^2 &= \frac{2m_1 m_2}{m_1 + m_2} (U + H_r), \\ \left( \frac{\delta V_r}{\delta r_0} \right)^2 + \frac{1}{r_0^2} \left( \frac{\delta V_r}{\delta \vartheta} \right)^2 &= \frac{2m_1 m_2}{m_1 + m_2} (U_0 + H_r); \end{aligned} \right\} \quad (Z^2.)$$

and if the first of the equations of this pair have its variation taken with respect to  $r$  and  $\vartheta$ , attention being paid to the dynamical meanings of the coefficients of the characteristic function, it will conduct (as in former instances) to the known differential equations of motion of the second order.

*On the undisturbed Motion of a Planet or Comet about the Sun: Dependence of the Characteristic Function of such Motion, on the chord and the sum of the Radii.*

15. To particularize still further, let

$$f(r) = \frac{1}{r}, \quad (101.)$$

that is, let us consider a binary system, such as a planet or comet and the sun, with the Newtonian law of attraction; and let us put, for abridgement,

$$m_1 + m_2 = \mu, \quad \frac{h^2}{\mu} = p, \quad \frac{-m_1 m_2}{2H} = a. \quad (102.)$$

The characteristic function  $V$  of relative motion may now be expressed as follows

$$V = \frac{m_1 m_2}{\sqrt{\mu}} \left( \vartheta \sqrt{p} + \int_{r_0}^r \pm \sqrt{\frac{2}{r} - \frac{1}{a} - \frac{p}{r^2}} dr \right); \quad (A^3.)$$

in which  $p$  is to be considered as a function of the extreme radii vectores  $r, r_0$ , and of their included angle  $\vartheta$ , involving also the quantity  $a$ , or the connected quantity  $H$ , and determined by the condition

$$\vartheta = \int_{r_0}^r \frac{\pm dr}{r^2 \sqrt{\frac{2}{rp} - \frac{1}{ap} - \frac{1}{r^2}}} \quad (B^3.)$$

that is, by the derivative of the formula (A<sup>3</sup>.), taken with respect to  $p$ ; the upper sign being taken in each expression when the distance  $r$  is increasing, and the lower sign when that distance is diminishing, and the quantity  $p$  being treated as constant in calculating the two definite integrals. It results from the foregoing remarks, that this quantity  $p$  is constant also in the sense of being independent of the time, so as not to vary in the course of the motion; and that the condition (B<sup>3</sup>.), connecting this constant with  $r, r_0, \vartheta, a$ , is the equation of the plane relative orbit; which is therefore (as it has long been known to be) an ellipse, hyperbola, or parabola, according as the constant  $a$  is positive, negative, or zero, the origin of  $r$  being always a focus of the curve, and  $p$  being the semiparameter. It results also, that the time of motion may be thus expressed:

$$t = \frac{\delta V}{\delta H} = \frac{2a^2}{m_1 m_2} \frac{\delta V}{\delta a}, \quad (C^3.)$$

and therefore thus:

$$t = \int_{r_0}^r \frac{\pm dr}{\sqrt{\frac{2\mu}{r} - \frac{\mu}{a} - \frac{\mu p}{r^2}}}; \quad (D^3.)$$

which latter is a known expression. Confining ourselves at present to the case  $a > 0$ , and introducing the known auxiliary quantities called excentricity and excentric anomaly, namely,

$$e = \sqrt{1 - \frac{p}{a}}, \quad (103.)$$

and

$$v = \cos^{-1} \left( \frac{a - r}{ae} \right), \quad (104.)$$

which give

$$\pm \sqrt{2ar - r^2 - pa} = ae \sin v, \quad (105.)$$

$v$  being considered as continually increasing with the time; and therefore, as is well known,

$$\begin{aligned} r &= a(1 - e \cos v), \quad r_0 = a(1 - e \cos v_0), \\ \vartheta &= 2 \tan^{-1} \left\{ \sqrt{\frac{1+e}{1-e}} \tan \frac{v}{2} \right\} - 2 \tan^{-1} \left\{ \sqrt{\frac{1+e}{1-e}} \tan \frac{v_0}{2} \right\}, \end{aligned} \quad (106.)$$

and

$$t = \sqrt{\frac{a^3}{\mu}} \cdot (v - v_0 - e \sin v + e \sin v_0); \quad (107.)$$

we find that this expression for the characteristic function of relative motion,

$$V_r = \frac{m_1 m_2}{\sqrt{\mu}} \int_{r_0}^r \frac{\pm \left( \frac{2}{r} - \frac{1}{a} \right) dr}{\sqrt{\frac{2}{r} - \frac{1}{a} - \frac{p}{r^2}}}, \quad (E^3.)$$

deduced from (A<sup>3</sup>.) and (B<sup>3</sup>.), may be transformed as follows:

$$V_r = m_1 m_2 \sqrt{\frac{a}{\mu}} (v - v_0 + e \sin v - e \sin v_0); \quad (F^3.)$$

in which the excentricity  $e$ , and the final and initial excentric anomalies  $v$ ,  $v_0$ , are to be considered as functions of the final and initial radii  $r$ ,  $r_0$ , and of the included angle  $\vartheta$ , determined by the equations (106.). The expression (F<sup>3</sup>.) may be thus written:

$$V_r = 2m_1 m_2 \sqrt{\frac{a}{\mu}} (v_r + e_r \sin v_r), \quad (G^3.)$$

if we put, for abridgement,

$$v_r = \frac{v - v_0}{2}, \quad e_r = e \cos \frac{v + v_0}{2}; \quad (108.)$$

for the complete determination of the characteristic function of the present relative motion, it remains therefore to determine the two variables  $v_r$  and  $e_r$ , as functions of  $r$ ,  $r_0$ ,  $\vartheta$ , or of some other set of quantities which mark the shape and size of the plane triangle bounded by the final and initial elliptic radii vectores and by the elliptic chord.

For this purpose it is convenient to introduce this elliptic chord itself, which we shall call  $\pm\tau$ , so that

$$\tau^2 = r^2 + r_0^2 - 2rr_0 \cos \vartheta; \quad (109.)$$

because this chord may be expressed as a function of the two variables  $v_t$ ,  $e_t$ , (involving also the mean distance  $a_t$ ,) as follows. The value (106.) for the angle  $\vartheta$ , that is, by (95.), for  $\theta - \theta_0$ , gives

$$\theta - 2 \tan^{-1} \left\{ \sqrt{\frac{1+e}{1-e}} \tan \frac{v}{2} \right\} = \theta_0 - 2 \tan^{-1} \left\{ \sqrt{\frac{1+e}{1-e}} \tan \frac{v_0}{2} \right\} = \varpi, \quad (110.)$$

$\varpi$  being a new constant independent of the time, namely, one of the values of the polar angle  $\theta$ , which correspond to the minimum of radius vector; and therefore, by (106.),

$$\left. \begin{aligned} r \cos(\theta - \varpi) &= a(\cos v - e), & r \sin(\theta - \varpi) &= a\sqrt{1-e^2} \sin v, \\ r_0 \cos(\theta_0 - \varpi) &= a(\cos v_0 - e), & r_0 \sin(\theta_0 - \varpi) &= a\sqrt{1-e^2} \sin v_0; \end{aligned} \right\} \quad (111.)$$

expressions which give the following value for the square of the elliptic chord:

$$\left. \begin{aligned} \tau^2 &= \{r \cos(\theta - \varpi) - r_0 \cos(\theta_0 - \varpi)\}^2 + \{r \sin(\theta - \varpi) - r_0 \sin(\theta_0 - \varpi)\}^2 \\ &= a^2 \{(\cos v - \cos v_0)^2 + (1 - e^2)(\sin v - \sin v_0)^2\} \\ &= 4a^2 \sin v_t^2 \left\{ \left( \sin \frac{v+v_0}{2} \right)^2 + (1 - e^2) \left( \cos \frac{v+v_0}{2} \right)^2 \right\} \\ &= 4a^2(1 - e_t^2) \sin v_t^2 : \end{aligned} \right\} \quad (112.)$$

we may also consider  $\tau$  as having the same sign with  $\sin v_t$ , if we consider it as alternately positive and negative, in the successive elliptic periods or revolutions, beginning with the initial position.

Besides, if we denote by  $\sigma$  the sum of the two elliptic radii vectores, final and initial, so that

$$\sigma = r + r_0, \quad (113.)$$

we shall have, with our present abridgements,

$$\sigma = 2a(1 - e, \cos v_t); \quad (114.)$$

the variables  $v$ ,  $e$ , are therefore functions of  $\sigma$ ,  $\tau$ ,  $a$ , and consequently the characteristic function  $V$ , is itself a function of those three quantities. We may therefore put

$$V_t = \frac{m_1 m_2 w}{m_1 + m_2}, \quad (\text{H}^3.)$$

$w$  being a function of  $\sigma$ ,  $\tau$ ,  $a$ , of which the form is to be determined by eliminating  $v$ ,  $e$ , between the three equations,

$$\left. \begin{aligned} w &= 2\sqrt{\mu a}(v_t + e_t \sin v_t), \\ \sigma &= 2a(1 - e_t \cos v_t), \\ \tau &= 2a(1 - e_t^2)^{\frac{1}{2}} \sin v_t; \end{aligned} \right\} \quad (\text{I}^3.)$$

and we may consider this new function  $w$  as itself a characteristic function of elliptic motion; the law of its variation being expressed as follows, in the notation of the present essay:

$$\delta w = \xi' \delta \xi - \alpha' \delta \alpha + \eta' \delta \eta - \beta' \delta \beta + \zeta' \delta \zeta - \gamma' \delta \gamma + \frac{t\mu \delta a}{2a^2}. \quad (\text{K}^3.)$$

In this expression  $\xi \eta \zeta$  are the relative coordinates of the point  $m_1$ , at the time  $t$ , referred to the other attracting point  $m_2$  as an origin, and to any three rectangular axes;  $\xi' \eta' \zeta'$  are their rates of increase, or the three rectangular components of final relative velocity;  $\alpha \beta \gamma$   $\alpha' \beta' \gamma'$  are the initial values, or values at the time zero, of these relative coordinates and components of relative velocity;  $a$  is a quantity independent of the time, namely, the mean distance of the two points  $m_1, m_2$ ; and  $\mu$  is the sum of their masses. And all the properties of the undisturbed elliptic motion of a planet or comet about the sun may be deduced in a new way, from the simplified characteristic function  $w$ , by comparing its variation ( $\text{K}^3.$ ) with the following other form,

$$\delta w = \frac{\delta w}{\delta \sigma} \delta \sigma + \frac{\delta w}{\delta \tau} \delta \tau + \frac{\delta w}{\delta a} \delta a; \quad (\text{L}^3.)$$

in which we are to observe that

$$\left. \begin{aligned} \sigma &= \sqrt{\xi^2 + \eta^2 + \zeta^2} + \sqrt{\alpha^2 + \beta^2 + \gamma^2}, \\ \tau &= \pm \sqrt{(\xi - \alpha)^2 + (\eta - \beta)^2 + (\zeta - \gamma)^2}. \end{aligned} \right\} \quad (\text{M}^3.)$$

By this comparison we are brought back to the general integral equations of the relative motion of a binary system, (89.) and (90.); but we have now the following particular values for the coefficients  $A, B, C$ :

$$A = \frac{1}{r} \frac{\delta w}{\delta \sigma} + \frac{1}{\tau} \frac{\delta w}{\delta \tau}, \quad B = \frac{1}{\tau} \frac{\delta w}{\delta \tau}, \quad C = \frac{1}{r_0} \frac{\delta w}{\delta \sigma} + \frac{1}{\tau} \frac{\delta w}{\delta \tau}; \quad (\text{N}^3.)$$

and with respect to the three partial differential coefficients,  $\frac{\delta w}{\delta \sigma}, \frac{\delta w}{\delta \tau}, \frac{\delta w}{\delta a}$ , we have the following relation between them:

$$a \frac{\delta w}{\delta a} + \sigma \frac{\delta w}{\delta \sigma} + \tau \frac{\delta w}{\delta \tau} = \frac{w}{2}, \quad (\text{O}^3.)$$

the function  $w$  being homogeneous of the dimension  $\frac{1}{2}$  with respect to the three quantities  $a, \sigma, \tau$ ; we have also, by (I<sup>3</sup>.),

$$\frac{\delta w}{\delta \sigma} = \sqrt{\frac{\mu}{a}} \cdot \frac{\sin v_r}{e_r - \cos v_r}, \quad \frac{\delta w}{\delta \tau} = \sqrt{\frac{\mu}{a}} \cdot \frac{\sqrt{1 - e_r^2}}{\cos v_r - e_r}, \quad (\text{P}^3.)$$

and therefore

$$\frac{\delta w}{\delta \sigma} \frac{\delta w}{\delta \tau} = \frac{-2\mu\tau}{\sigma^2 - \tau^2}, \quad \left( \frac{\delta w}{\delta \sigma} \right)^2 + \left( \frac{\delta w}{\delta \tau} \right)^2 + \frac{\mu}{a} = \frac{4\mu\sigma}{\sigma^2 - \tau^2}, \quad (\text{Q}^3.)$$

from which may be deduced the following remarkable expressions:

$$\left. \begin{aligned} \left( \frac{\delta w}{\delta \sigma} + \frac{\delta w}{\delta \tau} \right)^2 &= \frac{4\mu}{\sigma + \tau} - \frac{\mu}{a}, \\ \left( \frac{\delta w}{\delta \tau} - \frac{\delta w}{\delta \sigma} \right)^2 &= \frac{4\mu}{\sigma - \tau} - \frac{\mu}{a}. \end{aligned} \right\} \quad (\text{R}^3.)$$

These expressions will be found to be important in the application of the present method to the theory of elliptic motion.

16. We shall not enter, on this occasion, into any details of such application; but we may remark, that the circumstance of the characteristic function involving only the elliptic chord and the sum of the extreme radii, (besides the mean distance and the sum of the masses,) affords, by our general method, a new proof of the well-known theorem that the elliptic time also depends on the same chord and sum of radii; and gives a new expression for the law of this dependence, namely,

$$t = \frac{2a^2}{\mu} \frac{\delta w}{\delta a}. \quad (\text{S}^3.)$$

We may remark also, that the same form of the characteristic function of elliptic motion conducts, by our general method, to the following curious, but not novel property, of the ellipse, that if any two tangents be drawn to such a curve, from any common point outside, these tangents subtend equal angles at one focus; they subtend also equal angles at the other. Reciprocally, if any plane curve possess this property, when referred to a fixed point in its own plane, which may be taken as the origin of polar coordinates  $r, \theta$ , the curve must satisfy the following equation in mixed differences:

$$\cotan\left(\frac{\Delta\theta}{2}\right) \cdot \Delta \frac{1}{r} = (\Delta + 2) \frac{d}{d\theta} \frac{1}{r}, \quad (115.)$$

which may be brought to the following form,

$$\left( \frac{d}{d\theta} + \frac{d^3}{d\theta^3} \right) \frac{1}{r} = 0, \quad (116.)$$

and therefore gives, by integration,

$$r = \frac{p}{1 + e \cos(\theta - \varpi)}; \quad (117.)$$

the curve is, consequently, a conic section, and the fixed point is one of its foci.

The properties of parabolic are included as limiting cases in those of elliptic motion, and may be deduced from them by making

$$H_r = 0, \quad \text{or} \quad a = \infty; \quad (118.)$$

and therefore the characteristic function  $w$  and the time  $t$ , in parabolic as well as in elliptic motion, are functions of the chord and of the sum of the radii. By thus making a infinite in the foregoing expressions, we find, for parabolic motion, the partial differential equations

$$\left( \frac{\delta w}{\delta \sigma} + \frac{\delta w}{\delta \tau} \right)^2 = \frac{4\mu}{\sigma + \tau}, \quad \left( \frac{\delta w}{\delta \sigma} - \frac{\delta w}{\delta \tau} \right)^2 = \frac{4\mu}{\sigma - \tau}; \quad (\text{T}^3.)$$

and in fact the parabolic form of the simplified characteristic function  $w$  may easily be shown to be

$$w = 2\sqrt{\mu}(\sqrt{\sigma + \tau} \mp \sqrt{\sigma - \tau}), \quad (\text{U}^3.)$$

$\tau$  being, as before, the chord, and  $\sigma$  the sum of the radii; while the analogous limit of the expression (S<sup>3</sup>.), for the time, is

$$t = \frac{1}{6\sqrt{\mu}} \{(\sigma + \tau)^{\frac{3}{2}} \mp (\sigma - \tau)^{\frac{3}{2}}\} : \quad (\text{V}^3.)$$

which latter is a known expression.

The formulæ (K<sup>3</sup>.), and (L<sup>3</sup>.), to the comparison of which we have reduced the study of elliptic motion, extend to hyperbolic motion also; and in any binary system, with NEWTON's law of attraction, the simplified characteristic function  $w$  may be expressed by the definite integral

$$w = \int_{-\tau}^{\tau} \sqrt{\frac{\mu}{\sigma + \tau} - \frac{\mu}{4a}} \cdot d\tau, \quad (\text{W}^3.)$$

this function  $w$  being still connected with the relative action  $V$ , by the equation (H<sup>3</sup>.); while the time  $t$ , which may always be deduced from this function, by the law of varying action, is represented by this other connected integral,

$$t = \frac{1}{4} \int_{-\tau}^{\tau} \left( \frac{\mu}{\sigma + \tau} - \frac{\mu}{4a} \right)^{-\frac{1}{2}} d\tau : \quad (\text{X}^3.)$$

provided that, within the extent of these integrations, the radical does not vanish nor become infinite. When this condition is not satisfied, we may still express the simplified characteristic function  $w$ , and the time  $t$ , by the following analogous integrals:

$$w = \int_{\tau_r}^{\sigma_r} \pm \sqrt{\frac{2\mu}{\sigma_r} - \frac{\mu}{a}} d\sigma_r, \quad (\text{Y}^3.)$$

and

$$t = \int_{\tau_r}^{\sigma_r} \pm \left( \frac{2\mu}{\sigma_r} - \frac{\mu}{a} \right)^{-\frac{1}{2}} d\sigma_r, \quad (\text{Z}^3.)$$

in which we have put for abridgement

$$\sigma_r = \frac{\sigma + \tau}{2}, \quad \tau_r = \frac{\sigma - \tau}{2}, \quad (119.)$$

and in which it is easy to determine the signs of the radicals. But to treat fully of these various transformations would carry us too far at present, for it is time to consider the properties of systems with more points than two.

#### *On Systems of three Points, in general; and on their Characteristic Functions.*

17. For any system of three points, the known differential equations of motion of the 2nd order are included in the following formula:

$$\left. \begin{aligned} m_1(x_1'' \delta x_1 + y_1'' \delta y_1 + z_1'' \delta z_1) + m_2(x_2'' \delta x_2 + y_2'' \delta y_2 + z_2'' \delta z_2) \\ + m_3(x_3'' \delta x_3 + y_3'' \delta y_3 + z_3'' \delta z_3) = \delta U, \end{aligned} \right\} \quad (120.)$$

the known force-function  $U$  having the form

$$U = m_1 m_2 f^{(1,2)} + m_1 m_3 f^{(1,3)} + m_2 m_3 f^{(2,3)}, \quad (121.)$$

in which  $f^{(1,2)}$ ,  $f^{(1,3)}$ ,  $f^{(2,3)}$ , are functions respectively of the three following mutual distances of the points of the system:

$$\left. \begin{aligned} r^{(1,2)} &= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 - (z_1 - z_2)^2}, \\ r^{(1,3)} &= \sqrt{(x_1 - x_3)^2 + (y_1 - y_3)^2 - (z_1 - z_3)^2}, \\ r^{(2,3)} &= \sqrt{(x_2 - x_3)^2 + (y_2 - y_3)^2 - (z_2 - z_3)^2} : \end{aligned} \right\} \quad (122.)$$

the known differential equations of motion are therefore, separately, for the point  $m_1$ ,

$$\left. \begin{aligned} x_1'' &= m_2 \frac{\delta f^{(1,2)}}{\delta x_1} + m_3 \frac{\delta f^{(1,3)}}{\delta x_1}, \\ y_1'' &= m_2 \frac{\delta f^{(1,2)}}{\delta y_1} + m_3 \frac{\delta f^{(1,3)}}{\delta y_1}, \\ z_1'' &= m_2 \frac{\delta f^{(1,2)}}{\delta z_1} + m_3 \frac{\delta f^{(1,3)}}{\delta z_1}, \end{aligned} \right\} \quad (123.)$$

with six other analogous equations for the points  $m_2$  and  $m_3$ :  $x_2''$ , &c., denoting the component accelerations of the three points  $m_1$   $m_2$   $m_3$ , or the second differential coefficients of their coordinates, taken with respect to the time. To integrate these equations is to assign, by their means, nine relations between the time  $t$ , the three masses  $m_1$   $m_2$   $m_3$ , the nine varying coordinates  $x_1$   $y_1$   $z_1$   $x_2$   $y_2$   $z_2$   $x_3$   $y_3$   $z_3$ , and their nine initial values and nine initial rates of increase, which may be thus denoted,  $a_1$   $b_1$   $c_1$   $a_2$   $b_2$   $c_2$   $a_3$   $b_3$   $c_3$   $a'_1$   $b'_1$   $c'_1$   $a'_2$   $b'_2$   $c'_2$   $a'_3$   $b'_3$   $c'_3$ . The known intermediate integral containing the law of living force, namely,

$$\left. \begin{aligned} \frac{1}{2}m_1(x_1'^2 + y_1'^2 + z_1'^2) + \frac{1}{2}m_2(x_2'^2 + y_2'^2 + z_2'^2) + \frac{1}{2}m_3(x_3'^2 + y_3'^2 + z_3'^2) \\ = m_1 m_2 f^{(1,2)} + m_1 m_3 f^{(1,3)} + m_2 m_3 f^{(2,3)} + H, \end{aligned} \right\} \quad (124.)$$

gives the following initial relation:

$$\left. \begin{aligned} \frac{1}{2}m_1(a_1'^2 + b_1'^2 + c_1'^2) + \frac{1}{2}m_2(a_2'^2 + b_2'^2 + c_2'^2) + \frac{1}{2}m_3(a_3'^2 + b_3'^2 + c_3'^2) \\ = m_1 m_2 f_0^{(1,2)} + m_1 m_3 f_0^{(1,3)} + m_2 m_3 f_0^{(2,3)} + H, \end{aligned} \right\} \quad (125.)$$

in which  $f_0^{(1,2)}$ ,  $f_0^{(1,3)}$ ,  $f_0^{(2,3)}$  are composed of the initial coordinates, in the same manner as  $f^{(1,2)}$ ,  $f^{(1,3)}$ ,  $f^{(2,3)}$  are composed of the final coordinates. If then we knew the nine final integrals of the equations of motion of this ternary system, and combined them with the initial form (125.) of the law of living force, we should have ten relations to determine the ten quantities  $t$   $a'_1$   $b'_1$   $c'_1$   $a'_2$   $b'_2$   $c'_2$   $a'_3$   $b'_3$   $c'_3$ , namely, the time and the nine initial components of the velocities of the three points, as functions of the nine final and nine initial coordinates, and of the quantity  $H$ , involving also the masses; we could therefore determine whatever else

depends on the manner and time of motion of the system, from its initial to its final position, as a function of the same extreme coordinates, and of  $H$ . In particular, we could determine the action  $V$ , or the accumulated living force of the system, namely,

$$V = m_1 \int_0^t (x_1'^2 + y_1'^2 + z_1'^2) dt + m_2 \int_0^t (x_2'^2 + y_2'^2 + z_2'^2) dt + m_3 \int_0^t (x_3'^2 + y_3'^2 + z_3'^2) dt, \quad (\text{A}^4.)$$

as a function of these nineteen quantities,  $x_1 y_1 z_1 x_2 y_2 z_2 x_3 y_3 z_3 a_1 b_1 c_1 a_2 b_2 c_2 a_3 b_3 c_3 H$ ; and might then calculate the variation of this function,

$$\left. \begin{aligned} \delta V &= \frac{\delta V}{\delta x_1} \delta x_1 + \frac{\delta V}{\delta y_1} \delta y_1 + \frac{\delta V}{\delta z_1} \delta z_1 + \frac{\delta V}{\delta a_1} \delta a_1 + \frac{\delta V}{\delta b_1} \delta b_1 + \frac{\delta V}{\delta c_1} \delta c_1 \\ &+ \frac{\delta V}{\delta x_2} \delta x_2 + \frac{\delta V}{\delta y_2} \delta y_2 + \frac{\delta V}{\delta z_2} \delta z_2 + \frac{\delta V}{\delta a_2} \delta a_2 + \frac{\delta V}{\delta b_2} \delta b_2 + \frac{\delta V}{\delta c_2} \delta c_2 \\ &+ \frac{\delta V}{\delta x_3} \delta x_3 + \frac{\delta V}{\delta y_3} \delta y_3 + \frac{\delta V}{\delta z_3} \delta z_3 + \frac{\delta V}{\delta a_3} \delta a_3 + \frac{\delta V}{\delta b_3} \delta b_3 + \frac{\delta V}{\delta c_3} \delta c_3 \\ &+ \frac{\delta V}{\delta H} \delta H. \end{aligned} \right\} \quad (\text{B}^4.)$$

But the law of varying action gives, *previously*, the following expression for this variation:

$$\left. \begin{aligned} \delta V &= m_1(x'_1 \delta x_1 - a'_1 \delta a_1 + y'_1 \delta y_1 - b'_1 \delta b_1 + z'_1 \delta z_1 - c'_1 \delta c_1) \\ &+ m_2(x'_2 \delta x_2 - a'_2 \delta a_2 + y'_2 \delta y_2 - b'_2 \delta b_2 + z'_2 \delta z_2 - c'_2 \delta c_2) \\ &+ m_3(x'_3 \delta x_3 - a'_3 \delta a_3 + y'_3 \delta y_3 - b'_3 \delta b_3 + z'_3 \delta z_3 - c'_3 \delta c_3) \\ &+ t \delta H; \end{aligned} \right\} \quad (\text{C}^4.)$$

and shows, therefore, that the research of all the intermediate and all the final integral equations, of motion of the system, may be reduced, reciprocally, to the search and differentiation of this one characteristic function  $V$ ; because if we knew this one function, we should have the nine intermediate integrals of the known differential equations, under the forms

$$\left. \begin{aligned} \frac{\delta V}{\delta x_1} &= m_1 x'_1, & \frac{\delta V}{\delta y_1} &= m_1 y'_1, & \frac{\delta V}{\delta z_1} &= m_1 z'_1, \\ \frac{\delta V}{\delta x_2} &= m_2 x'_2, & \frac{\delta V}{\delta y_2} &= m_2 y'_2, & \frac{\delta V}{\delta z_2} &= m_2 z'_2, \\ \frac{\delta V}{\delta x_3} &= m_3 x'_3, & \frac{\delta V}{\delta y_3} &= m_3 y'_3, & \frac{\delta V}{\delta z_3} &= m_3 z'_3, \end{aligned} \right\} \quad (\text{D}^4.)$$

and the nine final integrals under the forms

$$\left. \begin{aligned} \frac{\delta V}{\delta a_1} &= -m_1 a'_1, & \frac{\delta V}{\delta b_1} &= -m_1 b'_1, & \frac{\delta V}{\delta c_1} &= -m_1 c'_1, \\ \frac{\delta V}{\delta a_2} &= -m_2 a'_2, & \frac{\delta V}{\delta b_2} &= -m_2 b'_2, & \frac{\delta V}{\delta c_2} &= -m_2 c'_2, \\ \frac{\delta V}{\delta a_3} &= -m_3 a'_3, & \frac{\delta V}{\delta b_3} &= -m_3 b'_3, & \frac{\delta V}{\delta c_3} &= -m_3 c'_3, \end{aligned} \right\} \quad (\text{E}^4.)$$

the auxiliary constant  $H$  being to be eliminated, and the time  $t$  introduced, by this other equation, which has often occurred in this essay,

$$t = \frac{\delta V}{\delta H}. \quad (\text{E.})$$

The same law of varying action suggests also a method of investigating the form of this characteristic function  $V$ , not requiring the previous integration of the known equations of motion; namely, the integration of a pair of partial differential equations connected with the law of living force; which are,

$$\left. \begin{aligned} & \frac{1}{2m_1} \left\{ \left( \frac{\delta V}{\delta x_1} \right)^2 + \left( \frac{\delta V}{\delta y_1} \right)^2 + \left( \frac{\delta V}{\delta z_1} \right)^2 \right\} + \frac{1}{2m_2} \left\{ \left( \frac{\delta V}{\delta x_2} \right)^2 + \left( \frac{\delta V}{\delta y_2} \right)^2 + \left( \frac{\delta V}{\delta z_2} \right)^2 \right\} \\ & + \frac{1}{2m_3} \left\{ \left( \frac{\delta V}{\delta x_3} \right)^2 + \left( \frac{\delta V}{\delta y_3} \right)^2 + \left( \frac{\delta V}{\delta z_3} \right)^2 \right\} \\ & = m_1 m_2 f^{(1,2)} + m_1 m_3 f^{(1,3)} + m_2 m_3 f^{(2,3)} + H, \end{aligned} \right\} \quad (\text{F}^4.)$$

and

$$\left. \begin{aligned} & \frac{1}{2m_1} \left\{ \left( \frac{\delta V}{\delta a_1} \right)^2 + \left( \frac{\delta V}{\delta b_1} \right)^2 + \left( \frac{\delta V}{\delta c_1} \right)^2 \right\} + \frac{1}{2m_2} \left\{ \left( \frac{\delta V}{\delta a_2} \right)^2 + \left( \frac{\delta V}{\delta b_2} \right)^2 + \left( \frac{\delta V}{\delta c_2} \right)^2 \right\} \\ & + \frac{1}{2m_3} \left\{ \left( \frac{\delta V}{\delta a_3} \right)^2 + \left( \frac{\delta V}{\delta b_3} \right)^2 + \left( \frac{\delta V}{\delta c_3} \right)^2 \right\} \\ & = m_1 m_2 f_0^{(1,2)} + m_1 m_3 f_0^{(1,3)} + m_2 m_3 f_0^{(2,3)} + H. \end{aligned} \right\} \quad (\text{G}^4.)$$

And to diminish the difficulty of thus determining the function  $V$ , which depends on 18 coordinates, we may separate it, by principles already explained, into a part  $V_{II}$  depending only on the motion of the centre of gravity of the system, and determined by the formula (H<sup>1.</sup>), and another part  $V_I$ , depending only on the relative motions of the points of the system about this internal centre, and equal to the accumulated living force, connected with this relative motion only. In this manner the difficulty is reduced to determining the relative action  $V_I$ ; and if we introduce the relative coordinates

$$\left. \begin{aligned} \xi_1 &= x_1 - x_3, & \eta_1 &= y_1 - y_3, & \zeta_1 &= z_1 - z_3, \\ \xi_2 &= x_2 - x_3, & \eta_2 &= y_2 - y_3, & \zeta_2 &= z_2 - z_3, \end{aligned} \right\} \quad (126.)$$

and

$$\left. \begin{aligned} \alpha_1 &= a_1 - a_3, & \beta_1 &= b_1 - b_3, & \gamma_1 &= c_1 - c_3, \\ \alpha_2 &= a_2 - a_3, & \beta_2 &= b_2 - b_3, & \gamma_2 &= c_2 - c_3, \end{aligned} \right\} \quad (127.)$$

we easily find, by the principles of the tenth and following numbers, that the function  $V_I$  may be considered as depending only on these relative coordinates and on a quantity  $H$ ,

analogous to  $H$  (besides the masses of the system); and that it must satisfy two partial differential equations, analogous to (F<sup>4</sup>.) and (G<sup>4</sup>.), namely

$$\left. \begin{aligned} & \frac{1}{2m_1} \left\{ \left( \frac{\delta V_r}{\delta \xi_1} \right)^2 + \left( \frac{\delta V_r}{\delta \eta_1} \right)^2 + \left( \frac{\delta V_r}{\delta \zeta_1} \right)^2 \right\} + \frac{1}{2m_2} \left\{ \left( \frac{\delta V_r}{\delta \xi_2} \right)^2 + \left( \frac{\delta V_r}{\delta \eta_2} \right)^2 + \left( \frac{\delta V_r}{\delta \zeta_2} \right)^2 \right\} \\ & + \frac{1}{2m_3} \left\{ \left( \frac{\delta V_r}{\delta \xi_1} + \frac{\delta V_r}{\delta \xi_2} \right)^2 + \left( \frac{\delta V_r}{\delta \eta_1} + \frac{\delta V_r}{\delta \eta_2} \right)^2 + \left( \frac{\delta V_r}{\delta \zeta_1} + \frac{\delta V_r}{\delta \zeta_2} \right)^2 \right\} \\ & = m_1 m_2 f^{(1,2)} + m_1 m_3 f^{(1,3)} + m_2 m_3 f^{(2,3)} + H; \end{aligned} \right\} \quad (\text{H}^4.)$$

and

$$\left. \begin{aligned} & \frac{1}{2m_1} \left\{ \left( \frac{\delta V_r}{\delta \alpha_1} \right)^2 + \left( \frac{\delta V_r}{\delta \beta_1} \right)^2 + \left( \frac{\delta V_r}{\delta \gamma_1} \right)^2 \right\} + \frac{1}{2m_2} \left\{ \left( \frac{\delta V_r}{\delta \alpha_2} \right)^2 + \left( \frac{\delta V_r}{\delta \beta_2} \right)^2 + \left( \frac{\delta V_r}{\delta \gamma_2} \right)^2 \right\} \\ & + \frac{1}{2m_3} \left\{ \left( \frac{\delta V_r}{\delta \alpha_1} + \frac{\delta V_r}{\delta \alpha_2} \right)^2 + \left( \frac{\delta V_r}{\delta \beta_1} + \frac{\delta V_r}{\delta \beta_2} \right)^2 + \left( \frac{\delta V_r}{\delta \gamma_1} + \frac{\delta V_r}{\delta \gamma_2} \right)^2 \right\} \\ & = m_1 m_2 f_0^{(1,2)} + m_1 m_3 f_0^{(1,3)} + m_2 m_3 f_0^{(2,3)} + H; \end{aligned} \right\} \quad (\text{I}^4.)$$

the law of the variation of this function being, by (Z<sup>1</sup>.),

$$\left. \begin{aligned} \delta V_r &= t \delta H, + m_1 (\xi'_1 \delta \xi_1 - \alpha'_1 \delta \alpha_1 + \eta'_1 \delta \eta_1 - \beta'_1 \delta \beta_1 + \zeta'_1 \delta \zeta_1 - \gamma'_1 \delta \gamma_1) \\ &+ m_2 (\xi'_2 \delta \xi_2 - \alpha'_2 \delta \alpha_2 + \eta'_2 \delta \eta_2 - \beta'_2 \delta \beta_2 + \zeta'_2 \delta \zeta_2 - \gamma'_2 \delta \gamma_2) \\ &- \frac{1}{m_1 + m_2 + m_3} \left\{ \begin{array}{l} (m_1 \xi'_1 + m_2 \xi'_2)(m_1 \delta \xi_1 + m_2 \delta \xi_2) \\ -(m_1 \alpha'_1 + m_2 \alpha'_2)(m_1 \delta \alpha_1 + m_2 \delta \alpha_2) \\ +(m_1 \eta'_1 + m_2 \eta'_2)(m_1 \delta \eta_1 + m_2 \delta \eta_2) \\ -(m_1 \beta'_1 + m_2 \beta'_2)(m_1 \delta \beta_1 + m_2 \delta \beta_2) \\ +(m_1 \zeta'_1 + m_2 \zeta'_2)(m_1 \delta \zeta_1 + m_2 \delta \zeta_2) \\ -(m_1 \gamma'_1 + m_2 \gamma'_2)(m_1 \delta \gamma_1 + m_2 \delta \gamma_2) \end{array} \right\} \end{aligned} \right\} \quad (\text{K}^4.)$$

which resolves itself in the same manner as before into the six intermediate and six final integrals of relative motion, namely, into the following equations:

$$\left. \begin{aligned} \frac{1}{m_1} \frac{\delta V_r}{\delta \xi_1} &= \xi'_1 - \frac{m_1 \xi'_1 + m_2 \xi'_2}{m_1 + m_2 + m_3}; & \frac{1}{m_2} \frac{\delta V_r}{\delta \xi_2} &= \xi'_2 - \frac{m_1 \xi'_1 + m_2 \xi'_2}{m_1 + m_2 + m_3}; \\ \frac{1}{m_1} \frac{\delta V_r}{\delta \eta_1} &= \eta'_1 - \frac{m_1 \eta'_1 + m_2 \eta'_2}{m_1 + m_2 + m_3}; & \frac{1}{m_2} \frac{\delta V_r}{\delta \eta_2} &= \eta'_2 - \frac{m_1 \eta'_1 + m_2 \eta'_2}{m_1 + m_2 + m_3}; \\ \frac{1}{m_1} \frac{\delta V_r}{\delta \zeta_1} &= \zeta'_1 - \frac{m_1 \zeta'_1 + m_2 \zeta'_2}{m_1 + m_2 + m_3}; & \frac{1}{m_2} \frac{\delta V_r}{\delta \zeta_2} &= \zeta'_2 - \frac{m_1 \zeta'_1 + m_2 \zeta'_2}{m_1 + m_2 + m_3}; \end{aligned} \right\} \quad (\text{L}^4.)$$

and

$$\left. \begin{aligned} \frac{-1}{m_1} \frac{\delta V_I}{\delta \alpha_1} &= \alpha'_1 - \frac{m_1 \alpha'_1 + m_2 \alpha'_2}{m_1 + m_2 + m_3}; & \frac{-1}{m_2} \frac{\delta V_I}{\delta \alpha_2} &= \alpha'_2 - \frac{m_1 \alpha'_1 + m_2 \alpha'_2}{m_1 + m_2 + m_3}; \\ \frac{-1}{m_1} \frac{\delta V_I}{\delta \beta_1} &= \beta'_1 - \frac{m_1 \beta'_1 + m_2 \beta'_2}{m_1 + m_2 + m_3}; & \frac{-1}{m_2} \frac{\delta V_I}{\delta \beta_2} &= \beta'_2 - \frac{m_1 \beta'_1 + m_2 \beta'_2}{m_1 + m_2 + m_3}; \\ \frac{-1}{m_1} \frac{\delta V_I}{\delta \gamma_1} &= \gamma'_1 - \frac{m_1 \gamma'_1 + m_2 \gamma'_2}{m_1 + m_2 + m_3}; & \frac{-1}{m_2} \frac{\delta V_I}{\delta \gamma_2} &= \gamma'_2 - \frac{m_1 \gamma'_1 + m_2 \gamma'_2}{m_1 + m_2 + m_3}; \end{aligned} \right\} \quad (\text{M}^4.)$$

which must be combined with our old formula,

$$\frac{\delta V_I}{\delta H} = t. \quad (\text{O}^1.)$$

18. The quantity  $H$ , in  $V_I$ , and the analogous quantity  $H_{II}$ , in  $V_{II}$ , are indeed independent of the time, and do not vary in the course of the motion; but it is required by the spirit of our method, that in deducing the absolute action or original characteristic function  $V$  from the two parts  $V_I$  and  $V_{II}$ , we should consider these two parts  $H_I$  and  $H_{II}$  of the original quantity  $H$ , as functions involving each the nine initial and nine final coordinates of the points of the ternary system; the forms of these two functions, of the eighteen coordinates and of  $H$ , being determined by the two conditions,

$$\frac{\delta V_I}{\delta H_I} = \frac{\delta V_{II}}{\delta H_{II}}, \quad H_I + H_{II} = H. \quad (\text{N}^4.)$$

However it results from these conditions, that in taking the variation of the whole original function  $V$ , of the first order, with respect to the eighteen coordinates, we may treat the two auxiliary quantities  $H_I$  and  $H_{II}$  as constant; and therefore that we have the following expressions for the partial differential coefficients of the first order of  $V$ , taken with respect to the coordinates parallel to  $x$ ,

$$\left. \begin{aligned} \frac{\delta V}{\delta x_1} &= \frac{\delta V_I}{\delta \xi_1} + \frac{m_1}{m_1 + m_2 + m_3} \frac{\delta V_{II}}{\delta x_{II}}, & \frac{\delta V}{\delta a_1} &= \frac{\delta V_I}{\delta \alpha_1} + \frac{m_1}{m_1 + m_2 + m_3} \frac{\delta V_{II}}{\delta a_{II}}, \\ \frac{\delta V}{\delta x_2} &= \frac{\delta V_I}{\delta \xi_2} + \frac{m_2}{m_1 + m_2 + m_3} \frac{\delta V_{II}}{\delta x_{II}}, & \frac{\delta V}{\delta a_2} &= \frac{\delta V_I}{\delta \alpha_2} + \frac{m_2}{m_1 + m_2 + m_3} \frac{\delta V_{II}}{\delta a_{II}}, \\ \frac{\delta V}{\delta x_3} &= -\frac{\delta V_I}{\delta \xi_1} - \frac{\delta V_I}{\delta \xi_2} + \frac{m_3}{m_1 + m_2 + m_3} \frac{\delta V_{II}}{\delta x_{II}}, & \frac{\delta V}{\delta a_3} &= -\frac{\delta V_I}{\delta \alpha_1} - \frac{\delta V_I}{\delta \alpha_2} + \frac{m_3}{m_1 + m_2 + m_3} \frac{\delta V_{II}}{\delta a_{II}}, \end{aligned} \right\} \quad (\text{O}^4.)$$

together with analogous expressions for the partial differential coefficients of the same order taken with respect to the other coordinates. Substituting these expressions in the equations of the form (O.), namely, in the following,

$$\left. \begin{aligned} \frac{\delta V}{\delta x_1} + \frac{\delta V}{\delta x_2} + \frac{\delta V}{\delta x_3} + \frac{\delta V}{\delta a_1} + \frac{\delta V}{\delta a_2} + \frac{\delta V}{\delta a_3} &= 0, \\ \frac{\delta V}{\delta y_1} + \frac{\delta V}{\delta y_2} + \frac{\delta V}{\delta y_3} + \frac{\delta V}{\delta b_1} + \frac{\delta V}{\delta b_2} + \frac{\delta V}{\delta b_3} &= 0, \\ \frac{\delta V}{\delta z_1} + \frac{\delta V}{\delta z_2} + \frac{\delta V}{\delta z_3} + \frac{\delta V}{\delta c_1} + \frac{\delta V}{\delta c_2} + \frac{\delta V}{\delta c_3} &= 0, \end{aligned} \right\} \quad (\text{P}^4.)$$

we find that these equations become identical, because

$$\frac{\delta V_{\prime\prime}}{\delta x_{\prime\prime}} + \frac{\delta V_{\prime\prime}}{\delta a_{\prime\prime}} = 0, \quad \frac{\delta V_{\prime\prime}}{\delta y_{\prime\prime}} + \frac{\delta V_{\prime\prime}}{\delta b_{\prime\prime}} = 0, \quad \frac{\delta V_{\prime\prime}}{\delta z_{\prime\prime}} + \frac{\delta V_{\prime\prime}}{\delta c_{\prime\prime}} = 0, \quad (\text{Q}^4.)$$

But substituting, in like manner, the expressions (O<sup>4</sup>.) in the equations of the form (P.), of which the first is, for a ternary system,

$$\left. \begin{aligned} & x_1 \frac{\delta V}{\delta y_1} - y_1 \frac{\delta V}{\delta x_1} + x_2 \frac{\delta V}{\delta y_2} - y_2 \frac{\delta V}{\delta x_2} + x_3 \frac{\delta V}{\delta y_3} - y_3 \frac{\delta V}{\delta x_3} \\ & + a_1 \frac{\delta V}{\delta b_1} - b_1 \frac{\delta V}{\delta a_1} + a_2 \frac{\delta V}{\delta b_2} - b_2 \frac{\delta V}{\delta a_2} + a_3 \frac{\delta V}{\delta b_3} - b_3 \frac{\delta V}{\delta a_3}; \end{aligned} \right\} \quad (\text{R}^4.)$$

and observing that we have

$$x_{\prime\prime} \frac{\delta V_{\prime\prime}}{\delta y_{\prime\prime}} - y_{\prime\prime} \frac{\delta V_{\prime\prime}}{\delta x_{\prime\prime}} + a_{\prime\prime} \frac{\delta V_{\prime\prime}}{\delta b_{\prime\prime}} - b_{\prime\prime} \frac{\delta V_{\prime\prime}}{\delta a_{\prime\prime}} = 0, \quad (\text{S}^4.)$$

along with two other analogous conditions, we find that the part  $V_r$ , or the characteristic function of relative motion of the ternary system, must satisfy the three following conditions, involving its partial differential coefficients of the first order and in the first degree,

$$\left. \begin{aligned} 0 &= \xi_1 \frac{\delta V_r}{\delta \eta_1} - \eta_1 \frac{\delta V_r}{\delta \xi_1} + \xi_2 \frac{\delta V_r}{\delta \eta_2} - \eta_2 \frac{\delta V_r}{\delta \xi_2} + \alpha_1 \frac{\delta V_r}{\delta \beta_1} - \beta_1 \frac{\delta V_r}{\delta \alpha_1} + \alpha_2 \frac{\delta V_r}{\delta \beta_2} - \beta_2 \frac{\delta V_r}{\delta \alpha_2}, \\ 0 &= \eta_1 \frac{\delta V_r}{\delta \zeta_1} - \zeta_1 \frac{\delta V_r}{\delta \eta_1} + \eta_2 \frac{\delta V_r}{\delta \zeta_2} - \zeta_2 \frac{\delta V_r}{\delta \eta_2} + \beta_1 \frac{\delta V_r}{\delta \gamma_1} - \gamma_1 \frac{\delta V_r}{\delta \beta_1} + \beta_2 \frac{\delta V_r}{\delta \gamma_2} - \gamma_2 \frac{\delta V_r}{\delta \beta_2}, \\ 0 &= \zeta_1 \frac{\delta V_r}{\delta \xi_1} - \xi_1 \frac{\delta V_r}{\delta \zeta_1} + \zeta_2 \frac{\delta V_r}{\delta \xi_2} - \xi_2 \frac{\delta V_r}{\delta \zeta_2} + \gamma_1 \frac{\delta V_r}{\delta \alpha_1} - \alpha_1 \frac{\delta V_r}{\delta \gamma_1} + \gamma_2 \frac{\delta V_r}{\delta \alpha_2} - \alpha_2 \frac{\delta V_r}{\delta \gamma_2}, \end{aligned} \right\} \quad (\text{T}^4.)$$

which show that this function can depend only on the shape and size of a pentagon, not generally plane, formed by the point  $m_3$  considered as fixed, and by the initial and final positions of the other two points  $m_1$  and  $m_2$ ; for example, the pentagon, of which the corners are, in order,  $m_3(m_1) m_2 m_2 m_1$ ; ( $m_1$ ) and ( $m_2$ ) denoting the initial positions of the points  $m_1$  and  $m_2$ , referred to  $m_3$  as a fixed origin. The shape and size of this pentagon may be determined by the ten mutual distances of its five points, that is, by the five sides and five diagonals, which may be thus denoted:

$$\left. \begin{aligned} m_3(m_1) &= \sqrt{s_1}, \quad (m_1)(m_2) = \sqrt{s_2}, \quad (m_2)m_2 = \sqrt{s_3}, \quad m_2m_1 = \sqrt{s_4}, \quad m_1m_3 = \sqrt{s_5}, \\ m_3(m_2) &= \sqrt{d_1}, \quad (m_1)m_2 = \sqrt{d_2}, \quad (m_2)m_1 = \sqrt{d_3}, \quad m_2m_3 = \sqrt{d_4}, \quad m_1(m_1) = \sqrt{d_5}; \end{aligned} \right\} \quad (128.)$$

the values of  $s_1 \dots d_5$  as functions of the twelve relative coordinates being

$$\left. \begin{aligned} s_1 &= \alpha_1^2 + \beta_1^2 + \gamma_1^2, & s_2 &= (\alpha_2 - \alpha_1)^2 + (\beta_2 - \beta_1)^2 + (\gamma_2 - \gamma_1)^2, \\ & & s_3 &= (\xi_2 - \alpha_2)^2 + (\eta_2 - \beta_2)^2 + (\zeta_2 - \gamma_2)^2, \\ s_5 &= \xi_1^2 + \eta_1^2 + \zeta_1^2, & s_4 &= (\xi_1 - \xi_2)^2 + (\eta_1 - \eta_2)^2 + (\zeta_1 - \zeta_2)^2, \\ d_1 &= \alpha_2^2 + \beta_2^2 + \gamma_2^2, & d_2 &= (\xi_2 - \alpha_1)^2 + (\eta_2 - \beta_1)^2 + (\zeta_2 - \gamma_1)^2, \\ & & d_3 &= (\xi_1 - \alpha_2)^2 + (\eta_1 - \beta_2)^2 + (\zeta_1 - \gamma_2)^2, \\ d_4 &= \xi_2^2 + \eta_2^2 + \zeta_2^2, & d_5 &= (\xi_1 - \alpha_1)^2 + (\eta_1 - \beta_1)^2 + (\zeta_1 - \gamma_1)^2. \end{aligned} \right\} \quad (129.)$$

These ten distances  $\sqrt{s_1}$ , &c., are not, however, all independent, but are connected by one equation of condition, namely,

$$\begin{aligned}
0 = & \left. \begin{array}{ccccc} s_1^2 s_3^2 & +s_2^2 s_4^2 & +s_3^2 s_5^2 & +s_4^2 s_1^2 & +s_5^2 s_2^2 \\ +s_1^2 d_3^2 & +s_2^2 d_4^2 & +s_3^2 d_5^2 & +s_4^2 d_1^2 & +s_5^2 d_2^2 \\ +d_1^2 d_2^2 & +d_2^2 d_3^2 & +d_3^2 d_4^2 & +d_4^2 d_5^2 & +d_5^2 d_1^2 \\ -2s_1^2 s_3 s_4 & -2s_2^2 s_4 s_5 & -2s_3^2 s_5 s_1 & -2s_4^2 s_1 s_2 & -2s_5^2 s_2 s_3 \\ -2s_1^2 s_3 d_3 & -2s_2^2 s_4 d_4 & -2s_3^2 s_5 d_5 & -2s_4^2 s_1 d_1 & -2s_5^2 s_2 d_2 \\ -2s_1^2 s_4 d_3 & -2s_1^2 s_5 d_4 & -2s_1^2 s_1 d_5 & -2s_1^2 s_2 d_1 & -2s_1^2 s_3 d_2 \\ -2s_1 d_2 d_3^2 & -2s_2 d_3 d_4^2 & -2s_3 d_4 d_5^2 & -2s_4 d_5 d_1^2 & -2s_5 d_1 d_2^2 \\ -2s_1 d_3^2 d_4 & -2s_2 d_4^2 d_5 & -2s_3 d_5^2 d_1 & -2s_4 d_1^2 d_2 & -2s_5 d_2^2 d_3 \\ -2d_1 d_2^2 d_3 & -2d_2 d_3^2 d_4 & -2d_3 d_4^2 d_5 & -2d_4 d_5^2 d_1 & -2d_5 d_1^2 d_2 \\ -4s_1 s_3 s_4 d_3 & -4s_2 s_4 s_5 d_4 & -4s_3 s_5 s_1 d_5 & -4s_4 s_1 s_2 d_1 & -4s_5 s_2 s_3 d_2 \\ -4s_1 d_2 d_3 d_4 & -4s_2 d_3 d_4 d_5 & -4s_3 d_4 d_5 d_1 & -4s_4 d_5 d_1 d_2 & -4s_5 d_1 d_2 d_3 \\ -2s_1 s_2 s_3 d_4 & -2s_2 s_3 s_4 d_5 & -2s_3 s_4 s_5 d_1 & -2s_4 s_5 s_1 d_2 & -2s_5 s_1 s_2 d_3 \\ -2s_1 s_3 d_1 d_2 & -2s_2 s_4 d_2 d_3 & -2s_3 s_5 d_3 d_4 & -2s_4 s_1 d_4 d_5 & -2s_5 s_2 d_5 d_1 \\ -2s_1 d_1 d_3 d_5 & -2s_2 d_2 d_4 d_1 & -2s_3 d_3 d_5 d_2 & -2s_4 d_4 d_1 d_3 & -2s_5 d_5 d_2 d_4 \\ +2s_1 s_2 s_3 s_4 & +2s_2 s_3 s_4 s_5 & +2s_3 s_4 s_5 s_1 & +2s_4 s_5 s_1 s_2 & +2s_5 s_1 s_2 s_3 \\ +2s_1 s_2 s_4 d_3 & +2s_2 s_3 s_5 d_4 & +2s_3 s_4 s_1 d_5 & +2s_4 s_5 s_2 d_1 & +2s_5 s_1 s_3 d_2 \\ +2s_1 s_3 s_4 d_1 & +2s_2 s_4 s_5 d_2 & +2s_3 s_5 s_1 d_3 & +2s_4 s_1 s_2 d_4 & +2s_5 s_2 s_3 d_5 \\ +2s_1 s_2 d_3 d_4 & +2s_2 s_3 d_4 d_5 & +2s_3 s_4 d_5 d_1 & +2s_4 s_5 d_1 d_2 & +2s_5 s_1 d_2 d_3 \\ +2s_1 s_3 d_2 d_3 & +2s_2 s_4 d_3 d_4 & +2s_3 s_5 d_4 d_5 & +2s_4 s_1 d_5 d_1 & +2s_5 s_2 d_1 d_2 \\ +2s_1 s_4 d_1 d_2 & +2s_2 s_5 d_2 d_3 & +2s_3 s_1 d_3 d_4 & +2s_4 s_2 d_4 d_5 & +2s_5 s_3 d_5 d_1 \\ +2s_1 s_4 d_1 d_3 & +2s_2 s_5 d_2 d_4 & +2s_3 s_1 d_3 d_5 & +2s_4 s_2 d_4 d_1 & +2s_5 s_3 d_5 d_2 \\ +2s_1 s_4 d_2 d_3 & +2s_2 s_5 d_3 d_4 & +2s_3 s_1 d_4 d_5 & +2s_4 s_2 d_5 d_1 & +2s_5 s_3 d_1 d_2 \\ +2s_1 s_4 d_3 d_4 & +2s_2 s_5 d_4 d_5 & +2s_3 s_1 d_5 d_1 & +2s_4 s_2 d_1 d_2 & +2s_5 s_3 d_2 d_3 \\ +2s_1 d_1 d_2 d_3 & +2s_2 d_2 d_3 d_4 & +2s_3 d_3 d_4 d_5 & +2s_4 d_4 d_5 d_1 & +2s_5 d_5 d_1 d_2 \\ +2s_1 d_3 d_4 d_5 & +2s_2 d_4 d_5 d_1 & +2s_3 d_5 d_1 d_2 & +2s_4 d_1 d_2 d_3 & +2s_5 d_2 d_3 d_4 \\ +2d_1 d_2 d_3 d_4 & +2d_2 d_3 d_4 d_5 & +2d_3 d_4 d_5 d_1 & +2d_4 d_5 d_1 d_2 & +2d_5 d_1 d_2 d_3; \end{array} \right\} \\
& (130.)
\end{aligned}$$

they may therefore be expressed as functions of nine independent quantities; for example, of four lines and five angles,  $r^{(1)} r_0^{(1)} r^{(2)} r_0^{(2)}$ ,  $\theta^{(1)} \theta_0^{(1)} \theta^{(2)} \theta_0^{(2)} \iota$ , on which they depend as follows:

$$\begin{aligned}
s_1 &= r_0^{(1)2}, \\
s_2 &= r_0^{(1)2} + r_0^{(2)2} - 2r_0^{(1)} r_0^{(2)} (\cos \theta_0^{(1)} \cos \theta_0^{(2)} + \sin \theta_0^{(1)} \sin \theta_0^{(2)} \cos \iota), \\
s_3 &= r_0^{(2)2} + r_0^{(2)2} - 2r_0^{(2)} r_0^{(2)} \cos(\theta^{(2)} - \theta_0^{(2)}), \\
s_4 &= r^{(2)2} + r^{(1)2} - 2r^{(2)} r^{(1)} (\cos \theta^{(1)} \cos \theta^{(2)} + \sin \theta^{(1)} \sin \theta^{(2)} \cos \iota), \\
s_5 &= r^{(1)2}, \\
d_1 &= r_0^{(2)2}, \\
d_2 &= r^{(2)2} + r_0^{(1)2} - 2r^{(2)} r_0^{(1)} (\cos \theta^{(2)} \cos \theta_0^{(1)} + \sin \theta^{(2)} \sin \theta_0^{(1)} \cos \iota), \\
d_3 &= r_0^{(2)2} + r^{(1)2} - 2r_0^{(2)} r^{(1)} (\cos \theta_0^{(2)} \cos \theta^{(1)} + \sin \theta_0^{(2)} \sin \theta^{(1)} \cos \iota), \\
d_4 &= r^{(2)2}, \\
d_5 &= r^{(1)2} + r_0^{(1)2} - 2r^{(1)} r_0^{(1)} \cos(\theta^{(1)} - \theta_0^{(1)}), \quad \left. \right\} \\
& (131.)
\end{aligned}$$

the two line-symbols  $r^{(1)} r^{(2)}$  denoting, for abridgement, the same two final radii vectores which were before denoted by  $r^{(1,3)} r^{(2,3)}$ , and  $r_0^{(1)} r_0^{(2)}$  representing the initial values of these radii; while  $\theta^{(1)} \theta^{(2)} \theta_0^{(1)} \theta_0^{(2)}$  are angles made by these four radii, with the line of intersection of the two planes  $r_0^{(1)} r^{(1)}, r_0^{(2)} r^{(2)}$ ; and  $\iota$  is the inclination of these two planes to each other. We may therefore consider the characteristic function  $V$ , of relative motion, for any ternary system, as depending only on these latter lines and angles, along with the quantity  $H$ .

The reasoning which it has been thought useful to develope here, for any system of three points, attracting or repelling one another according to any functions of their distances, was alluded to, under a more general form, in the twelth number of this essay; and shows, for example, that the characteristic function of relative motion in a system of four such points, depends on the shape and size of a heptagon, and therefore only on the mutual distances of its seven corners, which are in number  $\left(\frac{7 \times 6}{2} =\right) 21$ , but are connected by six equations of condition, leaving only fifteen independent. It is easy to extend these remarks to any multiple system.

*General method of improving an approximate expression for the Characteristic Function of motion of a System in any problem of Dynamics.*

19. The partial differential equation (F.), which the characteristic function  $V$  must satisfy, in every dynamical question, may receive some useful general transformations, by the separation of this function  $V$  into any two parts

$$V_1 + V_2 = V. \quad (\text{U}^4.)$$

For if we establish, for abridgement, the two following equations of definition,

$$\begin{aligned} T_1 &= \Sigma \cdot \frac{1}{2m} \left( \left( \frac{\delta V_1}{\delta x} \right)^2 + \left( \frac{\delta V_1}{\delta y} \right)^2 + \left( \frac{\delta V_1}{\delta z} \right)^2 \right), \\ T_2 &= \Sigma \cdot \frac{1}{2m} \left( \left( \frac{\delta V_2}{\delta x} \right)^2 + \left( \frac{\delta V_2}{\delta y} \right)^2 + \left( \frac{\delta V_2}{\delta z} \right)^2 \right), \end{aligned} \quad (\text{V}^4.)$$

analogous to the relation

$$T = \Sigma \cdot \frac{1}{2m} \left( \left( \frac{\delta V}{\delta x} \right)^2 + \left( \frac{\delta V}{\delta y} \right)^2 + \left( \frac{\delta V}{\delta z} \right)^2 \right), \quad (\text{W}^4.)$$

which served to transform the law of living force into the partial differential equation (F.); we shall have, by (U<sup>4</sup>.),

$$T = T_1 + T_2 + \Sigma \cdot \frac{1}{m} \left( \frac{\delta V_1}{\delta x} \frac{\delta V_2}{\delta x} + \frac{\delta V_1}{\delta y} \frac{\delta V_2}{\delta y} + \frac{\delta V_1}{\delta z} \frac{\delta V_2}{\delta z} \right); \quad (\text{X}^4.)$$

and this expression may be further transformed by the help of the formula (C.), or by the law of varying action. For that law gives the following symbolic equation,

$$\Sigma \cdot \frac{1}{m} \left( \frac{\delta V}{\delta x} \frac{\delta}{\delta x} + \frac{\delta V}{\delta x} \frac{\delta}{\delta y} + \frac{\delta V}{\delta x} \frac{\delta}{\delta z} \right) = \frac{d}{dt}, \quad (\text{Y}^4.)$$

the symbols in both members being prefixed to any one function of the varying coordinates of a system, not expressly involving the time; it gives therefore by (U<sup>4</sup>.), (V<sup>4</sup>.),

$$\Sigma \cdot \frac{1}{m} \left( \frac{\delta V_1}{\delta x} \frac{\delta V_2}{\delta x} + \frac{\delta V_1}{\delta y} \frac{\delta V_2}{\delta y} + \frac{\delta V_1}{\delta z} \frac{\delta V_2}{\delta z} \right) = \frac{dV_2}{dt} - 2T_2. \quad (\text{Z}^4.)$$

In this manner we find the following general and rigorous transformation of the equation (F.),

$$\frac{dV_2}{dt} = T - T_1 + T_2; \quad (\text{A}^5.)$$

$T$  being here retained for the sake of symmetry and conciseness, instead of the equal expression  $U + H$ . And if we suppose, as we may, that the part  $V_1$ , like the whole function  $V$ , is chosen so as to vanish with the time, then the other part  $V_2$  will also have that property, and may be expressed by the definite integral,

$$V_2 = \int_0^t (T - T_1 + T_2) dt. \quad (\text{B}^5.)$$

More generally, if we employ the principles of the seventh number, and introduce any  $3n$  marks  $\eta_1, \eta_2, \dots, \eta_{3n}$ , of the varying positions of the  $n$  points of any system, (whether they be the rectangular coordinates themselves, or any functions of them,) we shall have

$$T = F \left( \frac{\delta V}{\delta \eta_1}, \frac{\delta V}{\delta \eta_2}, \dots, \frac{\delta V}{\delta \eta_{3n}} \right), \quad (\text{C}^5.)$$

and may establish by analogy the two following equations of definition,

$$\begin{aligned} T_1 &= F \left( \frac{\delta V_1}{\delta \eta_1}, \frac{\delta V_1}{\delta \eta_2}, \dots, \frac{\delta V_1}{\delta \eta_{3n}} \right), \\ T_2 &= F \left( \frac{\delta V_2}{\delta \eta_1}, \frac{\delta V_2}{\delta \eta_2}, \dots, \frac{\delta V_2}{\delta \eta_{3n}} \right), \end{aligned} \quad (\text{D}^5.)$$

the function  $F$  being always rational and integer, and homogeneous of the second dimension; and being therefore such that (besides other properties)

$$T = T_1 + T_2 + \frac{\delta T_1}{\delta \frac{\delta V_1}{\delta \eta_1}} \frac{\delta V_2}{\delta \eta_1} + \frac{\delta T_1}{\delta \frac{\delta V_1}{\delta \eta_2}} \frac{\delta V_2}{\delta \eta_2} + \dots + \frac{\delta T_1}{\delta \frac{\delta V_1}{\delta \eta_{3n}}} \frac{\delta V_2}{\delta \eta_{3n}}, \quad (\text{E}^5.)$$

$$\frac{\delta T}{\delta \frac{\delta V}{\delta \eta_1}} = \frac{\delta T_1}{\delta \frac{\delta V_1}{\delta \eta_1}} + \frac{\delta T_2}{\delta \frac{\delta V_2}{\delta \eta_1}}, \dots \quad \frac{\delta T}{\delta \frac{\delta V}{\delta \eta_{3n}}} = \frac{\delta T_1}{\delta \frac{\delta V_1}{\delta \eta_{3n}}} + \frac{\delta T_2}{\delta \frac{\delta V_2}{\delta \eta_{3n}}}, \quad (\text{F}^5.)$$

and

$$\frac{\delta T_2}{\delta \frac{\delta V_2}{\delta \eta_1}} \frac{\delta V_2}{\delta \eta_1} + \frac{\delta T_2}{\delta \frac{\delta V_2}{\delta \eta_2}} \frac{\delta V_2}{\delta \eta_2} + \dots + \frac{\delta T_2}{\delta \frac{\delta V_2}{\delta \eta_{3n}}} \frac{\delta V_2}{\delta \eta_{3n}} = 2T_2. \quad (\text{G}^5.)$$

By the principles of the eighth number, we have also,

$$\frac{\delta T}{\delta \frac{\delta V}{\delta \eta_1}} = \eta'_1, \quad \frac{\delta T}{\delta \frac{\delta V}{\delta \eta_2}} = \eta'_2, \quad \dots \quad \frac{\delta T}{\delta \frac{\delta V}{\delta \eta_{3n}}} = \eta'_{3n}; \quad (\text{H}^5.)$$

and since the meanings of  $\eta'_1, \dots, \eta'_{3n}$  give evidently the symbolical equation,

$$\eta'_1 \frac{\delta}{\delta \eta_1} + \eta'_2 \frac{\delta}{\delta \eta_2} + \dots + \eta'_{3n} \frac{\delta}{\delta \eta_{3n}} = \frac{d}{dt}, \quad (\text{I}^5.)$$

we see that the equation (A<sup>5</sup>.) still holds with the present more general marks of position of a moving system, and gives still the expression (B<sup>5</sup>.), supposing only, as before, that the two parts of the whole characteristic function are chosen so as to vanish with the time.

It may not at first sight appear, that this rigorous transformation (B<sup>5</sup>.), of the partial differential equation (F.), or of the analogous equation (T.) with coordinates not rectangular, is likely to assist much in discovering the form of the part  $V_2$  of the characteristic function  $V$ , (the other part  $V_1$  being supposed to have been previously assumed;) because it involves under the sign of integration, in the term  $T_2$ , the partial differential coefficients of the sought part  $V_2$ . But if we observe that these unknown coefficients enter only by their squares and products, we shall perceive that it offers a general method of improving an approximation in any problem of dynamics. For if the first part  $V_1$  be an approximate value of the whole sought function  $V$ , the second part  $V_2$  will be small, and the term  $T_2$  will not only be also small, but will be in general of a higher order of smallness; we shall therefore in general improve an approximate value  $V_1$  of the characteristic function  $V$ , by adding to it the definite integral,

$$V_2 = \int_0^t (T - T_1) dt; \quad (\text{K}^5.)$$

though this is not, like (B<sup>5</sup>.), a perfectly rigorous expression for the remaining part of the function. And in calculating this integral (K<sup>5</sup>.), for the improvement of an approximation  $V_1$ , we may employ the following analogous approximations to the rigorous formulæ (D.) and (E.),

$$\left. \begin{aligned} \frac{\delta V_1}{\delta a_1} &= -m_1 a'_1; & \frac{\delta V_1}{\delta a_2} &= -m_2 a'_2; & \dots & \frac{\delta V_1}{\delta a_n} &= -m_n a'_n; \\ \frac{\delta V_1}{\delta b_1} &= -m_1 b'_1; & \frac{\delta V_1}{\delta b_2} &= -m_2 b'_2; & \dots & \frac{\delta V_1}{\delta b_n} &= -m_n b'_n; \\ \frac{\delta V_1}{\delta c_1} &= -m_1 c'_1; & \frac{\delta V_1}{\delta c_2} &= -m_2 c'_2; & \dots & \frac{\delta V_1}{\delta c_n} &= -m_n c'_n; \end{aligned} \right\} \quad (\text{L}^5.)$$

and

$$\frac{\delta V_1}{\delta H} = t; \quad (\text{M}^5.)$$

or with any other marks of final and initial position, (instead of rectangular coordinates,) the following approximate forms of the rigorous equations (S.),

$$\frac{\delta V_1}{\delta e_1} = -\frac{\delta T_0}{\delta e'_1}, \quad \frac{\delta V_1}{\delta e_2} = -\frac{\delta T_0}{\delta e'_2}, \quad \dots \quad \frac{\delta V_1}{\delta e_{3n}} = -\frac{\delta T_0}{\delta e'_{3n}}, \quad (\text{N}^5.)$$

together with the formula (M<sup>5.</sup>); by which new formulæ the manner of motion of the system is approximately though not rigorously expressed.

It is easy to extend these remarks to problems of relative motion, and to show that in such problems we have the rigorous transformation

$$V_{r2} = \int_0^t (T_r - T_{r1} + T_{r2}) dt, \quad (\text{O}^5.)$$

and the approximate expression

$$V_{r2} = \int_0^t (T_r - T_{r1}) dt, \quad (\text{P}^5.)$$

$V_{r1}$  being any approximate value of the function  $V_r$  of relative motion, and  $V_{r2}$  being the correction of this value; and  $T_{r1}$ ,  $T_{r2}$ , being homogeneous functions of the second dimension, composed of the partial differential coefficients of these two parts  $V_{r1}$ ,  $V_{r2}$ , in the same way as  $T_r$  is composed of the coefficients of the whole function  $V_r$ . These general remarks may usefully be illustrated by a particular but extensive application.

*Application of the foregoing method to the case of a Ternary or Multiple System, with any laws of attraction or repulsion, and with one predominant mass.*

20. The value (68.), for the relative living force  $2T_r$  of a system, reduces itself successively to the following parts,  $2T_r^{(1)}$ ,  $2T_r^{(2)}$ , ...,  $2T_r^{(n-1)}$ , when we suppose that all the  $n-1$  first masses vanish, with the exception of each successively; namely, to the part

$$2T_r^{(1)} = \frac{m_1 m_n}{m_1 + m_n} (\xi_1'^2 + \eta_1'^2 + \zeta_1'^2), \quad (132.)$$

when only  $m_1$ ,  $m_n$ , do not vanish; the part

$$2T_r^{(2)} = \frac{m_2 m_n}{m_2 + m_n} (\xi_2'^2 + \eta_2'^2 + \zeta_2'^2), \quad (133.)$$

when all but  $m_2$ ,  $m_n$ , vanish; and so on, as far as the part

$$2T_r^{(n-1)} = \frac{m_{n-1} m_n}{m_{n-1} + m_n} (\xi_{n-1}'^2 + \eta_{n-1}'^2 + \zeta_{n-1}'^2), \quad (134.)$$

which remains, when only the two last masses are retained. The sum of these  $n - 1$  parts is not, in general, equal to the whole relative living force  $2T_r$  of the system, with all the  $n$  masses retained; but it differs little from that whole when the first  $n - 1$  masses are small in comparison with the last mass  $m_n$ ; for the rigorous value of this difference is, by (68.), and by (132.) (133.) (134.),

$$\left. \begin{aligned} & 2T_r - 2T_r^{(1)} - 2T_r^{(2)} - \dots - 2T_r^{(n-1)} \\ &= \frac{2m_1}{m_n}(T_r^{(1)} - T_r) + \frac{2m_2}{m_n}(T_r^{(2)} - T_r) + \dots + \frac{2m_{n-1}}{m_n}(T_r^{(n-1)} - T_r) \\ &+ \frac{1}{m_n} \sum m_i m_k \{(\xi'_i - \xi'_k)^2 + (\eta'_i - \eta'_k)^2 + (\zeta'_i - \zeta'_k)^2\} : \end{aligned} \right\} \quad (135.)$$

an expression which is small of the second order when the  $n - 1$  first masses are small of the first order. If, then, we denote by  $V_r^{(1)}, V_r^{(2)}, \dots, V_r^{(n-1)}$ , the relative actions, or accumulated relative living forces, such as they would be in the  $n - 1$  binary systems,  $(m_1 m_n), (m_2 m_n), \dots, (m_{n-1} m_n)$ , without the perturbations of the other small masses of the entire multiple system of  $n$  points; so that

$$V_r^{(1)} = \int_0^t 2T_r^{(1)} dt, \quad V_r^{(2)} = \int_0^t 2T_r^{(2)} dt, \quad \dots \quad V_r^{(n-1)} = \int_0^t 2T_r^{(n-1)} dt, \quad (Q^5.)$$

the perturbations being neglected in calculating these  $n - 1$  definite integrals; we shall have, as an approximate value for the whole relative action  $V_r$  of the system, the sum  $V_{r1}$  of its values for these separate binary systems,

$$V_{r1} = V_r^{(1)} + V_r^{(2)} + \dots + V_r^{(n-1)}. \quad (R^5.)$$

This sum, by our theory of binary systems, may be otherwise expressed as follows:

$$V_{r1} = \frac{m_1 m_n w^{(1)}}{m_1 + m_n} + \frac{m_2 m_n w^{(2)}}{m_2 + m_n} + \dots + \frac{m_{n-1} m_n w^{(n-1)}}{m_{n-1} + m_n}, \quad (S^5.)$$

if we put for abridgement

$$\left. \begin{aligned} w^{(1)} &= h^{(1)} \vartheta^{(1)} + \int_{r_0^{(1)}}^{r^{(1)}} r'^{(1)} dr^{(1)}, \\ w^{(2)} &= h^{(2)} \vartheta^{(2)} + \int_{r_0^{(2)}}^{r^{(2)}} r'^{(2)} dr^{(2)}, \\ &\dots \\ w^{(n-1)} &= h^{(n-1)} \vartheta^{(n-1)} + \int_{r_0^{(n-1)}}^{r^{(n-1)}} r'^{(n-1)} dr^{(n-1)}. \end{aligned} \right\} \quad (T^5.)$$

In this expression,

$$\left. \begin{aligned} r'^{(1)} &= \pm \sqrt{2(m_1 + m_n) f^{(1)} + 2g^{(1)} - \frac{h^{(1)2}}{r^{(1)2}}}, \\ &\dots \\ r'^{(n-1)} &= \pm \sqrt{2(m_{n-1} + m_n) f^{(n-1)} + 2g^{(n-1)} - \frac{h^{(n-1)2}}{r^{(n-1)2}}}, \end{aligned} \right\} \quad (U^5.)$$

$r^{(1)}, \dots r^{(n-1)}$  being abridged expressions for the distances  $r^{(1,n)}, \dots r^{(n-1,n)}$ , and  $f^{(1)}, \dots f^{(n-1)}$  being abridgements for the functions  $f^{(1,n)}, \dots f^{(n-1,n)}$ , of these distances, of which the derivatives, according as they are negative or positive, express the laws of attraction or repulsion: we have also introduced  $2n - 2$  auxiliary quantities  $h^{(1)} g^{(1)} \dots h^{(n-1)} g^{(n-1)}$ , to be eliminated or determined by the following equations of condition:

$$\left. \begin{aligned} 0 &= \vartheta^{(1)} + \int_{r_0^{(1)}}^{r^{(1)}} \frac{\delta r'^{(1)}}{\delta h^{(1)}} dr^{(1)}, \\ 0 &= \vartheta^{(2)} + \int_{r_0^{(2)}}^{r^{(2)}} \frac{\delta r'^{(2)}}{\delta h^{(2)}} dr^{(2)}, \\ &\dots \\ 0 &= \vartheta^{(n-1)} + \int_{r_0^{(n-1)}}^{r^{(n-1)}} \frac{\delta r'^{(n-1)}}{\delta h^{(n-1)}} dr^{(n-1)}, \end{aligned} \right\} \quad (\text{V}^5.)$$

and

$$\int_{r_0^{(1)}}^{r^{(1)}} \frac{dr^{(1)}}{r'^{(1)}} = \int_{r_0^{(2)}}^{r^{(2)}} \frac{dr^{(2)}}{r'^{(2)}} = \dots = \int_{r_0^{(n-1)}}^{r^{(n-1)}} \frac{dr^{(n-1)}}{r'^{(n-1)}}, \quad (\text{W}^5.)$$

or

$$\frac{\delta w^{(1)}}{\delta g^{(1)}} = \frac{\delta w^{(2)}}{\delta g^{(2)}} = \dots = \frac{\delta w^{(n-1)}}{\delta g^{(n-1)}}, \quad (\text{X}^5.)$$

along with this last condition,

$$\frac{m_1 g^{(1)}}{m_1 + m_n} + \frac{m_2 g^{(2)}}{m_2 + m_n} + \frac{m_3 g^{(3)}}{m_3 + m_n} + \dots + \frac{m_{n-1} g^{(n-1)}}{m_{n-1} + m_n} = \frac{H}{m_n}; \quad (\text{Y}^5.)$$

and we have denoted by  $\vartheta^{(1)}, \dots \vartheta^{(n-1)}$ , the angles which the final distances  $r^{(1)}, \dots r^{(n-1)}$ , of the first  $n - 1$  points from the last or  $n$ th point of the system, make respectively with the initial distances corresponding, namely,  $r_0^{(1)}, \dots r_0^{(n-1)}$ . The variation of the sum  $V_{11}$  is, by (S<sup>5</sup>.),

$$\delta V_{11} = \frac{m_1 m_n \delta w^{(1)}}{m_1 + m_n} + \frac{m_2 m_n \delta w^{(2)}}{m_2 + m_n} + \dots + \frac{m_{n-1} m_n \delta w^{(n-1)}}{m_{n-1} + m_n}; \quad (\text{Z}^5.)$$

in which, by the equations of condition, we may treat all the auxiliary quantities  $h^{(1)} g^{(1)} \dots h^{(n-1)} g^{(n-1)}$  as constant, if  $H$ , be considered as given: so that the part of this variation  $\delta V_{11}$ , which depends on the variations of the final relative coordinates, may be put under the form,

$$\left. \begin{aligned} \delta_{\xi, \eta, \zeta} V_{11} &= \frac{m_1 m_n}{m_1 + m_n} \left( \frac{\delta w^{(1)}}{\delta \xi_1} \delta \xi_1 + \frac{\delta w^{(1)}}{\delta \eta_1} \delta \eta_1 + \frac{\delta w^{(1)}}{\delta \zeta_1} \delta \zeta_1 \right) \\ &+ \frac{m_2 m_n}{m_2 + m_n} \left( \frac{\delta w^{(2)}}{\delta \xi_2} \delta \xi_2 + \frac{\delta w^{(2)}}{\delta \eta_2} \delta \eta_2 + \frac{\delta w^{(2)}}{\delta \zeta_2} \delta \zeta_2 \right) \\ &+ \dots \\ &+ \frac{m_{n-1} m_n}{m_{n-1} + m_n} \left( \frac{\delta w^{(n-1)}}{\delta \xi_{n-1}} \delta \xi_{n-1} + \frac{\delta w^{(n-1)}}{\delta \eta_{n-1}} \delta \eta_{n-1} + \frac{\delta w^{(n-1)}}{\delta \zeta_{n-1}} \delta \zeta_{n-1} \right). \end{aligned} \right\} \quad (\text{A}^6.)$$

By the equations (T<sup>5</sup>.) (U<sup>5</sup>.), or by the theory of binary systems, we have, rigorously,

$$\left. \begin{aligned} \left( \frac{\delta w^{(1)}}{\delta \xi_1} \right)^2 + \left( \frac{\delta w^{(1)}}{\delta \eta_1} \right)^2 + \left( \frac{\delta w^{(1)}}{\delta \zeta_1} \right)^2 &= 2(m_1 + m_n)f^{(1)} + 2g^{(1)}; \\ \left( \frac{\delta w^{(2)}}{\delta \xi_2} \right)^2 + \left( \frac{\delta w^{(2)}}{\delta \eta_2} \right)^2 + \left( \frac{\delta w^{(2)}}{\delta \zeta_2} \right)^2 &= 2(m_2 + m_n)f^{(2)} + 2g^{(2)}; \\ \dots \\ \left( \frac{\delta w^{(n-1)}}{\delta \xi_{n-1}} \right)^2 + \left( \frac{\delta w^{(n-1)}}{\delta \eta_{n-1}} \right)^2 + \left( \frac{\delta w^{(n-1)}}{\delta \zeta_{n-1}} \right)^2 &= 2(m_{n-1} + m_n)f^{(n-1)} + 2g^{(n-1)}; \end{aligned} \right\} \quad (\text{B}^6.)$$

and the rigorous law of relative living force for the whole multiple system, is

$$T_t = U + H_t, \quad (50.)$$

in which

$$U = m_n(m_1f^{(1)} + m_2f^{(2)} + \dots + m_{n-1}f^{(n-1)}) + \sum_i m_i m_k f^{(i,k)}, \quad (\text{C}^6.)$$

and

$$\left. \begin{aligned} T_t &= \frac{1}{2} \left( \frac{1}{m_1} + \frac{1}{m_n} \right) \left\{ \left( \frac{\delta V_t}{\delta \xi_1} \right)^2 + \left( \frac{\delta V_t}{\delta \eta_1} \right)^2 + \left( \frac{\delta V_t}{\delta \zeta_1} \right)^2 \right\} \\ &\quad + \frac{1}{2} \left( \frac{1}{m_2} + \frac{1}{m_n} \right) \left\{ \left( \frac{\delta V_t}{\delta \xi_2} \right)^2 + \left( \frac{\delta V_t}{\delta \eta_2} \right)^2 + \left( \frac{\delta V_t}{\delta \zeta_2} \right)^2 \right\} \\ &\quad + \dots \\ &\quad + \frac{1}{2} \left( \frac{1}{m_{n-1}} + \frac{1}{m_n} \right) \left\{ \left( \frac{\delta V_t}{\delta \xi_{n-1}} \right)^2 + \left( \frac{\delta V_t}{\delta \eta_{n-1}} \right)^2 + \left( \frac{\delta V_t}{\delta \zeta_{n-1}} \right)^2 \right\} \\ &\quad + \frac{1}{m_n} \sum_i \left( \frac{\delta V_t}{\delta \xi_i} \frac{\delta V_t}{\delta \xi_k} + \frac{\delta V_t}{\delta \eta_i} \frac{\delta V_t}{\delta \eta_k} + \frac{\delta V_t}{\delta \zeta_i} \frac{\delta V_t}{\delta \zeta_k} \right). \end{aligned} \right\} \quad (\text{D}^6.)$$

We have therefore, by changing in this last expression the coefficients of the characteristic function  $V_t$  to those of its first part  $V_{t1}$ , and by attending to the foregoing equations,

$$\left. \begin{aligned} T_{t1} &= m_n \sum_i m_i f^{(i)} + H, \\ &\quad + m_n \sum_i \frac{m_i}{m_n + m_i} \frac{m_k}{m_n + m_k} \left( \frac{\delta w^{(i)}}{\delta \xi_i} \frac{\delta w^{(k)}}{\delta \xi_k} + \frac{\delta w^{(i)}}{\delta \eta_i} \frac{\delta w^{(k)}}{\delta \eta_k} + \frac{\delta w^{(i)}}{\delta \zeta_i} \frac{\delta w^{(k)}}{\delta \zeta_k} \right); \end{aligned} \right\} \quad (\text{E}^6.)$$

and consequently

$$\left. \begin{aligned} T_t - T_{t1} &= \sum_i m_i m_k \left\{ f^{(i,k)} \right. \\ &\quad \left. - \frac{m_n}{(m_n + m_i)(m_n + m_k)} \left( \frac{\delta w^{(i)}}{\delta \xi_i} \frac{\delta w^{(k)}}{\delta \xi_k} + \frac{\delta w^{(i)}}{\delta \eta_i} \frac{\delta w^{(k)}}{\delta \eta_k} + \frac{\delta w^{(i)}}{\delta \zeta_i} \frac{\delta w^{(k)}}{\delta \zeta_k} \right) \right\}. \end{aligned} \right\} \quad (\text{F}^6.)$$

The general transformation of the foregoing number gives therefore, rigorously, for the remaining part  $V_{r2}$  of the characteristic function  $V$ , of relative motion of the multiple system, the equation

$$V_{r2} = \int_0^t T_{r2} dt + \Sigma_i . m_i m_k \int_0^t \left\{ f^{(i,k)} - \frac{\frac{\delta w^{(i)}}{\delta \xi_i} \frac{\delta w^{(k)}}{\delta \xi_k} + \frac{\delta w^{(i)}}{\delta \eta_i} \frac{\delta w^{(k)}}{\delta \eta_k} + \frac{\delta w^{(i)}}{\delta \zeta_i} \frac{\delta w^{(k)}}{\delta \zeta_k}}{\frac{1}{m_n} (m_n + m_i)(m_n + m_k)} \right\} dt; \quad (\text{G}^6.)$$

and, approximately, the expression

$$V_{r2} = \Sigma_i . m_i m_k \int_0^t \left\{ f^{(i,k)} - \frac{1}{m_n} (\xi'_i \xi'_k + \eta'_i \eta'_k + \zeta'_i \zeta'_k) \right\} dt : \quad (\text{H}^6.)$$

with which last expression we may combine the following approximate formulae belonging in rigour to binary systems only,

$$\xi'_i = \frac{\delta w^{(i)}}{\delta \xi_i}, \quad \eta'_i = \frac{\delta w^{(i)}}{\delta \eta_i}, \quad \zeta'_i = \frac{\delta w^{(i)}}{\delta \zeta_i}, \quad (\text{I}^6.)$$

$$\alpha'_i = -\frac{\delta w^{(i)}}{\delta \alpha_i}, \quad \beta'_i = -\frac{\delta w^{(i)}}{\delta \beta_i}, \quad \gamma'_i = -\frac{\delta w^{(i)}}{\delta \gamma_i}, \quad (\text{K}^6.)$$

and

$$t = \frac{\delta w^{(i)}}{\delta g^{(i)}}. \quad (\text{L}^6.)$$

We have also, rigorously, for binary systems, the following differential equations of motion of the second order,

$$\xi''_i = (m_n + m_i) \frac{\delta f^{(i)}}{\delta \xi_i}; \quad \eta''_i = (m_n + m_i) \frac{\delta f^{(i)}}{\delta \eta_i}; \quad \zeta''_i = (m_n + m_i) \frac{\delta f^{(i)}}{\delta \zeta_i}; \quad (\text{M}^6.)$$

which enable us to transform in various ways the approximate expression (H<sup>6</sup>.). Thus, in the case of a ternary system, with any laws of attraction or repulsion, but with one predominant mass  $m_3$ , the *disturbing part*  $V_{r2}$  of the characteristic function  $V$ , of relative motion, may be put under the form

$$V_{r2} = m_1 m_2 W, \quad (\text{N}^6.)$$

in which the coefficient  $W$  may be approximately be expressed as follows:

$$W = \int_0^t \left\{ f^{(1,2)} - \frac{1}{m_3} (\xi'_1 \xi'_2 + \eta'_1 \eta'_2 + \zeta'_1 \zeta'_2) \right\} dt, \quad (\text{O}^6.)$$

or thus:

$$\left. \begin{aligned} W &= \int_0^t \left( f^{(1,2)} + \xi_2 \frac{\delta f^{(1)}}{\delta \xi_1} + \eta_2 \frac{\delta f^{(1)}}{\delta \eta_1} + \zeta_2 \frac{\delta f^{(1)}}{\delta \zeta_1} \right) dt \\ &\quad - \frac{1}{m_3} \left( \xi_2 \frac{\delta w^{(1)}}{\delta \xi_1} + \eta_2 \frac{\delta w^{(1)}}{\delta \eta_1} + \zeta_2 \frac{\delta w^{(1)}}{\delta \zeta_1} + \alpha_2 \frac{\delta w^{(1)}}{\delta \alpha_1} + \beta_2 \frac{\delta w^{(1)}}{\delta \beta_1} + \gamma_2 \frac{\delta w^{(1)}}{\delta \gamma_1} \right), \end{aligned} \right\} \quad (\text{P}^6.)$$

or finally,

$$\left. \begin{aligned} W &= \int_0^t \left( f^{(1,2)} + \xi_1 \frac{\delta f^{(2)}}{\delta \xi_2} + \eta_1 \frac{\delta f^{(2)}}{\delta \eta_2} + \zeta_1 \frac{\delta f^{(2)}}{\delta \zeta_2} \right) dt \\ &\quad - \frac{1}{m_3} \left( \xi_1 \frac{\delta w^{(2)}}{\delta \xi_2} + \eta_1 \frac{\delta w^{(2)}}{\delta \eta_2} + \zeta_1 \frac{\delta w^{(2)}}{\delta \zeta_2} + \alpha_1 \frac{\delta w^{(2)}}{\delta \alpha_2} + \beta_1 \frac{\delta w^{(2)}}{\delta \beta_2} + \gamma_1 \frac{\delta w^{(2)}}{\delta \gamma_2} \right). \end{aligned} \right\} \quad (\text{Q}^6.)$$

In general, for a multiple system, we may put

$$V_{i2} = \Sigma_i m_i m_k W^{(i,k)}; \quad (\text{R}^6.)$$

and approximately,

$$\left. \begin{aligned} W^{(i,k)} &= \int_0^t \left( f^{(i,k)} + \xi_k \frac{\delta f^{(i)}}{\delta \xi_i} + \eta_k \frac{\delta f^{(i)}}{\delta \eta_i} + \zeta_k \frac{\delta f^{(i)}}{\delta \zeta_i} \right) dt \\ &\quad - \frac{1}{m_n} \left( \xi_k \frac{\delta w^{(i)}}{\delta \xi_i} + \eta_k \frac{\delta w^{(i)}}{\delta \eta_i} + \zeta_k \frac{\delta w^{(i)}}{\delta \zeta_i} + \alpha_k \frac{\delta w^{(i)}}{\delta \alpha_i} + \beta_k \frac{\delta w^{(i)}}{\delta \beta_i} + \gamma_k \frac{\delta w^{(i)}}{\delta \gamma_i} \right), \end{aligned} \right\} \quad (\text{S}^6.)$$

or

$$\left. \begin{aligned} W^{(i,k)} &= \int_0^t \left( f^{(i,k)} + \xi_i \frac{\delta f^{(k)}}{\delta \xi_k} + \eta_i \frac{\delta f^{(k)}}{\delta \eta_k} + \zeta_i \frac{\delta f^{(k)}}{\delta \zeta_k} \right) dt \\ &\quad - \frac{1}{m_n} \left( \xi_i \frac{\delta w^{(k)}}{\delta \xi_k} + \eta_i \frac{\delta w^{(k)}}{\delta \eta_k} + \zeta_i \frac{\delta w^{(k)}}{\delta \zeta_k} + \alpha_i \frac{\delta w^{(k)}}{\delta \alpha_k} + \beta_i \frac{\delta w^{(k)}}{\delta \beta_k} + \gamma_i \frac{\delta w^{(k)}}{\delta \gamma_k} \right). \end{aligned} \right\} \quad (\text{T}^6.)$$

*Rigorous transition from the theory of Binary to that of Multiple Systems, by means of the disturbing part of the whole Characteristic Function; and approximate expressions for the perturbations.*

21. The three equations (K<sup>6</sup>.) when the auxiliary constant  $g^{(i)}$  is eliminated by the formula (L<sup>6</sup>.) are rigorously (by our theory) the three final integrals of the three known equations of the second order (M<sup>6</sup>.), for the relative motion of the binary system ( $m_i m_n$ ); and give, for such a system, the three varying relative coordinates  $\xi_i \eta_i \zeta_i$ , as functions of their initial values and initial rates of increase  $\alpha_i \beta_i \gamma_i \alpha'_i \beta'_i \gamma'_i$ , and of the time  $t$ . In like manner the three equations (I<sup>6</sup>.), when  $g^{(i)}$  is eliminated by (L<sup>6</sup>.), are rigorously the three intermediate integrals of the same known differential equations of motion of the same binary system. These integrals, however, cease to be rigorous when we introduce the perturbations of the relative motion of this partial or binary system ( $m_i m_n$ ), arising from the attractions or repulsions of

the other points  $m_k$ , of the whole proposed multiple system; but they may be corrected and rendered rigorous by employing the remaining part  $V_{r2}$  of the whole characteristic function of relative motion  $V_r$ , along with the principal part or approximate value  $V_{r1}$ .

The equations (X<sup>1</sup>.), (Y<sup>1</sup>. ) of the twelfth number, give rigorously

$$\xi'_i = \frac{1}{m_i} \frac{\delta V_r}{\delta \xi_i} + \frac{1}{m_n} \Sigma' \frac{\delta V_r}{\delta \xi_i}, \quad \eta'_i = \frac{1}{m_i} \frac{\delta V_r}{\delta \eta_i} + \frac{1}{m_n} \Sigma' \frac{\delta V_r}{\delta \eta_i}, \quad \zeta'_i = \frac{1}{m_i} \frac{\delta V_r}{\delta \zeta_i} + \frac{1}{m_n} \Sigma' \frac{\delta V_r}{\delta \zeta_i}, \quad (\text{U}^6.)$$

and

$$-\alpha'_i = \frac{1}{m_i} \frac{\delta V_r}{\delta \alpha_i} + \frac{1}{m_n} \Sigma' \frac{\delta V_r}{\delta \alpha_i}, \quad -\beta'_i = \frac{1}{m_i} \frac{\delta V_r}{\delta \beta_i} + \frac{1}{m_n} \Sigma' \frac{\delta V_r}{\delta \beta_i}, \quad -\gamma'_i = \frac{1}{m_i} \frac{\delta V_r}{\delta \gamma_i} + \frac{1}{m_n} \Sigma' \frac{\delta V_r}{\delta \gamma_i}, \quad (\text{V}^6.)$$

and therefore, by (A<sup>6</sup>.),

$$\left. \begin{aligned} \frac{\delta w^{(i)}}{\delta \xi_i} &= \xi'_i - \Sigma'' \cdot \frac{m_k}{m_k + m_n} \frac{\delta w^{(k)}}{\delta \xi_k} - \frac{1}{m_i} \frac{\delta V_{r2}}{\delta \xi_i} - \frac{1}{m_n} \Sigma' \frac{\delta V_{r2}}{\delta \xi_i}, \\ \frac{\delta w^{(i)}}{\delta \eta_i} &= \eta'_i - \Sigma'' \cdot \frac{m_k}{m_k + m_n} \frac{\delta w^{(k)}}{\delta \eta_k} - \frac{1}{m_i} \frac{\delta V_{r2}}{\delta \eta_i} - \frac{1}{m_n} \Sigma' \frac{\delta V_{r2}}{\delta \eta_i}, \\ \frac{\delta w^{(i)}}{\delta \zeta_i} &= \zeta'_i - \Sigma'' \cdot \frac{m_k}{m_k + m_n} \frac{\delta w^{(k)}}{\delta \zeta_k} - \frac{1}{m_i} \frac{\delta V_{r2}}{\delta \zeta_i} - \frac{1}{m_n} \Sigma' \frac{\delta V_{r2}}{\delta \zeta_i}, \end{aligned} \right\} \quad (\text{W}^6.)$$

and similarly

$$\left. \begin{aligned} -\frac{\delta w^{(i)}}{\delta \alpha_i} &= \alpha'_i + \Sigma'' \cdot \frac{m_k}{m_k + m_n} \frac{\delta w^{(k)}}{\delta \alpha_k} + \frac{1}{m_i} \frac{\delta V_{r2}}{\delta \alpha_i} + \frac{1}{m_n} \Sigma' \frac{\delta V_{r2}}{\delta \alpha_i}, \\ -\frac{\delta w^{(i)}}{\delta \beta_i} &= \beta'_i + \Sigma'' \cdot \frac{m_k}{m_k + m_n} \frac{\delta w^{(k)}}{\delta \beta_k} + \frac{1}{m_i} \frac{\delta V_{r2}}{\delta \beta_i} + \frac{1}{m_n} \Sigma' \frac{\delta V_{r2}}{\delta \beta_i}, \\ -\frac{\delta w^{(i)}}{\delta \gamma_i} &= \gamma'_i + \Sigma'' \cdot \frac{m_k}{m_k + m_n} \frac{\delta w^{(k)}}{\delta \gamma_k} + \frac{1}{m_i} \frac{\delta V_{r2}}{\delta \gamma_i} + \frac{1}{m_n} \Sigma' \frac{\delta V_{r2}}{\delta \gamma_i}, \end{aligned} \right\} \quad (\text{X}^6.)$$

the sign of summation  $\Sigma''$ , referring only to the disturbing masses  $m_k$ , to the exclusion of  $m_i$  and  $m_n$ ; and these equations (W<sup>6</sup>.), (X<sup>6</sup>. ) are the rigorous formulæ, corresponding to the approximate relations (I<sup>6</sup>.), (K<sup>6</sup>.). In like manner, the formula (L<sup>6</sup>. ) for the time of motion in a binary system, which is only an approximation when the system is considered as multiple, may be rigorously corrected for perturbation by adding to it an analogous term deduced from the disturbing part  $V_{r2}$  of the whole characteristic function; that is, by changing it to the following:

$$t = \frac{\delta w^{(i)}}{\delta g^{(i)}} + \frac{\delta V_{r2}}{\delta H_r}, \quad (\text{Y}^6.)$$

which gives, for this other coefficient of  $w^{(i)}$ , the corrected and rigorous expression

$$\frac{\delta w^{(i)}}{\delta g^{(i)}} = t - \frac{\delta V_{r2}}{\delta H_r} : \quad (\text{Z}^6.)$$

$V_{r2}$  being here supposed so chosen as to be rigorously the correction to  $V_{r1}$ . If therefore, by the theory of binary systems, or by eliminating  $g^{(i)}$  between the four equations (K<sup>6</sup>.) (L<sup>6</sup>.), we have deduced expressions for the three varying relative coordinates  $\xi_i \eta_i \zeta_i$  as functions of the time  $t$ , and of the six initial quantities  $\alpha_i \beta_i \gamma_i \alpha'_i \beta'_i \gamma'_i$ , which may be thus denoted,

$$\left. \begin{aligned} \xi_i &= \phi_1(\alpha_i, \beta_i, \gamma_i, \alpha'_i, \beta'_i, \gamma'_i, t), \\ \eta_i &= \phi_2(\alpha_i, \beta_i, \gamma_i, \alpha'_i, \beta'_i, \gamma'_i, t), \\ \zeta_i &= \phi_3(\alpha_i, \beta_i, \gamma_i, \alpha'_i, \beta'_i, \gamma'_i, t); \end{aligned} \right\} \quad (\text{A}^7.)$$

we shall know that the following relations are rigorously and *identically* true,

$$\left. \begin{aligned} \xi_i &= \phi_1 \left( \alpha_i, \beta_i, \gamma_i, -\frac{\delta w^{(i)}}{\delta \alpha_i}, -\frac{\delta w^{(i)}}{\delta \beta_i}, -\frac{\delta w^{(i)}}{\delta \gamma_i}, \frac{\delta w^{(i)}}{\delta g^{(i)}} \right), \\ \eta_i &= \phi_2 \left( \alpha_i, \beta_i, \gamma_i, -\frac{\delta w^{(i)}}{\delta \alpha_i}, -\frac{\delta w^{(i)}}{\delta \beta_i}, -\frac{\delta w^{(i)}}{\delta \gamma_i}, \frac{\delta w^{(i)}}{\delta g^{(i)}} \right), \\ \zeta_i &= \phi_3 \left( \alpha_i, \beta_i, \gamma_i, -\frac{\delta w^{(i)}}{\delta \alpha_i}, -\frac{\delta w^{(i)}}{\delta \beta_i}, -\frac{\delta w^{(i)}}{\delta \gamma_i}, \frac{\delta w^{(i)}}{\delta g^{(i)}} \right), \end{aligned} \right\} \quad (\text{B}^7.)$$

and consequently that these relations will still be rigorously true when we substitute for the four coefficients of  $w^{(i)}$  their rigorous values (X<sup>6</sup>.) and (Z<sup>6</sup>.) for the case of a multiple system. We may thus retain in rigour for any multiple system the final integrals (A<sup>7</sup>.) of the motion of a binary system, if only we add to the initial components  $\alpha'_i \beta'_i \gamma'_i$  of relative velocity, and to the time  $t$ , the following perturbational terms:

$$\left. \begin{aligned} \Delta \alpha'_i &= \Sigma'' \cdot \frac{m_k}{m_k + m_n} \frac{\delta w^{(k)}}{\delta \alpha_k} + \frac{1}{m_i} \frac{\delta V_{r2}}{\delta \alpha_i} + \frac{1}{m_n} \Sigma' \frac{\delta V_{r2}}{\delta \alpha_i}, \\ \Delta \beta'_i &= \Sigma'' \cdot \frac{m_k}{m_k + m_n} \frac{\delta w^{(k)}}{\delta \beta_k} + \frac{1}{m_i} \frac{\delta V_{r2}}{\delta \beta_i} + \frac{1}{m_n} \Sigma' \frac{\delta V_{r2}}{\delta \beta_i}, \\ \Delta \gamma'_i &= \Sigma'' \cdot \frac{m_k}{m_k + m_n} \frac{\delta w^{(k)}}{\delta \gamma_k} + \frac{1}{m_i} \frac{\delta V_{r2}}{\delta \gamma_i} + \frac{1}{m_n} \Sigma' \frac{\delta V_{r2}}{\delta \gamma_i}, \end{aligned} \right\} \quad (\text{C}^7.)$$

and

$$\Delta t = -\frac{\delta V_{r2}}{\delta H}. \quad (\text{D}^7.)$$

In the same way, if the theory of binary systems, or the elimination of  $g^{(i)}$  between the four equations (I<sup>6</sup>.) (L<sup>6</sup>.), has given three intermediate integrals, of the form

$$\left. \begin{aligned} \xi'_i &= \psi_1(\xi_i, \eta_i, \zeta_i, \alpha_i, \beta_i, \gamma_i, t), \\ \eta'_i &= \psi_2(\xi_i, \eta_i, \zeta_i, \alpha_i, \beta_i, \gamma_i, t), \\ \zeta'_i &= \psi_3(\xi_i, \eta_i, \zeta_i, \alpha_i, \beta_i, \gamma_i, t), \end{aligned} \right\} \quad (\text{E}^7.)$$

we can conclude that the following equations are rigorous and identical,

$$\left. \begin{aligned} \frac{\delta w^{(i)}}{\delta \xi_i} &= \psi_1 \left( \xi_i, \eta_i, \zeta_i, \alpha_i, \beta_i, \gamma_i, \frac{\delta w^{(i)}}{\delta g^{(i)}} \right), \\ \frac{\delta w^{(i)}}{\delta \eta_i} &= \psi_2 \left( \xi_i, \eta_i, \zeta_i, \alpha_i, \beta_i, \gamma_i, \frac{\delta w^{(i)}}{\delta g^{(i)}} \right), \\ \frac{\delta w^{(i)}}{\delta \zeta_i} &= \psi_3 \left( \xi_i, \eta_i, \zeta_i, \alpha_i, \beta_i, \gamma_i, \frac{\delta w^{(i)}}{\delta g^{(i)}} \right), \end{aligned} \right\} \quad (\text{F}^7.)$$

and must therefore be still true, when, in passing to a multiple system, we change the coefficients of  $w^{(i)}$  to their rigorous values (W<sup>6</sup>.) (Z<sup>6</sup>.). The three intermediate integrals (E<sup>7</sup>.) of the motion of a binary system may therefore be adapted rigorously to the case of a multiple system, by first adding to the time  $t$  the perturbational term (D<sup>7</sup>.), and afterwards adding to the resulting values of the final components of relative velocity the terms

$$\left. \begin{aligned} \Delta \xi'_i &= \Sigma'' \cdot \frac{m_k}{m_k + m_n} \frac{\delta w^{(k)}}{\delta \xi_k} + \frac{1}{m_i} \frac{\delta V_{r2}}{\delta \xi_i} + \frac{1}{m_n} \Sigma' \frac{\delta V_{r2}}{\delta \xi_i}, \\ \Delta \eta'_i &= \Sigma'' \cdot \frac{m_k}{m_k + m_n} \frac{\delta w^{(k)}}{\delta \eta_k} + \frac{1}{m_i} \frac{\delta V_{r2}}{\delta \eta_i} + \frac{1}{m_n} \Sigma' \frac{\delta V_{r2}}{\delta \eta_i}, \\ \Delta \zeta'_i &= \Sigma'' \cdot \frac{m_k}{m_k + m_n} \frac{\delta w^{(k)}}{\delta \zeta_k} + \frac{1}{m_i} \frac{\delta V_{r2}}{\delta \zeta_i} + \frac{1}{m_n} \Sigma' \frac{\delta V_{r2}}{\delta \zeta_i}. \end{aligned} \right\} \quad (\text{G}^7.)$$

22. To derive now, from these rigorous results, some useful approximate expressions, we shall neglect, in the perturbations, the terms which are of the second order, with respect to the small masses of the system, and with respect to the constant  $2H$ , of relative living force, which is easily seen to be small of the same order as the masses: and then the perturbations of these coordinates, deduced by the method that has been explained, become

$$\left. \begin{aligned} \Delta \xi_i &= \frac{\delta \xi_i}{\delta \alpha'_i} \Delta \alpha'_i + \frac{\delta \xi_i}{\delta \beta'_i} \Delta \beta'_i + \frac{\delta \xi_i}{\delta \gamma'_i} \Delta \gamma'_i + \frac{\delta \xi_i}{\delta t} \Delta t, \\ \Delta \eta_i &= \frac{\delta \eta_i}{\delta \alpha'_i} \Delta \alpha'_i + \frac{\delta \eta_i}{\delta \beta'_i} \Delta \beta'_i + \frac{\delta \eta_i}{\delta \gamma'_i} \Delta \gamma'_i + \frac{\delta \eta_i}{\delta t} \Delta t, \\ \Delta \zeta_i &= \frac{\delta \zeta_i}{\delta \alpha'_i} \Delta \alpha'_i + \frac{\delta \zeta_i}{\delta \beta'_i} \Delta \beta'_i + \frac{\delta \zeta_i}{\delta \gamma'_i} \Delta \gamma'_i + \frac{\delta \zeta_i}{\delta t} \Delta t, \end{aligned} \right\} \quad (\text{H}^7.)$$

in which we may employ, instead of the rigorous values (C<sup>7</sup>.) for  $\Delta \alpha'_i, \Delta \beta'_i, \Delta \gamma'_i$ , the following approximate values:

$$\left. \begin{aligned} \Delta \alpha'_i &= \Sigma'' \frac{m_k}{m_n} \frac{\delta w^{(k)}}{\delta \alpha_k} + \frac{1}{m_i} \frac{\delta V_{r2}}{\delta \alpha_i}, \\ \Delta \beta'_i &= \Sigma'' \frac{m_k}{m_n} \frac{\delta w^{(k)}}{\delta \beta_k} + \frac{1}{m_i} \frac{\delta V_{r2}}{\delta \beta_i}, \\ \Delta \gamma'_i &= \Sigma'' \frac{m_k}{m_n} \frac{\delta w^{(k)}}{\delta \gamma_k} + \frac{1}{m_i} \frac{\delta V_{r2}}{\delta \gamma_i}. \end{aligned} \right\} \quad (\text{I}^7.)$$

To calculate the four coefficients

$$\frac{\delta V_{r2}}{\delta \alpha_i}, \quad \frac{\delta V_{r2}}{\delta \beta_i}, \quad \frac{\delta V_{r2}}{\delta \gamma_i}, \quad \frac{\delta V_{r2}}{\delta H},$$

which enter into the values (I<sup>7</sup>.) (D<sup>7</sup>.), we may consider  $V_{r2}$ , by (R<sup>6</sup>.) (T<sup>6</sup>.), and by the theory of binary systems, as a function of the initial and final relative coordinates, and initial components of relative velocities, involving also expressly the time  $t$  and the  $n - 2$  auxiliary quantities of the form  $g^{(k)}$ ; and then we are to consider those initial components and auxiliary quantities and the time, as depending themselves on the initial and final coordinates, and on  $H_r$ . But it is not difficult to prove, by the foregoing principles, that when  $t$  and  $g^{(k)}$  are thus considered, their variations are, in the present order of approximation,

$$\delta t = \frac{\Sigma' . m \left( \frac{\delta^2 w}{\delta g^2} \right)^{-1} \delta, \frac{\delta w}{\delta g} + \delta H,}{\Sigma' . m \left( \frac{\delta^2 w}{\delta g^2} \right)^{-1}} \quad (\text{K}^7.)$$

and

$$\delta g^{(k)} = \left( \frac{\delta^2 w^{(k)}}{\delta g^{(k)2}} \right)^{-1} \left( \delta t - \delta, \frac{\delta w^{(k)}}{\delta g^{(k)}} \right), \quad (\text{L}^7.)$$

the sign of variation  $\delta$ , referring only to the initial and final coordinates; and also that

$$\frac{\delta^2 w^{(i)}}{\delta g^{(i)2}} \frac{\delta \xi_i}{\delta t} = \frac{\delta^2 w^{(i)}}{\delta \alpha_i \delta g^{(i)}} \frac{\delta \xi_i}{\delta \alpha'_i} + \frac{\delta^2 w^{(i)}}{\delta \beta_i \delta g^{(i)}} \frac{\delta \xi_i}{\delta \beta'_i} + \frac{\delta^2 w^{(i)}}{\delta \gamma_i \delta g^{(i)}} \frac{\delta \xi_i}{\delta \gamma'_i}, \quad (\text{M}^7.)$$

along with two other analogous relations between the coefficients of the two other coordinates  $\eta_i$ ,  $\zeta_i$ ; from which it follows that  $t$  and  $g^{(k)}$ , and therefore  $\alpha'_k$   $\beta'_k$   $\gamma'_k$ , may be treated as constant, in taking the variation of the disturbing part  $V_{r2}$ , for the purpose of calculating the perturbations (H<sup>7</sup>.): and that the terms involving  $\Delta t$  are destroyed by other terms. We may therefore put simply

$$\left. \begin{aligned} \Delta \xi_i &= \frac{\delta \xi_i}{\delta \alpha'_i} \Delta \alpha'_i + \frac{\delta \xi_i}{\delta \beta'_i} \Delta \beta'_i + \frac{\delta \xi_i}{\delta \gamma'_i} \Delta \gamma'_i, \\ \Delta \eta_i &= \frac{\delta \eta_i}{\delta \alpha'_i} \Delta \alpha'_i + \frac{\delta \eta_i}{\delta \beta'_i} \Delta \beta'_i + \frac{\delta \eta_i}{\delta \gamma'_i} \Delta \gamma'_i, \\ \Delta \zeta_i &= \frac{\delta \zeta_i}{\delta \alpha'_i} \Delta \alpha'_i + \frac{\delta \zeta_i}{\delta \beta'_i} \Delta \beta'_i + \frac{\delta \zeta_i}{\delta \gamma'_i} \Delta \gamma'_i, \end{aligned} \right\} \quad (\text{N}^7.)$$

employing for  $\Delta \alpha'_i$  the following new expression,

$$\Delta \alpha'_i = \Sigma'' . m_k \left\{ \int_0^t \frac{\delta R^{(i,k)}}{\delta \alpha_i} dt + \frac{\delta \alpha'_i}{\delta \alpha_i} \int_0^t \frac{\delta R^{(i,k)}}{\delta \alpha'_i} dt \right. \\ \left. + \frac{\delta \beta'_i}{\delta \alpha_i} \int_0^t \frac{\delta R^{(i,k)}}{\delta \beta'_i} dt + \frac{\delta \gamma'_i}{\delta \alpha_i} \int_0^t \frac{\delta R^{(i,k)}}{\delta \gamma'_i} dt \right\} \quad (\text{O}^7.)$$

together with analogous expressions for  $\Delta\beta'_i$ ,  $\Delta\gamma'_i$ , in which the sign of summation  $\Sigma_{II}$  refers to the disturbing masses, and in which the quantity

$$R^{(i,k)} = f^{(i,k)} + \xi_i \frac{\delta f^{(k)}}{\delta \xi_k} + \eta_i \frac{\delta f^{(k)}}{\delta \eta_k} + \zeta_i \frac{\delta f^{(k)}}{\delta \zeta_k} \quad (\text{P}^7.)$$

is considered as depending on  $\alpha_i \beta_i \gamma_i \alpha'_i \beta'_i \gamma'_i \alpha_k \beta_k \gamma_k \alpha'_k \beta'_k \gamma'_k t$  by the theory of binary systems, while  $\alpha'_i \beta'_i \gamma'_i$ , are considered as depending, by the same rules, on  $\alpha_i \beta_i \gamma_i \xi_i \eta_i \zeta_i$  and  $t$ .

It may also be easily shown, that

$$\frac{\delta \xi_i}{\delta \alpha'_i} \frac{\delta \alpha'_i}{\delta \alpha_i} + \frac{\delta \xi_i}{\delta \beta'_i} \frac{\delta \beta'_i}{\delta \beta_i} + \frac{\delta \xi_i}{\delta \gamma'_i} \frac{\delta \gamma'_i}{\delta \gamma_i} = -\frac{\delta \xi_i}{\delta \alpha_i}; \quad (\text{Q}7.)$$

with other analogous equations: the perturbation of the coordinates  $\xi_i$  may therefore be thus expressed,

$$\left. \begin{aligned} \Delta \xi_i &= \Sigma_{II} \cdot m_k \left\{ \frac{\delta \xi_i}{\delta \alpha'_i} \int_0^t \frac{\delta R^{(i,k)}}{\delta \alpha_i} dt - \frac{\delta \xi_i}{\delta \alpha_i} \int_0^t \frac{\delta R^{(i,k)}}{\delta \alpha'_i} dt \right. \\ &\quad \left. + \frac{\delta \xi_i}{\delta \beta'_i} \int_0^t \frac{\delta R^{(i,k)}}{\delta \beta_i} dt - \frac{\delta \xi_i}{\delta \beta_i} \int_0^t \frac{\delta R^{(i,k)}}{\delta \beta'_i} dt \right. \\ &\quad \left. + \frac{\delta \xi_i}{\delta \gamma'_i} \int_0^t \frac{\delta R^{(i,k)}}{\delta \gamma_i} dt - \frac{\delta \xi_i}{\delta \gamma_i} \int_0^t \frac{\delta R^{(i,k)}}{\delta \gamma'_i} dt \right\}, \end{aligned} \right\} \quad (\text{R}7.)$$

and the perturbations of the two other coordinates may be expressed in an analogous manner.

It results from the same principles, that in taking the first differentials of these perturbations (R<sup>7</sup>.), the integrals may be treated as constant; and therefore that we may either represent the change of place of the disturbed point  $m_i$ , in its relative orbit about  $m_n$ , by altering a little the initial components of velocity without altering the initial position, and then employing the rules of binary systems; or calculate at once the perturbations of place and of velocity, by employing the same rules, and altering at once the initial position and initial velocity. If we adopt the former of these two methods, we are to employ the expressions (O<sup>7</sup>.), which may be thus summed up,

$$\left. \begin{aligned} \Delta \alpha'_i &= \Sigma_{II} \cdot m_k \frac{\delta}{\delta \alpha_i} \int_0^t R^{(i,k)} dt, \\ \Delta \beta'_i &= \Sigma_{II} \cdot m_k \frac{\delta}{\delta \beta_i} \int_0^t R^{(i,k)} dt, \\ \Delta \gamma'_i &= \Sigma_{II} \cdot m_k \frac{\delta}{\delta \gamma_i} \int_0^t R^{(i,k)} dt; \end{aligned} \right\} \quad (\text{S}7.)$$

and if we adopt the latter method, we are to make,

$$\left. \begin{aligned} \Delta \alpha'_i &= \Sigma_{II} \cdot m_k \int_0^t \frac{\delta R^{(i,k)}}{\delta \alpha_i} dt, \quad \Delta \alpha_i = -\Sigma_{II} \cdot m_k \int_0^t \frac{\delta R^{(i,k)}}{\delta \alpha'_i} dt, \\ \Delta \beta'_i &= \Sigma_{II} \cdot m_k \int_0^t \frac{\delta R^{(i,k)}}{\delta \beta_i} dt, \quad \Delta \beta_i = -\Sigma_{II} \cdot m_k \int_0^t \frac{\delta R^{(i,k)}}{\delta \beta'_i} dt, \\ \Delta \gamma'_i &= \Sigma_{II} \cdot m_k \int_0^t \frac{\delta R^{(i,k)}}{\delta \gamma_i} dt, \quad \Delta \gamma_i = -\Sigma_{II} \cdot m_k \int_0^t \frac{\delta R^{(i,k)}}{\delta \gamma'_i} dt. \end{aligned} \right\} \quad (\text{T}7.)$$

The latter was the method of LAGRANGE: the former is suggested more immediately by the principles of the present essay.

*General introduction of the Time into the expression of the Characteristic Function in any dynamical problem.*

23. Before we conclude this sketch of our general method in dynamics, it will be proper to notice briefly a transformation of the characteristic function, which may be used in all applications. This transformation consists in putting, generally,

$$V = tH + S, \quad (\text{U}^7.)$$

and considering the part  $S$ , namely, the definite integral

$$S = \int_0^t (T + U) dt, \quad (\text{V}^7.)$$

as a function of the initial and final coordinates and of the time, of which the variation is, by our law of varying action,

$$\delta S = -H \delta t + \Sigma .m(x' \delta x - a' \delta a + y' \delta y - b' \delta b + z' \delta z - c' \delta c). \quad (\text{W}^7.)$$

The partial differential coefficients of the first order of this auxiliary function  $S$ , are hence,

$$\frac{\delta S}{\delta t} = -H; \quad (\text{X}^7.)$$

$$\frac{\delta S}{\delta x_i} = m_i x'_i, \quad \frac{\delta S}{\delta y_i} = m_i y'_i, \quad \frac{\delta S}{\delta z_i} = m_i z'_i; \quad (\text{Y}^7.)$$

and

$$\frac{\delta S}{\delta a_i} = -m_i a'_i, \quad \frac{\delta S}{\delta b_i} = -m_i b'_i, \quad \frac{\delta S}{\delta c_i} = -m_i c'_i. \quad (\text{Z}^7.)$$

These last expressions ( $\text{Z}^7.$ ) are forms for the final integrals of motion of any system, corresponding to the result of elimination of  $H$  between the equations ( $\text{D}.$ ) and ( $\text{E}.$ ); and the expressions ( $\text{Y}^7.$ ) are forms for the intermediate integrals, more convenient in many respects than the forms already employed.

24. The limits of the present essay do not permit us here to develope the consequences of these new expressions. We can only observe, that the auxiliary function  $S$  must satisfy the two following equations, in partial differentials of the first order, analogous to, and deduced from, the equations ( $\text{F}.$ ) and ( $\text{G}.$ ):

$$\frac{\delta S}{\delta t} + \Sigma \cdot \frac{1}{2m} \left\{ \left( \frac{\delta S}{\delta x} \right)^2 + \left( \frac{\delta S}{\delta y} \right)^2 + \left( \frac{\delta S}{\delta z} \right)^2 \right\} = U, \quad (\text{A}^8.)$$

and

$$\frac{\delta S}{\delta t} + \Sigma \cdot \frac{1}{2m} \left\{ \left( \frac{\delta S}{\delta a} \right)^2 + \left( \frac{\delta S}{\delta b} \right)^2 + \left( \frac{\delta S}{\delta c} \right)^2 \right\} = U_0; \quad (\text{B}^8.)$$

and that to correct an approximate value  $S_1$  of  $S$ , in the integration of these equations, or to find the remaining part  $S_2$ , if

$$S = S_1 + S_2, \quad (\text{C}^8.)$$

we may employ the symbolic equation

$$\frac{d}{dt} = \frac{\delta}{\delta t} + \Sigma \cdot \frac{1}{m} \left( \frac{\delta S}{\delta x} \frac{\delta}{\delta x} + \frac{\delta S}{\delta y} \frac{\delta}{\delta y} + \frac{\delta S}{\delta z} \frac{\delta}{\delta z} \right); \quad (\text{D}^8.)$$

which gives, rigorously,

$$\frac{dS_2}{dt} = U - U_1 + \Sigma \cdot \frac{1}{2m} \left\{ \left( \frac{\delta S_2}{\delta x} \right)^2 + \left( \frac{\delta S_2}{\delta y} \right)^2 + \left( \frac{\delta S_2}{\delta z} \right)^2 \right\} \quad (\text{E}^8.)$$

if we establish by analogy the definition

$$U_1 = \frac{\delta S_1}{\delta t} + \Sigma \cdot \frac{1}{2m} \left\{ \left( \frac{\delta S_1}{\delta x} \right)^2 + \left( \frac{\delta S_1}{\delta y} \right)^2 + \left( \frac{\delta S_1}{\delta z} \right)^2 \right\}; \quad (\text{F}^8.)$$

and therefore approximately

$$S_2 = \int_0^t (U - U_1) dt, \quad (\text{G}^8.)$$

the parts  $S_1$   $S_2$  being chosen so as to vanish with the time. These remarks may all be extended easily, so as to embrace relative and polar coordinates, and other marks of position, and offer a new and better way of investigating the orbits and perturbations of a system, by a new and better form of the function and method of this Essay.

*March 29, 1834.*



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.* Leipzig 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 aether.

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  $\delta x$ ,  $\delta y$ ,  $\delta z$  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  $\delta x$  and  $\delta y$  are independent,

$$\left. \begin{aligned} X &= \frac{d}{dt} (C p^2 u + C p q v), \\ Y &= \frac{d}{dt} (C p q u + C q^2 v). \end{aligned} \right\} \dots \dots \dots \dots \dots \quad (1)$$

We may call  $C p^2 u + C p q v$  the momentum of C referred to A, and  $C p q u + C q^2 v$  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

$$L = \Sigma (C p^2), \quad M = \Sigma (C p q), \quad \text{and } N = \Sigma (C q^2),$$

\* 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

$$Lu + Mv,$$

and referred to B,

$$Mu + Nv,$$

and we shall have

$$\left. \begin{aligned} X &= \frac{d}{dt}(Lu + Mv), \\ Y &= \frac{d}{dt}(Mu + Nv), \end{aligned} \right\} \dots \dots \dots \dots \dots \quad (2)$$

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  $\xi$  and  $\eta$  are the forces on A and B

$$\left. \begin{aligned} \xi &= X + Ru = Ru + \frac{d}{dt}(Lu + Mv), \\ \eta &= Y + Sv = Sv + \frac{d}{dt}(Mu + Nv) \end{aligned} \right\} \dots \dots \dots \dots \dots \quad (3)$$

If the velocity of A be increased at the rate  $\frac{du}{dt}$ , then in order to prevent B from moving a force,  $\eta = \frac{d}{dt}(Mu)$  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  $\xi$  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  $\xi = Rx$  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

$$Lx + My,$$

and that belonging to B,

$$Mx + Ny,$$

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

$$\xi = Rx + \frac{d}{dt}(Lx + My), \dots \dots \dots \dots \dots \quad (4)$$

and that of  $y$  in B

$$\eta = Sy + \frac{d}{dt}(Mx + Ny), \dots \dots \dots \dots \dots \quad (5)$$

where  $\xi$  and  $\eta$  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

$$Sy + \frac{d}{dt}(Mx + Ny) = 0,$$

whence

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

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

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

Here  $\xi x$  is the work done in unit of time by the electromotive force  $\xi$  acting on the current  $x$  and maintaining it, and  $\eta y$  is the work done by the electromotive force  $\eta$ . 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,

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

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

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

*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

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

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,

$$\frac{1}{2} \frac{dL}{dt} x^2 + \frac{dM}{dt} xy + \frac{1}{2} \frac{dN}{dt} y^2 = W \dots \dots \dots \dots \quad (11)$$

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

$$(M' - M)xy,$$

and the force in the direction of  $ds$  will be

$$\frac{dM}{ds} xy, \dots \dots \dots \dots \dots \dots \dots \quad (12)$$

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  $\xi$ , is

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

When  $\xi$  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

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

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

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

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  $\frac{1}{2}\frac{\xi}{R(R+r)}(2R+r)$ , 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  $\varrho^2 = R^2 + L^2 p^2$ , 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

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

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  $t$ ,

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

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

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

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  $\xi$  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

$$(LN - M^2)n^2 + (RN + LS)n + RS = 0, \dots \dots \dots \quad (16)$$

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

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

$$y = \frac{cM}{S} \frac{n_1 n_2}{n_1 - n_2} \{ e^{n_1 t} - e^{n_2 t} \}. \dots \dots \dots \dots \dots \dots \dots \quad (18)$$

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

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

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

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

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

for a time

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

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

$$\xi \int x dt = c^2 (Rt - L).$$

Work done in overcoming resistance and producing heat,

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

Energy remaining in the system,

$$= \frac{1}{2} c^2 L.$$

(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

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

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

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

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  $2 \frac{N}{S}$ .

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

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

and

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

The effect of the presence of the circuit S in the neighbourhood is to alter the value of A and  $\alpha$ , to that which they would be if R became

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

and L became

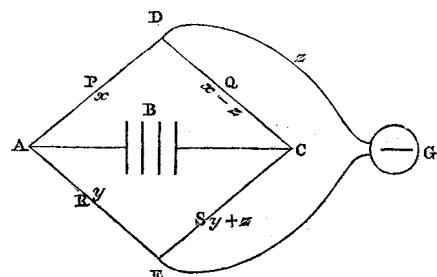
$$L - p^2 \frac{MN}{S^2 + p^2 N^2}.$$

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

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



Solving these equations for  $z$ , we find

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

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

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

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.

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

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

$$\left. \begin{aligned} P X &+ (p+h)x + (k+l)y = \int A dt - \int D dt, \\ Q(X-Z) + (h+q)x + (m+n)y &= \int D dt - \int C dt, \\ R Y &+ (k+m)x + (r+o)y = \int A dt - \int E dt, \\ S(Y+Z) + (l+n)x + (o+s)y &= \int E dt - \int C dt, \\ G Z &= \int D dt - \int E dt. \end{aligned} \right\} \quad \dots \quad (24)$$

Solving these equations for  $Z$ , we find

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

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

Let the permanent deflection produced by making the ratio of  $PS$  to  $QR$ ,  $\rho$  instead of unity, be  $\theta$ .

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

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

In determining  $\tau$  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  $\rho$  from 1 to 1.01 can be accurately measured.

We observe ( $\alpha$ ) 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 ( $\beta$ ) the greatest deflection produced by the permanent current when the resistance of one of the arms is increased in the ratio of 1 to  $\rho$ , 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  $\rho$  till  $\beta = 2\alpha$  nearly; then

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

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  $\tau$  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  $\mu$  is the same in whatever direction the lines of force pass through the substance, but in crystallized, strained, and organized solids the value of  $\mu$  may depend on the direction of the lines of force with respect to the axes of crystallization, strain, or growth.

In all bodies  $\mu$  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  $\gamma$  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  $\gamma$  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

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

#### *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

$$F = \int P dt.$$

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

$$P = -\frac{dF}{dt}, \quad Q = -\frac{dG}{dt}, \quad R = -\frac{dH}{dt}. \quad \dots \dots \dots \quad (29)$$

#### *Electromagnetic Momentum of a Circuit.*

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

$$\int \left( F \frac{dx}{ds} + G \frac{dy}{ds} + H \frac{dz}{ds} \right) ds. \quad \dots \dots \dots \quad (30)$$

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

$$\left( \frac{dH}{dy} - \frac{dG}{dz} \right) dy dz,$$

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

#### *Magnetic Force ( $\alpha, \beta, \gamma$ ).*

(59) Let  $\alpha, \beta, \gamma$  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 ( $\mu$ ).*

(60) Let  $\mu$  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  $\mu\alpha$  ( $\mu$  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.*

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

*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,

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

a complete differential of  $\phi$ , the magnetic potential.

The quantity  $\phi$  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  $\phi$  corresponding to a passage completely round a current of strength  $c$  being  $4\pi c$ .

Hence if there is no electric current,

$$\frac{dy}{dx} - \frac{d\beta}{d\gamma} = 0;$$

but if there is a current  $p'$ ,

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

Similarly,

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

We may call these the Equations of Currents.

*Electromotive Force in a Circuit.*

(63) Let  $\xi$  be the electromotive force acting round the circuit A, then

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

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

$$\int \left( F \frac{dx}{ds} + G \frac{dy}{ds} + H \frac{dz}{ds} \right) ds = Lu + Mv, \dots \dots \dots \dots \dots \quad (33)$$

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

$$\xi = -\frac{d}{dt}(Lu + Mv). \dots \dots \dots \dots \dots \dots \dots \dots \quad (34)$$

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

$$\left. \begin{aligned} P &= -\frac{dF}{dt} - \frac{d\Psi}{dx}, \\ Q &= -\frac{dG}{dt} - \frac{d\Psi}{dy}, \\ R &= -\frac{dH}{dt} - \frac{d\Psi}{dz}, \end{aligned} \right\} \dots \dots \dots \dots \dots \quad (35)$$

where  $\Psi$  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  $\Psi$  can always, however, be determined in any particular case when we know the actual conditions of the question. The physical interpretation of  $\Psi$  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

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

will be increased by the following increments,

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

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

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

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

*Equations of Electromotive Force.*

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

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  $\Psi$ . 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,*

$$\left. \begin{array}{l} P = kf, \\ Q = kg, \\ R = kh. \end{array} \right\} \dots \dots \dots \dots \dots \quad (E)$$

*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  $\rho$  is the specific resistance referred to unit of volume, we may write the

*Equations of Electric Resistance,*

$$\left. \begin{array}{l} P = -\rho p, \\ Q = -\rho q, \\ R = -\rho r. \end{array} \right\} \dots \dots \dots \dots \dots \quad (F)$$

*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,*

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

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

*Equation of Continuity,*

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

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

quantities, namely,

For Electromagnetic Momentum . . . . .	F	G	H
„ Magnetic Intensity . . . . .	$\alpha$	$\beta$	$\gamma$
„ 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 . . . . .	$\Psi$		

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

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

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

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

Integrating by parts, and remembering that  $\alpha, \beta, \gamma$  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

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

where  $\alpha, \beta, \gamma$  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  $\mu$  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

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

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

$$\int P df$$

from  $P=0$  to the given value of  $P$ .

Since  $P=kf$ , equation (E), this quantity becomes

$$\int k f df = \frac{1}{2} k f^2 = \frac{1}{2} Pf.$$

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

$$\frac{1}{2} \Sigma (Pf + Qg + Rh) dV.$$

The total energy existing in the field is therefore

$$E = \Sigma \left\{ \frac{1}{8\pi} (\alpha \mu \alpha + \beta \mu \beta + \gamma \mu \gamma) + \frac{1}{2} (Pf + Qg + Rh) \right\} dV. \dots \dots \quad (I)$$

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

$$a \left( \frac{dF}{dx} \frac{dx}{ds} + \frac{dG}{dx} \frac{dy}{ds} + \frac{dH}{dx} \frac{dz}{ds} \right) \delta x.$$

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

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

The total increment is

$$a \delta x \left\{ \frac{dy}{ds} \left( \frac{dG}{dx} - \frac{dF}{dy} \right) - \frac{dz}{ds} \left( \frac{dF}{dz} - \frac{dH}{dx} \right) \right\},$$

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  $Xa \delta x$ .

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

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

or

$$X = \mu\gamma q' - \mu\beta r'.)$$

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,

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

Substituting these values of  $\alpha$ ,  $\beta$ ,  $\gamma$  in the expression (equation 38) for the intrinsic energy of the field, and integrating by parts, it becomes

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

The expression

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

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  $4\pi$  lines belonging to it, since it produces unit of magnetic intensity at unit of distance over a sphere whose surface is  $4\pi$ .

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

$$E = -\sum (\frac{1}{2} \phi m) dV. \quad \dots \quad (40)$$

If there are two magnetic poles  $m_1$  and  $m_2$  producing potentials  $\phi_1$  and  $\phi_2$  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

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

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

Since the distribution  $\phi_1$  is determined by  $m_1$ , and  $\phi_2$  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

$$X = m_2 \frac{d\phi_1}{dx} = m_2 \alpha_1,$$

Similarly,  $\mathbf{Y} = m_2 \beta_1$ ,

$$Z = m_2 \gamma_1.$$

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  $\phi$  is

where  $m_i$  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

$$m_2 \frac{d\varphi_1}{dr} = \frac{m_1 m_2}{\mu r^2}. \quad . . . . . \quad (42)$$

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

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

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

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

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  $\Psi_1$ , has for components

$$\left. \begin{aligned} X &= e_2 \frac{d\Psi_1}{dx} = -P_1 e_2, \\ Y &= e_2 \frac{d\Psi_1}{dy} = -Q_1 e_2, \\ Z &= e_2 \frac{d\Psi_1}{dz} = -R_1 e_2. \end{aligned} \right\} \quad \dots \dots \dots \dots \dots \dots \quad (D)$$

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 electricity, the only solution of  $\Psi_1$  is

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

where  $r$  is the distance from the electrified body.

The repulsion between two electrified bodies  $e_1, e_2$  is therefore

$$e_2 \frac{d\Psi_1}{dr} = \frac{k}{4\pi} \frac{e_1 e_2}{r^2}. \quad \dots \dots \dots \dots \dots \quad (44)$$

#### *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  $\eta_1, \eta_2$  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  $\eta_1 = v e_1$  and  $\eta_2 = v e_2$ , and this repulsion becomes

$$v^2 \frac{e_1 e_2}{r^2} = \frac{k}{4\pi} \frac{e_1 e_2}{r^2} \text{ by equation (44), } \dots \dots \dots \dots \quad (45)$$

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

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

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 electro-motive 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  $\delta x$ , and  $X'$  the magnetic force,

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

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

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

or

$$E = C - E'$$

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

where  $\alpha, \beta, \gamma$  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

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

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  $\Psi_1$ , and on the other side  $\Psi_1 + 1$ , and within its substance

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

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

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

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

$$a_1 k_1 + a_2 k_2 + \&c. = ak, \dots \dots \dots \dots \quad (50)$$

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  $\Psi_1$ , and the electricity per unit of surface  $e_1$ .

Let the corresponding quantities at the boundary of the first and second surface be  $\Psi_2$  and  $e_2$ , and so on. Then by equations (G) and (H),

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

But by equations (E) and (F),

$$\left. \begin{array}{l} \Psi_1 - \Psi_2 = a_1 k_1 f_1 = -r_1 p_1, \\ \Psi_2 - \Psi_3 = a_2 k_2 f_2 = -r_2 p_2, \\ & \text{&c.} \end{array} \right\} \dots \dots \dots \dots \dots \dots \quad (52)$$

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

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

where  $\Psi$  is the total difference of potentials between the extreme layers. We have then

$$\left. \begin{array}{l} f_1 = -\frac{\Psi}{r} \frac{r_1}{a_1 k_1}, \quad f_2 = -\frac{\Psi}{r} \frac{r_2}{a_2 k_2}, \text{ &c.} \\ \text{and} \\ e_1 = \frac{\Psi}{r} \frac{r_1}{a_1 k_1}, \quad e_2 = \frac{\Psi}{r} \left( \frac{r_2}{a_2 k_2} - \frac{r_1}{a_1 k_1} \right), \text{ &c.} \end{array} \right\} \dots \dots \dots \dots \quad (53)$$

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

$$\Psi' = a_1 k_1 e'_1 + a_2 k_2 (e'_1 + e'_2) + a_3 k_3 (e'_1 + e'_2 + e'_3), \text{ &c.} = 0, \dots \dots \dots \quad (54)$$

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

$$e'_1 = \frac{\Psi}{r} \frac{r_1}{a_1 k_1} - \frac{\Psi}{ak} = e_1 - \frac{\Psi}{ak}. \dots \dots \dots \quad (55)$$

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  $\Psi'$  be the difference of potential of the extreme surfaces at any time  $t$ ; then

$$\Psi' = a_1 k_1 f_1 + a_2 k_2 f_2 + \text{&c.}; \dots \dots \dots \quad (56)$$

but

$$a_1 k_1 f_1 = -r_1 \frac{df_1}{dt},$$

$$a_2 k_2 f_2 = -r_2 \frac{df_2}{dt}.$$

Hence  $f_1 = A_1 e^{-\frac{a_1 k_1}{r_1} t}$ ,  $f_2 = A_2 e^{-\frac{a_2 k_2}{r_2} t}$ , &c.; and by referring to the values of  $e'_1$ ,  $e_2$ , &c., we find

$$\left. \begin{aligned} A_1 &= \frac{\Psi}{r} \frac{r_1}{a_1 k_1} - \frac{\Psi}{ak}, \\ A_2 &= \frac{\Psi}{r} \frac{r_2}{a_2 k_2} - \frac{\Psi}{ak}, \\ &\text{&c.}; \end{aligned} \right\} \quad \dots \dots \dots \dots \dots \dots \quad (57)$$

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

$$\Psi' = \Psi \left\{ \left( \frac{r_1}{r} - \frac{a_1 k_1}{ak} \right) e^{-\frac{a_1 k_1}{r_1} t} + \left( \frac{r_2}{r} - \frac{a_2 k_2}{ak} \right) e^{-\frac{a_2 k_2}{r_2} t} + \text{&c.} \right\}. \quad \dots \dots \dots \quad (58)$$

(89) It appears from this result that if all the layers are made of the same substance,  $\Psi'$  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  $\Psi'$  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,

$$\Psi' = p_1 r_1 + p_2 r_2 + \text{&c.} = pR. \quad \dots \dots \dots \quad (59)$$

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

$$q_1 r_1 + q_2 r_2 + \text{&c.} = qR. \quad \dots \dots \dots \quad (60)$$

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

$$qR = \frac{\Psi}{r} \left( \frac{r_1^2}{a_1 k_1} + \frac{r_2^2}{a_2 k_2} + \text{&c.} \right) - \frac{\Psi r}{ak},$$

or

$$q = \frac{\Psi}{akrR} \left\{ a_1 k_1 a_2 k_2 \left( \frac{r_1}{a_1 k_1} - \frac{r_2}{a_2 k_2} \right)^2 + a_2 k_2 a_3 k_3 \left( \frac{r_2}{a_2 k_2} - \frac{r_3}{a_3 k_3} \right)^2 + \text{&c.} \right\}, \quad \dots \dots \quad (61)$$

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

$$w = lx + my + nz - Vt.$$

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

$$\mu\alpha = m \frac{dH}{dw} - n \frac{dG}{dw},$$

$$\mu\beta = n \frac{dF}{dw} - l \frac{dH}{dw},$$

$$\mu\gamma = l \frac{dG}{dw} - m \frac{dF}{dw}.$$

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

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

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

$$\frac{dF}{dx} + \frac{dG}{dy} + \frac{dH}{dz} = J, \text{ and } \frac{d^2}{dx^2} + \frac{d^2}{dy^2} + \frac{d^2}{dz^2} = \nabla^2, \dots \dots \dots \quad (63)$$

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

\* 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),

$$p' = \frac{df}{dt}, \quad q' = \frac{dg}{dt}, \quad r' = \frac{dh}{dt}. \quad \dots \dots \dots \quad (65)$$

But these electric displacements are caused by electromotive forces, and by the equations of Electric Elasticity (E),

$$P = kf, \quad Q = kg, \quad R = kh. \quad \dots \dots \dots \quad (66)$$

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

$$\left. \begin{aligned} P &= -\frac{dF}{dt} - \frac{d\Psi}{dx}, \\ Q &= -\frac{dG}{dt} - \frac{d\Psi}{dy}, \\ R &= -\frac{dH}{dt} - \frac{d\Psi}{dz}. \end{aligned} \right\} \quad \dots \dots \dots \quad (67)$$

(94) Combining these equations, we obtain the following:—

$$\left. \begin{aligned} k \left( \frac{dJ}{dx} - \nabla^2 F \right) + 4\pi\mu \left( \frac{d^2 F}{dt^2} + \frac{d^2 \Psi}{dxdt} \right) &= 0, \\ k \left( \frac{dJ}{dy} - \nabla^2 G \right) + 4\pi\mu \left( \frac{d^2 G}{dt^2} + \frac{d^2 \Psi}{dydt} \right) &= 0, \\ k \left( \frac{dJ}{dz} - \nabla^2 H \right) + 4\pi\mu \left( \frac{d^2 H}{dt^2} + \frac{d^2 \Psi}{dzdt} \right) &= 0. \end{aligned} \right\} \quad \dots \dots \dots \quad (68)$$

If we differentiate the third of these equations with respect to  $y$ , and the second with respect to  $z$ , and subtract,  $J$  and  $\Psi$  disappear, and by remembering the equations (B) of magnetic force, the results may be written

$$\left. \begin{aligned} k\nabla^2\mu\alpha &= 4\pi\mu \frac{d^2}{dt^2}\mu\alpha, \\ k\nabla^2\mu\beta &= 4\pi\mu \frac{d^2}{dt^2}\mu\beta, \\ k\nabla^2\mu\gamma &= 4\pi\mu \frac{d^2}{dt^2}\mu\gamma. \end{aligned} \right\} \quad \dots \dots \dots \quad (69)$$

(95) If we assume that  $\alpha$ ,  $\beta$ ,  $\gamma$  are functions of  $lx+my+nz-Vt=w$ , the first equation becomes

$$k\mu \frac{d^2\alpha}{dw^2} = 4\pi\mu^2 V^2 \frac{d^2\alpha}{dw^2}, \quad \dots \dots \dots \quad (70)$$

or

$$V = \pm \sqrt{\frac{k}{4\pi\mu}}. \quad \dots \dots \dots \quad (71)$$

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),

$$V=v. \dots \dots \dots \dots \dots \dots \dots \quad (72)$$

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  $\Psi$  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  $\chi$  from the equation

$$\nabla^2\chi = \frac{d^2\chi}{dx^2} + \frac{d^2\chi}{dy^2} + \frac{d^2\chi}{dz^2} = J, \quad \dots \dots \dots \dots \dots \dots \quad (73)$$

and  $F'$ ,  $G'$ ,  $H'$  from the equations

$$F' = F - \frac{d\chi}{dx}, \quad G' = G - \frac{d\chi}{dy}, \quad H' = H - \frac{d\chi}{dz}, \quad \dots \dots \dots \quad (74)$$

then

$$\frac{dF'}{dx} + \frac{dG'}{dy} + \frac{dH'}{dz} = 0, \quad \dots \dots \dots \dots \dots \dots \quad (75)$$

and the equations in (94) become of the form

$$k\nabla^2F' = 4\pi\mu \left( \frac{d^2F'}{dt^2} + \frac{d}{dxdt} \left( \Psi + \frac{d\chi}{dt} \right) \right). \quad \dots \dots \dots \quad (76)$$

Differentiating the three equations with respect to  $x$ ,  $y$ , and  $z$ , and adding, we find that

$$\Psi = -\frac{d\chi}{dt} + \phi(x, y, z), \quad \dots \dots \dots \dots \dots \quad (77)$$

and that

$$\left. \begin{aligned} k\nabla^2F' &= 4\pi\mu \frac{d^2F'}{dt^2}, \\ k\nabla^2G' &= 4\pi\mu \frac{d^2G'}{dt^2}, \\ k\nabla^2H' &= 4\pi\mu \frac{d^2H'}{dt^2}. \end{aligned} \right\} \quad \dots \dots \dots \dots \dots \quad (78)$$

Hence the disturbances indicated by  $F'$ ,  $G'$ ,  $H'$  are propagated with the velocity  $V = \sqrt{\frac{k}{4\pi\mu}}$  through the field; and since

$$\frac{dF'}{dx} + \frac{dG'}{dy} + \frac{dH'}{dz} = 0,$$

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  $\chi$ , 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  $\nabla^2$  on this equation, it becomes

$$ke = \frac{dJ}{dt} - k\nabla^2\phi(x, y, z). \quad \dots \dots \dots \dots \dots \quad (79)$$

Since the medium is a perfect insulator,  $e$ , the free electricity, is immovable, and therefore  $\frac{dJ}{dt}$  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  $D = \frac{i^2}{\mu}, \dots \dots \dots \dots \dots \dots \dots \quad (80)$

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  $\mu$  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  $\lambda, \mu, \nu$ , then the equations of magnetic force ( $B$ ) become

$$\left. \begin{aligned} \lambda\alpha &= \frac{dH}{dy} - \frac{dG}{dz}, \\ \mu\beta &= \frac{dF}{dz} - \frac{dH}{dx}, \\ \nu\gamma &= \frac{dG}{dx} - \frac{dF}{dy}. \end{aligned} \right\} \dots \dots \dots \dots \dots \dots \dots \quad (81)$$

The equations of electric currents (C) remain as before.

The equations of electric elasticity (E) will be

$$\left. \begin{aligned} P &= 4\pi a^2 f, \\ Q &= 4\pi b^2 g, \\ R &= 4\pi c^2 h, \end{aligned} \right\} \quad \dots \dots \dots \dots \dots \quad (82)$$

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

$$\frac{1}{\mu\nu} \left( \lambda \frac{d^2 F}{dx^2} + \mu \frac{d^2 F}{dy^2} + \nu \frac{d^2 F}{dz^2} \right) - \frac{1}{\mu\nu} \frac{d}{dx} \left( \lambda \frac{dF}{dx} + \mu \frac{dG}{dy} + \nu \frac{dH}{dz} \right) = \frac{1}{a^2} \left( \frac{d^2 F}{dt^2} + \frac{d^2 \Psi}{dxdt} \right). \quad \dots \quad (83)$$

(104) If  $l$ ,  $m$ ,  $n$  are the direction-cosines of the wave, and  $V$  its velocity, and if

$$lx + my + nz - Vt = w, \quad \dots \dots \dots \quad (84)$$

then  $F$ ,  $G$ ,  $H$ , and  $\Psi$  will be functions of  $w$ ; and if we put  $F'$ ,  $G'$ ,  $H'$ ,  $\Psi'$  for the second differentials of these quantities with respect to  $w$ , the equations will be

$$\left. \begin{aligned} \left( V^2 - a^2 \left( \frac{m^2}{\nu} + \frac{n^2}{\mu} \right) \right) F' + \frac{a^2 lm}{\nu} G' + \frac{a^2 ln}{\mu} H' - lV\Psi' &= 0, \\ \left( V^2 - b^2 \left( \frac{n^2}{\lambda} + \frac{l^2}{\nu} \right) \right) G' + \frac{b^2 mn}{\lambda} H' + \frac{b^2 ml}{\nu} F' - mV\Psi' &= 0, \\ \left( V^2 - c^2 \left( \frac{l^2}{\mu} + \frac{m^2}{\lambda} \right) \right) H' + \frac{c^2 nl}{\mu} F' + \frac{c^2 nm}{\lambda} G' - nV\Psi' &= 0. \end{aligned} \right\} \quad \dots \dots \quad (85)$$

If we now put

$$\left. \begin{aligned} V^4 - V^2 \frac{1}{\lambda\mu\nu} \left\{ l^2 \lambda (b^2 \mu + c^2 \nu) + m^2 \mu (c^2 \nu + a^2 \lambda) + n^2 \nu (a^2 \lambda + b^2 \mu) \right\} \\ + \frac{a^2 b^2 c^2}{\lambda\mu\nu} \left( \frac{l^2}{a^2} + \frac{m^2}{b^2} + \frac{n^2}{c^2} \right) (l^2 \lambda + m^2 \mu + n^2 \nu) = U, \end{aligned} \right\} \quad \dots \quad (86)$$

we shall find

$$F'V^2U - l\Psi'VU = 0, \quad \dots \dots \quad (87)$$

with two similar equations for  $G'$  and  $H'$ . Hence either

$$V = 0, \quad \dots \dots \quad (88)$$

$$U = 0, \quad \dots \dots \quad (89)$$

or

$$VF' = l\Psi', \quad VG' = m\Psi', \quad VH' = n\Psi'. \quad \dots \quad (90)$$

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  $\Psi'$  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

$$\frac{l}{a^2} F' + \frac{m}{b^2} G' + \frac{n}{c^2} H' = 0, \dots \dots \dots \dots \dots \dots \quad (91)$$

$$\frac{a^2 l \lambda}{F'} (b^2 \mu - c^2 \nu) + \frac{b^2 m \mu}{G'} (c^2 \nu - a^2 \lambda) + \frac{c^2 n \nu}{H'} (a^2 \lambda - b^2 \mu) = 0, \dots \dots \dots \quad (92)$$

(105) The velocities along the axes are as follows:—

Direction of propagation . . . . .

Direction of the electric displacements

	$x$	$y$	$z$
$x$		$\frac{a^2}{\nu}$	$\frac{a^2}{\mu}$
$y$	$\frac{b^2}{\nu}$		$\frac{b^2}{\lambda}$
$z$	$\frac{c^2}{\mu}$	$\frac{c^2}{\lambda}$	

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

$$a^2 = b^2 = c^2. \dots \dots \dots \dots \dots \dots \quad (93)$$

If, on the contrary, the electric displacements are perpendicular to the plane of polarization,

$$\lambda = \mu = \nu. \dots \dots \dots \dots \dots \dots \quad (94)$$

We know, from the magnetic experiments of FARADAY, PLÜCKER, &c., that in many crystals  $\lambda$ ,  $\mu$ ,  $\nu$  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  $\lambda$ ,  $\mu$ ,  $\nu$  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  $\rho$ , 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

$$\begin{aligned}\frac{d^2G}{dx^2} &= -4\pi\mu(q') \\ &= -4\pi\mu\left(\frac{df}{dt} + q\right) \text{ by (A),} \\ \frac{d^2G}{dx^2} &= +4\pi\mu\left(\frac{1}{k}\frac{d^2G}{dt^2} - \frac{1}{\rho}\frac{dG}{dt}\right) \text{ by (E) and (F). . . . .} \quad (95)\end{aligned}$$

If  $G$  is of the form

$$G = e^{-px} \cos(qx + nt), \quad . . . . . \quad (96)$$

we find that

$$p = \frac{2\pi\mu}{\rho} \frac{n}{q} = \frac{2\pi\mu}{\rho} \frac{V}{i}, \quad . . . . . \quad (97)$$

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

$$e^{-2px}. \quad . . . . . \quad (98)$$

Let  $R$  be the resistance in electromagnetic measure of a plate of the substance whose thickness is  $x$ , breadth  $b$ , and length  $l$ , then

$$\begin{aligned}R &= \frac{l\rho}{bx}, \\ 2px &= 4\pi\mu \frac{V}{i} \frac{l}{bR}. \quad . . . . . \quad (99)\end{aligned}$$

(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

$$P = \sqrt{8\pi\mu V W},$$

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

$$13,000 \text{ DANIELL'S cells per metre.}$$

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

$$M = \int \left( F \frac{dx}{ds'} + G \frac{dy}{ds'} + H \frac{dz}{ds'} \right) ds',$$

where F, G, H are the components of electromagnetic momentum 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)

$$\frac{d^2 F}{dx^2} + \frac{d^2 F}{dy^2} + \frac{d^2 F}{dz^2} = -4\pi\mu p',$$

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

$$p' = \frac{dx}{ds} ds, \quad q' = \frac{dy}{ds} ds, \quad r' = \frac{dz}{ds} ds,$$

and the solution of the equation gives

$$F = \frac{\mu}{\rho} \frac{dx}{ds} ds, \quad G = \frac{\mu}{\rho} \frac{dy}{ds} ds, \quad H = \frac{\mu}{\rho} \frac{dz}{ds} ds,$$

where  $\rho$  is the distance of any point from  $ds$ . Hence

$$\begin{aligned} M &= \iint \frac{\mu}{\rho} \left( \frac{dx}{ds} \frac{dx}{ds'} + \frac{dy}{ds} \frac{dy}{ds'} + \frac{dz}{ds} \frac{dz}{ds'} \right) ds ds' \\ &= \iint \frac{\mu}{\rho} \cos \theta ds ds', \end{aligned}$$

where  $\theta$  is the angle between the directions of the two elements  $ds$ ,  $ds'$ , and  $\rho$  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

$$M = \Sigma (\mu \alpha l + \mu \beta m + \mu \gamma n) dS',$$

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

$$M = \mu \sum \frac{1}{\rho^2} \sin \theta \sin \theta' \sin \phi dS' ds,$$

where  $dS'$  is an element of the surface bounded by  $B$ ,  $ds$  is an element of the circuit  $A$ ,  $\rho$  is the distance between them,  $\theta$  and  $\theta'$  are the angles between  $\rho$  and  $ds$  and between  $\rho$  and the normal to  $dS'$  respectively, and  $\phi$  is the angle between the planes in which  $\theta$  and  $\theta'$  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  $\alpha, \beta, \gamma$  be the components of magnetic intensity at any point due to the first circuit,  $\alpha', \beta', \gamma'$  the same for the second circuit; then the intrinsic energy of the element of volume  $dV$  of the field is

$$\frac{\mu}{8\pi} ((\alpha + \alpha')^2 + (\beta + \beta')^2 + (\gamma + \gamma')^2) dV.$$

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

$$\frac{\mu}{4\pi} \sum \mu I I' \cos \theta dV$$

over all space, where  $\theta$  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

$$\frac{d^2M}{da^2} + \frac{d^2M}{db^2} + \frac{1}{a} \frac{dM}{da} = 0.$$

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:—

$$M = 4\pi a \log \frac{8a}{r} \left\{ 1 + \frac{1}{2} \frac{a-a_1}{a} + \frac{1}{16} \frac{3b^2 + (a_1-a)^2}{a^2} - \frac{1}{32} \frac{(3b^2 + (a-a_1)^2)(a-a_1)}{a^3} + \text{&c.} \right\}$$

$$- 4\pi a \left\{ 2 + \frac{1}{2} \frac{a-a_1}{a} + \frac{1}{16} \frac{b^2 - 3(a-a_1)^2}{a^2} - \frac{1}{48} \frac{(6b^2 - (a-a_1)^2)(a-a_1)}{a^3} + \text{&c.} \right\}.$$

(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,

$$L = \frac{n^2}{b^2 c^2} \iiint M(xy x'y') dx dy dx' dy',$$

where 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.

$$L = 4\pi n^2 a \left\{ \log \frac{8a}{r} + \frac{1}{12} - \frac{4}{3} \left( \theta - \frac{\pi}{4} \right) \cot 2\theta - \frac{\pi}{3} \cos 2\theta - \frac{1}{6} \cot^2 \theta \log \cos \theta - \frac{1}{6} \tan^2 \theta \log \sin \theta \right\}$$

$$+ \frac{\pi n^2 r^2}{24a} \left\{ \log \frac{8a}{r} (2 \sin^2 \theta + 1) + 3.45 + 27.475 \cos^2 \theta - 3.2 \left( \frac{\pi}{2} - \theta \right) \frac{\sin^3 \theta}{\cos \theta} + \frac{1}{5} \frac{\cos^4 \theta}{\sin^2 \theta} \log \cos \theta \right.$$

$$\left. + \frac{13}{3} \frac{\sin^4 \theta}{\cos^2 \theta} \log \sin \theta \right\} + \text{&c.}$$

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

$$A, B, C, A+B, B+C, A+B+C,$$

and  $n=1$  in each case.

Calling the results

$$L(A), L(B), L(C), \text{ &c.,}$$

we calculate the coefficient of mutual induction M(AC) of the two coils thus,

$$2ACM(AC) = (A+B+C)^2 L(A+B+C) - (A+B)^2 L(A+B) - (B+C)^2 L(B+C) + B^2 L(B).$$

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

$$L = n_1^2 L(A) + 2n_1 n_2 L(AC) + n_2^2 L(B).$$

(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

$$+0.0236.$$

For the sixteen in the next row

$$+0.00083.$$

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  $\rho$ , we have

$$p\rho = P - \frac{dF}{dt},$$

where  $F$  is got from the equation

$$\frac{d^2F}{dr^2} + \frac{1}{r} \frac{dF}{dr} = -4\pi\mu p,$$

$r$  being the distance from the axis of the cylinder.

Let one term of the value of  $F$  be of the form  $T r^n$ , where  $T$  is a function of the time, then the term of  $p$  which produced it is of the form

$$-\frac{1}{4\pi\mu} n^2 T r^{n-2}.$$

Hence if we write

$$F = T + \frac{\mu\pi}{\rho} \left( -P + \frac{dT}{dt} \right) r^2 + \frac{\mu\pi}{\rho} \left[ \frac{1}{1^2 \cdot 2^2} \frac{dT^2}{dt^2} r^4 + \dots \right] + \text{&c.}$$

$$p\rho = \left( P + \frac{dT}{dt} \right) - \frac{\mu\pi}{\rho} \frac{d^2T}{dt^2} r^2 - \frac{\mu\pi}{\rho} \left[ \frac{1}{1^2 \cdot 2^2} \frac{d^3T}{dt^3} r^4 - \dots \right] + \text{&c.}$$

The total counter current of self-induction at any point is

$$\int \left( \frac{P}{\rho} - p \right) dt = \frac{1}{\rho} T + \frac{\mu\pi}{\rho^2} \frac{dT}{dt} r^2 + \frac{\mu^2\pi^2}{\rho^3} \frac{1}{1^2 2^2} \frac{d^2T}{dt^2} r^4 + \text{&c.}$$

from  $t=0$  to  $t=\infty$ .

$$\text{When } t=0, p=0, \quad \therefore \left( \frac{dT}{dt} \right)_0 = P, \quad \left( \frac{d^2T}{dt^2} \right)_0 = 0, \quad \text{&c.}$$

$$\text{When } t=\infty, p=\frac{P}{\rho}, \quad \therefore \left( \frac{dT}{dt} \right)_\infty = 0, \quad \left( \frac{d^2T}{dt^2} \right)_\infty = 0, \quad \text{&c.}$$

$$\int_0^\infty \int_0^r 2\pi \left( \frac{P}{\rho} - p \right) r dr dt = \frac{1}{\rho} T \pi r^2 + \frac{1}{2} \frac{\mu\pi^2}{\rho^2} \frac{dT}{dt} r^4 + \frac{\mu^2\pi^3}{\rho^3} \frac{1}{1^2 \cdot 2^2 \cdot 3} \frac{d^2T}{dt^2} r^6 + \text{&c.}$$

from  $t=0$  to  $=\infty$ .

$$\text{When } t=0, p=0 \text{ throughout the section, } \quad \therefore \left( \frac{dT}{dt} \right)_0 = P, \quad \left( \frac{d^2T}{dt^2} \right)_0 = 0, \quad \text{&c.}$$

$$\text{When } t=\infty, p=0 \text{ throughout } \dots \quad \therefore \left( \frac{dT}{dt} \right)_\infty = 0, \quad \left( \frac{d^2T}{dt^2} \right)_\infty = 0, \quad \text{&c.}$$

Also if  $l$  be the length of the wire, and  $R$  its resistance,

$$R = \frac{\rho l}{\pi r^2};$$

and if  $C$  be the current when established in the wire,  $C = \frac{Pl}{R}$ .

The total counter current may be written

$$\frac{l}{R} (T_\infty - T_0) - \frac{1}{2} \mu \frac{l}{R} C = -\frac{LC}{R} \text{ by } \S (35).$$

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  $\gamma$  is the current in the wire at any instant, and the total countercurrent would have been

$$\int_0^\infty \int_0^r \frac{1}{\rho} \frac{dF}{dt} 2\pi r dr = \frac{l}{R} (T_\infty - T_0) - \frac{3}{4} \mu \frac{l}{R} C = - \frac{L' C}{R}, \text{ say.}$$

Hence

$$L = L' - \frac{1}{4} \mu l,$$

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  $\mu$  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 . . . . .	$=a=158194$
Depth of each coil . . . . .	$=b=01608$
Breadth of each coil . . . . .	$=c=01841$
Distance between the coils . . . . .	$=02010$
Number of windings . . . . .	$n=313$
Diameter of wire . . . . .	$=00126$

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	} + 44997
per unit of length . . . . .	
Correction of eight neighbouring wires . . . . .	+ 0236
For sixteen wires next to these . . . . .	+ 0008
Correction for variation of current in different parts of section	<u>— 2500</u>
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.

THE  
AMERICAN JOURNAL OF SCIENCE.

[THIRD SERIES.]

---

ART. XXXVI.—*On the Relative Motion of the Earth and the Luminiferous Ether*; by ALBERT A. MICHELSON and EDWARD W. MORLEY.\*

THE discovery of the aberration of light was soon followed by an explanation according to the emission theory. The effect was attributed to a simple composition of the velocity of light with the velocity of the earth in its orbit. The difficulties in this apparently sufficient explanation were overlooked until after an explanation on the undulatory theory of light was proposed. This new explanation was at first almost as simple as the former. But it failed to account for the fact proved by experiment that the aberration was unchanged when observations were made with a telescope filled with water. For if the tangent of the angle of aberration is the ratio of the velocity of the earth to the velocity of light, then, since the latter velocity in water is three-fourths its velocity in a vacuum, the aberration observed with a water telescope should be four-thirds of its true value.†

\* This research was carried out with the aid of the Bache Fund.

† It may be noticed that most writers admit the sufficiency of the explanation according to the emission theory of light; while in fact the difficulty is even greater than according to the undulatory theory. For on the emission theory the velocity of light must be greater in the water telescope, and therefore the angle of aberration should be less; hence, in order to reduce it to its true value, we must make the absurd hypothesis that the motion of the water in the telescope carries the ray of light in the opposite direction!

On the undulatory theory, according to Fresnel, first, the ether is supposed to be at rest except in the interior of transparent media, in which secondly, it is supposed to move with a velocity less than the velocity of the medium in the ratio  $\frac{n^2 - 1}{n^2}$ , where  $n$  is the index of refraction. These two hypotheses give a complete and satisfactory explanation of aberration. The second hypothesis, notwithstanding its seeming improbability, must be considered as fully proved, first, by the celebrated experiment of Fizeau,\* and secondly, by the ample confirmation of our own work.† The experimental trial of the first hypothesis forms the subject of the present paper.

If the earth were a transparent body, it might perhaps be conceded, in view of the experiments just cited, that the intermolecular ether was at rest in space, notwithstanding the motion of the earth in its orbit; but we have no right to extend the conclusion from these experiments to opaque bodies. But there can hardly be question that the ether can and does pass through metals. Lorentz cites the illustration of a metallic barometer tube. When the tube is inclined the ether in the space above the mercury is certainly forced out, for it is incompressible.‡ But again we have no right to assume that it makes its escape with perfect freedom, and if there be any resistance, however slight, we certainly could not assume an opaque body such as the whole earth to offer free passage through its entire mass. But as Lorentz aptly remarks: "quoi qui l'en soit, on fera bien, à mon avis, de ne pas se laisser guider, dans une question aussi importante, par des considérations sur le degré de probabilité ou de simplicité de l'une ou de l'autre hypothèse, mais de s'adresser à l'expérience pour apprendre à connaître l'état, de repos ou de mouvement, dans lequel se trouve l'éther à la surface terrestre."§

In April, 1881, a method was proposed and carried out for testing the question experimentally.||

In deducing the formula for the quantity to be measured, the effect of the motion of the earth through the ether on the path of the ray at right angles to this motion was overlooked.¶

\* Comptes Rendus, xxxiii, 349, 1851; Pogg. Ann. Ergänzungsband, iii, 457, 1853; Ann. Chim. Phys., III, lvii, 385, 1859.

† Influence of Motion of the Medium on the Velocity of Light. This Journal, III, xxxi, 377, 1886.

‡ It may be objected that it may escape by the space between the mercury and the walls; but this could be prevented by amalgamating the walls.

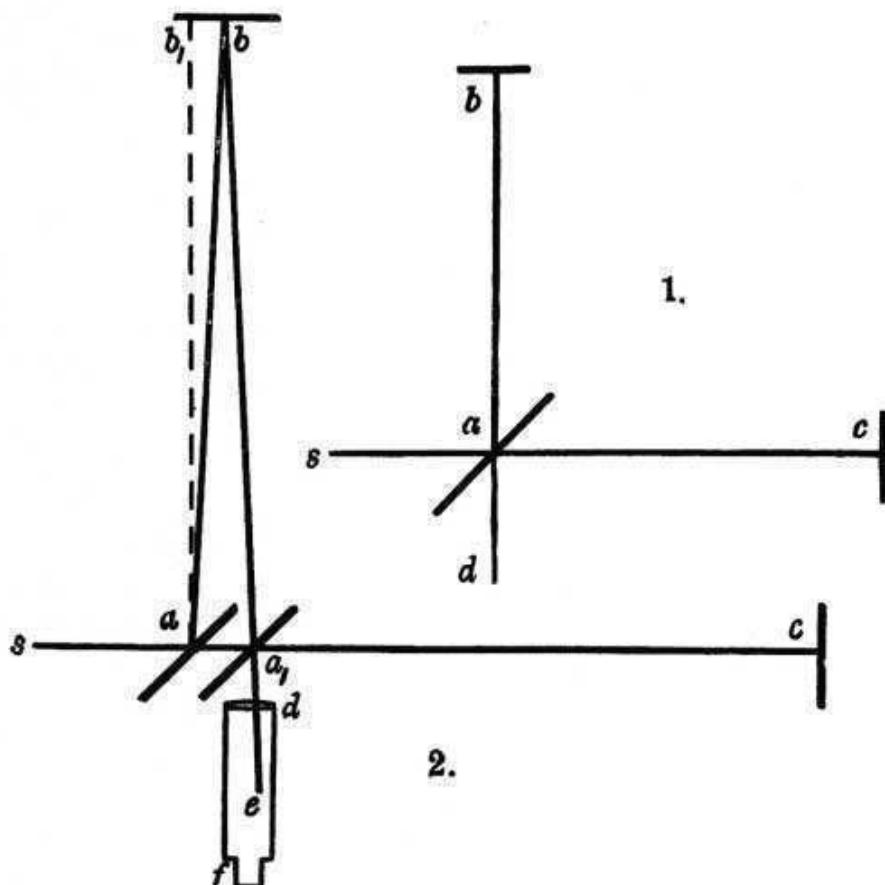
§ Archives Néerlandaises, xxi, 2<sup>me</sup> livr.

|| The relative motion of the earth and the luminiferous ether, by Albert A. Michelson, this Jour., III, xxii, 120.

¶ It may be mentioned here that the error was pointed out to the author of the former paper by M. A. Potier, of Paris, in the winter of 1881.

The discussion of this oversight and of the entire experiment forms the subject of a very searching analysis by H. A. Lorentz,\* who finds that this effect can by no means be disregarded. In consequence, the quantity to be measured had in fact but one-half the value supposed, and as it was already barely beyond the limits of errors of experiment, the conclusion drawn from the result of the experiment might well be questioned; since, however, the main portion of the theory remains unquestioned, it was decided to repeat the experiment with such modifications as would insure a theoretical result much too large to be masked by experimental errors. The theory of the method may be briefly stated as follows:

Let  $sa$ , fig. 1, be a ray of light which is partly reflected in  $ab$ , and partly transmitted in  $ac$ , being returned by the mirrors  $b$  and  $c$ , along  $ba$  and  $ca$ .  $ba$  is partly transmitted along  $ad$ ,



and  $ca$  is partly reflected along  $ad$ . If then the paths  $ab$  and  $ac$  are equal, the two rays interfere along  $ad$ . Suppose now, the ether being at rest, that the whole apparatus moves in the direction  $sc$ , with the velocity of the earth in its orbit, the direc-

\* De l'Influence du Mouvement de la Terre sur les Phen. Lum. Archives Néerlandaises, xxi, 2<sup>me</sup> livr., 1886.

tions and distances traversed by the rays will be altered thus:—The ray  $sa$  is reflected along  $ab$ , fig. 2; the angle  $bab$ , being equal to the aberration  $=\alpha$ , is returned along  $ba$ , ( $aba = 2\alpha$ ), and goes to the focus of the telescope, whose direction is unaltered. The transmitted ray goes along  $ac$ , is returned along  $ca$ , and is reflected at  $a$ , making  $ca_e$  equal  $90 - \alpha$ , and therefore still coinciding with the first ray. It may be remarked that the rays  $ba$ , and  $ca$ , do not now meet exactly in the same point  $a$ , though the difference is of the second order; this does not affect the validity of the reasoning. Let it now be required to find the difference in the two paths  $aba$ , and  $aca$ .

Let  $V$ =velocity of light.

$v$ =velocity of the earth in its orbit.

$D$ =distance  $ab$  or  $ac$ , fig. 1.

$T$ =time light occupies to pass from  $a$  to  $c$ .

$T'$ =time light occupies to return from  $c$  to  $a$ , (fig. 2.)

Then  $T = \frac{D}{V-v}$ ,  $T' = \frac{D}{V+v}$ . The whole time of going and coming is  $T+T' = 2D \frac{V}{V^2-v^2}$ , and the distance traveled in this time is  $2D \frac{V^2}{V^2-v^2} = 2D \left(1 + \frac{v^2}{V^2}\right)$ , neglecting terms of the fourth order.

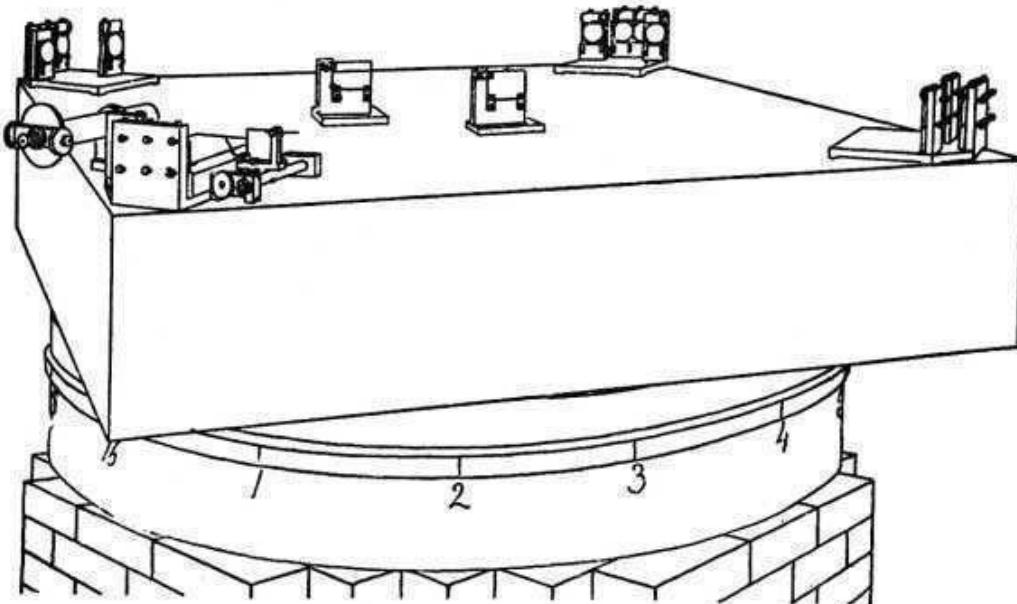
The length of the other path is evidently  $2D \sqrt{1 + \frac{v^2}{V^2}}$ , or to the same degree of accuracy,  $2D \left(1 + \frac{v^2}{2V^2}\right)$ . The difference is therefore  $D \frac{v^2}{V^2}$ . If now the whole apparatus be turned through  $90^\circ$ , the difference will be in the opposite direction, hence the displacement of the interference fringes should be  $2D \frac{v^2}{V^2}$ . Considering only the velocity of the earth in its orbit, this would be  $2D \times 10^{-6}$ . If, as was the case in the first experiment,  $D=2 \times 10^6$  waves of yellow light, the displacement to be expected would be 0.04 of the distance between the interference fringes.

In the first experiment one of the principal difficulties encountered was that of revolving the apparatus without producing distortion; and another was its extreme sensitiveness to vibration. This was so great that it was impossible to see the interference fringes except at brief intervals when working in the city, even at two o'clock in the morning. Finally, as before remarked, the quantity to be observed, namely, a displacement of something less than a twentieth of the distance between the interference fringes may have been too small to be detected when masked by experimental errors.

The first named difficulties were entirely overcome by mounting the apparatus on a massive stone floating on mercury; and the second by increasing, by repeated reflection, the path of the light to about ten times its former value.

The apparatus is represented in perspective in fig. 3, in plan in fig. 4, and in vertical section in fig. 5. The stone *a* (fig. 5) is about 1·5 meter square and 0·3 meter thick. It rests on an annular wooden float *bb*, 1·5 meter outside diameter, 0·7 meter inside diameter, and 0·25 meter thick. The float rests on mercury contained in the cast-iron trough *cc*, 1·5 centimeter thick, and of such dimensions as to leave a clearance of about one centimeter around the float. A pin *d*, guided by arms *gggg*, fits into a socket *e* attached to the float. The pin may be pushed into the socket or be withdrawn, by a lever pivoted at *f*. This pin keeps the float concentric with the trough, but does not bear any part of the weight of the stone. The annular iron trough rests on a bed of cement on a low brick pier built in the form of a hollow octagon.

3.

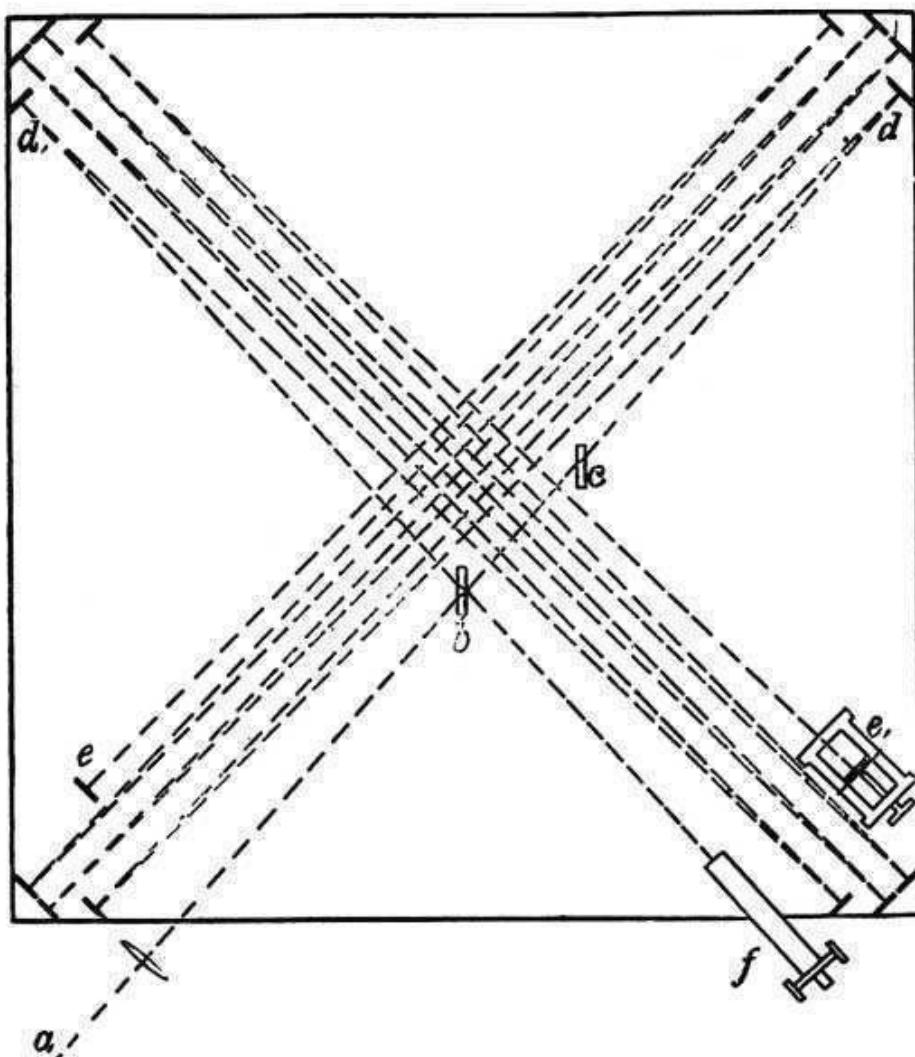


At each corner of the stone were placed four mirrors *dd ee* (fig. 4). Near the center of the stone was a plane-parallel glass *b*. These were so disposed that light from an argand burner *a*, passing through a lens, fell on *b* so as to be in part reflected to *d*; the two pencils followed the paths indicated in the figure, *bdedbf* and *bd,e,d,bf* respectively, and were observed by the telescope *f*. Both *f* and *a* revolved with the stone. The mirrors were of speculum metal carefully worked to optically plane surfaces five centimeters in diameter, and the glasses *b* and *c* were plane-parallel and of the same thickness, 1·25 centimeter;

their surfaces measured 5·0 by 7·5 centimeters. The second of these was placed in the path of one of the pencils to compensate for the passage of the other through the same thickness of glass. The whole of the optical portion of the apparatus was kept covered with a wooden cover to prevent air currents and rapid changes of temperature.

The adjustment was effected as follows: The mirrors having been adjusted by screws in the castings which held the

## 4.



mirrors, against which they were pressed by springs, till light from both pencils could be seen in the telescope, the lengths of the two paths were measured by a light wooden rod reaching diagonally from mirror to mirror, the distance being read from a small steel scale to tenths of millimeters. The difference in the lengths of the two paths was then annulled by moving the mirror  $e$ . This mirror had three adjustments; it had an adjustment in altitude and one in azimuth, like all the other mirrors,

but finer; it also had an adjustment in the direction of the incident ray, sliding forward or backward, but keeping very accurately parallel to its former plane. The three adjustments of this mirror could be made with the wooden cover in position.

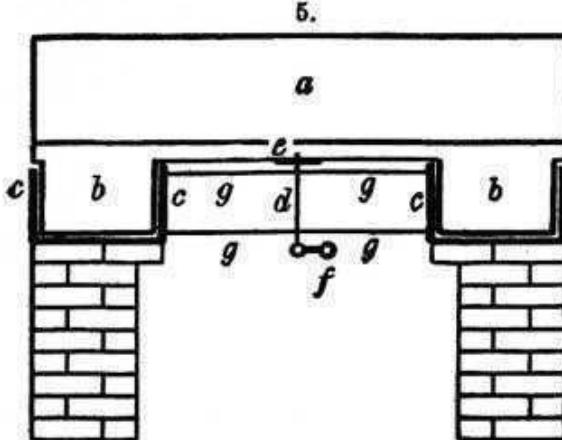
The paths being now approximately equal, the two images of the source of light or of some well-defined object placed in front of the condensing lens, were made to coincide, the telescope was now adjusted for distinct vision of the expected interference bands, and sodium light was substituted for white light, when the interference bands appeared. These were now made as clear as possible by adjusting the mirror  $e$ ; then white light was restored, the screw altering the length of path was very slowly moved (one turn of a screw of one hundred threads to the

inch altering the path nearly 1000 wave-lengths) till the colored interference fringes reappeared in white light. These were now given a convenient width and position, and the apparatus was ready for observation.

The observations were conducted as follows: Around the cast-iron

trough were sixteen equidistant marks. The apparatus was revolved very slowly (one turn in six minutes) and after a few minutes the cross wire of the micrometer was set on the clearest of the interference fringes at the instant of passing one of the marks. The motion was so slow that this could be done readily and accurately. The reading of the screw-head on the micrometer was noted, and a very slight and gradual impulse was given to keep up the motion of the stone; on passing the second mark, the same process was repeated, and this was continued till the apparatus had completed six revolutions. It was found that by keeping the apparatus in slow uniform motion, the results were much more uniform and consistent than when the stone was brought to rest for every observation; for the effects of strains could be noted for at least half a minute after the stone came to rest, and during this time effects of change of temperature came into action.

The following tables give the means of the six readings; the first, for observations made near noon, the second, those near six o'clock in the evening. The readings are divisions of the screw-heads. The width of the fringes varied from 40 to 60 divisions, the mean value being near 50, so that one division



means 0·02 wave-length. The rotation in the observations at noon was contrary to, and in the evening observations, with, that of the hands of a watch.

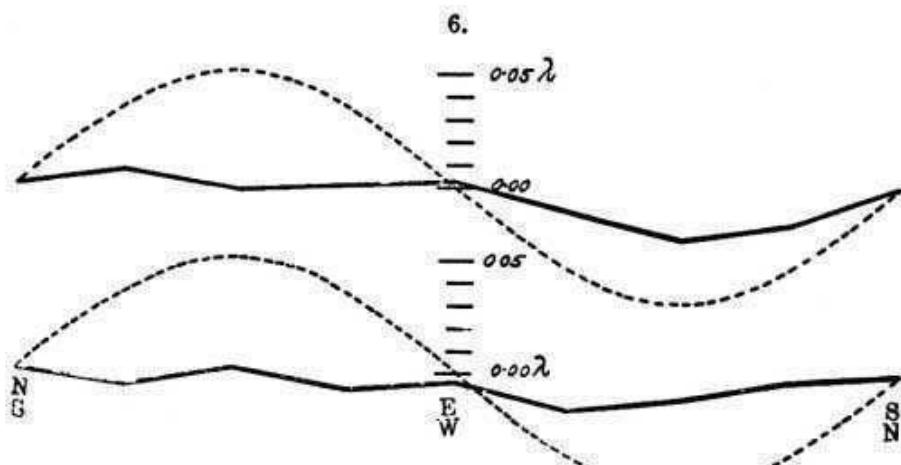
## NOON OBSERVATIONS.

	16.	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.
July 8 .....	44·7	44·0	43·5	39·7	35·2	34·7	34·3	32·5	28·2	26·2	23·8	23·2	20·3	18·7	17·5	16·8	13·7
July 9 .....	57·4	57·3	58·2	59·2	58·7	60·2	60·8	62·0	61·5	63·3	65·8	67·3	69·7	70·7	73·0	70·2	72·2
July 11 .....	27·3	23·5	22·0	19·3	19·2	19·3	18·7	18·8	16·2	14·3	13·3	12·8	13·3	12·3	10·2	7·3	6·5
Mean.....	43·1	41·6	41·2	39·4	37·7	38·1	37·9	37·8	35·3	34·6	34·3	34·4	34·4	33·9	33·6	31·4	30·8
Mean in w. l.	·802	·832	·824	·788	·754	·762	·758	·756	·706	·692	·686	·688	·688	·678	·672	·628	·616
Final mean.	·784	·762	·755	·738	·721	·720	·715	·692	·661								

## P. M. OBSERVATIONS.

July 8 ... .	61·2	63·3	63·3	68·2	67·7	69·3	70·3	69·8	69·0	71·3	71·3	70·5	71·2	71·2	70·5	72·5	75·7
July 9 .....	26·0	26·0	28·2	29·2	31·5	32·0	31·3	31·7	33·0	35·8	36·5	37·3	38·8	41·0	42·7	43·7	44·0
July 12 .....	66·8	66·5	66·0	64·3	62·2	61·0	61·3	59·7	58·2	55·7	53·7	54·7	55·0	58·2	58·5	57·0	56·0
Mean .....	51·3	51·9	52·5	53·9	53·8	54·1	54·3	53·7	53·4	54·3	53·8	54·2	55·0	56·8	57·2	57·7	58·6
Mean in w. l.	1·026	1·038	1·050	1·078	1·076	1·082	1·086	1·074	1·068	1·086	1·076	1·084	1·100	1·136	1·144	1·154	1·172
	1·068	1·086	1·076	1·084	1·100	1·136	1·144	1·154	1·172								
Final mean.	1·047	1·062	1·063	1·081	1·088	1·109	1·115	1·114	1·120								

The results of the observations are expressed graphically in fig. 6. The upper is the curve for the observations at noon, and the lower that for the evening observations. The dotted curves represent *one-eighth* of the theoretical displacements. It seems fair to conclude from the figure that if there is any dis-



placement due to the relative motion of the earth and the luminiferous ether, this cannot be much greater than 0·01 of the distance between the fringes.

Considering the motion of the earth in its orbit only, this

displacement should be  $2D \frac{v^2}{V^2} = 2D \times 10^{-8}$ . The distance D was about eleven meters, or  $2 \times 10^7$  wave-lengths of yellow light; hence the displacement to be expected was 0.4 fringe. The actual displacement was certainly less than the twentieth part of this, and probably less than the fortieth part. But since the displacement is proportional to the square of the velocity, the relative velocity of the earth and the ether is probably less than one sixth the earth's orbital velocity, and certainly less than one-fourth.

In what precedes, only the orbital motion of the earth is considered. If this is combined with the motion of the solar system, concerning which but little is known with certainty, the result would have to be modified; and it is just possible that the resultant velocity at the time of the observations was small though the chances are much against it. The experiment will therefore be repeated at intervals of three months, and thus all uncertainty will be avoided.

It appears, from all that precedes, reasonably certain that if there be any relative motion between the earth and the luminiferous ether, it must be small; quite small enough entirely to refute Fresnel's explanation of aberration. Stokes has given a theory of aberration which assumes the ether at the earth's surface to be at rest with regard to the latter, and only requires in addition that the relative velocity have a potential; but Lorentz shows that these conditions are incompatible. Lorentz then proposes a modification which combines some ideas of Stokes and Fresnel, and assumes the existence of a potential, together with Fresnel's coefficient. If now it were legitimate to conclude from the present work that the ether is at rest with regard to the earth's surface, according to Lorentz there could not be a velocity potential, and his own theory also fails.

#### *Supplement.*

It is obvious from what has gone before that it would be hopeless to attempt to solve the question of the motion of the solar system by observations of optical phenomena *at the surface of the earth*. But it is not impossible that at even moderate distances above the level of the sea, at the top of an isolated mountain peak, for instance, the relative motion might be perceptible in an apparatus like that used in these experiments. Perhaps if the experiment should ever be tried in these circumstances, the cover should be of glass, or should be removed.

It may be worth while to notice another method for multiplying the square of the aberration sufficiently to bring it within the range of observation, which has presented itself during the

preparation of this paper. This is founded on the fact that reflection from surfaces in motion varies from the ordinary laws of reflection.

Let  $ab$  (fig. 1) be a plane wave falling on the mirror  $mn$  at an incidence of  $45^\circ$ . If the mirror is at rest, the wave front after reflection will be  $ac$ .

Now suppose the mirror to move in a direction which makes an angle  $\alpha$  with its normal, with a velocity  $\omega$ . Let  $V$  be the velocity of light in the ether supposed stationary, and let  $cd$  be the increase in the distance the light has to travel to reach  $d$ .

In this time the mirror will have moved a distance  $\frac{cd}{\sqrt{2} \cos \alpha}$ .

We have  $\frac{cd}{ad} = \frac{\omega \sqrt{2} \cos \alpha}{V}$  which put  $= r$ , and  $\frac{ac}{ad} = 1 - r$ .

In order to find the new wave front, draw the arc  $fg$  with  $b$  as a center and  $ad$  as radius; the tangent to this arc from  $d$  will be the new wave front, and the normal to the tangent from  $b$  will be the new direction. This will differ from the direction  $ba$  by the angle  $\theta$  which it is required to find. From the equality of the triangles  $adb$  and  $edb$  it follows that  $\theta = 2\varphi$ ,  $ab = ac$ ,

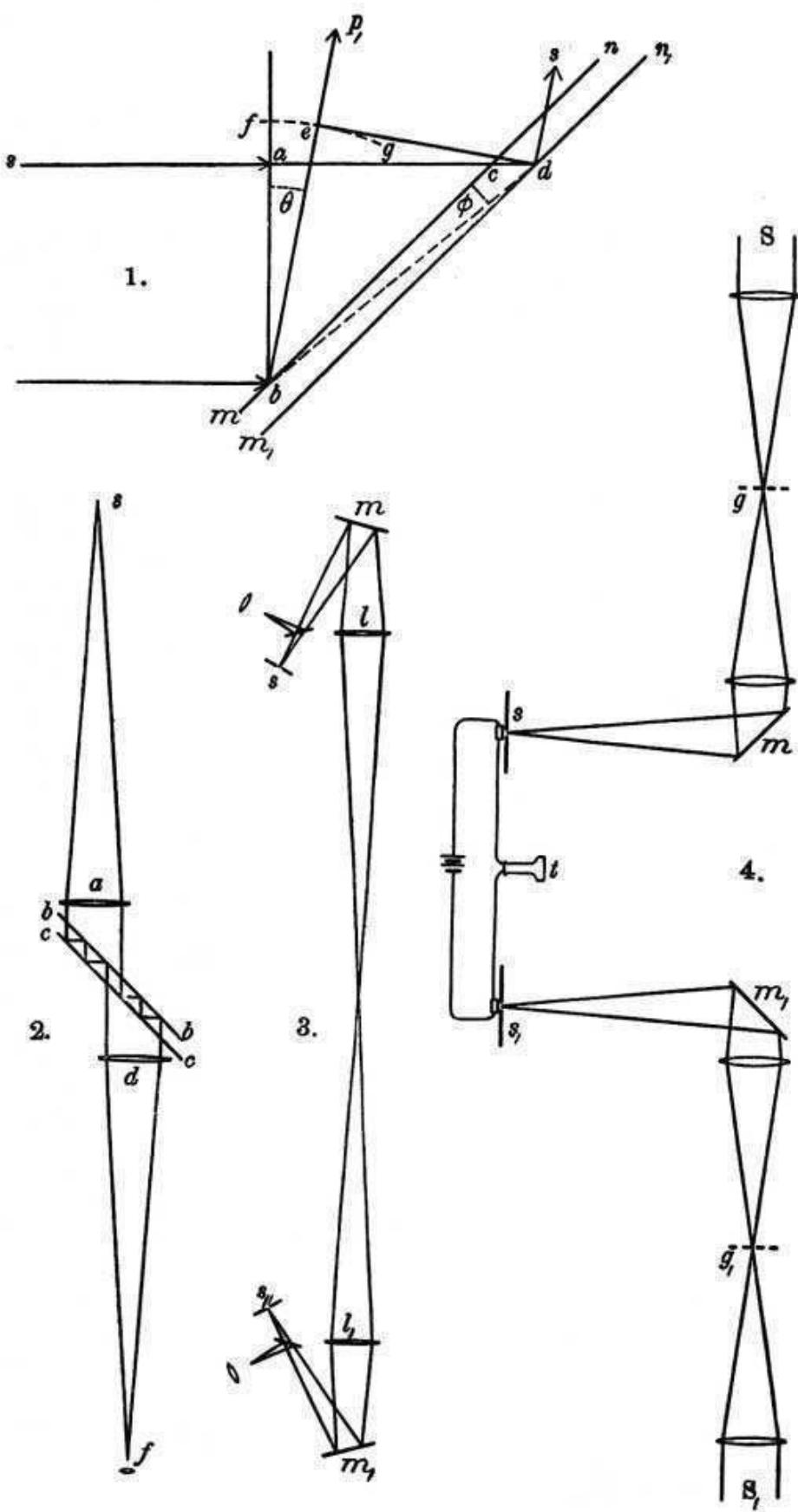
$$\tan adb = \tan\left(45^\circ - \frac{\theta}{2}\right) = \frac{1 - \tan \frac{\theta}{2}}{1 + \tan \frac{\theta}{2}} = \frac{ac}{ad} = 1 - r,$$

or neglecting terms of the order  $r^2$ ,

$$\theta = r + \frac{r^2}{2} = \frac{\sqrt{2}\omega \cos \alpha}{V} + \frac{\omega^2}{V^2} \cos^2 \alpha.$$

Now let the light fall on a parallel mirror facing the first, we should then have  $\theta_1 = \frac{-\sqrt{2}\omega \cos \alpha}{V} + \frac{\omega^2}{V^2} \cos^2 \alpha$ , and the total deviation would be  $\theta + \theta_1 = 2\rho^2 \cos^2 \alpha$  where  $\rho$  is the angle of aberration, if only the orbital motion of the earth is considered. The maximum displacement obtained by revolving the whole apparatus through  $90^\circ$  would be  $\Delta = 2\rho^2 = 0.004''$ . With fifty such couples the displacement would be  $0.2''$ . But astronomical observations in circumstances far less favorable than those in which these may be taken have been made to hundredths of a second; so that this new method bids fair to be at least as sensitive as the former.

The arrangement of apparatus might be as in fig. 2;  $s$  in the focus of the lens  $a$ , is a slit;  $bb' cc'$  are two glass mirrors optically plane and so silvered as to allow say one-twentieth of the light to pass through, and reflecting say ninety per cent. The intensity of the light falling on the observing telescope  $df$



would be about one-millionth of the original intensity, so that if sunlight or the electric arc were used it could still be readily seen. The mirrors *bb*, and *cc*, would differ from parallelism sufficiently to separate the successive images. Finally, the apparatus need not be mounted so as to revolve, as the earth's rotation would be sufficient.

If it were possible to measure with sufficient accuracy the velocity of light without returning the ray to its starting point, the problem of measuring the first power of the relative velocity of the earth with respect to the ether would be solved. This may not be as hopeless as might appear at first sight, since the difficulties are entirely mechanical and may possibly be surmounted in the course of time.

For example, suppose (fig. 3) *m* and *m*, two mirrors revolving with equal velocity in opposite directions. It is evident that light from *s* will form a stationary image at *s*, and similarly light from *s*, will form a stationary image at *s*. If now the velocity of the mirrors be increased sufficiently, their phases still being exactly the same, both images will be deflected from *s* and *s*, in inverse proportion to the velocities of light in the two directions; or, if the two deflections are made equal, and the difference of phase of the mirrors be simultaneously measured, this will evidently be proportional to the difference of velocity in the two directions. The only real difficulty lies in this measurement. The following is perhaps a possible solution: *gg*, (fig. 4) are two gratings on which sunlight is concentrated. These are placed so that after falling on the revolving mirrors *m* and *m*, the light forms images of the gratings at *s* and *s*, two very sensitive selenium cells in circuit with a battery and a telephone. If everything be symmetrical, the sound in the telephone will be a maximum. If now one of the slits *s* be displaced through half the distance between the image of the grating bars, there will be silence. Suppose now that the two deflections having been made exactly equal, the slit is adjusted for silence. Then if the experiment be repeated when the earth's rotation has turned the whole apparatus through  $180^\circ$ , and the deflections are again made equal, there will no longer be silence, and the angular distance through which *s* must be moved to restore silence will measure the required difference in phase.

There remain three other methods, all astronomical, for attacking the problem of the motion of the solar system through space.

1. The telescopic observation of the proper motions of the stars. This has given us a highly probable determination of the direction of this motion, but only a guess as to its amount.

2. The spectroscopic observation of the motion of stars in the line of sight. This could furnish data for the relative

motions only, though it seems likely that by the immense improvements in the photography of stellar spectra, the information thus obtained will be far more accurate than any other.

3. Finally there remains the determination of the velocity of light by observations of the eclipses of Jupiter's satellites. If the improved photometric methods practiced at the Harvard observatory make it possible to observe these with sufficient accuracy, the difference in the results found for the velocity of light when Jupiter is nearest to and farthest from the line of motion will give, not merely the motion of the solar system with reference to the stars, but with reference to the luminiferous ether itself.

---

---



### On the Invisible Radiations Emitted by the Salts of Uranium

H. Becquerel,  
(Received 1896)

#### *1. Action on Electrified Bodies*

At one of the sessions of the Academy, I announced that the invisible radiations emitted by the salts of uranium possessed the property of discharging electrified bodies. I have continued the study of this phenomenon by the use of Hurmuzescu's electroscope and I have been able to establish, in another way than [the work] I have done by photography, that the radiations in question penetrate various opaque substances, in particular, aluminium and copper. Platinum exhibits an absorption considerably greater than that of the two preceding metals.

If we follow the progressive approach of the gold leaves of the electroscope during the discharge, we discover that for deflections which do not exceed  $30^{\circ}$  the angular variations are very clearly proportional to the time, so that the speed of approach, or the fraction of a degree through which the gold leaves approach each other in one second, can give an idea of the relative intensity of the active radiations. I shall report here only the numbers relating to the absorption through a plate of quartz, perpendicular to the axis., 5 mm thick. The speeds are expressed in seconds of arc and in seconds of time.

A lamella of the double sulfate of uranium and potassium placed below the gold leaves dissipated the charge of the electroscope with a speed represented by 22.50. The interpositions of the plate of quartz reduced the speed to 5.43. The ratio of the two numbers is 4.15.

I have investigated whether the radiations emanating from the phosphorescent wall of a Crookes tube were weakened by the same quartz plate in a

ratio of the same order of magnitude. A Crookes tube was arranged alongside the electroscope, opposite one of the faces of the lantern in which the glass had been replaced by a sheet of aluminium 0.12 mm thick. In front of this sheet was placed a screen of copper pierced with a circular hole 15 mm in diameter. The radiations which passed through the copper were sufficiently weakened to give a negligible effect in the present experiment. When the Crookes tube was excited by an induction coil, the gold leaves of the electroscope rapidly approached each other, about  $1^\circ$  in 1.4 sec., corresponding to a speed of 2571.4 if expressed in the units adopted above.

When the quartz plate closed the circular opening, the speed of the fall of the gold leaves became 163.63, or 15.7 times smaller.

The weakened is nearly four times greater in the second case than in the first, but it is of the same order of magnitude. That is the only point which this experiment demonstrates. The observations is not contrary to the probable hypothesis which would attribute the difference to this: that the rays emitted by the uranium salt and the rays emitted by the tube or by the phosphorescent glass do not have the same wave lengths; but the different conditions of the two experiments permit no assurance of this heterogeneity.

The electroscope has also made it possible to demonstrate the small difference between the emission from a lamella of uranium salt kept eleven days in the dark and the emission from the same lamella brilliantly illuminated by magnesium. In the first case, the speed of the fall of the leaves was 20.69, and after the excitation by light it became 23.08.

It is not known what becomes of the electric charges dissipated in this way, as through dielectrics became conductors while being traversed by these radiations. Experience has been shown that a crystalline lamella, suitably insulated, did not become charged as it discharged the electrometer. In addition, a lamella standing for a long time in the presence of the apparatus communicated no charge to it.

## *2. Emission by Various Salts of Uranium. Persistence. Excitation.*

If the phenomenon of the emission of invisible radiations which we are studying is a phenomenon of phosphorescence, it should be possible to demonstrate its excitation by definite radiations. This study is made very difficult by the prodigious persistence of the emission when the substances are kept in darkness, shielded from luminous radiations or from the invisible

radiations whose nature we know. At the end of more than fifteen days, the salts of uranium still radiations which are nearly as intense as on the first day. When on the same photographic plate, above black paper, we arrange a lamella held for a long time in darkness and another which has just been exposed to daylight, the impression of the silhouette of the second is a little stronger than that of the first. The light of magnesium under the same conditions produced only an imperceptible effect. If lamella of the double sulfate of uranium and potassium are vigorously illuminated by the electric arc or by brilliant sparks from the discharge of a Leyden jar, the impressions are considerably blacker. The phenomenon then appears really to be a phenomenon of invisible phosphorescence but does not seem to be intimately linked with visible phosphorescence or fluorescence. Indeed, although the salts of the sesquioxide of uranium are very fluorescent, we know that the green uranous salts, whose curious properties of absorption I have had occasion to study, are neither phosphorescent nor fluorescent. Now, uranous sulfate behaves like uranic sulfate, and emits invisible radiations which are equally intense.

I shall report still another interesting experiment. It is known that uranium nitrate ceases to be either phosphorescent or fluorescent when is solution or melted in its water of crystallization. I then took a crystal of this salt and, after having placed it in a little tube closed by a thin plate of glass, I warmed it in darkness, managing to avoid even the radiations from the alcohol lamp that heated it. The salt melted, I let it crystallize in darkness and I then placed it on a photographic plate wrapped in black paper, protecting the salt always from the action of light. One might have expected to observe no effect, since all luminous excitation had been avoided from the moment when the substance ceased to be phosphorescent. Nevertheless the impression was as strong as for salts exposed to the light, and, at the points where the salt adhered to the glass plate, the impression was stronger than that of a fragment of uranic sulfate placed for comparison on the same plate.

On this same photographic plate there were other crystals of uranium nitrate resting with different faces on glass slips, for which the effects were substantially the same.

I also arranged some continuous surfaces, of uranic sulfate and also of the double uranium-potassium sulfate, and on these surfaces I projected the spectrum of the electric arc through an apparatus of quartz. The ultraviolet excitation bands were very sharply traced by fluorescence, but, when I reproduced the silhouette of these surface on a photographic plate, the silhouette had become almost uniformly black, indicating either that the characteristic emission of the substance masked the weak differences which

might have been observed for the different regions excitation, or that the excitation did not take place in the region of the spectrum projected on the surface studied.

### *3. Absorption by Various Substances*

It is very easy to make a qualitative study of the absorption by various substance of the radiations with which we are concerned by arranging on the same photographic plate, sheets of these substances or little flat tubes full of liquids and by covering these with a lamella of the double uranium-potassium sulfate or by any other salt of uranium.

Using various substance at thicknesses differing little from 2 mm, I discovered that water was very transparent; most of the solutions, even solutions of metallic salts, copper nitrate, gold chloride, uranium nitrate, an alcoholic solution of chlorophyll, behaved as sufficiently transparent; it was the same with paraffin and modeling wax; uranium glass was more opaque, as was a red-colored glass, aluminium at this thickness is hardly transparent, tin is more opaque, and a blue cobalt glass was more opaque than the preceding metals.

In another series of experiments I arranged various crystals, and various optical combinations, intended to exhibit the phenomena of double refraction and polarization. The images obtained were too weak for me to give any results today; nevertheless, it was discovered that quartz absorbs more of these invisible radiations than does Iceland spar; native sulfur behaved as if transparent.

Finally, the experiments in air and in the rarefied air I mentioned at the end of my last Note, without giving very notable differences, show plainly that the negatives in the rarefied air are a little stronger, which would demonstrate an absorption by the air.

### *4. Refraction*

The facts I mentioned in my last Note provide evidence for refraction by glass. To these experiments the following may be added: on one of the faces of a crown glass prism, a few millimeters from the edge and parallel to it, a little tube of very thin glass was fastened, about 1 mm in diameter and filled with crystalline uranium nitrate, forming a line source for the emission

of invisible radiation.

The other face of the prism was set on a photographic plate. When the plate was developed three days later, a diffuse impression was seen below the base of the prism, an impression separated from the trace of the edge by a white line, whose displacement is of the order of magnitude of that obtained under the same conditions with light. The tremendous decrease in luminous intensity when the sources were removed a little from the photographic plate has so far not permitted any measurements of the index of refraction.

### *5. Anomalies Presented by Various Substance*

Uranium salts emit invisible radiations with a remarkable constancy, but it is not the same with other phosphorescent substances.

I had obtained with calcium sulfide results of the same order as those which the salts of uranium give, and in my last note I mentioned a negative of remarkable intensity made through 2 mm of aluminium. The same phosphorescent material placed on a second photographic plate under the same conditions displayed no activity, and since then I have not succeeded in obtaining any images with calcium sulfide. I have had the same lack of success with pieces of hexagonal blonde of various origins. I then attempted to transmit a new activity to these substances by various known procedures. I heated them in the presence of the photographic plate without heating the latter, and I obtained no impression. In another series of experiments, the various substances were chilled to  $-20^{\circ}$ , excited by daylight and the light of magnesium, then placed on the photographic plate; only the salts of uranium gave any images.

Finally, I excited the sulfides and the hexagonal blonde by sparks from the discharge of a battery, and the substances (which become vividly phosphorescent) still displayed no activity through black paper. In the course of these experiments I learned that our eminent colleague, M. Troost, had observed a similar effect. Very old specimens of hexagonal blonde, which had given him energetic results at first, had later given progressively decreasing results, then had become inactive. Here is a curious fact for which further experiments will perhaps give us the explanation.



**Cathode Rays.**

J.J. Thomson,  
(Received October, 1897)

The experiments<sup>1</sup> discussed in this paper were undertaken in the hope of gaining, some information; is 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 authorial, 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 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.

---

<sup>1</sup>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).

### *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 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

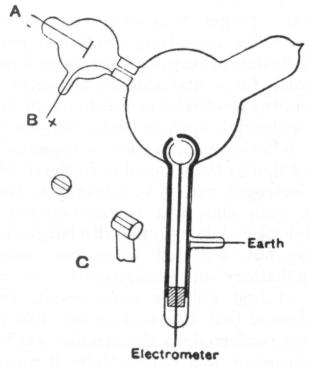


Figure 1:

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.

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,

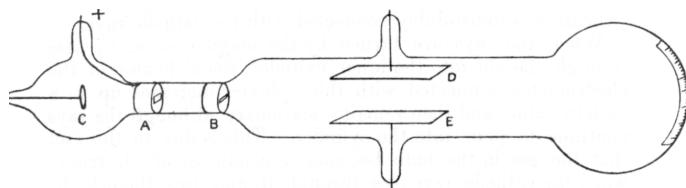


Figure 2:

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.

### *Conductivity of a Gas through which Cathode Rays are passing.*

The conductivity of the gas was investigated by means of the apparatus shown in fig. 2. The upper plate D was connected with one terminal of a battery of small storage-cells, the other terminal of which was connected with the earth; the other plate E was connected with one of the coatings of a condenser of one microfarad capacity, the other coating of which was to earth; one pair of quadrants of an electrometer was also connected with E, the other pair of quadrants being to earth. When the cathode rays are passing between the plates the two pairs of quadrants of the electrometer are first connected with each other, and then the connexion between them was broken. If the space between the plates were a non-conductor, the potential of the pair of quadrants not connected with the earth would remain zero and the needle of the electrometer would not move; if, however, the space between the plates were a conductor, then the potential of the lower plate would approach that of the upper, and the needle of the electrometer would be deflected. There is always a deflexion of the electrometer, showing that a current passes between the plates. The magnitude of the current depends very greatly upon the pressure of the gas; so much so, indeed, that it is difficult, to obtain consistent readings in consequence of the changes which always occur in the pressure when the discharge passes through the tube.

We shall first take the case when the pressure is only just low enough to allow the phosphorescent patch to appear at the end of the tube; in this case the relation between the current between the plates and the initial difference of potential is represented by the curve shown in fig. 3. In this figure the abscissæ represent the initial difference of potential between the plates, each division representing two volts, and the ordinates the rise in potential of the lower plate in one minute each division again representing two volts. The quantity of electricity which has passed between the plates in one minute is the quantity required to raise 1 microfarad to the potential-difference shown by the curve. The upper and lower curve relates to the case when the upper plate is connected with the negative and positive pole, respectively of the battery.

Even when there is no initial difference of potential between the plates the lower plate acquires a negative charge from the impact on it of some of the cathode rays.

We see from the curve that the current between the plates soon reaches

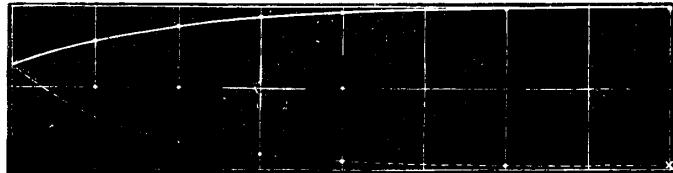


Figure 3:

a value where it is only slightly affected by an increase in the potential-difference between the plates; this is a feature common to conduction through gases traversed by Röntgen rays, by uranium rays, by ultra-violet light, and, as we now see, by cathode rays. The rate of leak is not greatly different whether the upper plate be initially positively or negatively electrified.

The current between the plates only lasts for a short time; it ceases long before the potential of the lower plate approaches that of the upper. Thus, for example, when the potential of the upper plate was about 400 volts above that of the earth, the potential of the lower plate never rose above 6 volts: similarly, if the upper plate were connected with the negative pole of the battery, the fall in potential of the lower plate was very small in comparison with the potential-difference between the upper plate and the earth.

These results are what we should expect if the gas between the plates and the plug B (fig. 2) were a very much better conductor than the gas between the plates, for the lower plate will be in a steady state when the current coming to it from the upper plate is equal to the current going from it to the plug: now if the conductivity of the gas between the plate and the plug is much greater than that between the plates, a small difference of potential between the lower plate and the plug will be consistent with a large potential-difference between the plates.

So far we have been considering the case when the pressure is as high as is consistent with the cathode rays reaching the end of the tube; we shall now go to the other extreme and consider the case when the pressure is as low as is consistent with the passage of a discharge through the bulb. In this case, when the plates are not connected with the battery we got a negative charge communicated to the lower plate, but only very slowly in comparison with the effect in the previous case. When the upper plate is connected with the negative pole of a battery, this current to the lower plate is only slightly increased even when the difference of potential is as much as 400 volts: a small potential-difference of about 20 volts seems slightly to decrease the

rate of leak. Potential-differences much exceeding 400 volts cannot be used, as though the dielectric between the plates is able to sustain them for some little time, yet after a time an intensely bright arc flashes across between the plates and liberates so much gas as to spoil the vacuum. The lines in the spectrum of this glare are chiefly mercury lines; its passage leaves very peculiar markings on the aluminium plates.

If the upper plate was charged positively, then the negative charge communicated to the lower plate was diminished, and stopped when the potential-difference between the plates was about 20 volts; but at the lowest pressure, however great (up to 400 volts) the potential-difference, there was no leak of positive electricity to the lower plate at all comparable with the leak of negative electricity to this plate when the two plates were disconnected from the battery. In fact at this very low pressure all the facts are consistent with the view that the effects are due to the negatively electrified particles travelling along the cathode rays, the rest of the gas possessing little conductivity. Some experiments were made with a tube similar to that shown in fig. 2, with the exception that the second plug B was absent, so that a much greater number of cathode rays passed between the plates. When the upper plate was connected with the positive pole of the battery a luminous discharge with well-marked striations passed between the upper plate and the earth-connected plug through which the cathode rays were streaming; this occurred even though the potential-difference between the plate and the plug did not exceed 20 volts. Thus it seems that if we supply cathode rays from an external source to the cathode a small potential-difference is sufficient to produce the characteristic discharge through a gas.

### *Magnetic Deflexion of the Cathode Rays in Different Gases.*

The deflexion of the cathode rays by the magnetic field was studied with the aid of the apparatus shown in fig. 4. The cathode was placed in a side-tube fastened on to a bell-jar; the opening between this tube and the bell-jar was closed by a metallic plug with a slit in it; this plug was connected with the earth and was used as the anode. The cathode rays passed through the slit in this plug into the bell-jar, passing in front of a vertical plate of glass ruled into small squares. The bell-jar was placed between two large parallel coils arranged as a Helmholtz galvanometer. The course of the rays was determined by taking photographs of the bell-jar when the cathode rays

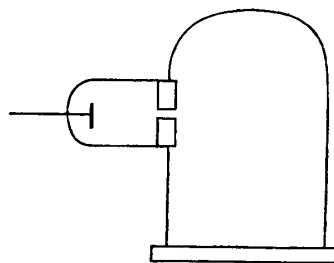


Figure 4:

were passing through it; the divisions on the plate enabled the path of the rays to be determined. Under the action of the magnetic field the narrow beam of cathode rays spreads out into a broad fan-shaped luminosity in the gas. The luminosity in this fan is not uniformly distributed, but is condensed along certain lines. The phosphorescence on the glass is also not uniformly distributed; it is much spread out, showing that the beam consists of rays which are not all deflected to the same extent by the magnet. The luminosity on the glass is crossed by bands along which the luminosity is very much greater than in the adjacent parts. These bright and dark bands are called by Birkeland, who first observed them, the magnetic spectrum. The brightest spots on the glass are by no means always the terminations of the brightest streaks of luminosity in the gas; in fact, in some cases a very bright spot on the glass is not connected with the cathode by any appreciable luminosity, though there may be plenty of luminosity in other parts of the gas. One very interesting point brought out by the photographs is that in a given magnetic field, and with a given mean potential-difference between the terminals, the path of the rays is independent of the nature of the gas. Photographs were taken of the discharge in hydrogen, air, carbonic acid, methyl iodide, i.e., in gases whose densities range from 1 to 70, and yet, not only were the paths of the most deflected rays the same in all cases, but even the details, such as the distribution of the bright and dark spaces, were the same; in fact, the photographs could hardly be distinguished from each other. It is to be noted that the pressures were not the same; the pressures in the different gases were adjusted so that the mean potential-differences between the cathode and the anode were the same in all the gases. When the pressure of a gas is lowered, the potential-difference between the terminal increases, and deflexion of the rays produced by a magnet diminishes, or at any rate the deflexion of the rays when the

phosphorescence is a maximum diminishes. If an air-break is inserted an effect of the same kind is produced.

In the experiments with different gases, the pressures were as high as was consistent with the appearance of the phosphorescence on the glass, so as to ensure having as much as possible of the gas under consideration in the tube.

As the cathode rays carry a charge of negative electricity, are deflected by an electrostatic force as if they were negatively electrified, and are acted on by a magnetic force in just the way in which this force would act on a negatively electrified body moving along the path of these rays, I can see no escape from the conclusion that they are charges of negative electricity carried by particles of matter. The question next arise, What are these particles? are they atoms, or molecules, or matter in a still finer state of subdivision? To throw some light on this point, I have made a series of measurements of the ratio of the mass of these particles to the charge carried by it. To determine this quantity, I have used two independent methods. The first of these is as follows : — Suppose we consider a bundle of homogeneous cathode rays. Let  $m$  be the mass of each of the particles,  $e$  the charge carried by it. Let  $N$  be the number of particles passing across any section of the beam in a given time; then  $Q$  the quantity of electricity carried by these particles is given by the equation

$$Ne = Q.$$

We can measure  $Q$  if we receive the cathode rays in the inside of a vessel connected with an electrometer. When these rays strike against a solid body, the temperature of the body is raised; the kinetic energy of the moving particles being converted into heat; if we suppose that all this energy is converted into heat, then if we measure the increase in the temperature of a body of known thermal capacity caused by the impact of these rays, we can determine  $W$ , the kinetic energy of the particles, and if  $v$  is the velocity of the particles,

$$\frac{1}{2} Nmv^2 = W.$$

If  $\rho$  the radius of curvature of the path of these rays in a uniform magnetic field  $H$ , then

$$\frac{mv}{e} = H\rho = I,$$

where  $I$  is written for  $H\rho$  for the sake of brevity. From these equations we get.

$$\frac{1}{2} \frac{m}{e} v^2 = \frac{W}{Q}.$$

$$v = \frac{2W}{QI},$$

$$\frac{m}{e} = \frac{I^2 Q}{2W}.$$

Thus, if we know the values of  $Q$ ,  $W$ , and  $I$ , we can deduce the values of  $v$  and  $m/e$ .

To measure these quantities, I have used tubes of three different types. The first I tried is like that represented in fig. 2, except that the plates E and D are absent, and two coaxial cylinders are fastened to the end of the tube. The rays from the cathode C fall on the metal plug B, which is connected with the earth, and serves for the anode; a horizontal slit is cut in this plug. The cathode rays pass through this slit, and then strike against the two coaxial cylinders at the end of the tube; slits are cut in these cylinders, so that the cathode rays pass into the inside of the inner cylinder. The outer cylinder is connected with the earth, the inner cylinder, which is insulated from the outer one, is connected with an electrometer, the deflexion of which measures  $Q$ , the quantity of electricity brought into the inner cylinder by the rays. A thermo-electric couple is placed behind the slit in the inner cylinder; this couple is made of very thin strips of iron and copper fastened to very fine iron and copper wires. These wires passed through the cylinders, being insulated from them, and through the glass to the outside of the tube, where they were connected with a low-resistance galvanometer, the deflexion of which gave data for calculating the rise of temperature of the junction produced by the impact against it of the cathode rays. The strips of iron and copper were large enough to ensure that every cathode ray which entered the inner cylinder struck against the junction. In some of the tubes the strips of iron and copper were placed end to end, so that some of the rays struck against the iron, and others against the copper; in others, the strip of one metal was placed in front of the other; no difference, however, could be detected between the results got with these two arrangements. The strips of iron and copper were weighed, and the thermal capacity of the junction calculated. In one set of junctions this capacity was  $5 \times 10^{-3}$ , in another  $3 \times 10^{-3}$ . If we assume that the cathode rays which strike against the junction give their energy up to it, the deflexion of the galvanometer gives us  $W$  or  $\frac{1}{2} Nmv^2$ .

The value of  $I$ , i. e.,  $H\rho$ , where  $\rho$  is the curvature of this path of the rays in a magnetic field of strength  $H$  was found as follows:— The tube was fixed between two large circular coils placed parallel to each other, and separated by a distance equal to the radius of either; these coils produce a

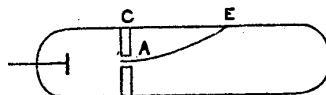


Figure 5:

uniform magnetic field, the strength of which is got by measuring with an ammeter the strength of the current passing through them. The cathode rays are thus in a uniform field, so that their path is circular. Suppose that the rays, when deflected by a magnet, strike against the glass of the tube at E

(fig. 5), then, if  $\rho$  is the radius of the circular path of the rays,

$$2\rho = \frac{CE^2}{AC} + AC;$$

thus, if we measure  $CE$  and  $AC$  we have the means of determining the radius of curvature of the path of the rays.

The determination of  $\rho$  is rendered to some extent uncertain, in consequence of the pencil of rays spreading out under the action of the magnetic field, so that the phosphorescent patch at  $E$  is several millimetres long; thus values of  $\rho$  differing appreciably from each other will be got by taking  $E$  at different points of this phosphorescent patch. Part of this patch was, however, generally considerably brighter than the rest; when this was the case,  $E$  was taken as the brightest point; when such a point of maximum brightness did not exist, the middle of the patch was taken for  $E$ . The uncertainty in the value of  $\rho$  thus introduced amounted sometimes to about 20 per cent.; by this I mean that if we took  $E$  first at one extremity of the patch and then at the other, we should get values of  $\rho$  differing by this amount.

The measurement of  $Q$ , the quantity of electricity which enters the inner cylinder, is complicated by the cathode rays making the gas through which they pass a conductor, so that though the insulation of the inner cylinder was perfect when the rays were off, it was not so when they were passing through the space between the cylinders; this caused some of the charge communicated to the inner cylinder to leak away so that the actual charge given to the cylinder by the cathode rays was larger than that indicated by the electrometer. To make the error from this cause as small as possible, the inner cylinder was connected to the largest capacity available, 1.5 microfarad, and the rays were only kept on for a short time, about 1 or 2 seconds, so that the alteration in potential of the inner cylinder was not large, ranging

in the various experiments from about 0.5 to 5 volts. Another reason why it is necessary to limit the duration of the rays to as short a time as possible, is to avoid the correction for the loss of heat from the thermo-electric junction by conduction along the wires; the rise in temperature of the junction was of the order  $2^{\circ}$  C; a series of experiments showed that with the same tube and the same gaseous pressure  $Q$  and  $W$  were proportional to each other when the rays were not kept on too long.

Tubes of this kind gave satisfactory results, the chief drawback being that sometimes in consequence of the charging up of the glass of the tube, a secondary discharge started from the cylinder to the walls of the tube, and the cylinders were surrounded by glow; when this glow appeared, the readings were very irregular; the glow could, however, be got rid of by pumping and letting the tube rest for some time. The results got with this tube are given in the Table under the heading Tube 1.

The second type of tube was like that used for photographing the path of the rays (fig. 4); double cylinders with a thermoelectric junction like those used in the previous tube were placed in the line of fire of the rays, the inside of the bell-jar was lined with copper gauze connected with the earth. This tube gave very satisfactory results; we were never troubled with any glow round the cylinders, and the readings were most concordant; the only drawback was that as some of the connexions had to be made with sealing-wax, it was not possible to get the highest exhaustions with this tube, so that the range of pressure for this tube is less than that for tube 1. The results got with this tube are given in the Table under the heading Tube 2.

The third type of tube was similar to the first, except that the openings in the two cylinders were made very much smaller; in this tube the slits in the cylinders were replaced by small holes, about 1.5 millim. diameter. In consequence of the smallness of the openings, the magnitude of the effects was very much reduced; in order to get measurable results it was necessary to reduce the capacity of the condenser in connexion with the inner cylinder to 0.15 microfarad, and to make the galvanometer exceedingly sensitive, as the rise in temperature of the thermo-electric junction was in these experiments only about  $0.5^{\circ}$  C on the average. The results obtained in this tube are given in the Table under the heading Tube 3.

The results of a series of measurements with these tubes are given in the following Table:—

Gas	Value of $W/Q$	$I$	$m/e$	$v$
Tube 1.				
Air	$4.6 \times 10^{11}$	230	$0.57 \times 10^{-7}$	$4 \times 10^9$
Air	$1.8 \times 10^{12}$	350	$0.34 \times 10^{-7}$	$1 \times 10^{10}$
Air	$6.1 \times 10^{11}$	230	$0.43 \times 10^{-7}$	$5.4 \times 10^9$
Air	$2.5 \times 10^{12}$	400	$0.32 \times 10^{-7}$	$1.2 \times 10^{10}$
Air	$5.5 \times 10^{11}$	230	$0.48 \times 10^{-7}$	$4.8 \times 10^9$
Air	$1 \times 10^{12}$	285	$0.4 \times 10^{-7}$	$7 \times 10^9$
Air	$1 \times 10^{12}$	285	$0.4 \times 10^{-7}$	$7 \times 10^9$
Hydrogen	$6 \times 10^{12}$	205	$0.35 \times 10^{-7}$	$6 \times 10^9$
Hydrogen	$2.1 \times 10^{12}$	460	$0.5 \times 10^{-7}$	$9.2 \times 10^9$
Carbonic acid	$8.4 \times 10^{11}$	260	$0.4 \times 10^{-7}$	$7.5 \times 10^9$
Carbonic acid	$1.47 \times 10^{12}$	340	$0.4 \times 10^{-7}$	$8.5 \times 10^9$
Carbonic acid	$3.0 \times 10^{12}$	480	$0.39 \times 10^{-7}$	$1.3 \times 10^{10}$
Tube 2.				
Air	$2.8 \times 10^{11}$	175	$0.53 \times 10^{-7}$	$3.3 \times 10^9$
Air	$2.8 \times 10^{11}$	175	$0.53 \times 10^{-7}$	$4.1 \times 10^9$
Air	$3.5 \times 10^{11}$	181	$0.47 \times 10^{-7}$	$3.8 \times 10^9$
Hydrogen	$2.8 \times 10^{11}$	175	$0.53 \times 10^{-7}$	$3.3 \times 10^9$
Air	$2.5 \times 10^{11}$	160	$0.51 \times 10^{-7}$	$3.1 \times 10^9$
Carbonic acid	$2 \times 10^{11}$	148	$0.54 \times 10^{-7}$	$2.5 \times 10^9$
Air	$1.8 \times 10^{11}$	151	$0.63 \times 10^{-7}$	$2.3 \times 10^9$
Hydrogen	$2.8 \times 10^{11}$	175	$0.53 \times 10^{-7}$	$3.3 \times 10^9$
Hydrogen	$4.4 \times 10^{11}$	201	$0.46 \times 10^{-7}$	$4.4 \times 10^9$
Air	$2.5 \times 10^{11}$	176	$0.61 \times 10^{-7}$	$2.8 \times 10^9$
Air	$4.2 \times 10^{11}$	200	$0.48 \times 10^{-7}$	$4.1 \times 10^9$
Tube 3.				
Air	$2.5 \times 10^{11}$	220	$0.9 \times 10^{-7}$	$2.4 \times 10^9$
Air	$3.5 \times 10^{11}$	225	$0.7 \times 10^{-7}$	$3.2 \times 10^9$
Hydrogen	$3 \times 10^{11}$	250	$1.0 \times 10^{-7}$	$2.5 \times 10^9$

It will be noticed that the value of  $m/e$  is considerably greater for Tube 3, where the opening is a small hole, than for Tubes 1 and 2, where the opening is a slit of much greater area. I am of opinion that the values of  $w/e$  got from Tubes 1 and 2 are too small, in consequence of the leakage from the inner cylinder to the outer by the gas being rendered a conductor by the passage of the cathode rays.

It will be seen from these tables that the value of  $m/e$  is independent of the nature of the gas. Thus, for the first tube the mean for air is  $0.40 \times 10^{-7}$ , for hydrogen  $0.42 \times 10^{-7}$ , and for carbonic acid gas  $0.4 \times 10^{-7}$ ; for the second tube the mean for air is  $0.52 \times 10^{-7}$ , for hydrogen  $0.50 \times 10^{-7}$ , and for carbonic acid gas  $0.54 \times 10^{-7}$ .

Experiments were tried with electrodes made of iron instead of aluminium; this altered the appearance of the discharge and the value of  $v$  at the same pressure, the values of  $m/e$  were, however, the same in the two tubes; the effect produced by different metals on the appearance of the discharge will be described later on.

In all the preceding experiments, the cathode rays were first deflected from the cylinder by a magnet, and it was then found that there was no deflexion either of the electrometer or the galvanometer, so that the deflexions observed were entirely due to the cathode rays; when the glow mentioned previously surrounded the cylinders there was a deflexion of the electrometer even when the cathode rays were deflected from the cylinder.

Before proceeding to discuss the results of these measurements I shall describe another method of measuring the quantities  $m/e$  and  $v$  of an entirely different kind from the preceding; this method is based upon the deflexion of the cathode rays in an electrostatic field. If we measure the deflexion experienced by the rays when traversing a given length under a uniform electric intensity, and the deflexion of the rays when they traverse a given distance under a uniform magnetic field, we can find the values of  $m/e$  and  $v$  in the following way:-

Let the space passed over by the rays under a uniform electric intensity  $F$  be  $l$ , the time taken for the rays to traverse this space is  $l/v$ , the velocity in the direction of  $F$  is therefore

$$\frac{Fe}{m} = \frac{l}{v},$$

so that  $\theta$ , the angle through which the rays are deflected when they leave the electric field and enter a region free from electric force, is given by the equation

$$\theta = \frac{Fe}{m} \frac{l}{v^2}.$$

If, instead of the electric intensity, the rays are acted on by a magnetic force  $H$  at right angles to the rays, and extending across the distance  $l$ , the velocity at right angles to the original path of the rays is

$$\frac{Hev}{m} = \frac{l}{v},$$

so that  $\phi$ , the angle through which the rays are deflected when they leave the magnetic field, is given by the equation

$$\phi = \frac{He}{m} \frac{l}{v}.$$

From these equations we get

$$v = \frac{\phi}{\theta} \frac{F}{H}$$

and

$$\frac{m}{e} = \frac{H^2 \theta l}{F \phi^2}.$$

In the actual experiments  $H$  was adjusted so that  $\phi = \theta$ ; in this case the equations become

$$v = \frac{F}{H},$$

$$\frac{m}{e} = \frac{H^2 l}{F \theta}.$$

The apparatus used to measure  $v$  and  $m/e$  by this means is that represented in fig. 2. The electric field was produced by connecting the two aluminium plates to the terminals of a battery of storage-cells. The phosphorescent patch at the end of the tube was deflected, and the deflexion measured by a scale pasted to the end of the tube. As it was necessary to darken the room to see the phosphorescent patch, a needle coated with luminous paint was placed so that by a screw it could be moved up and down the scale; this needle could be seen when the room was darkened, and it was moved until it coincided with the phosphorescent patch. Thus, when light was admitted, the deflexion of the phosphorescent patch could be measured.

The magnetic field was produced by placing outside the tube two coils whose diameter was equal to the length of the plates; the coils were placed so that they covered the space occupied by the plates, the distance between the coils was equal to the radius of either. The mean value of the magnetic force over the length  $l$  was determined in the following way: a narrow coil

$C$  whose length was  $l$ , connected with a ballistic galvanometer, was placed between the coils; the plane of the windings of  $C$  was parallel to the planes of the coils; the cross section of the coil was a rectangle 5 cm. by 1 cm. A given current, was sent through the outer coils and the kick  $a$  of the galvanometer observed when this current was reversed. The coil  $C$  was then placed at the centre of two very large coils, so as to be in a field of uniform magnetic force: the current through the large coils was reversed and the kick  $\beta$  of the galvanometer again observed; by comparing  $a$  and  $\beta$  we can get the mean value of the magnetic force over a length  $l$  this was found to be

$$60 \times l,$$

where  $l$  is the current flowing through the coils.

A series of experiments was made to see if the electrostatic deflexion was proportional to the electric intensity between the plates; this was found to be the case. In the following experiments the current through the coils was adjusted so that the electrostatic deflexion was the same as the magnetic:—

Gas	$\theta$	$H$	$F$	$l$	$m/e$	$v$
Air	8/110	5.5	$1.5 \times 10^{10}$	5	$1.3 \times 10^{-7}$	$2.8 \times 10^9$
Air	9.5/110	5.4	$1.5 \times 10^{10}$	5	$1.1 \times 10^{-7}$	$2.8 \times 10^9$
Air	13/110	6.6	$1.5 \times 10^{10}$	5	$1.2 \times 10^{-7}$	$2.3 \times 10^9$
Hydrogen	9/110	6.3	$1.5 \times 10^{10}$	5	$1.5 \times 10^{-7}$	$2.5 \times 10^9$
Carbonic acid	11/110	6.9	$1.5 \times 10^{10}$	5	$1.5 \times 10^{-7}$	$2.2 \times 10^9$
Air	6/110	5	$1.8 \times 10^{10}$	5	$1.3 \times 10^{-7}$	$3.6 \times 10^9$
Air	7/110	3.6	$1 \times 10^{10}$	5	$1.1 \times 10^{-7}$	$2.8 \times 10^9$

The cathode in the first five experiments was aluminium, in the last two experiments it was made of platinum; in the last experiment Sir William Crookes's method of getting rid of the mercury vapour by inserting tubes of powdered sulphur, sulphur iodide, and copper filings between the bulb and the pump was adopted. In the calculation of  $m/e$  and  $v$  no allowance has been made for the magnetic force due to the coil in the region outside the plates; in this region the magnetic force will be in the opposite direction to that between the plates, and will tend to bend the cathode rays in the

opposite direction: thus the effective value of  $H$  will be smaller than the value used in the equations, so that the values of  $m/e$  are larger, and those of  $v$  less than they would be if this correction were applied. This method of determining the values of  $m/e$  and  $v$  is much less laborious and probably more accurate than the former method; it cannot, however, be used over so wide a range of pressures.

From these determinations we see that the value of  $m/e$  is independent of the nature of the gas, and that its value  $10^{-7}$  is very small compared with the value  $10^{-4}$ , which is the smallest value of this quantity previously known, and which is the value for the hydrogen ion in electrolysis.

Thus for the carriers of the electricity in the cathode rays  $m/e$  is very small compared with its value in electrolysis. The smallness of  $m/e$  may be due to the smallness of  $m$  or the largeness of  $e$ , or to a combination of these two. That the carriers of the charges in the cathode rays are small compared with ordinary molecules is shown, I think, by Lenard's results as to the rate at which the brightness of the phosphorescence produced by these rays diminishes with the length of path travelled by the ray. If we regard this phosphorescence as due to the impact of the charged particles, the distance through which the rays must travel before the phosphorescence fades to a given fraction (say  $l/e$ , where  $e = 2.71$ ) of its original intensity, will be some moderate multiple of the mean free path. Now Lenard found that this distance depends solely upon the density of the medium, and not upon its chemical nature or physical state. In air at atmospheric pressure the distance was about half a centimetre, and this most be comparable with the mean free path of the carriers through air at atmospheric pressure. But the mean free path of the molecules of air is a quantity of quite a different order. The carrier, then, must be small compared with ordinary molecules.

The two fundamental points about these carriers seem to me to be (1) that these carriers are the same whatever the gas through which the discharge passes, (2) that the mean free paths depend upon nothing but the density of the medium traversed by these rays.

It might be supposed that the independence of the mass of the carriers of the gas through which the discharge passes was due to the mass concerned being the quasi mass which a charged body possesses in virtue of the electric field set up in its neighbourhood; moving the body involves the production of a varying electric field, and, therefore, of a certain amount of energy which is proportional to the square of the velocity. This causes the charged body to behave as if its mass were increased by a quantity, which for a charged sphere is  $1/5 e^2/\mu a$  ('Recent Researches in Electricity and Magnetism'), where  $e$  is the charge and  $a$  the radius of the sphere. If we assume that it is this mass

which we are concerned with in the cathode rays, since  $m/e$  would vary as  $e/a$ , it affords no clue to the experiment of either of the properties (1 and 2) of these rays. This is not by any means the only objection to this hypothesis, which I only mention to show that it has not been overlooked.

The experiment which seems to me to account in the most simple and straightforward manner for the facts is founded on a view of the constitution of the chemical elements which has been favourably entertained by many chemists: this view is that the atoms of the different chemical elements are different aggregations of atoms of the same kind. In the form in which this hypothesis was enunciated by Prout, the atoms of the different elements were hydrogen atoms; in this precise form the hypothesis is not tenable, but if we substitute for hydrogen some unknown primordial substance  $X$ , there is nothing known which is inconsistent with this hypothesis, which is one that has been recently supported by Sir Norman Lockyer for reasons derived from the study of the stellar spectra.

If, in the very intense electric field in the neighbourhood of the cathode, the molecules of the gas are dissociated and are split up, not into the ordinary chemical atoms, but into these primordial atoms, which we shall for brevity call corpuscles; and if these corpuscles are charged with electricity and projected from the cathode by the electric field, they would behave exactly like the cathode rays. They would evidently give a. value of  $m/e$  which is independent of the nature of the gas and its pressure, for the carriers are the same whatever the gas may be; again, the mean free paths of these corpuscles would depend solely upon the density of the medium through which they pass. For the molecules of the medium are composed of a number of such corpuscles separated by considerable spaces; now the collision between a single corpuscle and the molecule will not be between the corpuscles and the molecule as a whole, but between this corpuscle and the individual corpuscles which form the molecule; thus the number of collisions the particle makes as it moves through a crowd of these molecules will be proportional, not to the number of the molecules in the crowd, but to the number of the individual corpuscles. The mean free path is inversely proportional to the number of collisions in unit time, and so is inversely proportional to the number of corpuscles in unit volume; now as these corpuscles are all of the same mass, the number of corpuscles in unit volume will be proportional to the mass of unit volume, that is the mean free path will be inversely proportional to the density of the gas. We see, too, that so long as the distance between neighbouring corpuscles is large compared with the linear dimensions of a corpuscle the mean free path will be independent of the way they are arranged, provided the number in unit volume remains constant,

that is the mean free path will depend only on the density of the medium traversed by the corpuscles, and will be independent of its chemical nature and physical state: this from Lenard's very remarkable measurements of the absorption of the cathode rays by various media, must be a property possessed by the carriers of the charges in the cathode rays.

Thus on this view we have in the cathode rays matter in a new state, a state in which the subdivision of matter is carried very much further than in the ordinary gaseous state: a state in which all matter – that is, matter derived from different sources such as hydrogen, oxygen, &c. – is of one and the same kind; this matter being the substance from which all the chemical elements are built up.

With appliances of ordinary magnitude, the quantity of matter produced by means of the dissociation at the cathode is so small as to almost to preclude the possibility of any direct chemical investigation of its properties. Thus the coil I used would, I calculate, if kept going uninterruptedly night and day for a year, produce only about one three-millionth part of a gramme of this substance.

The smallness of the value of  $m/e$  is, I think, due to the largeness of  $e$  as well as the smallness of  $m$ . There seems to me to be some evidence that. the charges carried by the corpuscles in the atom arc large compared with those carried by the ions of an electrolyte. In the molecule of HCl, for example, I picture the components of the hydrogen atoms as held together by a great number of tubes of electrostatic force; the components of the chlorine atom are similarly held together, while only one stray tube binds the hydrogen atom to the chlorine atom. The reason for attributing this high charge to the constituents of the atom is derived from the values of the specific inductive capacity of gases: we may imagine that the specific inductive capacity of a gas is due to the setting in the electric field of the electric doublet formed by the two oppositely electrified atoms which form the molecule of the gas. The measurements of the specific inductive capacity show, however, that this is very approximately an additive quantity: that is, that we can assign a certain value to each element, and find the specific inductive capacity of HCl by adding the value for hydrogen to the value for chlorine; the value of  $H_2O$  by adding twice the value for hydrogen to the value for oxygen, and so on. Now the electrical moment of the doublet formed by a positive charge on one atom of the molecule and a negative charge on the other atom would not be an additive property; it, however, each atom had a definite electrical moment, and this were large compared with the electrical moment of the two atoms in the molecule, then the electrical moment of any compound, and hence its specific inductive capacity, would be an additive property. For

the electrical moment of the atom, however, to be large compared with that of the molecule, the charge on the corpuscles would have to be very large compared with those on the ion.

If we regard the chemical atom as an aggregation of a number of primordial atoms, the problem of finding the configurations of stable equilibrium for a number of equal particles acting on each other according to some law of force — whether that of Boscovich, where the force between them is a repulsion when they are separated by less than a certain critical distance, and an attraction when they are separated by a greater distance, or even the simpler case of a number of mutually repellent particles held together by a central force — is of great interest in connexion with the relation between the properties of an element and its atomic weight. Unfortunately the equations which determine the stability of such a collection of particles increase so rapidly in complexity with the number of particles that a general mathematical investigation is scarcely possible. We can, however, obtain a good deal of insight into the general laws which govern such configurations by the use of models, the simplest of which is the floating magnets of Professor Mayer. In this model the magnets arrange themselves in equilibrium under their mutual repulsions and a central attraction caused by the pole of a large magnet placed above the floating magnets.

A study of the forms taken by these magnets seems to me to be suggestive in relation to the periodic law. Mayer showed that when the number of floating magnets did not exceed 5 they arranged themselves at the corners of a regular polygon — 5 at the corners of a pentagon, 4 at the corners of a square, and so on. When the number exceeds 5, however, this law no longer holds: thus 6 magnets do not arrange themselves at the corners of a hexagon, but divide into two systems, consisting of 1 in the middle surrounded by 5 at the corners of a pentagon. For 8 we have two in the inside and 6 outside; this arrangement in two systems, an inner and an outer, lasts up to 18 magnets. After this we have three systems: an inner, a middle, and an outer; for a still larger number of magnets we have four systems, and so on.

Mayor found the arrangement of magnets was as follows: —  
where, for example, 1. 6. 10. 12 means an arrangement with one magnet in the middle, then a ring of six, then a ring of ten, and a ring of twelve outside.

Now suppose that a certain property is associated with two magnets forming a group by themselves; we should have this property with 2 magnets, again with 8 and 9, again with 19 and 20, and again with 34, 35, and so on. If we regard the system of magnets as a model of an atom, the number of magnets being proportional to the atomic weight, we should have

1.	2.	3.	4.	5.
$\{ 1.5$	$\{ 2.6$	$\{ 3.7$	$\{ 4.8$	$5.9$
$1.6$	$2.7$	$3.8$	$4.9$	
$1.7$				
$\{ 1.5.9$	$\{ 2.7.10$	$\{ 3.7.10$	$\{ 4.8.12$	$\{ 5.9.12$
$1.6.9$	$2.8.10$	$3.7.11$	$4.8.13$	$5.9.13$
$\{ 1.6.10$	$2.7.11$	$\{ 3.8.10$	$\{ 4.9.12$	
$1.6.11$		$3.8.11$	$4.9.13$	
		$3.8.12$		
		$3.8.13$		
$\{ 1.5.9.12$	$\{ 2.7.10.15$	$\{ 3.7.12.13$	$\{ 4.9.13.14$	
$1.5.9.13$	$2.7.12.14$	$3.7.12.14$	$4.9.13.15$	
$1.6.9.12$		$3.7.13.14$	$4.9.14.15$	
$1.6.10.12$		$3.7.13.15$		
$1.6.10.13$				
$1.6.11.12$				
$1.6.11.13$				
$1.6.11.14$				
$1.6.11.15$				
$1.7.12.14$				

this property occurring in elements of atomic weight 2, (8, 9), 19, 20, (34, 35). Again, any property conferred by three magnets forming a system by themselves would occur with atomic weights 3, 10, and 11; 20, 21, 22, 23, and 24; 35, 36, 37 and 39; in fact, we should have something quite analogous to the periodic law, the first series corresponding to the arrangement of the magnets in a single group, the second series to the arrangement in two groups, the third series in three groups, and so on.

The velocity of the cathode rays is variable, depending upon the potential-difference between the cathode and anode, which is a function of the pressure of the gas — the velocity increases as the exhaustion improves; the measurements given above show, however, that at all the pressures at which experiments were made the velocity exceeded  $10^9$  cm./sec. This velocity is much greater than the value  $2 \times 10^7$  which I previously obtained (Phil. Mag. Oct. 1894) by measuring directly the interval which separated the appearance of luminosity at two places on the walls of the tube situated at different distances from the cathode.

In my earlier experiments the pressure was higher than in the experiments described in this paper, so that the velocity of the cathode rays would on this account be less. The difference between the two results is, however, too great to be wholly explained in this way, and I attribute the difference to the glass requiring to be bombarded by the rays for a finite time before becoming phosphorescent, this time depending upon the intensity of the bombardment. As this time diminishes with the intensity of bombard-

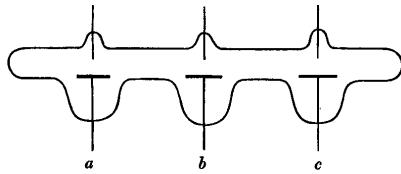


Figure 6:

ment, the appearance of phosphorescence at the piece of glass most removed from the cathode would be delayed beyond the time taken for the rays to pass from one place to the other by the difference in time taken by the glass to become luminous; the apparent velocity measured in this way would thus be less than the true velocity. In the former experiments endeavours were made to diminish this effect by making the rays strike the glass at the greater distance from the cathode less obliquely than they struck the glass nearer to the cathode; the obliquity was adjusted until the brightness of the phosphorescence was approximately equal in the two cases. In view, however, of the discrepancy between the results obtained in this way and those obtained by the later method, I think that it was not successful in eliminating the lag caused by the finite time required by the gas to light up.

### *Experiments with Electrodes of Different Materials.*

In the experiments described in this paper the electrodes were generally made of aluminium. Some experiments, however, were made with iron and platinum electrodes.

Though the value of  $m/e$  came out the same whatever the material of the electrode, the appearance of the discharge varied greatly; and as the measurements showed, the potential-difference between the cathode and anode depended greatly upon the metal used for the electrode; the pressure being the same in all cases.

To test this point further I used a tube like that shown in fig. 6, where  $a$ ,  $b$ ,  $c$  are cathodes made of different metals, the anodes being in all cases platinum wires. The cathodes were disks of aluminium, iron, lead, tin, copper, mercury, sodium amalgam, and silver chloride; the potential-difference between the cathode and anode was measured by Lord Kelvin's vertical voltmeter, and also by measuring the length of spark in air which, when placed

in parallel with the anode and cathode, seemed to allow the discharge to go as often through the spark-gap as through the tube. With this arrangement the pressures were the same for all the cathodes. The potential-difference between the anode and cathode and the equivalent spark-length depended greatly upon the nature of the cathode. The extent of the variation in potential may be estimated from the following table: —

Cathode	Mean Potential-Difference between Cathode and Anode
Aluminium	1800 volts
Lead	2100 volts
Tin	2400 volts
Copper	2600 volts
Iron	2900 volts

The potential-difference when the cathode was made of sodium amalgam or silver chloride was less even than that of aluminium.

The order of many of the metals changed about very capriciously, experiments made at intervals of a few minutes frequently giving quite different results. From the abrupt way in which these changes take place I am inclined to think that gas absorbed by the electrode has considerable influence on the passage of the discharge.

I have much pleasure in thanking Mr. Everitt for the assistance he has given me in the preceding investigation.

Cambridge, Aug. 7, 1897.

**On an Improvement of Wien's Equation for the Spectrum**

M. Planck

Berlin

(Received 1900)



English translation from "The Old Quantum Theory," ed. by D. ter Haar,  
Pergamon Press, 1967, p. 79.



The interesting result of long wave length spectral energy measurements which were communicated by Mr. Kurlbaum at today's meeting, and which were obtained by him and Mr. Rubens, confirm the statement by Mr. Lummer and Mr. Pringsheim, which was based on their observations that Wien's energy distribution law is not as generally valid, as many supposed up to now, but that this law at most has the character of a limiting case, the simple form of which was due only to a restriction to short wave lengths and low temperatures<sup>1</sup>. Since I myself even in this Society have expressed the opinion that Wien's law must be necessarily true, I may perhaps be permitted to explain briefly the relationship between the electromagnetic theory developed by me and the experimental data.

The energy distribution law is according to this theory determined as soon as the entropy  $S$  of a linear resonator which interacts with the radiation is known as function of the vibrational energy  $U$ . I have, however, already in my last paper on this subject[1] stated that the law of increase of by itself not yet sufficient to determine this function completely; my view that Wien's law would be of general validity, was brought about rather by special considerations, namely by the evaluation of an infinitesimal increase of the entropy of a system of  $n$  identical resonators in a stationary radiation field by two different methods which led to the equation

$$dU_n \cdot \Delta U_n \cdot f(U_n) = ndU \cdot \Delta U \cdot f(U),$$

---

<sup>1</sup>Mr. Paschen has written to me that he has also recently found appreciable deviations from Wien's law

where

$$U_n = nU \quad \text{and} \quad f(U) = -\frac{3}{5} \frac{d^2S}{dU^2}.$$

From this equation Wien's law follows in the form

$$\frac{d^2S}{dU^2} = \frac{\text{const.}}{U}.$$

The expression on the right-hand side of this functional equation is certainly the above-mentioned change in entropy since  $n$  identical processes occur independently, the entropy changes of which must simply add up. However, I consider the possibility, even if it would not be easily understandable and in any case would be difficult to prove, that the expression on the left-hand side would not have the general meaning which I attributed to it earlier, in other words: that the values of  $U_n$ ,  $dU_n$  and  $\Delta U_n$  are not by themselves sufficient to determine the change of entropy under consideration, but that  $U$  itself must also be known for this. Following this suggestion I have finally started to construct completely arbitrary expressions for the entropy which although they are more complicated than Wien's expression still seem to satisfy just as completely all requirements of the thermodynamic and electromagnetic theory.

I was especially attracted by one of the expressions thus constructed which is nearly as simple as Wien's expression and which deserves to be investigated since Wien's expression is not sufficient to cover all observations. We get this expression by putting<sup>2</sup>

$$\frac{d^2S}{dU^2} = \frac{\alpha}{U(\beta + U)}.$$

It is by far simplest of all expressions which lead to  $S$  as a logarithmic function of  $U$ —which is suggested from probability considerations—and which moreover reduces to Wien's expression for small values of  $U$ . Using the relation

$$\frac{dS}{dU} = \frac{1}{T}$$

and Wien's “displacement” law<sup>3</sup> one gets a radiation formula with two con-

---

<sup>2</sup>I use the second derivative of  $S$  with respect to  $U$  since this quantity has a simple physical meaning.

<sup>3</sup>The expression of Wien's displacement law is simply

$$S = f(U/\nu),$$

where  $\nu$  is the frequency of the resonator, as I shall show elsewhere.

stants:

$$E = \frac{C\lambda^{-5}}{e^{c/\lambda T} - 1},$$

which, as far as I can see at the moment, fits the observational data, published up to now, as satisfactory as the best equations put forward for the spectrum, namely those of Thiesen[2]<sup>4</sup> Lummer–Jahnke[4], and Lummer–Pringsheim[5]. (This was demonstrated by some numerical examples.) I should therefore be permitted to draw your attention to this new formula which I consider to be the simplest possible, apart from Wien’s expression, from the point of view of the electromagnetic theory of radiation.

## References

- [1] M. Planck Am. Physik. 1, 730 (1900);
- [2] M. Thiesen, Verh. D. Phys. Ges. Berlin 2, 67 (1900);
- [3] M. Planck, Ann. Physik 1, 719 (1900);
- [4] O. Lummer and E. Jahnke, Ann. Physik Lpz. 3, 288 (1900);
- [5] O. Lummer and E. Pringsheim, Verh. Dtsch. Phys. Ges. Berlin 2, 174 (1900)

---

<sup>4</sup>One can see there that Mr. Thiesen had put forward his formula before Mr. Lummer and Mr. Pringsheim had extended their measurements to longer wave lengths. I emphasise this point as I have made a statement to the contrary[3] before this paper was published.



## On the Theory of the Energy Distribution Law of the Normal Spectrum

M. Planck  
Berlin  
(Received 1900)

— — ◇ ◇ — —  
English translation from “The Old Quantum Theory,” ed. by D. ter Haar,  
Pergamon Press, 1967, p. 82.

— — ◇ ◇ — —

GENTLEMEN: when some weeks ago I had the honour to draw your attention to a new formula which seemed to me to be suited to express the law of the distribution of radiation energy over the whole range of the normal spectrum [1], I mentioned already then that in my opinion the usefulness of this equation was not based only on the apparently close agreement of the few numbers, which I could then communicate, with the available experimental data,<sup>1</sup> but mainly on the simple structure of the formula and especially on the fact that it gave a very simple logarithmic expression for the dependence of the entropy of an irradiated monochromatic vibrating resonator on its vibrational energy. This formula seemed to promise in any case the possibility of a general interpretation much rather than other equations which have been proposed, apart from Wien’s formula which, however, was not confirmed by experiment.

Entropy means disorder, and I thought that one should find this disorder in the irregularity with which even in a completely stationary radiation field the vibrations of the resonator change their amplitude and phase, as long as considers time intervals long compared to the period of one vibration, but short compared to the duration of a measurement. The constant energy of the stationary vibrating resonator can thus only be considered to be a time

---

<sup>1</sup>Verh. Dtsch. Phys. Ges. Berlin 2, 237 (1900)

average, or, put differently, to be an instantaneous average of the energies of a large number of identical resonators which are in the same stationary radiation field, but far enough from one another not to influence each other. Since the entropy of a resonator is thus determined by the way in which the energy is distributed at one time over many resonators, I suspected that one should evaluate this quantity in the electromagnetic radiation theory by introducing probability considerations, the importance of which for the second law of thermodynamics was first of all discovered by Mr. Boltzmann[3]. This suspicion has been confirmed; I have been able to derive deductively an expression for the entropy of a monochromatically vibrating resonator and thus for the energy distribution in a stationary radiation state, that is, in the normal spectrum. To do this it was only necessary to extend somewhat the interpretation of the hypothesis of "natural radiation" which is introduced in electromagnetic theory. Apart from this I have obtained other relations which seem to me to be of considerable importance for other branches of physics and also of chemistry.

I do not wish to give today this deduction – which is based on the laws of electromagnetic radiation, thermodynamics and probability calculus – systematically in all details, but rather to explain as clearly as possible the real core of the theory. This can be done most easily by describing to you a new, completely elementary treatment through which one can evaluate – without knowing anything about a spectral formula or about any theory – the distribution of a given amount of energy over the different colours of the normal spectrum using one constant of nature only and after that the value of the temperature of this energy radiation using a second constant of nature. You will find many points in the treatment to be presented arbitrary and complicated, but as I said a moment ago I do not want to pay attention to a proof of the necessity and the simple, practical details, but to the clarity and uniqueness of the given prescriptions for the solution of the problem.

Let us consider a large number of monochromatically vibrating resonator –  $N$  of frequency  $\nu$  (per second),  $N'$  of frequency  $\nu'$ ,  $N''$  of frequency  $\nu''$ , ..., with all  $N$  large number – which are at large distances apart and are enclosed in a diathermic medium with light velocity  $c$  and bounded by reflecting walls. Let the system contain a certain amount of energy, the total energy  $E_t$  (erg) which is present partly in the medium as travelling radiation and partly in the resonators as vibrational energy. The question is how in a stationary state this energy is distributed over the vibrations of the resonator and over the various of the radiation present in the medium, and what will be the temperature of the total system.

To answer this question we first of all consider the vibrations of the resonators and assign to them arbitrary definite energies, for instance, an energy  $E$  to the  $N$  resonators  $\nu$ ,  $E'$  to the  $N'$  resonators  $\nu'$ , . . . . The sum

$$E + E' + E'' + \dots = E_0$$

must, of course, be less than  $E_t$ . The remainder  $E_t - E_0$  pertains then to the radiation present in the medium. We must now give the distribution of the energy over the separate resonators of each group, first of all the distribution of the energy  $E$  over the  $N$  resonators of frequency  $\nu$ . If  $E$  considered to be continuously divisible quantity, this distribution is possible in infinitely many ways. We consider, however – this is the most essential point of the whole calculation –  $E$  to be composed of a very definite number of equal parts and use thereto the constant of nature  $h = 6.55 \times 10^{-27}$  erg · sec. This constant multiplied by the common frequency  $\nu$  of the resonators gives us the energy element  $\varepsilon$  in erg, and dividing  $E$  by  $\varepsilon$  we get the number  $P$  of energy elements which must be divided over the  $N$  resonators. If the ratio is not an integer, we take for  $P$  an integer in the neighbourhood.

It is clear that the distribution of  $P$  energy elements over  $N$  resonators can only take place in a finite, well-defined number of ways. Each of these ways of distribution we call a "complexion", using an expression introduced by Mr. Boltzmann for a similar quantity. If we denote the resonators by the numbers 1, 2, 3, . . . ,  $N$ , and write these in a row, and if we under each resonator put the number of its energy elements, we get for each complexion a symbol of the following form

1	2	3	4	5	6	7	8	9	10
7	38	11	0	9	2	20	4	4	5

We have taken here  $N = 10$ ,  $P = 100$ . The number of all possible complexions is clearly equal to the number of all possible sets of number which one can obtain for lower sequence for given  $N$  and  $P$ . To exclude all misunderstandings, we remark that two complexions must be considered to be different if the corresponding sequences contain the same numbers, but in different order. From the theory of permutations we get for the number of all possible complexions

$$\frac{N(N+1) \cdot (N+2) \dots (N+P-1)}{1 \cdot 2 \cdot 3 \dots P} = \frac{(N+P-1)!}{(N-1)!P!}$$

or to a sufficient approximations,

$$= \frac{(N+P)^{N+P}}{N^N P^P}.$$

We perform the same calculation for the resonators of the other groups, by determining for each group of resonators the number of possible complexions for the energy given to the group. The multiplication of all numbers obtained in this way gives us then the total number  $R$  of all possible complexions for the arbitrary assigned energy distribution over all resonators.

In the same way any other arbitrarily chosen energy distribution  $E, E', E'' \dots$  will correspond to a definite number  $R$  of all possible complexions which is evaluated in the above manner. Among all energy distributions which are possible for a constant  $E_0 = E + E' + E'' + \dots$  there is one well-defined one for which the number of possible complexions  $R_0$  is larger than for any other distribution. We look for this distribution, if necessary by trial, since this will just be the distribution taken up by the resonators in the stationary radiation field, if they together possess the energy  $E_0$ . This quantities  $E, E', E'', \dots$  can then be expressed in terms of  $E_0$ . Dividing  $E$  by  $N, E'$  by  $N', \dots$  we obtain the stationary value of the energy  $U_\nu, U'_\nu, U''_\nu, \dots$  of a single resonator of each group, and thus also the spatial density of the corresponding radiation energy in a diathermic medium in the spectral range  $\nu$  to  $\nu + d\nu$ ,

$$u_\nu d\nu = \frac{8\pi\nu^2}{c^3} \cdot U_\nu d\nu,$$

so that the energy of the medium is also determined.

Of all quantities which occur only  $E_0$  seems now still to be arbitrary. One sees easily, however, how one can finally evaluate  $E_0$  from the total energy  $E_t$ , since if the chosen value of  $E_0$  leads, for instance, to too large a value of  $E_t$ , we must decrease it, and the other way round.

After the stationary energy distribution is thus determined using a constant  $h$ , we can find the corresponding temperature  $\vartheta$  in degrees absolute<sup>2</sup> using a second constant of nature  $k = 1.346 \times 10^{-6}$  erg · degree<sup>-1</sup> through the equation

$$\frac{1}{\vartheta} = k \frac{d \ln R_0}{dE_0}.$$

The product  $k \ln R_0$  is the entropy of the system of resonators; it is the sum of the entropy of all separate resonators.

It would, to be sure, be very complicated to perform explicitly the above-mentioned calculations, although it would not be without some interest to test the truth of the attainable degree of approximation in a simple case. A more general calculation which is performed very simply, using the above

---

<sup>2</sup> The original states "degrees centigrade" which is clearly a slip [D. t. H.]

prescriptions shows much more directly that the normal energy distribution determined in this way for a medium containing radiation is given by expression

$$u_\nu d\nu = \frac{8\pi\nu^3}{c^3} \frac{d\nu}{e^{h\nu/k\vartheta} - 1}$$

which corresponds exactly to the spectral formula which I give earlier

$$E_\lambda d\lambda = \frac{c_1 \lambda^{-5}}{e^{c_2/\lambda\vartheta} - 1} d\lambda.$$

The formal differences are due to the differences in the definitions of  $u_\nu$  and  $E_\lambda$ . The first equation is somewhat more general insofar as it is valid for arbitrary diathermic medium with light velocity  $c$ . The numerical values of  $h$  and  $k$  which I mentioned were calculated from that equation using the measurements by F. Kurlbaum and by O. Lummer and E. Pringsheim.<sup>3</sup>

I shall now make a few short remarks about the question of the necessity of the above given deduction. The fact that the chosen energy element  $\varepsilon$  for a given group of resonators must be proportional to the frequency  $\nu$  follows immediately from the extremely important Wien displacement law. The relation between  $u$  and  $U$  is one of the basic equations of the electromagnetic theory of radiation. Apart from that, the whole deduction is based upon the theorem that the entropy of a system of resonators with given energy is proportional to the logarithm of the total number of possible complexions for the given energy. This theorem can be split into two other theorems: (1) The entropy of the system in a given state is proportional to the logarithm of the probability of that state, and (2) The probability of any state is proportional to the number of corresponding complexions, or, in other words, any definite complexion is equally probable as any other complexion. The first theorem is, as far as radiative phenomena are concerned, just a definition of the probability of the state, insofar as we have for energy radiation no other a priori way to define the probability that the definition of its entropy. We have here a distinction from the corresponding situation in the kinetic theory of gases. The second theorem is the core of the whole of the theory presented here: in the last resort its proof can only be given empirically. It can also be understood as a more detailed definition of the hypothesis of natural radiation which I have introduced. This hypothesis I have expressed before [6] only in the form that the energy of the radiation is completely "randomly" distributed over the various partial vibrations present in the

---

<sup>3</sup>F. Kurlbaum [4] gives  $S_{100} - S_0 = 0.0731 \text{ Watt cm}^{-2}$ , while O. Lummer and E. Pringsheim [5] give  $\lambda_m \vartheta = 2940 \mu \cdot \text{degree}$ .

radiation.<sup>4</sup> I plan to communicate elsewhere in detail the considerations, which have only been sketched here, with all calculations and with a survey of the development of the theory up to the present.

To conclude I may point to an important consequence of this theory which at the same time makes possible a further test of its reliability. Mr. Boltzmann [7] has shown that the entropy of a monatomic gas in equilibrium is equal to  $\omega R \ln P_0$ , where  $P_0$  is the number of possible complexions (the “permutability”) corresponding to the most probable velocity distribution,  $R$  being the well known gas constant ( $8.31 \times 10^7$  for  $O = 16$ ),  $\omega$  the ratio of the mass of a real molecule to the mass of a mole, which is the same for all substances. If there are any radiating resonators present in the gas, the entropy of the total system must according to the theory developed here be proportional to the logarithm of the number of all possible complexions, including both velocities and radiation. Since according to the electromagnetic theory of the radiation the velocities of the atoms are completely independent of the distribution of the radiation energy, the total number of complexions is simply equal to the product of the number relating to the velocities and the number relating to the radiation. For the total entropy we have thus

$$f \ln (P_0 R_0) = f \ln P_0 + f \ln R_0,$$

where  $f$  is a factor of proportionality. Comparing this with the earlier expressions we find

$$f = \omega R = k,$$

or

$$\omega = \frac{k}{R} = 1.62 \times 10^{-24},$$

that is, a real molecule is  $1.62 \times 10^{-24}$  of a mole or, a hydrogen atom weighs  $1.64 \times 10^{-24}$  g, since  $H = 1.01$ , or, in a mole of any substance there are  $1/\omega = 6.175 \times 10^{23}$  real molecules. Mr. O.E Mayer [8] gives for this number  $640 \times 10^{21}$  which agrees closely.

---

<sup>4</sup>When Mr. Wien in his Paris report about the theoretical radiation laws did not find my theory on the irreversible radiation phenomena satisfactory since it did not give the proof that the hypothesis of natural radiation is the only one which leads to irreversibility, he surely demanded, in my opinion, too much of this hypothesis. If one could prove the hypothesis, it would no longer be a hypothesis, and one did not have to formulate it. However, one could then not derive anything new from it. From the same point of view one should also declare the kinetic theory of gases to be unsatisfactory since nobody has yet proved that the atomistic hypothesis is the only which explains irreversibility. A similar objection could with more or less justice be raised against all inductively obtained theories.

Loschmidt's number  $L$ , that is, the number of gas molecules in 1 cm<sup>3</sup> at 0° C and 1 atm is

$$L = \frac{1\ 013\ 200}{R \cdot 273 \cdot \omega} = 2.76 \times 10^{19}.$$

Mr. Drude [9] finds  $L = 2.1 \times 10^{19}$ .

The Boltzmann-Drude constant  $\alpha$ , that is, the average kinetic energy of an atom at the absolute temperature 1 is

$$\alpha = \frac{3}{2} \omega R = \frac{3}{2} k = 2.02 \times 10^{-16}.$$

Mr. Drude [9] finds  $\alpha = 2.65 \times 10^{-16}$ .

The elementary quantum of electricity  $e$ , that is, the electrical charge of a positive monovalent ion or of an electron is, if  $\varepsilon$  is the known charge of a monovalent mole,

$$e = \varepsilon \omega = 4.69 \times 10^{-10} \text{c.s.u.}$$

Mr. F. Richarz [10] finds  $1.29 \times 10^{-10}$  and Mr. Thomson [11] recently  $6.5 \times 10^{-10}$ .

If the theory is at all correct, all these relations should be not approximately, but absolutely, valid. The accuracy of the calculated number is thus essentially the same as that of the relatively worst known, the radiation constant  $k$ , and is thus much better than all determinations up to now. To test it by more direct methods should be both an important and a necessary task for further research.

## References

- [1] M. Planck, Verh. D. Physik. Ges. Berlin 2, 202 (1900) (reprinted as Paper 1 on p. 79 in the present volume).
- [2] H. Rubens and F. Kurlbaum, S.B. Preuss. Akad. Wiss. p. 929 (1900);
- [3] L. Boltzmann, S.B. Akad. Wiss. Wien 76, 373 (1877);
- [4] F. Kurlbaum, Ann. Physik 65, 759 (1898);
- [5] O. Lummer and E. Pringsheim, Verh. D. Physik. Ges. Berlin 2, 176 (1900);
- [6] M. Planck, Ann. Physik 1, 73 (1900);

- [7] L. Boltzmann, S.B. Akad. Wiss. Wien 76, 428 (1877);
- [8] O.E. Mayer, Die Kinetische Theotie der Gase, 2nd ed., p. 337 ( 1899);
- [9] P. Drude, Ann. Physik 1, 578 (1900);
- [10] F. Richarz, Ann. Physik 52, 397 (1894);
- [11] J.J. Thomson, Phil. Mag. 46, 528 (1898).

**On the Law of the Energy Distribution  
in the Normal Spectrum**

M. Planck

(Received January 7, 1901)

In other form reported in the German Physical Society (Deutsche Physikalische Gesellschaft) in the meetings of October 19 and December 14, 1900, published in Verh. Dtsch. Phys. Ges. Berlin, 1900, 2, 202 and 237

— — ◇ ◇ — —

*Translated from German by Kuyanov Yu. V. [kuyanov@mx.ihp.su]*

— — ◇ ◇ — —

## Preface

The recent spectral measurements of O. Lummer and E. Pringsheim<sup>1</sup> and even more striking those of H. Rubens and F. Kurlbaum<sup>2</sup>, both confirming more recent results obtained by H. Beckmann<sup>3</sup>, would discover that the law of the energy distribution in the normal spectrum first stated by W. Wien from the molecular-kinetic consideration and later by me from the theory of electromagnetic radiation is not universally correct.

In any case an improvement on the theory is needed and I shall further try to carry through basing on the theory of electromagnetic radiation developed by me. First of all there is necessary for it to find an alterable link in the chain of reasons resulting in the Wien's energy distribution law. So one handles to remove this link from the chain and create a suitable substitute.

The fact that the physical ground of the electromagnetic radiation theory including the hypothesis of the "natural radiation", resists destructive criticism, is shown in my recent work<sup>4</sup>; and since the calculations are known to be error free, so the statement remains to be held that the energy distribution law of the normal spectrum is totally defined if one succeeds in calculation of

---

<sup>1</sup>O. Lummer, E. Pringsheim. Verhandl. Deutsch. Phys. Ges., 1900, 2, 163.

<sup>2</sup>H. Rubens, F. Kurlbaum. Sitzungsber. Akad. Wiss. Berlin, 1900, 929.

<sup>3</sup>H. Beckmann. Inaug-Dissert. Tübingen, 1898, see also: H. Rubens. Wied. Ann., 1899, 69, 582.

<sup>4</sup>M. Planck. Ann. Phys., 1900, 1, 719.

entropy  $S$  of irradiated monochromatic vibrating resonator as a function of its vibrational energy. So then from the relation  $dS/dU = 1/\vartheta$  one keeps the temperature  $\vartheta$  dependence on energy  $U$ , and since the energy  $U$ , on the other hand, is simply related<sup>5</sup> with a radiation density of appropriate number of vibrations, so the temperature dependence on this radiation density is also obtained. So the normal distribution of energy is one for which the radiation densities of any different numbers of vibrations have the same temperature.

Thus the total problem is self reduced to that of definition  $S$  as a function of  $U$ , and the essential part of the following research is devoted to the solution of this problem. In the first my work on this problem I have entered  $S$  directly by defining with no further substantiation, as the simple function of  $U$ , and have limited by showing that such form for the entropy satisfies to all requirements of the thermodynamics. Then I considered that it is alone possible and therefore the Wien's law, from it flowing out, necessarily is the universal one. In later, more particular research<sup>6</sup> it seemed to me, however, that it should be expressions, doing the same, and that in any case therefore one more condition is needed for anyone being able to calculate  $S$  uniquely. It seemed to me that I have found one such condition in the form of statement, immediately then considered by me as plausible, that by the infinitesimal irreversible alteration of the near thermal equilibrium being system of  $N$  uniform, just in stationary radiation field placed resonators, the bound up with it alteration of the total entropy  $S_N = NS$  depends only on their total energy  $U_N = NU$  and their alteration but not on the energy  $U$  of particular resonators. This statement leads again with necessity to the Wien's energy distribution law. But now however the later is not confirmed by experience, so the conclusion is forced that this statement in its universality also cannot be right and so from the theory is to be removed<sup>7</sup>.

Therefore yet another condition should be entered which enables the calculation of  $S$ , and for its realization the more detailed consideration of the entropy concept is needed. The direction of these deliberate thoughts is indicated by the consideration of the fragility of early made supposition. The path is below described, in which the new simple expression for entropy as well as the new formula for radiation are self found, both contradicting no fact established till now.

---

<sup>5</sup>See below equation (8)

<sup>6</sup>M. Planck. Ann. Phys., 1900, **1**, 730.

<sup>7</sup>One compares besides the criticism, to which this statement is exposed yet: W. Wien. Rapport für den Pariser Congress, 1900, **2**, 40; O. Lummer. Loc. cit., p. 92.

## I. The calculation of entropy of any resonator as a function of its energy

### § 1

An entropy is conditioned by disorder, and this disorder in accordance with electro-magnetic theory of radiation is based on monochromatic vibrations of any resonator if although it remains in a stable stationary field of radiation, on non-regularity by which it permanently changes its amplitude and its phase, since one clocks time intervals which are long compared with a time of vibration, but short compared with a measurement time. If the amplitude and the phase both are absolutely constant as well as vibrations are quite homogeneous, no entropy could exist and the vibrational energy should be quite free convertible into the work. A constant energy  $U$  of alone stationary vibrating resonator is therefore as an average by time to be perceived or what turns to quite the same result, as a simultaneous average of energies of large number  $N$  of uniform resonators, just into stationary radiation field placed, sufficiently removed from one another to have no affect to each other directly. In this sense in future we will speak about an average energy  $U$  of a separate resonator. Then a total energy

$$U_N = NU \quad (1)$$

of such system of  $N$  resonators is corresponded to certain total entropy

$$S_N = NS \quad (2)$$

of the same system where an average entropy of any separate resonator is represented by  $S$ , and this entropy  $S_N$  is based on a disorder with which the total energy  $U_N$  is distributed among particular resonators.

### § 2

Now we suppose an entropy  $S_N$  of a system with an arbitrary remaining additive constant to be proportional to logarithm of the probability  $W$  with which  $N$  resonators altogether possess an energy  $U_N$ ; therefore:

$$S_N = k \ln W + \text{const.} \quad (3)$$

In my opinion this supposition originates from the base of the definition of the probability  $W$  mentioned whereas in the premise, put on the ground of the electromagnetic theory of radiation, we have not any support, enabling to speak about such probability in a definite sense. For the expedience of so aimed supposition its simplicity as well as its neighbourhood with that of the kinetic theory of gases are standing for<sup>8</sup>.

### § 3

Now it is worth reminding to find the probability  $W$  of  $N$  resonators alltogether having a vibrational energy  $U_N$ . It is necessary for it to imagine  $U_N$  not as a continuous unlimited divided value, but as a discrete one, composed of integer number of finite equal parts. If we give a name energy element  $\varepsilon$  to such part, so one can suppose that

$$U_N = P \cdot \varepsilon, \quad (4)$$

where  $P$  is an integer, in general, large number, whereas the value for  $\varepsilon$  is till to be defined.

Now it is clear that the distribution  $P$  of energy elements among  $N$  resonators can happen by some limited quite definite number of manners. We give a name “complexion” to every such manner of distribution following L. Boltzmann who had used this name for an expression with a similar idea. Having numbered resonators by 1, 2, 3, ...,  $N$ , one writes them in a row each to another and under each resonator places a number of energy elements fallen to it in some arbitrary distribution, so for each complexion one obtains a symbol of the following form:

1	2	3	4	5	6	7	8	9	10
7	38	11	0	9	2	20	4	4	5

Here  $N = 10$ ,  $P = 100$  are considered. The number  $\mathfrak{R}$  of all possible complexions is obviously equal to one of all possible digital images which can be obtained in this manner for the lower row with definite  $N$  and  $P$ . For intelligibility it should be mentioned that two complexions are considered as different if corresponding digital images have the same numbers but in a different order.

---

<sup>8</sup>L. Boltzmann. Sitzungsber d. k. Akad. d. Wissensch. zu Wien (I), 1877, **76**, 428.

Following combinatory, the number of all possible complexions is

$$\mathfrak{R} = \frac{N \cdot (N+1) \cdot (N+2) \dots (N+P-1)}{1 \cdot 2 \cdot 3 \dots P} = \frac{(N+P-1)!}{(N-1)! \cdot P!}.$$

Here is in a first approximation according to Stirling offer:

$$N! = N^N;$$

therefore in appropriate approximation

$$\mathfrak{R} = \frac{(N+P)^{N+P}}{N^N \cdot P^P}.$$

#### § 4

The hypothesis, we now wish to put into the base of further calculation, is as follows: the probability of that  $N$  resonators altogether possess vibrational energy  $U_N$  is proportional to the number  $\mathfrak{R}$  of all possible complexions with energy  $U_N$  distributed among  $N$  resonators, or by other words: each certain complexion is as probable as either another one. It should in last line only by experience be proved whether this hypothesis virtually hit into nature. Instead however an opposite one should be possible: once an experience should judge in its favor, the validity of hypothesis will result in the further conclusions on the special nature of resonator's vibrations, namely on the character of meanwhile appearing "indifferent and in their value compared primary game spaces" by expression manner of J. v. Kries<sup>9</sup>. In a modern state of this question a further promotion of this idea should certainly appear as premature.

#### § 5

According to hypothesis introduced in relation with the equation (3), the entropy of considered system of resonators with suitable definition of additive constant is:

$$S_N = k \ln \mathfrak{R} = k \{(N+P) \ln (N+P) - N \ln N - P \ln P\}, \quad (5)$$

---

<sup>9</sup>Joh. v. Kries. Die Principien der Wahrscheinlichkeitsrechnung. Freiburg, 1886, p. 36.

and accepting (4) and (1):

$$S_N = kN \left\{ \left( 1 + \frac{U}{\varepsilon} \right) \ln \left( 1 + \frac{U}{\varepsilon} \right) - \frac{U}{\varepsilon} \ln \frac{U}{\varepsilon} \right\}.$$

Therefore according to (2), entropy  $S$  of a resonator as a function of its energy  $U$  is:

$$S = k \left\{ \left( 1 + \frac{U}{\varepsilon} \right) \ln \left( 1 + \frac{U}{\varepsilon} \right) - \frac{U}{\varepsilon} \ln \frac{U}{\varepsilon} \right\}. \quad (6)$$

## II. The deduction of the Wien's displacement law

### § 6

Following a Kirchhoff's law of proportionality of both emission- and absorbability, discovered by W. Wien<sup>10</sup> and called by his name so-called the displacement law, including, as a particular case, the law of Stefan–Boltzmann of full emittance dependence on temperature, builds the most valuable constituent in the well grounded foundation of the theory of heat radiation. In a fashion, given by M. Thiesen<sup>11</sup>, it announces:

$$E \cdot d\lambda = \vartheta^5 \psi(\lambda\vartheta) \cdot d\lambda,$$

where  $\lambda$  is a wavelength,  $E d\lambda$  is a volume density of a spectral slice between  $\lambda$  and  $\lambda + d\lambda$  belonging to “black” radiation<sup>12</sup>,  $\vartheta$  is a temperature and  $\psi(x)$  is a known function of a single argument  $x$ .

### § 7

Now we are coming to investigate what Wien's displacement law says about our resonator's entropy  $S$  dependence on its energy and its own period, that is in those general case that resonator itself is in an arbitrary

---

<sup>10</sup>W. Wien. Sitzungsber. Acad. Wissenschaft. Berlin, 1893, 55.

<sup>11</sup>M. Thiesen. Verhandl. Deutsch. Phys. Ges., 1900, 2, 66.

<sup>12</sup>One should perhaps more conveniently speak about “white” radiation, whose proper generalization is now understood as a “quite white light”.

diathermal medium. For this aim first of all let us generalize the Thiesen's form of the law on the radiation in an arbitrary diathermal medium with the velocity of light propagation  $c$ . Since we have to consider not a total radiation but monochromatic one, so when comparing different diathermal media, the number of vibrations  $\nu$  should necessarily be introduced instead of wavelength  $\lambda$ .

Thus the volume density of a spectral slice between  $\nu$  and  $\nu + d\nu$ , belonging to energy of radiation, is to be denoted as  $ud\nu$ , so one should write:  $ud\nu$  instead of  $Ed\nu$ ,  $c/\nu$  instead of  $\lambda$  and  $cd\nu/\nu^2$  instead of  $d\lambda$ . This results in:

$$\mathbf{u} = \vartheta^5 \cdot \frac{c}{\nu^2} \cdot \psi \left( \frac{c\vartheta}{\nu} \right).$$

Now according to known Kirchhoff-Clausius's law, the energy, emitted by black surface in a time unit into a diathermal medium, for defined temperature  $\vartheta$  and defined number of vibrations  $\nu$  is reverse proportional to the square of the velocity of propagation  $c^2$ ; thus the volume energy density  $\mathbf{u}$  is reverse proportional to  $c^3$ , and we obtain:

$$\mathbf{u} = \frac{\vartheta^5}{\nu^2 c^3} f \left( \frac{\vartheta}{\nu} \right),$$

where constants of the function  $f$  do not depend on  $c$ .

Instead of it we could also write when  $f$  every time, as in following, means a new function of a single argument:

$$\mathbf{u} = \frac{\nu^3}{c^3} f \left( \frac{\vartheta}{\nu} \right) \quad (7)$$

and by the way see that in a cube of a wavelength size a contained radiation energy with a certain temperature as well as a number of vibrations is known to be:  $\mathbf{u}\lambda^3$ , the same for all diathermal media.

## § 8

In order to pass from the volume density of radiation  $\mathbf{u}$  to the energy  $U$  of the resonator being in the radiation field and stationary vibrating with the same number of vibrations  $\nu$ , we shall use the relation, published in equation (34) of my work on non-reversible processes of radiation<sup>13</sup>:

$$\mathfrak{R} = \frac{\nu^2}{c^2} \cdot U$$

---

<sup>13</sup>M. Planck. Ann. Phys., 1900, **1**, 99.

( $\mathfrak{R}$  is the intensity of monochromatic line-polarized beam), which together with the known equation

$$\mathbf{u} = \frac{8\pi\mathfrak{R}}{c}$$

yields the relation:

$$\mathbf{u} = \frac{8\pi\nu^2}{c^3} U. \quad (8)$$

From here and (7) it follows:

$$U = \nu f\left(\frac{\vartheta}{\nu}\right),$$

where now  $c$  is not at all present. Instead of it we should also write:

$$\vartheta = \nu f\left(\frac{U}{\nu}\right).$$

## § 9

Finally introducing yet more the entropy of resonator  $S$ , we assign:

$$\frac{1}{\vartheta} = \frac{dS}{dU}. \quad (9)$$

Then it turns out:

$$\frac{dS}{dU} = \frac{1}{\nu} f\left(\frac{U}{\nu}\right)$$

and integrating, one obtains:

$$S = f\left(\frac{U}{\nu}\right), \quad (10)$$

i.e. the entropy of resonator, vibrating in an arbitrary diathermal medium, depends only on the single variable  $U/\nu$  and besides keeps only the universal constants. This, as I know, is the simplest representation of the Wien's displacement law.

## § 10

Applying the Wien's displacement law in its latter representation to the expression (6) for the entropy  $S$ , one can realize that the energy element  $\varepsilon$  should be proportional to the number of vibrations  $\nu$ , so:

$$\varepsilon = h \cdot \nu$$

and therefore:

$$S = k \left\{ \left( 1 + \frac{U}{h\nu} \right) \ln \left( 1 + \frac{U}{h\nu} \right) - \frac{U}{h\nu} \ln \frac{U}{h\nu} \right\}.$$

Here  $h$  and  $k$  are the universal constants.

By substitution into (9) one obtains:

$$\frac{1}{\vartheta} = \frac{k}{h\nu} \ln \left( 1 + \frac{h\nu}{U} \right), \quad (11)$$

$$U = \frac{h\nu}{e^{\frac{h\nu}{k\vartheta}} - 1}$$

and the energy distribution law searched then follows from (8):

$$\mathbf{u} = \frac{8\pi h\nu^3}{c^3} \cdot \frac{1}{e^{\frac{h\nu}{k\vartheta}} - 1}, \quad (12)$$

or also if one with in § 7 shown substitutions instead of the number of vibrations  $\nu$  introduces again the wavelength  $\lambda$ , that is:

$$E = \frac{8\pi ch}{\lambda^5} \cdot \frac{1}{e^{\frac{ch}{k\lambda\vartheta}} - 1}. \quad (13)$$

I suppose to show in the other place the expression for the intensity and one for the entropy of the in diathermal medium propagating radiation as well as the law of the increase of the total entropy in unstationary radiating process.

### III. The numeral values

#### § 11

The values of both natural constants  $h$  and  $k$  may be calculated well precisely with a help of measurements available. F. Kurlbaum<sup>14</sup> has found that if one designates by  $S_t$  the total energy, radiating into an air in 1 sec from the  $1 \text{ cm}^2$  surface of the black body exposed with  $t^\circ$ , then it is:

$$S_{100} - S_0 = 0.0731 \frac{\text{Watt}}{\text{cm}^2} = 7.31 \cdot 10^5 \frac{\text{erg}}{\text{cm}^2 \cdot \text{sec}}.$$

From here the volume density of the total radiation energy in the air for the absolute temperature of 1 turns out:

$$\frac{4 \cdot 7.31 \cdot 10^5}{3 \cdot 10^{10} \cdot (373^4 - 273^4)} = 7.061 \cdot 10^{-15} \frac{\text{erg}}{\text{cm}^2 \cdot \text{grad}^4}.$$

From the other hand, according to (12), the volume density of the total radiation energy for  $\vartheta = 1$  is as follows:

$$\begin{aligned} u &= \int_0^\infty \mathbf{u} d\nu = \frac{8\pi h}{c^3} \int_0^\infty \frac{\nu^3 d\nu}{e^{\frac{h\nu}{k}} - 1} \\ &= \frac{8\pi h}{c^3} \int_0^\infty \nu^3 \left( e^{-\frac{h\nu}{k}} + e^{-\frac{2h\nu}{k}} + e^{-\frac{3h\nu}{k}} + \dots \right) d\nu \end{aligned}$$

and by all terms integration it yields:

$$u = \frac{8\pi h}{c^3} \cdot 6 \left( \frac{k}{h} \right)^4 \left( 1 + \frac{1}{2^4} + \frac{1}{3^4} + \frac{1}{4^4} + \dots \right) = \frac{48\pi k^4}{c^3 h^3} \cdot 1.0823.$$

Assuming it to be equal to  $7.061 \cdot 10^{-15}$ , one obtains, since  $c = 3 \cdot 10^{10}$ ,

$$\frac{k^4}{h^3} = 1.1682 \cdot 10^{15}. \quad (14)$$

---

<sup>14</sup>F. Kurlbaum. Wied. Ann., 1898, **65**, 759.

## § 12

O. Lummer and E. Pringsheim<sup>15</sup> have determined the product  $\lambda_m \vartheta$ , where  $\lambda_m$  is the wavelength of the maximum of  $E$  in the air for the temperature  $\vartheta$ , having value up to 2940  $\mu\text{-grad}$ .

So in absolute units that is

$$\lambda_m \vartheta = 0.294 \text{ cm} \cdot \text{grad.}$$

From the other hand, if one assumes the partial derivative of  $E$  in respect to  $\lambda$  to be equal to zero, so that  $\lambda = \lambda_m$ , then it follows from (13):

$$\left(1 - \frac{ch}{5k\lambda_m \vartheta}\right) \cdot e^{\frac{ch}{k\lambda_m \vartheta}} = 1$$

and from this transcendental equation one obtains:

$$\lambda_m \vartheta = \frac{ch}{4.9651 \cdot k}.$$

Therefore:

$$\frac{h}{k} = \frac{4.9651 \cdot 0.294}{3 \cdot 10^{10}} = 4.866 \cdot 10^{-11}.$$

From here and from (14) the values for the universal constants turn out:

$$h = 6.55 \cdot 10^{-27} \text{ erg} \cdot \text{sec}, \quad (15)$$

$$k = 1.346 \cdot 10^{-16} \text{ erg/grad.} \quad (16)$$

These are just the same values that I have presented in my recent communication.

---

<sup>15</sup>O. Lummer, E. Pringsheim. Verhandl. Deutsch. Phys. Ges., 1900, **2**, 176.



## Concerning an Heuristic Point of View Toward the Emission and Transformation of Light

A. Einstein

Bern, 17 March 1905

(Received March 18, 1905)

Translation into English

American Journal of Physics, v. 33, n. 5, May 1965

— — ◇ ◇ — —

A profound formal distinction exists between the theoretical concepts which physicists have formed regarding gases and other ponderable bodies and the Maxwellian theory of electromagnetic processes in so-called empty space. While we consider the state of a body to be completely determined by the positions and velocities of a very large, yet finite, number of atoms and electrons, we make use of continuous spatial functions to describe the electromagnetic state of a given volume, and a finite number of parameters cannot be regarded as sufficient for the complete determination of such a state. According to the Maxwellian theory, energy is to be considered a continuous spatial function in the case of all purely electromagnetic phenomena including light, while the energy of a ponderable object should, according to the present conceptions of physicists, be represented as a sum carried over the atoms and electrons. The energy of a ponderable body cannot be subdivided into arbitrarily many or arbitrarily small parts, while the energy of a beam of light from a point source (according to the Maxwellian theory of light or, more generally, according to any wave theory) is continuously spread an ever increasing volume.

The wave theory of light, which operates with continuous spatial functions, has worked well in the representation of purely optical phenomena

and will probably never be replaced by another theory. It should be kept in mind, however, that the optical observations refer to time averages rather than instantaneous values. In spite of the complete experimental confirmation of the theory as applied to diffraction, reflection, refraction, dispersion, etc., it is still conceivable that the theory of light which operates with continuous spatial functions may lead to contradictions with experience when it is applied to the phenomena of emission and transformation of light.

It seems to me that the observations associated with blackbody radiation, fluorescence, the production of cathode rays by ultraviolet light, and other related phenomena connected with the emission or transformation of light are more readily understood if one assumes that the energy of light is discontinuously distributed in space. In accordance with the assumption to be considered here, the energy of a light ray spreading out from a point source is not continuously distributed over an increasing space but consists of a finite number of energy quanta which are localized at points in space, which move without dividing, and which can only be produced and absorbed as complete units.

In the following I wish to present the line of thought and the facts which have led me to this point of view, hoping that this approach may be useful to some investigators in their research.

## 1. Concerning a Difficulty with Regard to the Theory of Blackbody Radiation

We start first with the point of view taken in the Maxwellian and the electron theories and consider the following case. In a space enclosed by completely reflecting walls, let there be a number of gas molecules and electrons which are free to move and which exert conservative forces on each other on close approach: i.e. they can collide with each other like molecules in the kinetic theory of gases.<sup>1</sup> Furthermore, let there be a number of electrons which are bound to widely separated points by forces proportional to their distances from these points. The bound electrons are also to participate in conservative interactions with the free molecules and electrons when the latter come

---

<sup>1</sup>This assumption is equivalent to the supposition that the average kinetic energies of gas molecules and electrons are equal to each other at thermal equilibrium. It is well known that, with the help of this assumption, Herr Drude derived a theoretical expression for the ratio of thermal and electrical conductivities of metals.

very close. We call the bound electrons “oscillators”: they emit and absorb electromagnetic waves of definite periods.

According to the present view regarding the origin of light, the radiation in the space we are considering (radiation which is found for the case of dynamic equilibrium in accordance with the Maxwellian theory) must be identical with the blackbody radiation — at least if oscillators of all the relevant frequencies are considered to be present.

For the time being, we disregard the radiation emitted and absorbed by the oscillators and inquire into the condition of dynamical equilibrium associated with the interaction (or collision) of molecules and electrons. The kinetic theory of gases asserts that the average kinetic energy of an oscillator electron must be equal to the average kinetic energy of a translating gas molecule. If we separate the motion of an oscillator electron into three components at angles to each other, we find for the average energy  $\bar{E}$  of one of these linear components the expression

$$\bar{E} = (R/N) T,$$

where  $R$  denotes the universal gas constant.  $N$  denotes the number of “real molecules” in a gram equivalent, and  $T$  the absolute temperature. The energy  $\bar{E}$  is equal to two-thirds the kinetic energy of a free monatomic gas particle because of the equality the time average values of the kinetic and potential energies of the oscillator. If through any cause—in our case through radiation processes—it should occur that the energy of an oscillator takes on a time-average value greater or less than  $\bar{E}$ , then the collisions with the free electrons and molecules would lead to a gain or loss of energy by the gas, different on the average from zero. Therefore, in the case we are considering, dynamic equilibrium is possible only when each oscillator has the average energy  $\bar{E}$ .

We shall now proceed to present a similar argument regarding the interaction between the oscillators and the radiation present in the cavity. Herr Planck has derived<sup>2</sup> the condition for the dynamics equilibrium in this case under the supposition that the radiation can be considered a completely random process.<sup>3</sup> He found

$$(\bar{E}_\nu) = (L^3/8\pi\nu^2)\rho_\nu,$$

---

<sup>2</sup>M. Planck, Ann. Phys. 1, 99 (1900).

<sup>3</sup>This problem can be formulated in the following manner. We expand the  $Z$  component of the electrical force ( $Z$ ) at an arbitrary point during the time interval between  $t = 0$  and  $t = T$  in a Fourier series in which  $A_\nu \geq 0$  and  $0 \leq \alpha_\nu \leq 2\pi$ : the time  $T$  is taken to

where  $(\bar{E}_\nu)$  is the average energy (per degree of freedom) of an oscillator with eigenfrequency  $\nu$ ,  $L$  the velocity of light,  $\nu$  the frequency, and  $\rho_\nu d\nu$  the energy per unit volume of that portion of the radiation with frequency between  $\nu$  and  $\nu + d\nu$ .

If the radiation energy of frequency  $\nu$  is not continually increasing or decreasing, the following relations must obtain:

$$(R/N) T = \bar{E} = \bar{E}\nu = (L^3/8\pi\nu^2)\rho_\nu,$$

$$\rho_\nu = (R/N)(8\pi\nu^2/L^3) T.$$

These relations, found to be the conditions of dynamic equilibrium, not only fail to coincide with experiment, but also state that in our model there can be no talk of a definite energy distribution between ether and matter. The wider the range of wave numbers of the oscillators, the greater will be the radiation energy of the space, and in the limit we obtain

$$\int_0^\infty \rho_\nu d\nu = \frac{R}{N} \cdot \frac{8\pi}{L^3} \cdot T \int_0^\infty \nu^2 d\nu = \infty.$$

be very large relative to all the periods of oscillation that are present:

$$Z = \sum_{\nu=1}^{\nu=\infty} A_\nu \sin \left( 2\pi\nu \frac{t}{T} + \alpha_\nu \right),$$

If one imagines making this expansion arbitrary often at a given point in space at randomly chosen instants of time, one will obtain various sets of values of  $A_\nu$  and  $\alpha_\nu$ . There then exist for the frequency of occurrence of different sets of values of  $A_\nu$  and  $\alpha_\nu$  (statistical) probabilities  $dW$  of the form:

$$dW = f(a_1, A_2, \dots, \alpha_1, \alpha_2, \dots) dA_1 dA_2 \dots d\alpha_1 d\alpha_2 \dots,$$

The radiation is then as disordered as conceivable if

$$f(A_1, A_2, \dots, \alpha_1, \alpha_2, \dots) = F_1(A_1)F_2(A_2) \dots f_1(\alpha_1)f_2(\alpha_2) \dots,$$

i.e., if the probability of a particular value of  $A$  or  $\alpha$  is independent of other values of  $A$  or  $\alpha$ . The more closely this condition is fulfilled (namely, that the individual pairs of values of  $A_\nu$  and  $\alpha_\nu$  are dependent upon the emission and absorption processes of specific groups of oscillators) the more closely will radiation in the case being considered approximate a perfectly random state.

## 2. Concerning Planck's Determination of the Fundamental Constants

We wish to show in the following that Herr Planck's determination of the fundamental constants is, to a certain extent, independent of his theory of blackbody radiation.

Planck's formula,<sup>4</sup> which has proved adequate up to this point, gives for  $\rho_\nu$

$$\rho_\nu = \frac{\alpha\nu^3}{e^{\beta\nu/T} - 1},$$

$$\alpha = 6.10 \times 10^{-56},$$

$$\beta = 4.866 \times 10^{-11}.$$

For large values of  $T/\nu$ ; i.e. for large wavelengths and radiation densities, this equation takes the form

$$\rho_\nu = (\alpha/\beta) \nu^2 T.$$

It is evident that this equation is identical with the one obtained in Sec. 1 from the Maxwellian and electron theories. By equating the coefficients of both formulas one obtains

$$(R/N)(8\pi/L^3) = (\alpha/\beta)$$

or

$$N = (\beta/\alpha)(8\pi R/L^3) = 6.17 \times 10^{23}.$$

i.e., an atom of hydrogen weighs  $1/N$  grams =  $1.62 \times 10^{-24}$  g. This is exactly the value found by Herr Planck, which in turn agrees with values found by other methods.

We therefore arrive at the conclusion: the greater the energy density and the wavelength of a radiation, the more useful do the theoretical principles we have employed turn out to be: for small wavelengths and small radiation densities, however, these principles fail us completely.

In the following we shall consider the experimental facts concerning blackbody radiation without invoking a model for the emission and propagation of the radiation itself.

---

<sup>4</sup>M. Planck, Ann. Phys. 4, 561 (1901).

### 3. Concerning the Entropy of Radiation

The following treatment is to be found in a famous work by Herr W. Wien and is introduced here only for the sake of completeness.

Suppose we have radiation occupying a volume  $v$ . We assume that the observable properties of the radiation are completely determined when the radiation density  $\rho(\nu)$  is given for all frequencies.<sup>5</sup> Since radiation of different frequencies are to be considered independent of each other when there is no transfer of heat or work, the entropy of the radiation can be represented by

$$S = v \int_0^\infty \varphi(\rho, \nu) d\nu,$$

where  $\varphi$  is a function of the variables  $\rho$  and  $\nu$ .

$\varphi$  can be reduced to a function of a single variable through formulation of the condition that the entropy of the radiation is unaltered during adiabatic compression between reflecting walls. We shall not enter into this problem, however, but shall directly investigate the derivation of the function  $\varphi$  from the blackbody radiation law.

In the case of blackbody radiation,  $\rho$  is such a function of  $\nu$  that the entropy is maximum for a fixed value of energy; i.e.,

$$\delta \int_0^\infty \varphi(\rho, \nu) d\nu = 0,$$

providing

$$\delta \int_0^\infty \rho d\nu = 0.$$

From this it follows that for every choice of  $\delta\rho$  as a function of  $\nu$

$$\int_0^\infty \left( \frac{\partial \varphi}{\partial \rho} - \lambda \right) \delta \rho d\nu = 0,$$

where  $\lambda$  is independent of  $\nu$ . In the case of blackbody radiation, therefore,  $\partial\varphi/\partial\rho$  is independent of  $\nu$ .

---

<sup>5</sup>This assumption is an arbitrary one. One will naturally cling to this simplest assumption as long as it is not controverted experiment.

The following equation applies when the temperature of a unit volume of blackbody radiation increases by  $dT$

$$dS = \int_{\nu=0}^{\nu=\infty} \left( \frac{\partial \varphi}{\partial \rho} \right) d\rho d\nu,$$

or, since  $\partial\varphi/\partial\rho$  is independent of  $\nu$ .

$$dS = (\partial\varphi/\partial\rho) dE.$$

Since  $dE$  is equal to the heat added and since the process is reversible, the following statement also applies

$$dS = (1/T) dE.$$

By comparison one obtains

$$\partial\varphi/\partial\rho = 1/T.$$

This is the law of blackbody radiation. Therefore one can derive the law of blackbody radiation from the function  $\varphi$ , and, inversely, one can derive the function  $\varphi$  by integration, keeping in mind the fact that  $\varphi$  vanishes when  $\rho = 0$ .

#### 4. Asymptotic formula for the Entropy of Monochromatic Radiation at Low Radiation Density

From existing observations of the blackbody radiation, it is clear that the law originally postulated by Herr W. Wien,

$$\rho = \alpha \nu^3 e^{-\beta \nu/T},$$

is not exactly valid. It is, however, well confirmed experimentally for large values of  $\nu/T$ . We shall base our analysis on this formula, keeping in mind that our results are only valid within certain limits.

This formula gives immediately

$$(1/T) = -(1/\beta\nu) \ln (\rho/\alpha\nu^3)$$

and then, by using the relation obtained in the preceding section,

$$\varphi(\rho, \nu) = -\frac{\rho}{\beta\nu} \left[ \ln \left( \frac{\rho}{\alpha\nu^3} \right) - 1 \right].$$

Suppose that we have radiation of energy  $E$ , with frequency between  $\nu$  and  $\nu + d\nu$ , enclosed in volume  $v$ . The entropy of this radiation is:

$$S = v\varphi(\rho, \nu)d\nu = -\frac{E}{\beta\nu} \left[ \ln \left( \frac{E}{v\alpha\nu^3d\nu} \right) - 1 \right].$$

If we confine ourselves to investigating the dependence of the entropy on the volume occupied by the radiation, and if we denote by  $S_0$  the entropy of the radiation at volume  $v_0$ , we obtain

$$S - S_0 = (E/\beta\nu) \ln (v/v_0).$$

This equation shows that the entropy of a monochromatic radiation of sufficiently low density varies with the volume in the same manner as the entropy of an ideal gas or a dilute solution. In the following, this equation will be interpreted in accordance with the principle introduced into physics by Herr Boltzmann, namely that the entropy of a system is a function of the probability its state.

## 5. Molecular-Theoretic Investigation of the Dependence of the Entropy of Gases and Dilute solutions on the volume

In the calculation of entropy by molecular-theoretic methods we frequently use the word "probability" in a sense differing from that employed in the calculus of probabilities. In particular "gases of equal probability" have frequently been hypothetically established when one theoretical models being utilized are definite enough to permit a deduction rather than a conjecture. I will show in a separate paper that the so-called "statistical probability" is fully adequate for the treatment of thermal phenomena, and I hope that by doing so I will eliminate a logical difficulty that obstructs the application of Boltzmann's Principle. here, however, only a general formulation and application to very special cases will be given.

If it is reasonable to speak of the probability of the state of a system, and furthermore if every entropy increase can be understood as a transition to a state of higher probability, then the entropy  $S_1$  of a system is a function of  $W_1$ , the probability of its instantaneous state. If we have two noninteracting systems  $S_1$  and  $S_2$ , we can write

$$S_1 = \varphi_1(W_1),$$

$$S_2 = \varphi_2(W_2).$$

If one considers these two systems as a single system of entropy  $S$  and probability  $W$ , it follows that

$$S = S_1 + S_2 = \varphi(W)$$

and

$$W = W_1 \cdot W_2.$$

The last equation says that the states of the two systems are independent of each other.

From these equations it follows that

$$\varphi(W_1 \cdot W_2) = \varphi_1(W_1) + \varphi_2(W_2)$$

and finally

$$\varphi_1(W_1) = C \ln(W_1) + \text{const},$$

$$\varphi_2(W_2) = C \ln(W_2) + \text{const},$$

$$\varphi(W) = C \ln(W) + \text{const}.$$

The quantity  $C$  is therefore a universal constant; the kinetic theory of gases shows its value to be  $R/N$ , where the constants  $R$  and  $N$  have been defined above. If  $S_0$  denotes the entropy of a system in some initial state and  $W$  denotes the relative probability of a state of entropy  $S$ , we obtain in general

$$S - S_0 = (R/N) \ln W.$$

First we treat the following special case. We consider a number ( $n$ ) of movable points (e.g., molecules) confined in a volume  $v_0$ . Besides these points, there can be in the space any number of other movable points of any kind. We shall not assume anything concerning the law in accordance with which the points move in this space except that with regard to this motion, no part of the space (and no direction within it) can be distinguished from

any other. Further, we take the number of these movable points to be so small that we can disregard interactions between them.

This system, which, for example, can be an ideal gas or a dilute solution, possesses an entropy  $S_0$ . Let us imagine transferring all  $n$  movable points into a volume  $v$  (part of the volume  $v_0$ ) without anything else being changed in the system. This state obviously possesses a different entropy ( $S$ ), and now wish to evaluate the entropy difference with the help of the Boltzmann Principle.

We inquire: How large is the probability of the latter state relative to the original one? Or: How large is the probability that at a randomly chosen instant of time all  $n$  movable points in the given volume  $v_0$  will be found by chance in the volume  $v$ ?

For this probability, which is a “statistical probability”, one obviously obtains:

$$W = (v/v_0)^n;$$

By applying the Boltzmann Principle, one then obtains

$$S - S_0 = R \cdot (n/N) \ln (v/v_0).$$

It is noteworthy that in the derivation of this equation, from which one can easily obtain the law of Boyle and Gay-Lussac as well as the analogous law of osmotic pressure thermodynamically,<sup>6</sup> no assumption had to be made as to a law of motion of the molecules.

## 6. Interpretation of the Expression for the volume Dependence of the entropy of Monochromatic Radiation in Accordance with Boltzmann's Principle

In Sec. 4, we found the following expression for the dependence of the entropy of monochromatic radiation on the volume

$$S - S_0 = (E/\beta\nu) \ln (v/v_0).$$

---

<sup>6</sup>If  $E$  is the energy of the system, one obtains:

$$-d \cdot (E - TS) = pdv = TdS = RT \cdot (n/N) \cdot (dv/v);$$

therefore

$$pv = R \cdot (n/N) \cdot T.$$

If one writes this in the form

$$S - S_0 = (R/N) \ln \left[ (v/v_0)^{(N/R)(E/\beta\nu)} \right].$$

and if one compares this with the general formula for the Boltzmann principle

$$S - S_0 = (R/N) \ln W,$$

one arrives at the following conclusion:

If monochromatic radiation of frequency  $\nu$  and energy  $E$  is enclosed by reflecting walls in a volume  $v_0$ , the probability that the total radiation energy will be found in a volume  $v$  (part of the volume  $v_0$ ) at any randomly chosen instant is

$$W = (v/v_0)^{(N/R)(E/\beta\nu)}.$$

From this we further conclude that: Monochromatic radiation of low density (within the range of validity of Wien's radiation formula) behaves thermodynamically as though it consisted of a number of independent energy quanta of magnitude  $R\beta\nu/N$ .

We still wish to compare the average magnitude of the energy quanta of the blackbody radiation with the average translational kinetic energy of a molecule at the same temperature. The latter is  $3/2(R/N)T$ , while, according to the Wien formula, one obtains for the average magnitude of an energy quantum

$$\int_0^{\infty} \alpha\nu^3 e^{-\beta\nu/T} d\nu / \int_0^{\infty} \frac{N}{R\beta\nu} \alpha\nu^3 e^{-\beta\nu/T} d\nu = 3(RT/N).$$

If the entropy of monochromatic radiation depends on volume as though the radiation were a discontinuous medium consisting of energy quanta of magnitude  $R\beta\nu/N$ , the next obvious step is to investigate whether the laws of emission and transformation of light are also of such a nature that they can be interpreted or explained by considering light to consist of such energy quanta. We shall examine this question in the following.

## 7. Concerning Stokes's Rule

According to the result just obtained, let us assume that, when monochromatic light is transformed through photoluminescence into light of a different

frequency, both the incident and emitted light consist of energy quanta of magnitude  $R\beta\nu/N$ , where  $\nu$  denotes the relevant frequency. The transformation process is to be interpreted in the following manner. Each incident energy quantum of frequency  $\nu_1$  is absorbed and generates by itself—at least at sufficiently low densities of incident energy quanta—a light quantum of frequency  $\nu_2$ ; it is possible that the absorption of the incident light quanta can give rise to the simultaneous emission of light quanta of frequencies  $\nu_3, \nu_4$  etc., as well as to energy of other kinds, e.g., heat. It does not matter what intermediate processes give rise to this final result. If the fluorescent substance is not a perpetual source of energy, the principle of conservation of energy requires that the energy of an emitted energy quantum cannot be greater than that of the incident light quantum; it follows that

$$R\beta\nu_2/N \leq R\beta\nu_1/N$$

or

$$\nu_2 \leq \nu_1.$$

This is the well-known Stokes's Rule.

It should be strongly emphasized that according to our conception the quantity of light emitted under conditions of low illumination (other conditions remaining constant) must be proportional to the strength of the incident light, since each incident energy quantum will cause an elementary process of the postulated kind, independently of the action of other incident energy quanta. In particular, there will be no lower limit for the intensity of incident light necessary to excite the fluorescent effect.

According to the conception set forth above, deviations from Stokes's Rule are conceivable in the following cases:

1. when the number of simultaneously interacting energy quanta per unit volume is so large that an energy quantum of emitted light can receive its energy from several incident energy quanta;
2. when the incident (or emitted) light is not of such a composition that it corresponds to blackbody radiation within the range of validity of Wien's Law, that is to say, for example, when the incident light is produced by a body of such high temperature that for the wavelengths under consideration Wien's Law is no longer valid.

The last-mentioned possibility commands especial interest. According to the conception we have outlined, the possibility is not excluded that a “non-Wien radiation” of very low density can exhibit an energy behavior different from that of a blackbody radiation within the range of validity of Wien's Law.

## 8. Concerning the Emission of Cathode Rays Through Illumination of Solid Bodies

The usual conception that the energy of light is continuously distributed over the space through which it propagates, encounters very serious difficulties when one attempts to explain the photoelectric phenomena, as has been pointed out in Herr Lenard's pioneering paper.<sup>7</sup>

According to the concept that the incident light consists of energy quanta of magnitude  $R\beta\nu/N$ , however, one can conceive of the ejection of electrons by light in the following way. Energy quanta penetrate into the surface layer of the body, and their energy is transformed, at least in part, into kinetic energy of electrons. The simplest way to imagine this is that a light quantum delivers its entire energy to a single electron: we shall assume that this is what happens. The possibility should not be excluded, however, that electrons might receive their energy only in part from the light quantum.

An electron to which kinetic energy has been imparted in the interior of the body will have lost some of this energy by the time it reaches the surface. Furthermore, we shall assume that in leaving the body each electron must perform an amount of work  $P$  characteristic of the substance. The ejected electrons leaving the body with the largest normal velocity will be those that were directly at the surface. The kinetic energy of such electrons is given by

$$R\beta\nu/N - P.$$

In the body is charged to a positive potential  $\Pi$  and is surrounded by conductors at zero potential, and if  $\Pi$  is just large enough to prevent loss of electricity by the body, it follows that:

$$\Pi\epsilon = R\beta\nu/N - P$$

where  $\epsilon$  denotes the electronic charge, or

$$\Pi E = R\beta\nu - P'$$

where  $E$  is the charge of a gram equivalent of a monovalent ion and  $P'$  is the potential of this quantity of negative electricity relative to the body.<sup>8</sup>

<sup>7</sup>P. Lenard, Ann. Phys., 8, 169, 170 (1902).

<sup>8</sup>If one assumes that the individual electron is detached from a neutral molecule by light with the performance of a certain amount of work, nothing in the relation derived above need be changed; one can simply consider  $P'$  as the sum of two terms.

If one takes  $E = 9.6 \times 10^3$ , then  $\Pi \cdot 10^{-8}$  is the potential in volts which the body assumes when irradiated in a vacuum.

In order to see whether the derived relation yields an order of magnitude consistent with experience, we take  $P' = 0$ ,  $\nu = 1.03 \times 10^{15}$  (corresponding to the limit of the solar spectrum toward the ultraviolet) and  $\beta = 4.866 \times 10^{-11}$ . We obtain  $\Pi \cdot 10^7 = 4.3$  volts, a result agreeing in order magnitude with those of Herr Lenard.<sup>9</sup>

If the derived formula is correct, then  $\Pi$ , when represented in Cartesian coordinates as a function of the frequency of the incident light, must be a straight line whose slope is independent of the nature of the emitting substance.

As far as I can see, there is no contradiction between these conceptions and the properties of the photoelectric observed by Herr Lenard. If each energy quantum of the incident light, independently of everything else, delivers its energy of electrons, then the velocity distribution of the ejected electrons will be independent of the intensity of the incident light; on the other hand the number of electrons leaving the body will, if other conditions are kept constant, be proportional to the intensity of the incident light.<sup>10</sup>

Remarks similar to those made concerning hypothetical deviations from Stokes's Rule can be made with regard to hypothetical boundaries of validity of the law set forth above.

In the foregoing it has been assumed that the energy of at least some of the quanta of the incident light is delivered completely to individual electrons. If one does not make this obvious assumption, one obtains, in place of the last equation:

$$\Pi E + P' \leq R\beta\nu.$$

For fluorescence induced by cathode rays, which is the inverse process to the one discussed above, one obtains by analogous considerations:

$$\Pi E + P' \geq R\beta\nu.$$

In the case, of the substances investigated by Herr Lenard,  $PE$ <sup>11</sup> is always significantly greater than  $R\beta\nu$ , since the potential difference, which the cathode rays must traverse in order to produce visible light, amounts in some cases to hundreds and in others to thousands of volts.<sup>12</sup> It is therefore to

---

<sup>9</sup>P.Lenard, Ann. Phys. 8, pp. 163, 185, and Table I, Fig. 2 (1902).

<sup>10</sup>P. Lenard, Ref. 9, p. 150 and p. 166–168.

<sup>11</sup>Should be  $\Pi E$  (translator's note).

<sup>12</sup>P. Lenard, Ann. Phys., 12, 469 (1903).

be assumed that the kinetic energy of an electron goes into the production of many light energy quanta.

## 9. Concerning the Ionization of Gases by Ultraviolet Light Solid Bodies

We shall have to assume that, the ionization of a gas by ultraviolet light, an individual light energy quantum is used for the ionization of an individual gas molecule. From this follows immediately that the work of ionization (i.e., the work theoretically needed for ionization) of a molecule cannot be greater than the energy of an absorbed light quantum capable of producing this effect. If one denotes by  $J$  the (theoretical) work of ionization per gram equivalent, then it follows that:

$$R \beta\nu \geq J.$$

According to Lenard's measurements, however, the largest effective wavelength for air is approximately  $1.9 \times 10^{-5}$  cm: therefore:

$$R \beta\nu = 6.4 \cdot 10^{12} \text{ erg} \geq J.$$

An upper limit for the work of ionization can also be obtained from the ionization potentials of rarefied gases. according to J. Stark<sup>13</sup> the smallest observed ionization potentials for air (at platinum anodes) is about 10 V.<sup>14</sup> One therefore obtains  $9.6 \times 10^{12}$  as an upper limit for  $J$ , which is nearly equal to the value found above.

There is another consequence the experimental testing of which seems to me to be of great importance. If every absorbed light energy quantum ionizes a molecule, the following relation must obtain between the quantity of absorbed light  $L$  and the number of gram molecules of ionized gas  $j$ :

$$j = L/R\beta\nu.$$

If our conception is correct, this relationship must be valid for all gases which (at the relevant frequency) show no appreciable absorption without ionization.

---

<sup>13</sup>J. Stark, Die Electrizität in Gasen (Leipzig, 1902, p. 57)

<sup>14</sup>In the interior of gases the ionization potential for negative ions is, however, five times greater.



## On the Electrodynamics of Moving Bodies

A. Einstein

Received June 30, 1905

— — ◇ ◇ — —

Translation into English: H. Lorentz, A. Einstein, H. Minkowsky, "The Principle of Relativity," Methuen, London, (1923) 35.

— — ◇ ◇ — —

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.<sup>1</sup> 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 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.<sup>2</sup> 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

---

<sup>1</sup>The preceding memoir by Lorentz was not at this time known to the author.

<sup>2</sup>i.e. to the first approximation.

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." <sup>3</sup>

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. 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"  $t_A$  from *A* towards *B*, let it at the "*B* time" to be reflected at *B* in the direction of *A*, and arrive again at *A* at the "*A* time"  $t'_B$ .

In accordance with definition the two clocks synchronize if

$$t_B - t_A = t'_A - t_B.$$

---

<sup>3</sup>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 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

$$\frac{2AB}{t'_A - t_A} = c$$

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 coordinates 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 <sup>4</sup> $t_A$  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

$$t_B - t_A = \frac{\tau_{AB}}{c-v} \quad \text{and} \quad t'_A - t_B = \frac{\tau_{AB}}{c+v}$$

---

<sup>4</sup>"Time" here denotes "time of the stationary system" and also "position of hands of the moving clock situated at the place under discussion."

where  $\tau_{AB}$  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  $\xi, \eta, \zeta$  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  $\tau$  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  $\xi, \eta, \zeta, \tau$ , 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  $\tau$  as a function of  $x', y, z$ , and  $t$ . To do this we have to express in equations that  $\tau$  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  $\tau_0$  along the  $X$ -axis to  $x'$ , and at the time  $\tau_1$  be reflected thence to the origin of the co-ordinates, arriving there at the time  $\tau_2$ ; we then must have  $\frac{1}{2}(\tau_0 + \tau_2) = \tau_1$ , or, by inserting the arguments of the function  $\tau$  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

$$\frac{\partial \tau}{\partial y} = 0, \quad \frac{\partial \tau}{\partial z} = 0.$$

Since  $\tau$  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  $\xi, \eta, \zeta$  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  $\xi$

$$\xi = c\tau, \quad \text{or} \quad \xi = ac \left( t - \frac{v}{c^2 - v^2} x' \right)$$

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  $\xi$ , 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

$$\frac{y}{\sqrt{c^2 - v^2}} = t, \quad x' = 0;$$

Thus

$$\eta = a \frac{c}{\sqrt{c^2 - v^2}} y \quad \text{and} \quad \xi = a \frac{c}{\sqrt{c^2 - v^2}} z.$$

Substituting for  $x'$  its value, we obtain

$$\tau = \phi(v)\beta \left( t - vx/c^2 \right),$$

$$\xi = \phi(v)\beta(x - vt),$$

$$\eta = \phi(v)y,$$

$$\xi = \phi(v)z,$$

where

$$\beta = \frac{1}{\sqrt{1 - (v/c)^2}},$$

and  $\phi$  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  $\tau$ , 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.<sup>5</sup>

In the equations of transformation which have been developed there enters an unknown function  $\phi$  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  $t = x = y = z = 0$  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

$$\begin{aligned} t' &= \phi(-v) \beta(-v) (\tau + v\xi/c^2) = \phi(v)\phi(-v)t, \\ x' &= \phi(-v)\beta(-v)(\xi + v\tau) = \phi(v)\phi(-v)x, \\ y' &= \phi(-v)\eta = \phi(v)\phi(-v)y, \end{aligned}$$

---

<sup>5</sup>The equations of the Lorentz transformation may be more simply deduced directly from the condition that in virtue of those equations the relation  $x^2 + y^2 + z^2 = c^2 t^2$  shall have as its consequence the second relation  $\xi^2 + \eta^2 + \zeta^2 = c^2 \tau^2$ .

$$z' = \phi(-v)\zeta = \phi(v)\phi(-v)z.$$

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  $\xi = l$ ,  $\eta = 0$ ,  $\zeta = 0$ . This part of the axis of  $Y$  is a rod moving perpendicularly to its axis with velocity  $v$  relatively to system  $K$ . Its ends possess in  $K$  the co-ordinates

$$x_1 = vt, \quad y_1 = \frac{l}{\phi(v)}, \quad z_1 = 0$$

and

$$x_2 = vt, \quad y_2 = 0, \quad z_2 = 0.$$

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

$$\tau = \beta \left( t - vx/c^2 \right),$$

$$\xi = \beta (x - vt),$$

$$\eta = y, \quad \zeta = z,$$

where

$$\beta = 1/\sqrt{1 - (v/c)^2}.$$

## § 4. Physical Meaning of the Equations Obtained in Respect to Moving Rigid Bodies and Moving Clocks

We envisage a rigid sphere<sup>6</sup> 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

$$\xi^2 + \eta^2 + \zeta^2 = R^2.$$

The equation of this surface expressed in  $x, y, z$  at the time  $t = 0$  is

$$\frac{x^2}{(\sqrt{1 - (v/c)^2})^2} + y^2 + z^2 = R^2.$$

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

$$R\sqrt{(1 - v^2/c^2)}, \quad R, \quad R.$$

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 plain 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  $\tau$  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  $\tau$ . What is the rate of this clock, when viewed from the stationary system?

Between the quantities  $x, t$ , and  $\tau$ , which refer to the position of the clock, we have, evidently,  $x = vt$  and

$$\tau = \frac{1}{\sqrt{1 - (v/c)^2}} (t - vx/c^2)$$

---

<sup>6</sup>That is, a body possessing spherical form when examined at rest.

Therefore,

$$\tau = t \sqrt{1 - (v/c)^2} = t - \left(1 - \sqrt{1 - (v/c)^2}\right) t,$$

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  $1/2 v^2/c^2$ .

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  $1/2tv^2/c^2$  (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 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  $1/2tv^2/c^2$  second slow. Thence we conclude that a balance-clock <sup>7</sup> 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

$$\xi = w_\xi \tau, \quad \eta = w_\eta \tau, \quad \zeta = 0,$$

where  $w_\xi$  and  $w_\eta$  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

$$x = \frac{w_\xi + v}{1 + vw_\xi/c^2} t,$$

---

<sup>7</sup>Not a pendulum-clock, which is physically a system to which the Earth belongs. This case had to be excluded.

$$y = \frac{\sqrt{1 - (v/c)^2}}{1 + vw\xi/c^2} w_\eta t,$$

$$z = 0.$$

Thus the law of the parallelogram of velocities is valid according to our theory only to a first approximation. We set

$$V^2 = \left( \frac{dx}{dt} \right)^2 + \left( \frac{dy}{dt} \right)^2,$$

$$w^2 = w_\xi^2 + w_\eta^2,$$

$$\alpha = \tan^{-1} w_y/w_x,$$

$\alpha$  is then to be looked upon as the angle between the velocities  $v$  and  $w$ . After a simple calculation we obtain

$$V = \frac{\sqrt{[(v^2 + w^2 + 2vw \cos \alpha) - (vw \sin \alpha/c^2)^2]}}{1 + vw \cos \alpha/c^2}.$$

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  $v = c - \kappa$ ,  $w = c - \lambda$ ,  $\kappa$  and  $\lambda$  being positive and less than  $c$ , then

$$V = c \frac{2c - \kappa - \lambda}{2c - \kappa - \lambda + \kappa\lambda/c} < V.$$

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

$$\frac{v + w}{1 + cw/c^2};$$

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

$$\begin{aligned} \frac{1}{c} \frac{\partial X}{\partial t} &= \frac{\partial N}{\partial y} - \frac{\partial M}{\partial z}, & \frac{1}{c} \frac{\partial L}{\partial t} &= \frac{\partial Y}{\partial z} - \frac{\partial Z}{\partial y}, \\ \frac{1}{c} \frac{\partial Y}{\partial t} &= \frac{\partial L}{\partial z} - \frac{\partial N}{\partial x}, & \frac{1}{c} \frac{\partial M}{\partial t} &= \frac{\partial Z}{\partial x} - \frac{\partial X}{\partial z}, \\ \frac{1}{c} \frac{\partial Z}{\partial t} &= \frac{\partial M}{\partial x} - \frac{\partial L}{\partial y}, & \frac{1}{c} \frac{\partial N}{\partial t} &= \frac{\partial X}{\partial y} - \frac{\partial Y}{\partial x}. \end{aligned}$$

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

$$\begin{aligned} \frac{1}{c} \frac{\partial X}{\partial \tau} &= \frac{\partial}{\partial \eta} \left\{ \beta \left( N - \frac{v}{c} Y \right) \right\} - \frac{\partial}{\partial \zeta} \left\{ \beta \left( M + \frac{v}{c} Y \right) \right\}, \\ \frac{1}{c} \frac{\partial}{\partial \tau} \left\{ \beta \left( Y - \frac{v}{c} N \right) \right\} &= \frac{\partial L}{\partial \xi} - \frac{\partial}{\partial \zeta} \left\{ \beta \left( N - \frac{v}{c} Z \right) \right\}. \end{aligned}$$

$$\begin{aligned}\frac{1}{c} \frac{\partial}{\partial \tau} \left\{ \beta \left( Z - \frac{v}{c} M \right) \right\} &= \frac{\partial}{\partial \xi} \left\{ \beta \left( M - \frac{v}{c} Z \right) \right\} - \frac{\partial L}{\partial \eta}, \\ \frac{1}{c} \frac{\partial L}{\partial \tau} &= \frac{\partial}{\partial \zeta} \left\{ \beta \left( Y - \frac{v}{c} N \right) \right\} - \frac{\partial}{\partial \eta} \left\{ \beta \left( Z + \frac{v}{c} M \right) \right\}, \\ \frac{1}{c} \frac{\partial}{\partial \tau} \left\{ \beta \left( M + \frac{v}{c} Z \right) \right\} &= \frac{\partial}{\partial \xi} \left\{ \beta \left( Z + \frac{v}{c} M \right) \right\} - \frac{\partial X}{\partial \zeta}, \\ \frac{1}{c} \frac{\partial}{\partial \tau} \left\{ \beta \left( N - \frac{v}{c} Y \right) \right\} &= \frac{\partial X}{\partial \eta} - \frac{\partial}{\partial \xi} \left\{ \beta \left( Y - \frac{v}{c} N \right) \right\},\end{aligned}$$

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: –

$$\begin{aligned}\frac{1}{c} \frac{\partial X'}{\partial \tau} &= \frac{\partial N'}{\partial \eta} - \frac{\partial M'}{\partial \zeta}, & \frac{1}{c} \frac{\partial L'}{\partial \tau} &= \frac{\partial Y'}{\partial \zeta} - \frac{\partial Z'}{\partial \eta}, \\ \frac{1}{c} \frac{\partial Y'}{\partial \tau} &= \frac{\partial L'}{\partial \zeta} - \frac{\partial N'}{\partial \xi}, & \frac{1}{c} \frac{\partial M'}{\partial \tau} &= \frac{\partial Z'}{\partial \xi} - \frac{\partial X'}{\partial \zeta}, \\ \frac{1}{c} \frac{\partial Z'}{\partial \tau} &= \frac{\partial M'}{\partial \xi} - \frac{\partial L'}{\partial \eta}, & \frac{1}{c} \frac{\partial N'}{\partial \tau} &= \frac{\partial X'}{\partial \eta} - \frac{\partial Y'}{\partial \xi}.\end{aligned}$$

Evidently the two systems of equations found for system 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 the system of equations, and is independent of  $\xi, \eta, \zeta$  and  $\tau$  it 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, mostly by solving the equations just obtained, and secondly applying the equations

to the inverse transformation (from 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<sup>8</sup>  $\psi(v) = \psi(-v)$ , and therefore

$$\psi(v) = 1,$$

and our equations assume the form

$$\begin{aligned} X' &= X, & L' &= L, \\ Y' &= \beta(Y - \frac{v}{c}N), & M' &= \beta(M + \frac{v}{c}Z), \\ Z' &= \beta(Z + \frac{v}{c}M), & N' &= \beta(N - \frac{v}{c}Y). \end{aligned}$$

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, which owes its introduction to the circumstance that electric and

---

<sup>8</sup>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 numerical value,  $Y'$  must also change sign without changing its numerical???

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

$$X = X_0 \sin \Phi, \quad L = L_0 \sin \Phi,$$

$$Y = Y_0 \sin \Phi, \quad M = M_0 \sin \Phi,$$

$$Z = Z_0 \sin \Phi, \quad N = N_0 \sin \Phi,$$

where

$$\Phi = \omega \left\{ t - \frac{1}{c} (lx + my + nz) \right\}.$$

Here  $(X_0, Y_0, Z_0)$  and  $(L_0, M_0, N_0)$  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

$$\begin{aligned} X' &= X_0 \sin \Phi', & L' &= L_0 \sin \Phi', \\ Y' &= \beta (Y_0 - vN_0/c) \sin \Phi', & M' &= \beta (M_0 + vZ_0/c) \sin \Phi', \\ Z' &= \beta (Z_0 + vM_0/c) \sin \Phi', & N' &= \beta (N_0 - vY_0/c) \sin \Phi', \end{aligned}$$

$$\Phi' = \omega' \left\{ \tau - \frac{1}{c} (l'\xi + m'\eta + n'\zeta) \right\}$$

where

$$\omega' = \omega \beta (1 - lv/c),$$

$$l' = \frac{l - v/c}{1 - lv/c},$$

$$m' = \frac{m}{\beta(1 - lv/c)},$$

$$n' = \frac{n}{\beta(1 - lv/c)}.$$

From the equation for  $\omega'$  it follows that if an observer is moving with velocity  $v$  relatively to an infinitely distant source of light of frequency  $\nu$ , in such a way that the connecting line “sourceobserver” makes the angle ( $\phi$ ) 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  $\nu'$  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 “sourceobserver”  $\phi'$ , 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  $\phi = 1/2\pi$ , the equation becomes simply

$$\cos \phi' = -v,$$

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

$$A'^2 = A^2 \frac{(1 - \cos \phi \cdot v/c)^2}{1 - v^2/c^2}$$

which equation, if  $\phi = 0$ , simplifies into

$$A'^2 = A^2 \frac{1 - v/c}{1 + v/c}.$$

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  $A^2/8\pi$  equals the energy of light per unit of volume, we have to regard  $A'^2/8\pi$ , 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: –

$$(x - let)^2 + (y - mct)^2 + (z - nct)^2 = R^2.$$

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

$$(\beta\xi - l\beta\xi v/c)^2 + (\eta - m\beta\xi v/c)^2 + (\zeta - n\beta\xi v/c)^2 = R^2.$$

If  $S$  is the volume of the sphere, and  $S'$  that of this ellipsoid, then by a simple calculation

$$\frac{S'}{S} = \frac{\sqrt{1 - v^2/c^2}}{1 - \cos \phi \cdot v/c}.$$

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

$$\frac{E'}{E} = \frac{A'^2 S'}{A^2 S} = \frac{1 - \cos \phi \cdot v/c}{\sqrt{(1 - v^2/c^2)}},$$

and this formula, when  $\phi = 0$ , simplifies into

$$\frac{E'}{E} = \sqrt{\frac{1-v/c}{1+v/c}}.$$

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  $A, \cos \phi, \nu$  (referred to system  $K$ ). Viewed from  $k$  the corresponding quantities are

$$\begin{aligned} A' &= A \frac{1 - \cos \phi \cdot v/c}{\sqrt{(1 - v^2/c^2)}}, \\ \cos \phi' &= \frac{\cos \phi - v/c}{1 - \cos \phi \cdot v/c}, \\ \nu' &= \nu \frac{1 - \cos \phi \cdot v/c}{\sqrt{1 - v^2/c^2}}. \end{aligned}$$

For the reflected light, referring the process to system  $k$ , we obtain

$$\begin{aligned} A'' &= A' \\ \cos \phi'' &= -\cos \phi' \\ \nu'' &= \nu' \end{aligned}$$

Finally, by transforming back to the stationary system  $K$ , we obtain for the reflected light

$$\begin{aligned} A''' &= A'' \frac{1 + \cos \phi'' \cdot v/c}{\sqrt{(1 - v^2/c^2)}} = A \frac{1 - 2 \cos \phi \cdot v/c + v^2/c^2}{1 - v^2/c^2}, \\ \cos \phi''' &= \frac{\cos \phi'' + v/c}{1 + \cos \phi'' \cdot v/c} = -\frac{(1 + v^2/c^2) \cos \phi - 2v/c}{1 - 2 \cos \phi \cdot v/c + v^2/c^2} \\ \nu''' &= \nu'' \frac{1 + \cos \phi'' v/c}{\sqrt{(1 - v^2/c^2)}} = \nu \frac{1 - 2 \cos \phi \cdot v/c + v^2/c^2}{1 - v^2/c^2}. \end{aligned}$$

The energy (measured in the stationary system) which is incident upon unit area of the mirror in unit time is evidently  $A^2(c \cos \phi - v)/8\pi$ . The energy

leaving the unit of surface of the mirror in the unit of time is  $A'''^2(-c \cos \phi''' + v)/8\pi$ . 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

$$P = 2 \cdot \frac{A^2}{8\pi} \frac{(\cos \phi - v/c)^2}{1 - v^2/c^2}.$$

In agreement with experiment and with other theories, we obtain to a first approximation

$$P = 2 \cdot \frac{A^2}{8\pi} \cos^2 \phi.$$

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

$$\begin{aligned} \frac{1}{c} \left\{ u_x \rho + \frac{\partial X}{\partial t} \right\} &= \frac{\partial N}{\partial y} - \frac{\partial M}{\partial z}, & \frac{1}{c} \frac{\partial L}{\partial t} &= \frac{\partial Y}{\partial z} - \frac{\partial Z}{\partial y}, \\ \frac{1}{c} \left\{ u_y \rho + \frac{\partial Y}{\partial t} \right\} &= \frac{\partial L}{\partial z} - \frac{\partial M}{\partial x}, & \frac{1}{c} \frac{\partial M}{\partial t} &= \frac{\partial Z}{\partial x} - \frac{\partial X}{\partial z}, \\ \frac{1}{c} \left\{ u_z \rho + \frac{\partial Z}{\partial t} \right\} &= \frac{\partial M}{\partial x} - \frac{\partial L}{\partial y}, & \frac{1}{c} \frac{\partial N}{\partial t} &= \frac{\partial X}{\partial y} - \frac{\partial Y}{\partial x}, \end{aligned}$$

where

$$\rho = \frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} + \frac{\partial Z}{\partial z}$$

denotes  $4\pi$  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

$$\begin{aligned}\frac{1}{c} \left\{ u_\xi \rho' + \frac{\partial X'}{\partial \tau} \right\} &= \frac{\partial N'}{\partial \eta} - \frac{\partial M}{\partial \xi}, \quad \frac{1}{c} \frac{\partial L'}{\partial \tau} = \frac{\partial Y'}{\partial \zeta} - \frac{\partial Z'}{\partial \eta}, \\ \frac{1}{c} \left\{ u_\eta \rho' + \frac{\partial Y'}{\partial \tau} \right\} &= \frac{\partial L}{\partial \zeta} - \frac{\partial N'}{\partial \xi}, \quad \frac{1}{c} \frac{\partial M'}{\partial \tau} = \frac{\partial Z'}{\partial \xi} - \frac{\partial X'}{\partial \zeta}, \\ \frac{1}{c} \left\{ u_\zeta \rho' + \frac{\partial Z'}{\partial \tau} \right\} &= \frac{\partial M'}{\partial \xi} - \frac{\partial L'}{\partial \eta}, \quad \frac{1}{c} \frac{\partial N'}{\partial \tau} = \frac{\partial X'}{\partial \eta} - \frac{\partial Y'}{\partial \xi},\end{aligned}$$

where

$$\begin{aligned}u_\xi &= \frac{u_x - v}{1 - u_x v / c^2} \\ u_\eta &= \frac{u_y}{\beta(1 - u_x v / c^2)} \\ u_\zeta &= \frac{u_z}{\beta(1 - u_x v / c^2)},\end{aligned}$$

and

$$\rho' = \frac{\partial X'}{\partial \xi} + \frac{\partial Y'}{\partial \eta} + \frac{\partial Z'}{\partial \zeta} = \beta (1 - u_x v / c^2) \rho.$$

Sinceas follows from the theorem of addition of velocities (§5)the vector  $(u_\xi, u_\eta, u_\zeta)$  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 remainswhen 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

$$m \frac{d^2x}{dt^2} = \epsilon X$$

$$m \frac{d^2y}{dt^2} = \epsilon Y$$

$$m \frac{d^2z}{dt^2} = \epsilon Z$$

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

$$m \frac{d^2\xi}{d\tau^2} = \epsilon X',$$

$$m \frac{d^2\eta}{d\tau^2} = \epsilon Y',$$

$$m \frac{d^2\zeta}{d\tau^2} = \epsilon Z',$$

in which the symbols  $\xi, \eta, \tau, X', Y', Z'$  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

$$\xi = \beta (x - vt), \quad \eta = y, \quad \zeta = z, \quad \tau = \beta (t - vx/c^2)$$

$$X' = X, Y' = \beta (Y - vN/c), \quad Z' = \beta (Z + vM/c).$$

With the help of these equations we transform the above equations of motion from system  $k$  to system  $K$ , and obtain

$$\left. \begin{aligned} \frac{d^2x}{dt^2} &= \frac{\epsilon}{m\beta^3} X, \\ \frac{d^2y}{dt^2} &= \frac{\epsilon}{m\beta} \left( Y - \frac{v}{c} N \right), \\ \frac{d^2z}{dt^2} &= \frac{\epsilon}{m\beta} \left( Z + \frac{v}{c} M \right). \end{aligned} \right\} \quad (A)$$

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

$$\begin{aligned} m \beta^3 \frac{d^2x}{dt^2} &= \epsilon X = \epsilon X', \\ m \beta^2 \frac{d^2y}{dt^2} &= \epsilon \beta \left( Y - \frac{v}{c} N \right) = \epsilon Y', \\ m \beta^2 \frac{d^2z}{dt^2} &= \epsilon \beta \left( Z + \frac{v}{c} M \right) = \epsilon Z'. \end{aligned}$$

and remark firstly that  $\epsilon X'$ ,  $\epsilon Y'$ ,  $\epsilon Z'$  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,”<sup>9</sup> and maintain the equation - mass  $\times$  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

---

<sup>9</sup>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.

electron (in our sense of the word) by the addition of an electric charge, *no matter how small*.

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  $\int \epsilon X dx$ . 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

$$W = \int \epsilon X dx = \int_0^v \beta^3 \mu v dv = mc^2 \left\{ \frac{1}{\sqrt{1 - (v^2/c^2)}} - 1 \right\}.$$

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  $Y = Nv/c$ . 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

$$\frac{A_m}{A_e} = \frac{v}{c}.$$

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

$$P = \int X dx = \frac{m}{\epsilon} c^2 \left\{ \frac{1}{\sqrt{1 - (v^2/c^2)}} - 1 \right\}.$$

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

$$-\frac{d^2y}{dt^2} = \frac{v^2}{R} = \frac{\epsilon}{m} \frac{v}{c} N \sqrt{1 - \frac{v^2}{c^2}},$$

or

$$R = \frac{mc^2}{\epsilon} \cdot \frac{v/c}{\sqrt{1 - (v^2/c^2)}} \cdot \frac{1}{N}.$$

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.

## Does the Inertia of a Body Depend on its Energy Content?

A. Einstein  
Bern  
(Received 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<sup>1</sup> as my basis I deduced inter alia the following result:

Let a system of plane waves of light, referred to the system of coordinates  $(x, y, z)$ , possess the energy  $l$ ; let the direction of the ray (the wave-normal) make an angle  $\phi$  with the axis of  $x$  of the system. If we introduce a new system of co-ordinates  $(\xi, \eta, \zeta)$  moving in uniform parallel translation with respect to the system  $(x, y, z)$ , and having its origin of coordinates in motion along the axis of  $x$  with the velocity  $v$ , then this quantity of light-measured in the system  $(\xi, \eta, \zeta)$  — possesses the energy

$$l^* = l \cdot \frac{1 - \frac{v}{V} \cos \phi}{\sqrt{1 - (v/V)^2}},$$

where  $c$  denotes the velocity of light. We shall make use of this result in what follows.

---

<sup>1</sup>The principle of the constancy of the velocity of light is of course contained in Maxwell's equations.

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  $(\xi, \eta, \zeta)$ , moving as above with the velocity  $v$ , be  $H_0$ .

Let this body send out, in a direction making an angle  $\phi$  with the axis of  $x$ , plane waves of light, of energy  $L/2$  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 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  $(\xi, \eta, \zeta)$  respectively, then by employing the relation given above we obtain

$$E_0 = E_1 + \left( \frac{L}{2} + \frac{L}{2} \right),$$

$$H_0 = H_1 + \left[ \frac{L}{2} \cdot \frac{1 - \frac{v}{V} \cos \phi}{\sqrt{1 - (v/V)^2}} + \frac{L}{2} \cdot \frac{1 + \frac{v}{V} \cos \phi}{\sqrt{1 - (v/V)^2}} \right] = H_1 + \frac{L}{\sqrt{1 - (v/V)^2}}.$$

By subtraction we obtain from these equations

$$(H_0 - E_0) - (H_1 - E_1) = L \cdot \left\{ \frac{1}{\sqrt{1 - (v/V)^2}} - 1 \right\}.$$

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  $(\xi, \eta, \zeta)$ , 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

$$H_0 - E_0 = K_0 + C,$$

$$H_1 - E_1 = K_1 + C,$$

since  $C$  does not change during the emission of light. So we have

$$K_0 - K_1 = L \cdot \left\{ \frac{1}{\sqrt{1 - (v/V)^2}} - 1 \right\}.$$

The kinetic energy of the body with respect to  $(\xi, \eta, \zeta)$  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, depends on the velocity.

Neglecting magnitudes of fourth and higher orders we may place

$$K_0 - K_1 = \frac{L}{V^2} \cdot \frac{v^2}{2}.$$

From this equation it directly follows that:

If a body gives off the energy  $L$  in the form of radiation, its mass diminishes by  $L/V^2$ . 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  $L/9 \times 10^{20}$ , 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.



Doc. 47

ON THE RELATIVITY PRINCIPLE AND THE CONCLUSIONS DRAWN FROM IT

by A. Einstein

[*Jahrbuch der Radioaktivität und Elektronik* 4 (1907): 411-462]

Newton's equations of motion retain their form when one transforms to a new system of coordinates that is in uniform translational motion relative to the system used originally according to the equations

$$x' = x - vt$$

$$x' = y$$

$$z' = z .$$

As long as one believed that all of physics can be founded on Newton's equations of motion, one therefore could not doubt that the laws of nature are the same without regard to which of the coordinate systems moving uniformly (without acceleration) relative to each other they are referred. However, this independence from the state of motion of the system of coordinates used, which we will call "the principle of relativity," seemed to have been suddenly called into question by the brilliant confirmations of H. A. Lorentz's electrodynamics of moving bodies.<sup>1</sup> That theory is built on the presupposition of a resting, immovable, luminiferous ether; its basic equations are not such that they transform to equations of the same form when the above transformation equations are applied.

After the acceptance of that theory, one had to expect that one would succeed in demonstrating an effect of the terrestrial motion relative to the luminiferous ether on optical phenomena. It is true that in the study cited Lorentz proved that in optical experiments, as a consequence of his basic assumptions, an effect of that relative motion on the ray path is not to be expected as long as the calculation is limited to terms in which the ratio

<sup>1</sup>H. A. Lorentz, *Versuch einer Theorie der elektrischen und optischen Erscheinungen in bewegten Körpern*. [Attempt at a theory of electric and optical phenomena in moving bodies] Leiden, 1895. Reprinted Leipzig, 1906.

$$dE = F_x dx + F_y dy + F_z dz - pdV + TdS \quad (28) \quad [90]$$

$$F_x = \frac{dG_x}{dt}, \text{ etc.} \quad (29)$$

Keeping in mind that

$$F_x dx = F_x \dot{x} dt = \dot{x} dG = d(\dot{x} G_x) - G_x d\dot{x}, \text{ etc.} \quad [91]$$

and

$$Td\eta = d(T\eta) - \eta dT,$$

one obtains from the above equations the relation

$$d(-E + T\eta + qG) = G_x d\dot{x} + G_y d\dot{y} + G_z d\dot{z} + pdV + \eta dT.$$

Since the right-hand side of this equation must also be a total differential, and taking into account (29), it follows that

$$\begin{aligned} \frac{d}{dt} \left[ \frac{\partial H}{\partial \dot{x}} \right] &= F_x & \frac{d}{dt} \left[ \frac{\partial H}{\partial \dot{y}} \right] &= F_y & \frac{d}{dt} \left[ \frac{\partial H}{\partial \dot{z}} \right] &= F_z \\ \frac{\partial H}{\partial V} &= p & \frac{\partial H}{\partial T} &= \eta. \end{aligned}$$

But these are the equations derivable by means of the principle of least action which Mr. Planck had used as his starting point.

[92]

## V. PRINCIPLE OF RELATIVITY AND GRAVITATION

### §17. Accelerated reference system and gravitational field

So far we have applied the principle of relativity, i.e., the assumption that the physical laws are independent of the state of motion of the reference system, only to *nonaccelerated* reference systems. Is it conceivable that the principle of relativity also applies to systems that are accelerated relative to each other?

While this is not the place for a detailed discussion of this question, it will occur to anybody who has been following the applications of the principle of relativity. Therefore I will not refrain from taking a stand on this question here.

We consider two systems  $\Sigma_1$  and  $\Sigma_2$  in motion. Let  $\Sigma_1$  be accelerated in the direction of its  $X$ -axis, and let  $\gamma$  be the (temporally constant) magnitude of that acceleration.  $\Sigma_2$  shall be at rest, but it shall be located in a homogeneous gravitational field that imparts to all objects an acceleration  $-\gamma$  in the direction of the  $X$ -axis.

As far as we know, the physical laws with respect to  $\Sigma_1$  do not differ from those with respect to  $\Sigma_2$ ; this is based on the fact that all bodies are equally accelerated in the gravitational field. At our present state of experience we have thus no reason to assume that the systems  $\Sigma_1$  and  $\Sigma_2$  differ from each other in any respect, and in the discussion that follows, we shall therefore assume the complete physical equivalence of a gravitational field and a corresponding acceleration of the reference system.

This assumption extends the principle of relativity to the uniformly accelerated translational motion of the reference system. The heuristic value of this assumption rests on the fact that it permits the replacement of a homogeneous gravitational field by a uniformly accelerated reference system, the latter case being to some extent accessible to theoretical treatment.

### *§18. Space and time in a uniformly accelerated reference system*

We first consider a body whose individual material points, at a given time  $t$  of the nonaccelerated reference system  $S$ , possess no velocity relative to  $S$ , but a certain acceleration. What is the influence of this acceleration  $\gamma$  on the shape of the body with respect to  $S$ ?

If such an influence is present, it will consist of a constant-ratio dilatation in the direction of acceleration and possibly in the two directions perpendicular to it, since an effect of another kind is impossible for reasons of symmetry. The acceleration-caused dilatations (if such exist at all) must be even functions of  $\gamma$ ; hence they can be neglected if one restricts oneself to the case in which  $\gamma$  is so small that terms of the second or higher power

in  $\gamma$  may be neglected. Since we are going to restrict ourselves to that case, we do not have to assume that the acceleration has any influence on the shape of the body.

We now consider a reference system  $\Sigma$  that is uniformly accelerated relative to the nonaccelerated system  $S$  in the direction of the latter's  $X$ -axis. The clocks and measuring rods of  $\Sigma$ , examined at rest, shall be identical with the clocks and measuring rods of  $S$ . The coordinate origin of  $\Sigma$  shall move along the  $X$ -axis of  $S$ , and the axes of  $\Sigma$  shall be perpetually parallel to those of  $S$ . At any moment there exists a nonaccelerated reference system  $S'$  whose coordinate axes coincide with the coordinate axes of  $\Sigma$  at the moment in question (at a given time  $t'$  of  $S'$ ). If the coordinates of a point event occurring at this time  $t'$  are  $\xi$ ,  $\eta$ ,  $\zeta$  with respect to  $\Sigma$ , we will have

$$\left. \begin{array}{l} x' = \xi \\ y' = \eta \\ z' = \zeta \end{array} \right\},$$

because in accordance with what we said above, we are not to assume that acceleration affects the shape of the measuring instruments used for measuring  $\xi$ ,  $\eta$ ,  $\zeta$ . We shall also imagine that the clocks of  $\Sigma$  are set at time  $t'$  of  $S'$  such that their readings at that moment equal  $t'$ . What about the rate of the clocks in the next time element  $\tau$ ?

First of all, we have to bear in mind that a specific effect of acceleration on the rate of the clocks of  $\Sigma$  need not be taken into account, since it would have to be of the order  $\gamma^2$ . Furthermore, since the effect of the velocity attained during  $\tau$  on the rate of the clocks is negligible, and the distances traveled by the clocks during the time  $\tau$  relative to those traveled by  $S'$  are also of the order  $\tau^2$ , i.e., negligible, the readings of the clocks of  $\Sigma$  may be fully replaced by readings of the clocks of  $S'$  for the time element  $\tau$ .

[95]

From the foregoing it follows that, relative to  $\Sigma$ , light in vacuum is propagated during the time element  $\tau$  with the universal velocity  $c$  if we define simultaneity in the system  $S'$  which is momentarily at rest relative

to  $\Sigma$ , and if the clocks and measuring rods we use for measuring the time and length are identical with those used for the measurement of time and space in nonaccelerated systems. Thus the principle of constancy of the velocity of light can be used here too to define simultaneity if one restricts oneself to very short light paths.

We now imagine that the clocks of  $\Sigma$  are adjusted, in the way described, at that time  $t = 0$  of  $S$  at which  $\Sigma$  is instantaneously at rest relative to  $S$ . The totality of readings of the clocks of  $\Sigma$  adjusted in this way is called the "local time"  $\sigma$  of the system  $\Sigma$ . It is immediately evident that the physical meaning of the local time  $\sigma$  is as follows. If one uses the local time  $\sigma$  for the temporal evaluation of processes occurring in the individual space elements of  $\Sigma$ , then the laws obeyed by these processes cannot depend on the position of these space elements, i.e., on their coordinates, if not only the clocks, but also the other measuring tools used in the various space elements are identical.

However, we must not simply refer to the local time  $\sigma$  as the "time" of  $\Sigma$ , because according to the definition given above, two point events occurring at different points of  $\Sigma$  are not simultaneous when their local times  $\sigma$  are equal. For if at time  $t = 0$  two clocks of  $\Sigma$  are synchronous with respect to  $S$  and are subjected to the same motions, then they remain forever synchronous with respect to  $S$ . However, for this reason, in accordance with §4, they do not run synchronously with respect to a reference system  $S'$  instantaneously at rest relative to  $\Sigma$  but in motion relative to  $S$ , and hence according to our definition they do not run synchronously with respect to  $\Sigma$  either.

We now define the "time"  $\tau$  of the system  $\Sigma$  as the totality of those readings of the clock situated at the coordinate origin of  $\Sigma$  which are, according to the above definition, simultaneous with the events which are to be temporally evaluated.<sup>1</sup>

We shall now determine the relation between the time  $\tau$  and the local time  $\sigma$  of a point event. It follows from the first of equations (1) that

---

<sup>1</sup>Thus the symbol " $\tau$ " is used here in a different sense than above.

two events are simultaneous with respect to  $S'$ , and thus also with respect to  $\Sigma$ , if

$$t_1 - \frac{v}{c^2} x_1 = t_2 - \frac{v}{c^2} x_2 ,$$

where the subscripts refer to the one or to the other point event, respectively. We shall first confine ourselves to the consideration of times that are so short<sup>1</sup> that all terms containing the second or higher power of  $\tau$  or  $v$  can be omitted; taking (1) and (29) into account, we then have to put [98]

$$\begin{aligned} x_2 - x_1 &= x_2^1 - x_1^1 = \xi_2 - \xi_1 \\ t_1 &= \sigma_1 \quad t_2 = \sigma_2 \\ v &= \gamma t = \gamma \tau , \end{aligned}$$

[99]

so that we obtain from the above equation

$$\sigma_2 - \sigma_1 = \frac{\gamma \tau}{c^2} (\xi_2 - \xi_1) .$$

If we move the first point event to the coordinate origin, so that  $\sigma_1 = \tau$  and  $\xi_1 = 0$ , we obtain, omitting the subscript for the second point event,

$$\sigma = \tau \left[ 1 + \frac{\gamma \xi}{c^2} \right] . \quad (30)$$

This equation holds first of all if  $\tau$  and  $\xi$  lie below certain limits. It is obvious that it holds for arbitrarily large  $\tau$  if the acceleration  $\gamma$  is constant with respect to  $\Sigma$ , because the relation between  $\sigma$  and  $\tau$  must then be linear. Equation (30) does not hold for arbitrarily large  $\xi$ . From the fact that the choice of the coordinate origin must not affect the relation, one must conclude that, strictly speaking, equation (30) should be replaced by the equation

$$\sigma = \tau e^{\frac{\gamma \xi}{c^2}} .$$

Nevertheless, we shall maintain formula (30).

<sup>1</sup>In accordance with (1), we thereby also assume a certain restriction with respect to the values of  $\xi = x'$ .

According to §17, equation (30) is also applicable to a coordinate system in which a homogeneous gravitational field is acting. In that case we have to put  $\Phi = \gamma\xi$ , where  $\Phi$  is the gravitational potential, so that we obtain

$$\sigma = \tau \left[ 1 + \frac{\Phi}{c^2} \right]. \quad (30a)$$

We have defined two kinds of times for  $\Sigma$ . Which of the two definitions do we have to use in the various cases? Let us assume that at two locations of different gravitational potentials ( $\gamma\xi$ ) there exists one physical system each, and we want to compare their physical quantities. To do this, the most natural procedure might be as follows: First we take our measuring tools to the first physical system and carry out our measurements there; then we take our measuring tools to the second system to carry out the same measurement here. If the two sets of measurements give the same results, we shall denote the two physical systems as "equal." The measuring tools include a clock with which we measure local times  $\sigma$ . From this it follows that to define the physical quantities at some position of the gravitational field, it is natural to use the time  $\sigma$ .

However, if we deal with a phenomenon in which objects situated at positions with different gravitational potentials must be considered simultaneously, we have to use the time  $\tau$  in those terms in which time occurs explicitly (i.e., not only in the definition of physical quantities), because otherwise the simultaneity of the events would not be expressed by the equality of the time values of the two events. Since in the definition of the time  $\tau$  a clock situated in an arbitrarily chosen position is used, but not an arbitrarily chosen instant, when using time  $\tau$  the laws of nature can vary with position but not with time.

### *§19. The effect of the gravitational field on clocks*

If a clock showing local time is located in a point  $P$  of gravitational potential  $\Phi$ , then, according to (30a), its reading will be  $(1 + \frac{\Phi}{c^2})$  times greater than the time  $\tau$ , i.e., it runs  $(1 + \frac{\Phi}{c^2})$  times faster than an

identical clock located at the coordinate origin. Suppose an observer located somewhere in space perceives the indications of the two clocks in a certain way, e.g., optically. As the time  $\Delta\tau$  that elapses between the instants at which a clock indication occurs and at which this indication is perceived by the observer is independent of  $\tau$ , for an observer situated somewhere in space the clock in point  $P$  runs  $(1 + \frac{\Phi}{c^2})$  times faster than the clock at the coordinate origin. In this sense we may say that the process occurring in the clock, and, more generally, any physical process, proceeds faster the greater the gravitational potential at the position of the process taking place.

There exist "clocks" that are present at locations of different gravitational potentials and whose rates can be controlled with great precision; these are the producers of spectral lines. It can be concluded from the aforesaid<sup>1</sup> that the wave length of light coming from the sun's surface, which originates from such a producer, is larger by about one part in two millionth than that of light produced by the same substance on earth. [100]

## §20. *The effect of gravitation on electromagnetic phenomena*

If we refer an electromagnetic process at some point of time to a non-accelerated reference system  $S'$  that is instantaneously at rest relative to the reference system  $\Sigma$  accelerated as above, then the following equations will hold according to (5) and (6):

$$\frac{1}{c} \left[ \rho' u'_x + \frac{\partial X'}{\partial t'} \right] = \frac{\partial N'}{\partial y'} - \frac{\partial M'}{\partial z'}, \text{ etc.}$$

and

$$\frac{1}{c} \frac{\partial L'}{\partial t'} = \frac{\partial Y'}{\partial z'} - \frac{\partial Z'}{\partial y'}, \text{ etc.}$$

In accordance with the above, we may readily equate the  $S'$ -referred quantities  $\rho'$ ,  $u'$ ,  $X'$ ,  $L'$ ,  $x'$ , etc., with the corresponding  $\Sigma$ -referred

---

<sup>1</sup>While assuming that equation (30a) holds for an inhomogeneous gravitational field as well.

quantities  $\rho$ ,  $u$ ,  $X$ ,  $L$ ,  $\xi$ , etc., if we limit ourselves to an infinitesimally short period<sup>1</sup> that is infinitesimally close to the time of relative rest of  $S'$  and  $\Sigma$ . Further, we have to replace  $t'$  by the local time  $\sigma$ . However, we must not simply put

$$\frac{\partial}{\partial t'} = \frac{\partial}{\partial \sigma} ,$$

because a point which is at rest relative to  $\Sigma$ , and to which equations transformed to  $\Sigma$  should refer, changes its velocity relative to  $S'$  during the time element  $dt' = d\sigma$ , to which change, according to equations (7a) and (7b), there corresponds a temporal change of the  $\Sigma$ -related field component. Hence we have to put

$$\begin{array}{ll} \frac{\partial X'}{\partial t'} = \frac{\partial X}{\partial \sigma} & \frac{\partial L'}{\partial t'} = \frac{\partial L}{\partial \sigma} \\ \frac{\partial Y'}{\partial t'} = \frac{\partial Y}{\partial \sigma} + \frac{\gamma}{c} N & \frac{\partial M'}{\partial t'} = \frac{\partial M}{\partial \sigma} - \frac{\gamma}{c} Z \\ \frac{\partial Z'}{\partial t'} = \frac{\partial Z}{\partial \sigma} - \frac{\gamma}{c} M & \frac{\partial N'}{\partial t'} = \frac{\partial N}{\partial \sigma} + \frac{\gamma}{c} Y . \end{array}$$

Hence the  $\Sigma$ -referred electromagnetic equations are

$$\begin{aligned} \frac{1}{c} \left[ \rho u_\xi + \frac{\partial X}{\partial \sigma} \right] &= \frac{\partial N}{\partial \eta} - \frac{\partial M}{\partial \zeta} \\ \frac{1}{c} \left[ \rho u_\eta + \frac{\partial Y}{\partial \sigma} + \frac{\gamma}{c} N \right] &= \frac{\partial L}{\partial \zeta} - \frac{\partial M}{\partial \xi} \\ \frac{1}{c} \left[ \rho u_\xi + \frac{\partial Z}{\partial \sigma} - \frac{\gamma}{c} M \right] &= \frac{\partial M}{\partial \xi} - \frac{\partial L}{\partial \eta} \\ \frac{1}{c} \frac{\partial L}{\partial \sigma} &= \frac{\partial Y}{\partial \zeta} - \frac{\partial Z}{\partial \eta} \\ \frac{1}{c} \left[ \frac{\partial M}{\partial \sigma} - \frac{\gamma}{c} Z \right] &= \frac{\partial Z}{\partial \xi} - \frac{\partial X}{\partial \zeta} \\ \frac{1}{c} \left[ \frac{\partial N}{\partial \sigma} + \frac{\gamma}{c} Y \right] &= \frac{\partial X}{\partial \eta} - \frac{\partial Y}{\partial \xi} . \end{aligned}$$

---

<sup>1</sup>This restriction does not affect the range of validity of our results because inherently the laws to be derived cannot depend on the time.

We multiply these equations by  $\left[1 + \frac{\gamma\xi}{c^2}\right]$  and put for the sake of brevity

$$X^* = X\left[1 + \frac{\gamma\xi}{c^2}\right], \quad Y^* = Y\left[1 + \frac{\gamma\xi}{c^2}\right], \text{ etc.}$$

$$\rho^* = \rho\left[1 + \frac{\gamma\xi}{c^2}\right].$$

Neglecting terms of the second power in  $\gamma$ , we obtain the equations

$$\left. \begin{aligned} \frac{1}{c} \left[ \rho^* u_\xi + \frac{\partial X^*}{\partial \sigma} \right] &= \frac{\partial N^*}{\partial \eta} - \frac{\partial M^*}{\partial \zeta} \\ \frac{1}{c} \left[ \rho^* u_\eta + \frac{\partial Y^*}{\partial \sigma} \right] &= \frac{\partial L^*}{\partial \zeta} - \frac{\partial N^*}{\partial \xi} \\ \frac{1}{c} \left[ \rho^* u_\zeta + \frac{\partial Z^*}{\partial \sigma} \right] &= \frac{\partial M^*}{\partial \xi} - \frac{\partial L^*}{\partial \eta} \\ \frac{1}{c} \frac{\partial L^*}{\partial \sigma} &= \frac{\partial Y^*}{\partial \zeta} - \frac{\partial Z^*}{\partial \eta} \\ \frac{1}{c} \frac{\partial M^*}{\partial \sigma} &= \frac{\partial Z^*}{\partial \xi} - \frac{\partial X^*}{\partial \zeta} \\ \frac{1}{c} \frac{\partial N^*}{\partial \sigma} &= \frac{\partial X^*}{\partial \eta} - \frac{\partial Y^*}{\partial \xi} \end{aligned} \right\} \quad (31a)$$

$$\left. \begin{aligned} \frac{1}{c} \frac{\partial L^*}{\partial \sigma} &= \frac{\partial Y^*}{\partial \zeta} - \frac{\partial Z^*}{\partial \eta} \\ \frac{1}{c} \frac{\partial M^*}{\partial \sigma} &= \frac{\partial Z^*}{\partial \xi} - \frac{\partial X^*}{\partial \zeta} \\ \frac{1}{c} \frac{\partial N^*}{\partial \sigma} &= \frac{\partial X^*}{\partial \eta} - \frac{\partial Y^*}{\partial \xi} \end{aligned} \right\} \quad (32a)$$

These equations show first of all how the gravitational field affects the static and stationary phenomena. The same laws hold as in the gravitation-free field, except that the field components  $X$ , etc. are replaced by

$X\left[1 + \frac{\gamma\xi}{c^2}\right]$ , etc., and  $\rho$  is replaced by  $\rho\left[1 + \frac{\gamma\xi}{c^2}\right]$ .

Furthermore, to follow the development of nonstationary states, we make use of the time  $\tau$  in the terms differentiated with respect to time as well as in the definition of the velocity of electricity, i.e., we put according to (30)

$$\frac{\partial}{\partial \tau} = \left[1 + \frac{\gamma\xi}{c^2}\right] \frac{\partial}{\partial \tau} \quad [101]$$

and

$$u_\xi = \left[1 + \frac{\gamma\xi}{c^2}\right] u_\xi \quad [102]$$

We thus obtain

$$\frac{1}{c \left[ 1 + \frac{\gamma \xi}{c^2} \right]} \left[ \rho^* u_\xi + \frac{\partial X^*}{\partial \tau} \right] = \frac{\partial N^*}{\partial \eta} - \frac{\partial M^*}{\partial \zeta} \quad \text{etc.} \quad (31b)$$

and

$$[103] \quad \frac{1}{c \left[ 1 + \frac{\gamma \xi}{c^2} \right]} \frac{\partial L^*}{\partial \tau} = \frac{\partial Y^*}{\partial \zeta} - \frac{\partial Z^*}{\partial \eta} \quad \text{etc.} \quad (32b)$$

These equations too have the same form as the corresponding equations of the nonaccelerated or gravitation-free space; however,  $c$  is here replaced by the value

$$c \left[ 1 + \frac{\gamma \xi}{c^2} \right] = c \left[ 1 + \frac{\Phi}{c^2} \right].$$

From this it follows that those light rays that do not propagate along the  $\xi$ -axis are bent by the gravitational field; it can easily be seen that the change of direction amounts to  $\frac{1}{c^2} \sin \varphi$  per cm light path, where  $\varphi$  [104] denotes the angle between the direction of gravity and that of the light ray.

With the help of these equations and the equations relating the field strength and the electric current of one point, which are known from the optics of bodies at rest, we can calculate the effect of the gravitational field on optical phenomena in bodies at rest. One has to bear in mind, however, that the above-mentioned equations from the optics of bodies at rest hold for the local time  $\sigma$ . Unfortunately, the effect of the terrestrial gravitational field is so small according to our theory (because of the smallness of  $\frac{\gamma \xi}{c^2}$ ) that there is no prospect of a comparison of the results of [105] the theory with experience.

If we successively multiply equations (31a) and (32a) by  $\frac{X^*}{4\pi} \dots \frac{N^*}{4\pi}$  and integrate over infinite space, we obtain, using our earlier notation,

$$[106] \quad \int \left[ 1 + \frac{\gamma \xi}{c^2} \right]^2 \frac{\rho}{4\pi} (u_\xi X + u_\eta Y + u_\zeta Z) d\omega + \int \left[ 1 + \frac{\gamma \xi}{c^2} \right]^2 \cdot \frac{1}{8\pi} \frac{\partial}{\partial \sigma} (X^2 + Y^2 + \dots + N^2) d\omega = 0.$$

[107]  $\frac{\rho}{4\pi} (u_\xi X + u_\eta Y + u_\zeta Z)$  is the energy  $\eta_\sigma$  supplied to the matter per unit volume and unit local time  $\sigma$  if this energy is measured by measuring tools situated at the corresponding location. Hence, according to (30),

$\eta_\tau = \eta^\sigma \left[ 1 + \frac{\gamma\xi}{c^2} \right]$  is the (similarly measured) energy supplied to the matter per [108] unit volume and unit local time  $\tau$ ;  $\frac{1}{8\pi}(X^2 + Y^2 \dots + N^2)$  is the electromagnetic energy  $\epsilon$  per unit volume, measured the same way. If we take into account that according to (30) we have to set  $\frac{\partial}{\partial\sigma} = \left[ 1 - \frac{\gamma\xi}{c^2} \right] \frac{\partial}{\partial\tau}$ , we obtain

$$\int \left[ 1 + \frac{\gamma\xi}{c^2} \right] \eta_\tau d\omega + \frac{d}{d\tau} \left\{ \int \left[ 1 + \frac{\gamma\xi}{c^2} \right] \epsilon d\omega \right\} = 0.$$

This equation expresses the principle of conservation of energy and contains a very remarkable result. An energy, or energy input, that, measured locally, has the value  $E = \epsilon d\omega$  or  $E = \eta d\omega d\tau$ , respectively, contributes to the energy integral, in addition to the value  $E$  that corresponds to its magnitude, also a value  $\frac{E}{c^2} \gamma\xi = \frac{E}{c^2} \Phi$  that corresponds to its *position*. Thus, to each energy  $E$  in the gravitational field there corresponds an energy of position that equals the potential energy of a "ponderable" mass of magnitude  $\frac{E}{c^2}$ .

Thus the proposition derived in §11, that to an amount of energy  $E$  there corresponds a mass of magnitude  $\frac{E}{c^2}$ , holds not only for the *inertial* but also for the *gravitational* mass, if the assumption introduced in §17 is correct.

(Received on 4 December 1907)

## Doc. 49

CORRECTIONS TO THE PAPER: "ON THE RELATIVITY PRINCIPLE  
AND THE CONCLUSIONS DRAWN FROM IT"<sup>1</sup>

by A. Einstein

[*Jahrbuch der Radioaktivität und Elektronik* 5 (1908): 98-99]

[2] During the proofreading of the article cited I missed unfortunately several errors that have to be corrected because they impede the reading of the article.

Formula 15b (p. 435) should read

$$\frac{d}{dt} \left[ \int \frac{1}{4\pi c} (YN - ZM) d\omega \right] + \Sigma \frac{\mu \dot{x}}{\sqrt{1 - \frac{q^2}{c^2}}} = 0 .$$

The factor  $\frac{4}{3}$  in the second formula on p. 451 is in error: the formula should read

$$G = \frac{q}{\sqrt{1 - \frac{q^2}{c^2}}} \frac{E_0}{c^2} .$$

Formula 28 on p. 453 should read

$$dE = F_x dx + F_y dy + F_z dz - pdV + Td\eta .$$

A few lines further on, the subscript in  $G_x$  should be added. In the penultimate line on p. 455 it should read "replaceable" instead of "usable." [Translator's note: This correction does not apply to the translated version.]

On p. 451 it should read

---

[1] <sup>1</sup>This *Jahrbuch* 4 (1907): 411.

$$\frac{\partial}{\partial \tau} = \left[ 1 + \frac{\gamma \xi}{c^2} \right] \frac{\partial}{\partial \sigma}$$

and

$$w_\xi = \left[ 1 + \frac{\gamma \xi}{c^2} \right] u_\xi .$$

On p. 462 the subscripts in the quantities  $u_\xi$  and  $u_\zeta$  have to be added. Also, in about the middle of this page a mistake in sign should be corrected: the equation should read

$$\eta_\sigma = \eta_\tau \left[ 1 - \frac{\gamma \xi}{c^2} \right] .$$

A letter by Mr. Planck induced me to add the following supplementary remark so as to prevent a misunderstanding that could arise easily:

In the section "Principle of relativity and gravitation", a reference system at rest situated in a temporally constant, homogeneous gravitational field is treated as physically equivalent to a uniformly accelerated, gravitation-free reference system. The concept "uniformly accelerated" needs further clarification.

If—as in our case—one considers a rectilinear motion (of the system  $\Sigma$ ), the acceleration is given by the expression  $\frac{dv}{dt}$ , where  $v$  denotes the velocity. According to the kinematics in use up to now,  $\frac{dv}{dt}$  is independent of the state of motion of the (nonaccelerated) reference system, so that one might speak directly of (instantaneous) acceleration when the motion in a certain time element is given. According to the kinematics used by us,  $\frac{dv}{dt}$  does depend on the state of motion of the (nonaccelerated) reference system. But among all the values of acceleration that can be so obtained for a certain motion epoch, that one is distinguished which corresponds to a reference system with respect to which the body considered has the velocity  $v = 0$ . It is this value of acceleration which has to remain constant in our "uniformly [3] accelerated" system. The relation  $v = \gamma t$  used on p. 457 thus holds only in first approximation; however, this is sufficient, because only terms linear in  $t$  and  $\tau$ , respectively, have to be taken into account in these considerations.

(Received on 3 March 1908)

## On the Influence of Gravitation on the Propagation of Light

By A. Einstein.

Annalen der Physik, **35**, pp. 898-908, 1911

There are two translations of this paper that I know of:

1. In *The principle of relativity, a collection of original memoirs on the special and general theory of relativity*, by H. A. Lorentz, A. Einstein, H. Minkowski and H. Weyl, with notes by A. Sommerfeld, tr. by W. Perrett and G. B. Jeffery, Dover Publications, Inc., 1923. This paper appears on pp. 97-108.
2. In *The Collected Papers of Albert Einstein, Volume 3: The Swiss Years: Writings, 1909-1911*, Albert Einstein, vol. 3 edited by Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann, Series ed. John Stachel, Princeton University Press, 1987-.

Neither of these translations is readily available outside major research libraries. The Dover publication is available through AbeBooks.com (typical price \$4.) and Princeton vol. 3 is, I think, still in print. (The current AbeBooks price is \$168.) It appears that Princeton Press has no interest in making such material available on the web.

The original Annalen der Physik publication is available on the web at Gallica (<http://gallica.bnf.fr/ark:/12148/bpt6k15338w.image>) or on the Wiley website, which, however, requires subscription.

The two English translations appear to have been done independently, although much of the Princeton text is quite similar to the Dover text. Both just used “cut and paste” to produce the Figures. Neither translation seemed to me to be sufficiently accurate to fully convey what Einstein wrote. For these reasons I felt that a new, and freely available, translation would be helpful. In doing the translation I made substantial use of the Dover text, which seems somewhat better than the Princeton one. Where problems occurred I also checked the Princeton text. For anyone with reasonable familiarity with German it is, I think, still a good idea to download a copy of the original. Einstein was, as is well-known, a quite original writer. His exact choice of phrases and individual wording is often important. The purely technical content is, in any case, in the equations. Both translations almost always correctly transcribed the equations, although they both just transcribed typos of the math symbols in the text.

I have checked this text carefully, but it is likely that errors have been missed. Any corrections or suggestions for clarification would be welcome.

Michael D. Godfrey  
Information Systems Lab, Stanford University  
email: michaeldgodfrey@gmail.com

## On the Influence of Gravitation on the Propagation of Light

By A. Einstein.

---

In a contribution published four years ago\* I tried to answer the question whether the propagation of light is influenced by gravitation. I return to this theme because my previous presentation of the subject does not satisfy me, but even more because I now see that one of the most important consequences of my former treatment is capable of being tested experimentally. For it follows from the theory to be presented here, that light-rays passing close to the sun are deflected by its gravitational field so that the apparent angular distance between the sun and a visible fixed star near to it is increased by nearly a second of arc.

In the course of these investigations further results which relate to gravitation are shown. But, as the exposition of the entire group of considerations would be rather difficult to follow, only a few quite elementary investigations will be given in the following pages, from which the reader will readily be able to orient himself as to the direction and train of thought of the theory. The relations here deduced, even though the theoretical foundation is sound, are valid only to a first approximation.

### § 1. A Hypothesis as to the Physical Nature of the Gravitational Field

In a homogeneous gravitational field (acceleration of gravity  $\gamma$ ) let there be a stationary system of co-ordinates  $K$ , orientated so that the lines of force of the gravitational field run in the negative direction of the  $z$ -axis. In a space free of gravitational fields let there be a second system of co-ordinates  $K'$ , moving with uniform acceleration ( $\gamma$ ) in the positive direction of its  $z$ -axis. To avoid unnecessary complications, let us for the present disregard the theory of relativity, and regard both systems from the customary point of view of kinematics, and the movements occurring in them from that of ordinary mechanics.

Relative to  $K$ , as well as relative to  $K'$ , material points which are not subjected to the action of other material points, move according to the equations:

$$\frac{d^2x_\nu}{dt^2} = 0, \quad \frac{d^2y_\nu}{dt^2} = 0, \quad \frac{d^2z_\nu}{dt^2} = -\gamma.$$

For the accelerated system  $K'$  this follows directly from Galileo's principle, but for the system  $K$ , at rest in a homogeneous gravitational field, it follows from the experience that all bodies in such a field are equally and uniformly accelerated. This experience, of the equal falling of all bodies in the gravitational field, is one of the most universal which the observation of nature has yielded to us; but in spite of this, this law has found no place in the foundations of our world view (Weltbildes) of the physical universe.

But we arrive at a very satisfactory interpretation of this empirical law, if we assume that the systems  $K$  and  $K'$  are physically exactly equivalent, that is, if we

---

\* A. Einstein, Jahrbuch für Radioact, und Electronik, 4, 1907.

assume that we may just as well regard the system  $K'$  as being in a space free from gravitational fields; then we must regard  $K$  as uniformly accelerated. This assumption of exact physical equivalence makes it impossible for us to speak of the *absolute acceleration* of the system of reference, just as the usual theory of relativity forbids us to talk of the *absolute velocity* of a system;<sup>1</sup> This assumption also makes the equal falling of all bodies in a gravitational field seem obvious.

As long as we restrict ourselves to purely mechanical processes in the realm where Newton's mechanics is valid, we are certain of the equivalence of the systems  $K$  and  $K'$ . But our view of this will not have any deeper significance unless the systems  $K$  and  $K'$  are equivalent with respect to all physical processes, that is, unless the laws of nature with respect to  $K$  are in entire agreement with those with respect to  $K'$ . By assuming this to be so, we arrive at a principle which, if it is really true, has great heuristic importance. For by theoretical consideration of processes which take place relative to a system of reference with uniform acceleration, we obtain information as to the behavior of processes in a homogeneous gravitational field.<sup>2</sup> We shall now show, first of all from the standpoint of the ordinary theory of relativity, that our hypothesis has considerable probability.

## § 2. On the Gravitation of Energy

The theory of relativity shows that the inertial mass of a body increases with the energy it contains; if the increase of energy amounts to  $E$ , the increase in inertial mass is equal to  $E/c^2$ , where  $c$  denotes the velocity of light. Now, is there an increase of gravitational mass corresponding to this increase of inertial mass? If not, then a body would fall in the same gravitational field with varying acceleration according to the energy it contained. And then the highly satisfactory result of the theory of relativity, by which the law of the conservation of mass leads to the law of conservation of energy, could not be maintained, because it would compel us to abandon the law of the conservation of mass in its old form for *inertial* mass, but maintain it for gravitational mass.

This must be regarded as very improbable. On the other hand, the usual theory of relativity does not provide us with any argument from which to infer that the weight of a body depends on the energy contained in it. But we shall show that our hypothesis of the equivalence of the systems  $K$  and  $K'$  gives us gravitation of energy as a necessary consequence.

Let two material systems  $S_1$  and  $S_2$  (Fig. 1), each provided with measuring instruments, be situated on the  $z$ -axis of  $K$  at the distance  $h$  from each other,<sup>3</sup> so that the gravitational potential at  $S_2$  is greater than that at  $S_1$  by  $\gamma h$ . Let a definite quantity of

---

<sup>1</sup> Of course, we cannot replace an *arbitrary* gravitational field by a state of motion of a system without a gravitational field, just as we cannot transform to rest all the points of an arbitrarily moving medium by means of a relativistic transformation.

<sup>2</sup> It will be shown in a subsequent paper that the gravitational field considered here is homogeneous only to a first approximation.

<sup>3</sup>  $S_1$  and  $S_2$  are regarded as infinitely small in comparison with  $h$ .

energy  $E$  be emitted from  $S_2$  towards  $S_1$ . Let the quantities of energy in  $S_1$  and  $S_2$  be measured by devices which – brought to *one* location in the system  $z$  and there compared – are perfectly alike. As to the process of this energy transmission by radiation we can make no a priori assertion because we do not know the influence of the gravitational field on the radiation and the measuring instruments at  $S_1$  and  $S_2$ .

But by our postulate of the equivalence of  $K$  and  $K'$  we are able, in place of the system  $K$  in a homogeneous gravitational field, to set the gravitation-free system  $K'$ , which moves with uniform acceleration in the direction of positive  $z$ , and by the  $z$ -axis of which the material systems  $S_1$  and  $S_2$  are rigidly connected.

We consider the process of transmission of energy by radiation from  $S_2$  to  $S_1$  from a system  $K_0$ , which is free of acceleration. At the moment when the radiation energy  $E_2$  is emitted from  $S_2$  toward  $S_1$ , let the velocity of  $K'$  relative to  $K_0$  be zero. The radiation will arrive at  $S_1$  when the time  $h/c$  has elapsed (to a first approximation). But at this moment the velocity of  $S_1$  relative to  $K_0$  is  $\gamma h/c = v$ . Therefore by the ordinary theory of relativity the radiation arriving at  $S_1$  does not possess the energy  $E_2$ , but a greater energy  $E_1$ , which is related to  $E_2$ , to a first approximation, by the equation<sup>1</sup>:

$$(1) \quad E_1 = E_2 \left( 1 + \frac{v}{c} \right) = E_2 \left( 1 + \frac{\gamma h}{c^2} \right).$$

By our assumption exactly the same relation holds if the same process takes place in the system  $K$ , which is not accelerated, but is provided with a gravitational field. In this case we may replace  $\gamma h$  by the potential  $\Phi$  of the gravitation vector in  $S_2$ , if the arbitrary constant of  $\Phi$  in  $S_1$  is set to zero. We then have the equation:

$$(1a) \quad E_1 = E_2 + \frac{E_2}{c^2} \Phi.$$

This equation expresses the energy law for the process under observation. The energy  $E_1$  arriving at  $S_1$  is greater than the energy  $E_2$ , measured by the same means, which was emitted from  $S_2$ , the excess being the potential energy of the mass  $E_2/c^2$  in the gravitational field. This shows that in order to satisfy the energy principle we have to ascribe to the energy  $E$ , before its emission from  $S_2$ , a potential energy, due to gravity, which corresponds to the (gravitational) mass  $E/c^2$ . Our assumption of the equivalence of  $K$  and  $K'$  thus removes the difficulty mentioned at the beginning of this Section, which is left unsolved by the ordinary theory of relativity.

The meaning of this result is shown particularly clearly if we consider the following cycle of operations: –

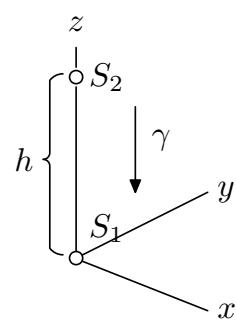


Fig. 1.

<sup>1</sup> A. Einstein, Ann. d. Phys. 17, p. 913 – 914. 1905.

1. The energy  $E$ , as measured in  $S_2$ , is emitted in the form of radiation from  $S_2$  towards  $S_1$ , where, by the result just obtained, the energy  $E(1 + \gamma h/c^2)$  (as measured in  $S_1$ ) is absorbed.
2. A body  $W$  of mass  $M$  is lowered from  $S_2$  to  $S_1$ , work  $M\gamma h$  thereby being done.
3. The energy  $E$  is transmitted from  $S_1$  to the body  $W$  while  $W$  is in  $S_1$ . The gravitational mass  $M$  is thereby changed so that it acquires the value  $M'$ .
4. Let  $W$  be again raised to  $S_2$ , work  $M'\gamma h$  being done as a result.
5. Let  $E$  be transmitted from  $W$  back to  $S_2$ .

The effect of this cycle is simply that  $S_1$  has undergone an energy increase of  $E(\gamma h/c^2)$ , and that the quantity of energy

$$M'\gamma h - M\gamma h$$

has been supplied to the system in the form of mechanical work. By the energy principle, we must therefore have

$$E \frac{\gamma h}{c^2} = M'\gamma h - M\gamma h$$

or

$$(1b) \quad M' - M = \frac{E}{c^2}.$$

The increase in *gravitational* mass is thus equal to  $E/c^2$ , and therefore equal to the increase in *inertial* mass as given by the theory of relativity.

This result emerges still more directly from the equivalence of the systems  $K$  and  $K'$ , according to which the *gravitational* mass of  $K$  is exactly equal to the *inertial* mass of  $K'$ ; energy must therefore possess a *gravitational* mass which is equal to its *inertial* mass. If a mass  $M_0$  be suspended on a spring balance in the system  $K'$  the balance will indicate the apparent weight  $M_0\gamma$  on account of the inertia of  $M_0$ . If the quantity of energy  $E$  be transmitted to  $M_0$ , the spring balance, by the law of the inertia of energy, will indicate  $\left(M_0 + \frac{E}{c^2}\right)\gamma$ . By reason of our fundamental assumption exactly the same thing must occur when the experiment is repeated in the system  $K$ , that is, in the gravitational field.

### § 3. Time and the Velocity of Light in the Gravitational Field

If the radiation emitted in the uniformly accelerated system  $K'$  in  $S_2$  toward  $S_1$  had the frequency  $\nu_2$  relative to the clock at  $S_2$ , then, relative to  $S_1$ , at its arrival at  $S_1$  it no longer has the frequency  $\nu_2$  relative to an identical clock at  $S_1$ , but a greater frequency  $\nu_1$ , such that, to a first approximation

$$(2) \quad \nu_1 = \nu_2 \left(1 + \gamma \frac{h}{c^2}\right).$$

If we again introduce the unaccelerated reference system  $K_0$ , relative to which at the time of the emission of light,  $K'$  has no velocity, then  $S_1$ , at the time of arrival of

the radiation at  $S_1$  has, relative to  $K_0$ , the velocity  $\gamma(h/c)$  from which, by Doppler's principle, the relation as given results immediately.

In agreement with our assumption of the equivalence of the systems  $K'$  and  $K$ , this equation also holds for a stationary system of co-ordinates  $K_0$  in a uniform gravitational field, if in it the transmission by radiation takes place as described. It follows, then, that a light-ray emitted from  $S_2$  with a definite gravitational potential, and possessing at its emission the frequency  $\nu_2$  – compared with a clock at  $S_2$  – will, at its arrival at  $S_1$ , possess a different frequency  $\nu_1$  measured by an identical clock at  $S_1$ . For  $\gamma h$  we substitute the gravitational potential  $\Phi$  of  $S_2$  – that of  $S_1$  being taken as zero – and assume that the relation which we have deduced for the *homogeneous* gravitational field also holds for other forms of field. Then

$$(2a) \quad \nu_1 = \nu_2 \left( 1 + \frac{\Phi}{c^2} \right)$$

This result (which by our derivation is valid to a first approximation) permits, first, the following application. Let  $\nu_0$  be the oscillation-number of an elementary light-generator, measured by a clock  $U$  at the same location. This oscillation-number is then independent of the locations of the light-generator and the clock. Let us imagine them both at a position on the surface of the Sun (where our  $S_2$  is located). Of the light emitted from there a portion reaches the Earth ( $S_1$ ), where we measure the frequency  $\nu$  of the arriving light with a clock  $U$  of exactly the same properties as the one just mentioned. Then by (2a),

$$\nu = \nu_0 \left( 1 + \frac{\Phi}{c^2} \right)$$

where  $\Phi$  is the (negative) difference of gravitational potential between the surface of the Sun and the Earth. Thus according to our view the spectral lines of sunlight, as compared with the corresponding spectral lines of terrestrial light sources, must be somewhat displaced toward the red, in fact by the relative amount

$$\frac{\nu_0 - \nu}{\nu_0} = \frac{-\Phi}{c^2} = 2 \times 10^{-6}.$$

If the conditions under which the solar lines arise were exactly known, this shifting would be susceptible of measurement. But as other influences (pressure, temperature) affect the position of the centers of the spectral lines, it is difficult to discover whether the inferred influence of the gravitational potential really exists.<sup>1</sup>

On superficial consideration equation (2) or (2a), respectively, seems to assert an absurdity. If there is constant transmission of light from  $S_2$  to  $S_1$ , how can any other number of periods per second arrive at  $S_1$  than is emitted from  $S_2$ ? But the answer is

---

<sup>1</sup> L. F. Jewell (Journ. de Phys., **6**, p. 84, 1897) and particularly Ch. Fabry and H. Boisson (Compt. rend. **148**, p. 688-690, 1909) have actually found such displacements of fine spectral lines toward the red end of the spectrum, of the order of magnitude here calculated, but have ascribed them to an effect of pressure in the absorbing layer.

simple. We cannot regard  $\nu_2$  or respectively  $\nu_1$  simply as frequencies (as the number of periods per second) since we have not yet determined a time in system  $K$ . What  $\nu_2$  denotes is the number of periods per second with reference to the time-unit of the clock  $U$  at  $S_2$ , while  $\nu_1$  denotes the number of periods per second with reference to the identical clock at  $S_1$ . Nothing compels us to assume that the clocks  $U$  in different gravitation potentials must be regarded as going at the same rate. On the contrary, we must certainly define the time in  $K$  in such a way that the number of wave crests and troughs between  $S_2$  and  $S_1$  is independent of the absolute value of time: for the process under observation is by nature a stationary one. If we did not satisfy this condition, we should arrive at a definition of time such that by its application time would enter explicitly into the laws of nature, and this would certainly be unnatural and inappropriate. Therefore the two clocks at  $S_1$  and  $S_2$  do not both give the "time" correctly. If we measure time at  $S_1$  with the clock  $U$ , then we must measure time at  $S_2$  with a clock which goes  $1 + \Phi/c^2$  times more slowly than the clock  $U$  when compared with  $U$  at one at the same location. For when measured by such a clock, the frequency of the light-ray which is considered above is at its emission from  $S_2$

$$\nu_2 \left( 1 + \frac{\Phi}{c^2} \right),$$

and is therefore, by (2a), equal to the frequency  $\nu_1$  of the same light-ray on its arrival at  $S_1$ .

This has a consequence which is of fundamental importance for our theory. For if we measure the velocity of light at different locations in the accelerated, gravitation-free system  $K'$ , employing clocks  $U$  of identical properties we obtain the same magnitude at all these locations. The same holds good, by our fundamental assumption, for the system  $K$  as well. But from what has just been said we must use clocks of unlike properties for measuring time at locations with differing gravitation potential. For measuring time at a location which, relative to the origin of the co-ordinates, has the gravitation potential  $\Phi$ , we must employ a clock which – when transferred to the co-ordinate origin – goes  $(1 + \Phi/c^2)$  times more slowly than the clock used for measuring time at the origin of co-ordinates. If we call the velocity of light at the origin of co-ordinates  $c_0$ , then the velocity of light  $c$  at a location with the gravitation potential  $\Phi$  will be given by the relation

$$(3) \quad c = c_0 \left( 1 + \frac{\Phi}{c^2} \right).$$

The principle of the constancy of the velocity of light holds good according to this theory in a different form from that which usually underlies the ordinary theory of relativity.

#### § 4. Bending of Light-Rays in the Gravitational Field

From the proposition which has just been proved, that the velocity of light in the gravitational field is a function of the location, we may easily infer, by means of Huygens's principle, that light-rays propagated across a gravitational field undergo

deflection. For let  $\varepsilon$  be a wave front of a plane light-wave at the time  $t$ , and let  $P_1$  and  $P_2$  be two points in that plane at unit distance from each other.  $P_1$  and  $P_2$  lie in the plane of the paper, which is chosen so that the differential coefficient of  $\Phi$ , taken in the direction of the normal to the plane, and therefore also that of  $c$ , vanishes. We obtain the corresponding wave front at time  $t+dt$ , or, rather, its intersection with the plane of the paper, by describing circles round the points  $P_1$  and  $P_2$  with radii  $c_1 dt$  and  $c_2 dt$  respectively, where  $c_1$  and  $c_2$  denote the velocity of light at the points  $P_1$  and  $P_2$  respectively, and by drawing the tangent to these circles. The angle through which the light-ray is deflected on the path  $cdt$  is therefore

$$\frac{(c_1 - c_2)dt}{1} = -\frac{\partial c}{\partial n'} dt,$$

if we calculate the angle positively when the ray is bent toward the side of increasing  $n'$ .

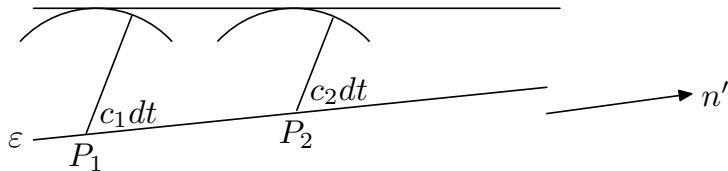


Fig. 2.

The angle of deflection per unit of path of the light-ray is thus

$$-\frac{1}{c} \frac{\partial c}{\partial n'}$$

or by (3) it is

$$-\frac{1}{c^2} \frac{\partial \Phi}{\partial n'}.$$

Finally, we obtain for the deflection  $\alpha$ , which a light-ray experiences toward the side  $n'$  on any path  $(s)$  the expression

$$(4) \quad \alpha = -\frac{1}{c^2} \int \frac{\partial \Phi}{\partial n'} ds.$$

We might have obtained the same result by directly considering the propagation of a light-ray in the uniformly accelerated system  $K'$ , and transferring the result to the system  $K$ , and thence to the case of a gravitational field of any form.

By equation (4) a light-ray passing by a heavenly body suffers a deflection to the side of the diminishing gravitational potential, that is, to the side directed toward the heavenly body, of the magnitude

$$\alpha = \frac{1}{c^2} \int_{\vartheta=-\frac{\pi}{2}}^{\vartheta=+\frac{\pi}{2}} \frac{kM}{r^2} \cos(\vartheta) ds = \frac{2kM}{c^2 \Delta},$$

where  $k$  denotes the constant of gravitation,  $M$  the mass of the heavenly body,  $\Delta$  the distance of the ray from the center of the body (and  $r$  and  $\vartheta$  are as shown in Fig. 3). *A light-ray going past the Sun would accordingly undergo deflection by the amount of  $4 \times 10^6 = 0.83$  seconds of arc.* The angular distance of the star from the center of the

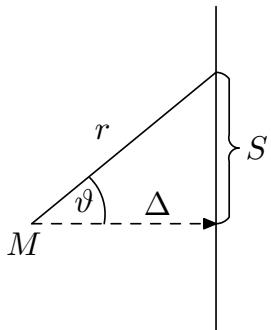


Fig. 3.

Sun appears to be increased by this amount. As the fixed stars in the parts of the sky near the Sun are visible during total eclipses of the Sun, this consequence of the theory may be compared with experimental evidence. With the planet Jupiter the displacement to be expected reaches to about 1/100 of the amount given. It would be urgently wished that astronomers take up the question here raised, even though the considerations presented above may seem insufficiently established or even bizarre. For, apart from any theory, there is the question whether it is possible with the equipment at present available to detect an influence of gravitational fields on the propagation of light.

Prague, June 1911.

(Submitted 21 June 1911.)

---



On an Expansion Apparatus for Making Visible the  
Tracks of Ionising Particles in Gases and Some Results  
Obtained by its Use.

T.R.. Wilson

(Received 1912)

In a recent communication<sup>1</sup> I described a method of making visible the tracks of ionising particles through a moist gas by condensing water upon the ions immediately after their liberation. At that time I had only succeeded in obtaining photographs of the clouds condensed on the ions produced along the tracks of  $\alpha$ -particles and of the corpuscles set free by the passage of X-rays through the gas. The interpretation of the photographs was complicated to a certain extent by distribution arising from the position which the camera occupied.

The expansion apparatus and the method of illuminating the clouds have both been improved in detail, and it has now been found possible to photograph the tracks of even the fastest  $\beta$ -particles, the individual ions being rendered visible. In the photographs of the X-rays clouds the drops in many of the tracks are also individually visible; the clouds found in the  $\alpha$ -ray tracks are generally too dense to be resolved into drops. The photographs are now free from distortion. The cloud chamber has been greatly increased in size; it is now wide enough to give ample room for the longest  $\alpha$ -ray, and high enough to admit of a horizontal beam of X-rays being sent through it without any risk of complications due to the proximity of the roof and floor.

---

<sup>1</sup>Roy. Soc. Proc.,' 1911, A, vol. 85, p. 285.

### *The Expansion Apparatus*

The essential features of the expansion apparatus are shown in fig. 1. The cylindrical cloud chamber A is 16.5 cm. in diameter and 3.4 cm. high; the roof, walls and floor are of glass, coated inside with gelatine, that on the floor being blackened by adding a little Indian ink. The plate glass floor is fixed on the top of a thin-walled brass cylinder (the "plunger"), 10 cm. high, open below, and sliding freely within an outer brass cylinder (the "expansion cylinder") of the same height and about 16 cm. in internal diameter. The expansion cylinder supports the walls of the cloud chamber and rests on a thin sheet of indiarubber lying on a thick brass disc, which forms the bottom of a shallow receptacle containing water to a depth of about 2 cm. The water separates completely the air in the cloud chamber from that below the plunger. The base plate rests on a wooden stand, not shown on the diagram.

The expansion is effected by opening the valve B and so putting the air space below the plunger in communication with the vacuum chamber C through wide glass connecting tubes of about 2 cm. in diameter. The floor of the cloud chamber, in consequence, drops suddenly until brought to a sudden stop, when the plunger strikes the indiarubber-covered base plate, against which it remains firmly fixed by the pressure of the air in the cloud chamber. To reduce the volume of air passing through the connecting tubes at each expansion the wooden cylinder D was inserted within the air space below the plunger.

The valve is opened by the fall of a weight W released by a trigger arrangement T (fig. 3). On closing the valve and opening communication with the atmosphere through the pinch-cock F, the plunger rises and so reduced the volume of the air in the cloud chamber. By means of the two pinch-cocks F and G (the latter on a tube communicating with the vacuum chamber), the plunger may be adjusted to give any desired initial volume  $v_1$  between the upper limit  $v_2$  – the maximum volume of the cloud chamber – and the lower limit reached when the pressure below the plunger is that of the atmosphere.

The final volume  $v_2$  is always the same (about 750 cc.), the expansion ratio  $v_2/v_1$  depending only on the initial volume. A scale attached to the side of the cloud chamber enables the position of the top of the plunger to be read, and hence the initial volume to be determined, the area of the cross-section of the plunger and the maximum volume  $v_2$  of the cloud chamber being known.

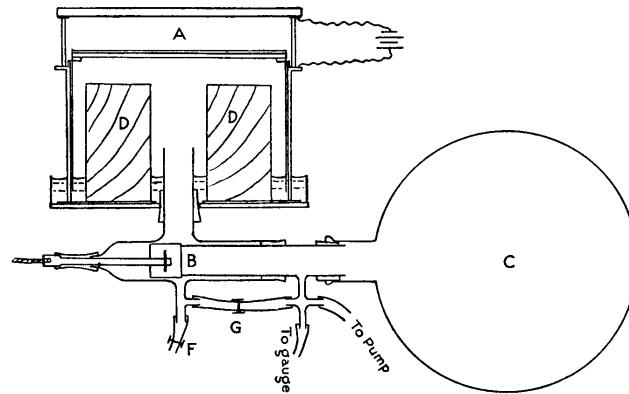


Figure 1:

In setting up the apparatus, the plunger is placed on the rubber-covered base plate, and the expansion cylinder slipped over it, a hole in the side of the cloud chamber being open at this stage to allow of the imprisoned air escaping. Then, by blowing in air through F, momentarily opened for the purpose, the plunger is driven up to a height sufficient to allow of the largest desired expansions being made. The aperture in the wall of the cloud chamber is then closed, and the mass of imprisoned air remains unchanged during subsequent operations.

The gelatine layer under the roof of the cloud chamber is connected, through a ring of tinfoil cemented between the cylindrical wall and the roof, with one terminal of a battery of cells of which the other terminal is connected, through the brass expansion cylinder and plunger, with the layer of blackened gelatine on the floor of the cloud chamber. An approximately uniform vertical electric field of any desired intensity may thus be maintained in the cloud chamber.

The gelatine lining of the roof and walls is formed by pouring into the cloud chamber, before attaching it to the expansion cylinder, a hot solution containing about 4 per cent. of gelatine and 0.1 per cent. of boracic acid and allowing the surplus to drain away by inverting the vessel. The thin coating of gelatine which remains is allowed to dry over calcium chloride. The cloud chamber is cemented to the expansion cylinder by means of gelatine.

A comparatively thick layer (about 1 mm) of a solution containing 15 per cent. of gelatine, 2 per cent. of boracic acid, and 3 per cent. of Indian ink is

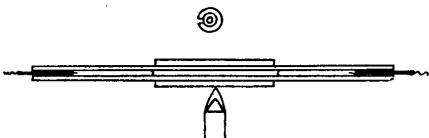


Figure 2:

poured on to the glass plate which forms the floor of the cloud chamber, the brass walls of the plunger being prolonged for about 1 mm. above the upper surface of the plate, thus forming a shallow receptacle for the gelatine and making an efficient electric contact with it. The blackened gelatine is not allowed to dry, but at once covered to prevent evaporation and to protect it from dust till ready for use. The gelatine is in all cases previously sterilised by heat.

#### *Method of Illuminating and Photographing the Clouds*

As in experiments described in my last paper, a Leyden jar discharge through mercury vapour at atmospheric pressure is used for the instantaneous illumination of the clouds resulting from the expansion. A horizontal silica tube about 15 cm. long, and having an internal diameter of about 1 mm., is filled with mercury and enclosed, for the central 4 cm. of its length, by a close-fitting silver tube about 2 mm. thick, and having a slot about 1 mm. wide extending from end to end. The silver tube when heated by a small flame serves to keep the enclosed position of the silica tube at a nearly uniform temperature high enough to vaporise the mercury, and thus from a mercury-vapour spark-gap. Connection with the Leyden jars is made through platinum wires fused through the ends of glass tubes filled with mercury and inserted into the ends of the silica tube.

The silica tube is first filled with mercury, the end pieces inserted, and a small flame placed under the silver tube. When the mercury occupying the portion of the silica tube which is surrounded by the silver jacket has all been vaporised (the excess of mercury escaping from the ends of the tube) no further change takes place and the spark-gap is ready for use. The very considerable capillary forces set up when the mercury is forced into

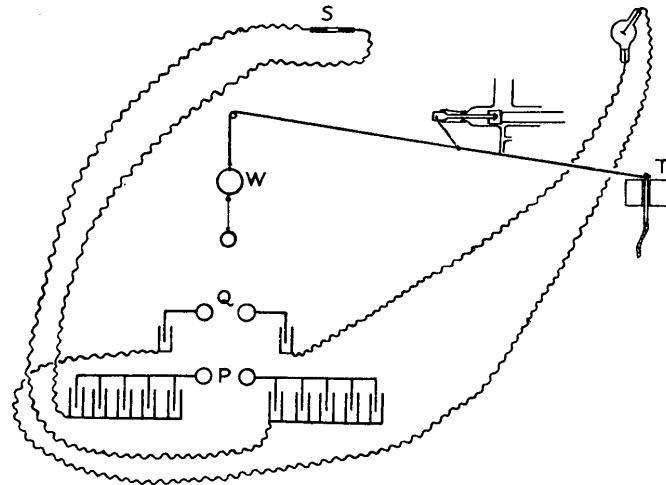


Figure 3:

the narrow space between the glass end pieces and the surrounding silica tube effectually prevent the violet oscillatory motion which are apt to be the principal source of trouble in the use of a mercury spark-gap of this type.

For firing the spark the arrangement used is essentially that which has generally been employed in instantaneous photography by the Leyden jar discharge. The outer coatings of two sets of 4 or 5 "gallon" Leyden jars, standing on the floor of the room, are connected to the terminals of the illuminating spark. The inner coatings are connected to the terminals of a Wimshurst machine and to two brass balls separated by a space of about 5 cm. which forms the primary spark-gap. The jars having been charged almost to sparking potential, a metal ball is allowed to fall between the terminals of the primary spark-gap, causing a spark to pass at both gaps. The ball whose fall causes the spark is hung by a fine thread, just strong enough to carry it, from the weight W which works the valve of the expansion apparatus.

The arrangements for firing the spark at a definite interval after the expansion are shown diagrammatically in fig. 3. The weight W is carried by a cord which passes through an iron ring in a firm support, and thence nearly horizontally to the trigger T, to which it is attached by a loop. A second string, slack at this stage, connects a point on the first cord with the valve

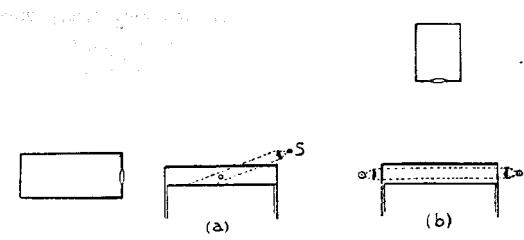


Figure 4:

of the expansion apparatus. On pulling the trigger the cord attached to it is released and the weight falls until the second string is stretched tight, when it is brought to a sudden stop, the valve being simultaneously opened and the expansion thereby effected. The thread breaks at this moment and the steel sphere continues to fall, finally passing through the primary spark-gap, P, and causing the illuminating spark to pass at S. The upper spark-gap Q, shown in the figure, was only employed in the experiments with X-rays.

In the experiments described in this paper, the camera lens has always occupied one of the two positions indicated diagrammatically in fig. 4, (a) and (b). In (a) the small circle represents a transverse section of a narrow horizontal beam of ionising rays crossing a diameter of the cloud chamber. The camera looks in a horizontal direction normal to the ionising beam. The mercury spark-gap is at S, at the principal focus of a cylindrical lens about 20 cm. long and 2 cm. wide, and having a focal length of about 3 cm. With this arrangement the whole of the cloud produced by a considerable length of the ionising beam is illuminated, while the direction of the incident light makes a comparatively small angle (about 25 degrees) with the axis of the camera.

Arrangement (b) has been used chiefly with the  $\alpha$ -rays, which give clouds of sufficient density to scatter a large amount of light at right angles to the illuminating beam. The camera lens is vertically over the centre of the cloud chamber; and by means of two similar mercury spark-gaps (arranged in series), each at the principal focus of a cylindrical lens like that used in (a), a horizontal stratum of about 2 cm. in vertical thickness and extending across the whole area of the vessel, is illuminated.

The lens which I have used is a Beck "isostigmar", the full aperture, marked F 5.8, being utilised; Ilford "Monarch" plates were employed.

### *Ionization by $\alpha$ -Rays (Plate 6, and fig. 1 of Plate 7.)*

Fig. 1 (Plate 6) is a typical photograph of the cloud obtained on expansion when a minute quantity of radium is placed on the tip of a wire projecting into the cloud chamber. A potential difference of 40 volts was maintained between the roof and floor, the roof being at the higher potential. The camera was placed with its axis vertical, and a horizontal section of the cloud chamber, about 2 cm. depth, was illuminated (arrangement (b) of (fig. 4). The  $\beta$ -rays are not visible in the photographs obtained with this mode of illumination.

The narrow, sharply defined rays of these photographs are clouds condensed along the tracks of  $\alpha$  particles which have traversed the supersaturated air after expansion, so that there has been very little time for the ions to diffuse before losing their mobility through condensation of water upon them. The diffuse rays are clouds condensed upon ions set free by  $\alpha$ -particles which have traversed the air before its expansion, so that there has been time for diffusion of the ions before the formation of the cloud. The weaker the electric field the greater is the maximum possible age, and consequent diffuseness, of the tracks which may be present; with a potential difference of only two or three volts, wide finger-like clouds are formed on expansion.

$\alpha$ -rays which pass after the expansion can only leave visible trails if the degree of supersaturation still remains sufficient to cause water to condense on the ions. In the immediate neighbourhood of the cloud already condensed on an older tracks, the supersaturation remaining may be insufficient to cause condensation, although elsewhere the  $\alpha$ -particle may leave a visible trail. This is doubtless the explanation of the fact that most of the sharply defined trails only seem to begin at some considerable distance from the radium, the diffuse cloud trails formed at the moment of expansion being so closely packed near the source of the rays that there is little chance of an  $\alpha$ -particle, ejected after the expansion, finding the supersaturation necessary for rendering its trail visible, until it has travelled for some distance.

Except in the case of photographs taken very soon after the insertion of the radium, the trails of  $\alpha$ -particles from the emanation and later radioactive products also appear. Fig. 4 is a photograph of the cloud formed by expansion after the radium-tipped wire had been for some days in the cloud chamber and had then been removed;  $\alpha$ -rays are seen running in all directions. A sharply defined trail may sometimes be observed crossing one or more diffuse ones, and is then frequently invisible for some little distance

on either side of the diffuse ray, the necessary supersaturation not being attained owing to previous condensation on the ions of the older trail.

For some purpose (it, for example, the ranges of the  $\alpha$ -particles were under investigation) it would be necessary to know definitely whether the  $\alpha$ -particle giving rise to any given trail had passed before or after the expansion, the density of the air traversed being different in the two cases. The trails of  $\alpha$ -particles passing previously to the expansion have their dimensions altered between the liberation of the ions and the deposition of water upon them; but, the displacement of the air being everywhere nearly in a vertical direction, the horizontal dimensions are almost unaffected. In the photographs it will be observed that the diffuse rays are shorter than the sharply defined rays in accordance with the greater density of the air at the moment of passage of the  $\alpha$ -particle.

There is no difficulty in securing that the particles whose tracks are being photographed shall have passed either all before or all after the expansion. It is only necessary to attach to the plunger a vertical plate (glass 2 mm. thick was used) immediately in front of the source of the rays, and with a horizontal slot so placed that it shall be at the level of the radiant point either before or after the expansion. Fig. 2 is a photograph taken under the latter condition; the diffuse tracks are now absent. This method is, of course, inapplicable to the study of the rays from the emanation within the cloud chamber.

As will be seen from the photographs, the  $\alpha$ -rays are generally straight over the greater part of their length, but they nearly all are bent, often abruptly, in the last 2 mm. of their course. abrupt bends through considerable angles are seen much earlier in the course of some of the rays.

In fig. 3 of Plate 6 is shown an enlargement of a particularly interesting trail. Here there are two absolutely abrupt bends – the first through about 10.5 degrees, the second through about 43 degrees. There is very well-marked spur at the second bend, which it is difficult to interpret otherwise than as being due to ionisation by the recoil of the atom., by collision with which the course of the  $\alpha$ -particle has been abruptly changed. (But for the spur this  $\alpha$ -ray shows an astonishingly close resemblance to one in a diagram constructed by Prof. Bragg<sup>2</sup> to illustrate what he considered to be likely forms of  $\alpha$ -ray paths.)

Apart from these sudden bends a certain amount of curvature is apparent in some of the tracks. In some cases, where the curvature occurs close to the walls of the cloud chamber, it is certainly a spurious effect, due to

---

<sup>2</sup>'Archives of the Röntgen Ray', April, 1911.

displacement of the tracks by air motions or to optical distortion arising from a thickening of the gelatine around the circumference of the roof. Where, however, it appears at no great distance from the centre of the cloud chamber it is probably genuine, indicating deviation of the  $\alpha$ -particle by repeated small deflections. There is generally unmistakably genuine curvature in the last millimetre of the track.

The photographs thus furnish evidence of two distinct ways in which  $\alpha$ -particles are "scattered" in passing through air, what Rutherford has called "single" and "compound" scattering respectively. And, as Rutherford<sup>3</sup> has contended, the scattering of [a] large amount is in the case of  $\alpha$ -particles mainly due to the former process, that is to say, it is the result of single deflection through considerable angles and not a cumulative effect due to a very large number of minute deviations.

When the  $\alpha$ -rays arise from the emanation it is possible to photograph the complete track of an  $\alpha$ -particle, including the beginning and end. The latter is at once recognisable by its characteristic bend or hook. In fig. 4 and 5 (Plate 6) are shown the tracks of two  $\alpha$ -particles, each of which completed its course in the illuminated layer; in both cases the beginning of the trail is seen to be marked by an enlarged head in which the cloud is of greater density than elsewhere. This may represent ionisation by the recoil of the atom from which the  $\alpha$ -particle has escaped. The same characteristic head appears at what is presumably the beginning of other, obviously foreshortened, tracks whose ends lie outside the illuminated layer.

Of two complete  $\alpha$ -ray tracks from the emanation one has a length (when reduced to 760 mm. and 15 degrees C.) of 4.3 cm., in good agreement with the usually accepted value of the range. The other is apparently somewhat shorter, about 3.8 cm., the low value being probably due to foreshortening.

Some photographs of  $\alpha$ -ray trails were obtained with the camera in the lateral position and with oblique illumination – arrangement (a) of fig. 4 (p. 281 ???). The radium tipped wire was surrounded by a glass tube about 1 mm. wide open at the end and projecting for about 1 cm. beyond the radium, the object being to confine the rays to a moderately narrow pencil with its axis in the plane for which the camera was focussed.

An example of one of the photographs obtained in this way is shown in Plate 7, fig. 1, which is an enlargement of the original negative. The track of an  $\alpha$ -particle is seen near the bottom of the picture. Some of the ions appear to have retained their mobility in the supersaturated atmosphere long enough to enable them to travel some distance under the action of the

---

<sup>3</sup>Phil. Mag.,' 1911, vol. 21, p. 669.

electric field before growing into drops, thus giving rise to a vertical sheet or curtain of drops. The effect is most marked above the main track, i.e. on the side to which the negative ions would travel.

I have not yet succeeded in obtaining photographs in which all the drops in a known length of the cloud trail left by an  $\alpha$ -particle could be counted. It would obviously be of interest to determine by a direct method of this kind the number of ions produced by an  $\alpha$ -particle.

### *Ionisation by $\beta$ -Rays (Plate 7.)*

When the camera is in the lateral position and oblique illumination is used the individual cloud particles, so long as they are not too close together for resolution, leave distinct images on the photographic plate. It is possible, therefore, to photograph the track of any ionising particle, however small the number of ions produced per centimetre of its path may be.

Some photographs of  $\beta$ -ray trails were obtained along with those of the  $\alpha$ -rays in the course of the experiments last described. Figs. 1, 3, and 4 of Plate 7 were obtained in this way; fig. 2 shows the result of passing a narrow beam of  $\gamma$ -rays through the cloud chamber; the tracks in this case are doubtless those of  $\beta$ -particles starting in the walls of the vessel.

The almost straight trail of figs. 3 and 4 (the actual length of trail photographed amounts to about 4 cm.) is evidently that of a  $\beta$ -particle in the earlier stage of its free existence while its velocity is still very high. This is indicated not only by the straightness of its path but also by the very small ionisation along it. The distribution of the ionisation along the path is interesting. Over considerable distances the ions occur mainly in pairs, but here and there 20 or 30 appear to have been liberated in a closely packed group. (A similar distribution appears in a second approximately straight ray which crosses the first.) The groups show a peculiarity which is also met with in the clouds condensed on the cathode rays produced by X-rays when a suitable expansion ratio is used; while the negative ions have given rise to a densely packed cluster of drops the positive ions have been drawn out by the electric field before losing their mobility, giving the appearance of a shower of drops falling from the negative cloud.

If we omit the clusters, the number of ions in the trail amounts to about 32, i.e. 16 pairs, per centimetre at atmospheric pressure. If we take into account the groups the number of ions per centimetre is roughly doubled,

giving a number not very much lower than the estimate of 48 pairs per centimetre obtained by Eve<sup>4</sup> by indirect methods.

The occurrence of the groups or clusters of ions may be interpreted as indicating that in certain cases the corpuscle liberated from an atom by a  $\beta$ -particle of high velocity may itself have energy enough to ionise for a very short distance. The  $\beta$ -rays of fig. 2 are obviously of smaller velocity, producing much more ionisation per centimetre and being much more readily deviated.

A still later stage of slower velocity has been reached by the particles giving the abruptly ending coiled up trails which appear in figs. 1 and 3. These  $\beta$ -rays ending are indistinguishable from the cathode rays, produced in air by Röntgen rays, such as are shown in the succeeding plates.

It will be noticed that the  $\beta$ -rays photographed do not show abrupt deflections like the  $\alpha$ -rays but, except while the velocity remains very high, they show gradual bending resulting in large deviations. The scattering of the  $\beta$ -rays is thus mainly or entirely of the cumulative or "compound" type due to a large number of successive deflections, each in itself inappreciable.

### *Ionisation by Röntgen Rays (Plate 8 and 9.)*

The X-ray bulb was excited by a Leyden jar discharge, in most cases so timed that the rays traversed the cloud chamber immediately after the expansion, while the gas was in the supersaturated condition. The ions had thus extremely little time in which to diffuse before being fixed by the condensation of water upon them.

The terminals of the upper stark gap Q of fig. 3 (p. 281 ???) were connected with the inner coatings of two Leyden jars, the outer coatings of which were connected to the Crookes tube. The inner coatings were also connected through glass tubes filled with water to the terminals of the Wimshurst machine. The steel ball in its fall first caused the X-ray discharge and afterwards the illuminating spark, the water tubes having a sufficiently high resistance to prevent the jars supplying the illuminating spark from discharging while the ball traversed the upper spark gap, but conducting sufficiently well to allow of both sets of jars being simultaneously charged by means of the Wimshurst machine.

---

<sup>4</sup>Phil. Mag.,' 1911, vol. 22, p. 551.

The moment of occurrence of the X-ray flash relative to the expansion was adjusted by varying the length of the thread suspending the steel ball. Tests were made of the length of thread required to make the X-ray discharge simultaneous with the completion of the expansion; it was then known that with a thread shorter than this the X-rays would pass after the expansion. For the purpose of making this test the X-ray tube was removed, and the wires supplying it were connected instead to the mercury spark-gap; two illuminating sparks thus traversed the mercury vapour during the fall of the ball, the interval between them being the same as that between the X-ray flash and the illuminating spark, for the same length of thread, in the ordinary use of the apparatus. A series of photographs was taken with different lengths of thread, the camera, placed horizontally, being focussed on a pointer attached to the plunger; a single image of the pointer on the photographic plate indicated that the expansion had been completed before the passage of the first spark, a double image resulting if the first spark passed before the expansion was completed. These photographs also furnished information regarding the rapidity of the expansion; it was found to be completed within about  $1/50$ th of a second.

The Crookes tube was fixed at a distance varying from 30 to 70 cm. from an aperture in the wall of the cloud chamber about 1.2 cm. in diameter; this was closed by a quartz plate 0.38 thick. The quartz window was used as it happened to be convenient for another purpose. The rays passed through a narrow cylindrical channel, in most cases of 2 mm. bore, in a lead block about 5 cm. thick placed close to the quartz window, lead screens being also inserted to shield the rest of the cloud chamber from the rays. The camera occupied position (a) of fig. 4. A horizontal beam of X-rays was in this way made to traverse the cloud chamber across its centre, and was at such a distance from the camera that the magnification was about 2.45 diameters. To avoid distortion by the cylindrical walls of the cloud chamber a portion of the cylinder 5 cm. in length was removed and replaced by a plane parallel glass plate.

Photographs of some typical X-ray clouds are shown in Plate 8. In all cases (with one exception, fig. 5) the rays traversed the supersaturated gas, the order of events being: (1) Production of the supersaturated condition by sudden expansion; (2) Leyden jar discharge through the Crookes tube, causing ionisation within the cloud chamber; (3) condensation of water upon the ions; (4) passage of the illuminating spark, giving a photograph of the cloud condensed on the ions.

The potential difference between the top and bottom of the cloud chamber was in some cases 40 volts, in others only 4 volts, the top being always

positive.

In most cases the expansion ratio was between 1.33 and 1.36; i. e., it considerably exceeded the minimum (approximately 1.31) required to cause condensation on the positive as well as the negative ions (the minimum for the latter is 1.25), but less than is required to give dense clouds in the absence of ions ( $v_2/v_1 = 1.38$ ). Under these conditions, as the photographs show, the tracks of the cathode - or  $\beta$ -particles produced in the gas by the X-rays are very sharply defined, the ions being fixed by condensation of water upon them before they have had time to diffuse, or travel under the action of the electric force, for any appreciable distance.

The following are among the more striking features of the photographs, of which figs. 1 to 4 of Plate 8 are a few examples out of a considerable number obtained under these conditions.

1. Cathode or  $\beta$ -rays are seen to start from within the track of the primary X-ray beam, many of them extending to some distance outside it.

2. There is no indication of any effect of the X-rays on the gas other than the production of the corpuscular radiation; the track of the primary X-ray beam is not distinguishable otherwise than as being the region in which the  $\beta$ -rays have their origin. In some photographs, it is true, there appear scattered throughout the region illuminated by the spark drops which might be taken to represent ions set free by the X-rays, but these show no concentration along the path of the primary beam, and, more over, they appear in equal numbers in comparison photographs taken under conditions otherwise identical but without any X-ray discharge. There is no doubt, I think, that these scattered drops are condensed upon uncharged nuclei similar in nature to those produced by weak ultra-violet light and certain metals, which require a similar expansion to catch them. They appear to be due to a chemical in which some trace of impurity plays an essential part, as they are much more numerous when the air in the apparatus has recently been renewed.

Ionisation by X-rays appears therefore to be, as Bragg has suggested, entirely a secondary process, except in so far as each cathode ray produced in the gas may be said to indicate the formation of one pair of ions by the X-radiation.

3. The number of cathode rays produced in air in a known length of a limited beam of X-rays can readily be counted by this method.

4. The X-radiation thus far used has been heterogeneous. It is to be expected therefore that the cathode rays should be of varying length. Reduced to atmosphere pressure, a frequent length, measured along the path, was from  $3/4$  to 1 cm., or measured in a straight line from beginning to end of

the path, about half these amounts. Tracks as long as 2 cm. were, however, met with.

5. The rays show two distinct kinds of deflection as a result of their encounters with the atoms of the gas – Rutherford's "single" and "compound" scattering. The gradual or cumulative deviation due to successive deflections of very small amount is evidently, however, in this case much the more important factor in causing scattering, all the rays showing a large amount of curvature, while quite a small proportion show abrupt bends. When abrupt deflections occur they are frequently through large angles, 90 degrees or more.

6. The rays tend to become more and bent as the end is approached, the actual, end of each cloud trail being also enlarged into a kind of head, possibly owing to the path of the corpuscle finally becoming extremely irregular in form.

7. In many of the photographs there are cloud trails sufficiently sharply in focus over at least a portion of their length to show the individual drops and, therefore, allow of the ions on which they have condensed being counted. An enlargement of one such tracks is shown in fig. 6 Plate 8; number per cm. of this trail amounts to about 278, the equivalent of 376 ions or 188 pairs of ions at atmospheric pressure. This number appears to be fairly typical for the middle portions of the tracks, i.e. about 5 mm. from the end. Out of 12 counts of this kind, the smallest number obtained is 150 pairs per cm. (at atmospheric pressure) – this is at the beginning of a ray – the largest 2160 pairs per cm. in the last  $\frac{1}{2}$  mm. of a ray.

8. The cathode rays appear to start in all directions. I have not yet attempted any systematical study such as would be required to determine the relative frequency of different initial directions of the rays with respect to the direction of propagation of the Röntgen radiation.

When the expansion ratio is less than about 1.33, the cathode-ray cloud-trails begin to lose their sharpness, as is illustrated by the photographs in Plate 9. With expansion ratios between 1.31 and 1.33, the positive ions are spread out by the electric field before becoming fixed, giving rise to what looks like a shower of drops falling from each trail, which is still marked by the negative ions. When expansion falls below that required to catch the positive ions, the negative ions begin to show a similar spreading out under the action of the field, and finally, while the expansion still considerably exceeds that required to catch negative ions, the clouds cease to give any picture of the original path of the corpuscle. To get the form of the path of an ionising particle as accurately as possible, the expansion ratio ought to exceed 1.33; but, on the other hand, for counting the ions, a smaller

expansion has advantages. An expansion just too small to catch positive ions is perhaps the best for counting the ions; it was only in photographs obtained under such conditions that the ionisation at the ends of the trails could be determined.

When the Röntgen rays are flashed through the cloud chamber before the expansion of the air, diffuse double tracks are obtained, the positive and negative ions being separated by the electric field, and a certain amount of diffusion of the ions occurring in both positive and negative trails (Plate 8, fig. 5). It would have been interesting to obtain by this very direct method a test as to whether the number of positive and negative ions set free is the same, or, in other words, whether the positive and negative ion carry equal charges – a question which has been raised by certain experiments by Townsend. Unfortunately, I have thus far only succeeded in obtaining on the negative a few very short portions of such double tracks, which are at the same time sharply in focus and free from complications due to overlapping with other tracks. These short portions do, however, show equality in the numbers of positive and negative ions; the positive and negative clouds, to take one example, containing each 30 to 31 drops, there being some uncertainty in one or two cases as to whether an image on the plate represents one drop or two.

These experiments have been carried out in the Cavendish Laboratory. I have to thank Mr. F. Lincoln and his assistants in the workshop for most efficient aid in the construction of the apparatus.

### Description of the Plates.

The pictures are photographs of clouds condensed on the ions set free in moist air by rays of different kinds. In what follows,  $\rho_1$  is the density of the air before expansion (relative to saturated air at 15° C. and 760 mm.),  $\rho_2$  the density after expansion,  $v_2/v_1$  the expansion ratio,  $V$  the potential difference between the roof and floor of the cloud chamber, and  $M$  the magnification. In all cases the roof of the cloud chamber was positive, so that positive ions travelled downwards, negative upwards.

### Plate 6.

Ionisation by  $\alpha$ -rays. Axis of camera vertical; horizontal layer of 2 cm. in depth illuminated by mercury spark.

Fig. 1       $\alpha$ -Rays from radium. Some of the  $\alpha$ -particles have traversed the air before the expansion, others after the expansion.

$$\rho_1 = 0.98, \quad v_2/v_1 = 1.36, \quad \rho_2 = 0.72, \quad V = 40 \text{ volts}, \quad M = 1/2.18$$

Fig. 2       $\alpha$ -Rays from radium. The  $\alpha$ -particles have all traversed the air after the expansion

$$\rho_1 = 0.97, \quad v_2/v_1 = 1.33, \quad \rho_2 = 0.73, \quad V = 40 \text{ volts}, \quad M = 1.05$$

Fig. 3       $\alpha$ -Rays from radium. Enlargement of a portion of fig. 2.

$$\rho_1 = 0.97, \quad v_2/v_1 = 1.33, \quad \rho_2 = 0.73, \quad V = 40 \text{ volts}, \quad M = 2.57$$

Fig. 4       $\alpha$ -Rays from radium emanation and active deposit.

$$\rho_1 = 1.00, \quad v_2/v_1 = 1.36, \quad \rho_2 = 0.74, \quad V = 40 \text{ volts}, \quad M = 1/1.24$$

Fig. 5      A complete  $\alpha$ -ray from radium emanation.

$$\rho_1 = 0.97, \quad v_2/v_1 = 1.36, \quad \rho_2 = 0.71, \quad V = 40 \text{ volts}, \quad M = 1.16.$$

### Plate 7.

Ionisation by  $\alpha$  - and  $\beta$ -rays. The source of the rays is on the right of the picture. Axis of camera horizontal (arrangement (a) of p. 281 ???).

Fig. 1       $\alpha$  - and  $\beta$ -Rays from radium.

$$\rho_1 = 0.98, \quad v_2/v_1 = 1.33, \quad \rho_2 = 0.74, \quad V = 30 \text{ volts}, \quad M = 6.0$$

Fig. 2         $\beta$ -Rays produced by  $\gamma$ -radiation.

$$\rho_1 = 1.00, \quad v_2/v_1 = 1.34, \quad \rho_2 = 0.75, \quad V = 40 \text{ volts}, \quad M = 6.0$$

Fig. 3         $\beta$ -Rays from radium.

$$\rho_1 = 0.99, \quad v_2/v_1 = 1.31, \quad \rho_2 = 0.76, \quad V = 40 \text{ volts}, \quad M = 2.45$$

Fig. 4         $\beta$ -Rays. Enlargement of a portion of fig. 3.

$$\rho_1 = 0.99, \quad v_2/v_1 = 1.31, \quad \rho_2 = 0.76, \quad V = 40 \text{ volts}, \quad M = 6.0$$

#### Plate 8.

Ionisation by Röntgen rays. Axis of camera horizontal, X-rays passing from right to left. In all cases except fig. 5 the X-rays traversed the air after the expansion.

Fig. 1        Ionisation by cylindrical X-ray beam about 2 mm. in diameter.

$$\rho_1 = 1.00, \quad v_2/v_1 = 1.35, \quad \rho_2 = 0.74, \quad V = 4 \text{ volts}, \quad M = 2.45$$

Fig. 2        Ionisation by X-ray beam about 2 mm. in diameter.

$$\rho_1 = 1.00, \quad v_2/v_1 = 1.34, \quad \rho_2 = 0.75, \quad V = 4 \text{ volts}, \quad M = 2.45$$

Fig. 3        Ionisation by X-ray beam about 2 mm. in diameter.

$$\rho_1 = 0.93, \quad v_2/v_1 = 1.33, \quad \rho_2 = 0.70, \quad V = 40 \text{ volts}, \quad M = 2.45$$

Fig. 4        Ionisation by X-ray beam about 5 mm. in diameter.

$$\rho_1 = 1.00, \quad v_2/v_1 = 1.36, \quad \rho_2 = 0.74, \quad V = 40 \text{ volts}, \quad M = 2.45$$

Fig. 5        Ionisation by X-ray beam about 5 mm. in diameter. The X-rays traversed the air before its expansion; the positive and negative ions have been separated by the electric field before losing their mobility by the condensation of water upon them.

$$\rho_1 = 1.00, \quad v_2/v_1 = 1.36, \quad \rho_2 = 0.74, \quad V = 40 \text{ volts}, \quad M = 2.45$$

Fig. 6        Portion of fig. 4 enlarged, showing the individual ions produced along a portion of one of the cathode-ray tracks. (The fig. has been turned through  $90^\circ$ .)

$$\rho_1 = 1.00, \quad v_2/v_1 = 1.36, \quad \rho_2 = 0.74, \quad V = 40 \text{ volts}, \quad M = 14.7$$

#### Plate 9.

Ionisation by Röntgen rays. Conditions as in Plate 8; X-ray beam about 2 mm. in diameter. Figs. 2, 3, and 4 belong to a series in which the expansion ratio  $v_2/v_1$  was varied while all other conditions were kept constant.

Fig. 1 Enlargement of a portion of fig. 1, Plate 8.

$$\rho_1 = 1.00, \quad v_2/v_1 = 1.35, \quad \rho_2 = 0.74, \quad V = 4 \text{ volts}, \quad M = 6.0$$

Fig. 2 Enlargement of a portion of fig. 1, Plate 8. In this, as in all the preceding X-ray pictures, the maximum supersaturation attained has been sufficient to cause the ions to lose their mobility immediately after being set free so that the cathode-ray particles leave sharply defined trails.

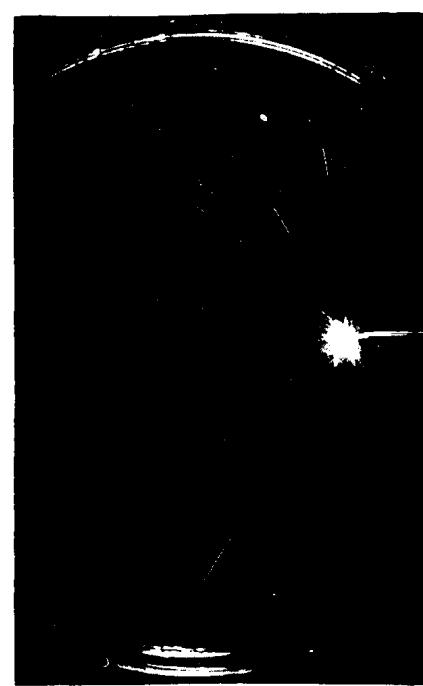
$$\rho_1 = 0.93, \quad v_2/v_1 = 1.33, \quad \rho_2 = 0.70, \quad V = 40 \text{ volts}, \quad M = 6.0$$

Fig. 3 The maximum supersaturation has only slightly exceeded that required to cause condensation on the positive ions, which have therefore travelled varying distance under the action of the electric field before becoming fixed, by condensation of water.

$$\rho_1 = 0.92, \quad v_2/v_1 = 1.31, \quad \rho_2 = 0.70, \quad V = 40 \text{ volts}, \quad M = 6.0$$

Fig. 4 Negative ions, which alone are caught with the maximum degree of supersaturation attained, have retained their mobility for varying lengths of time, the cathode-ray trails being therefore drawn out into diffuse sheets under the action of the electric field.

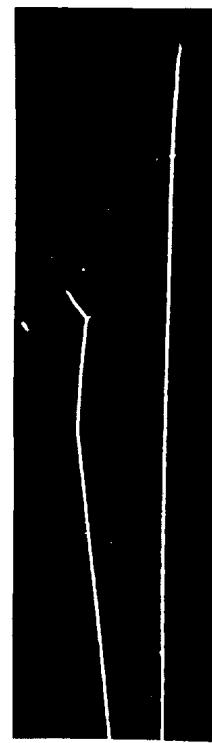
$$\rho_1 = 0.90, \quad v_2/v_1 = 1.28, \quad \rho_2 = 0.70, \quad V = 40 \text{ volts}, \quad M = 6.0$$



1



2



3

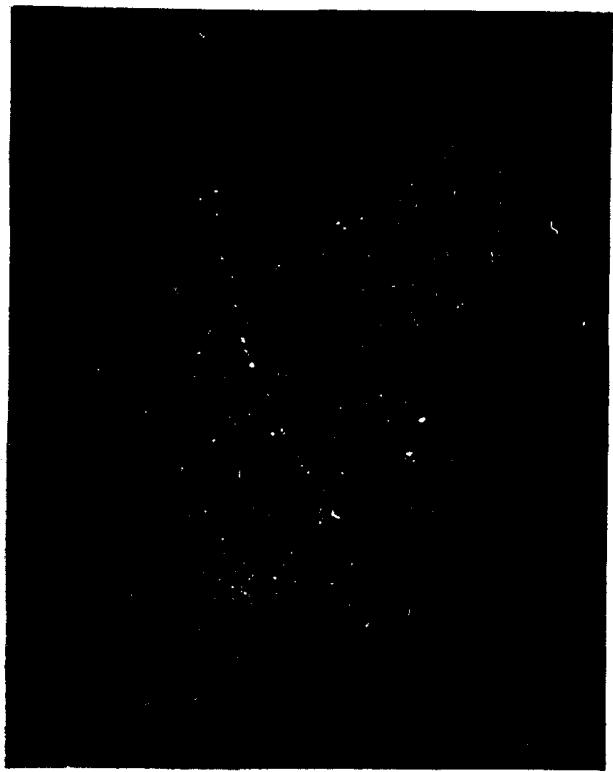


4

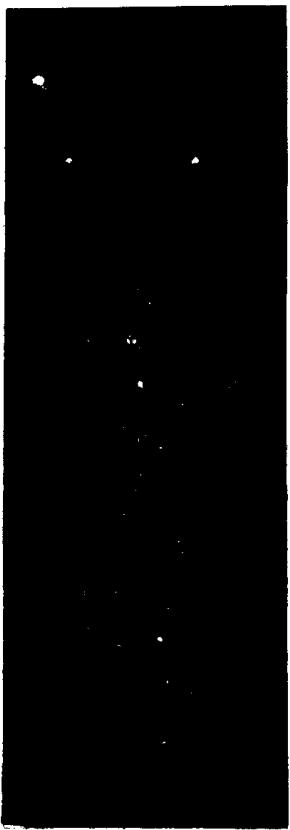


5

Figure 5:



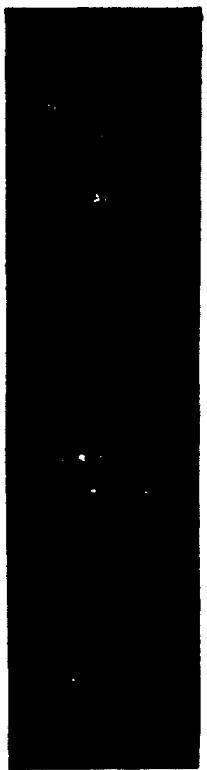
2



4



1



3

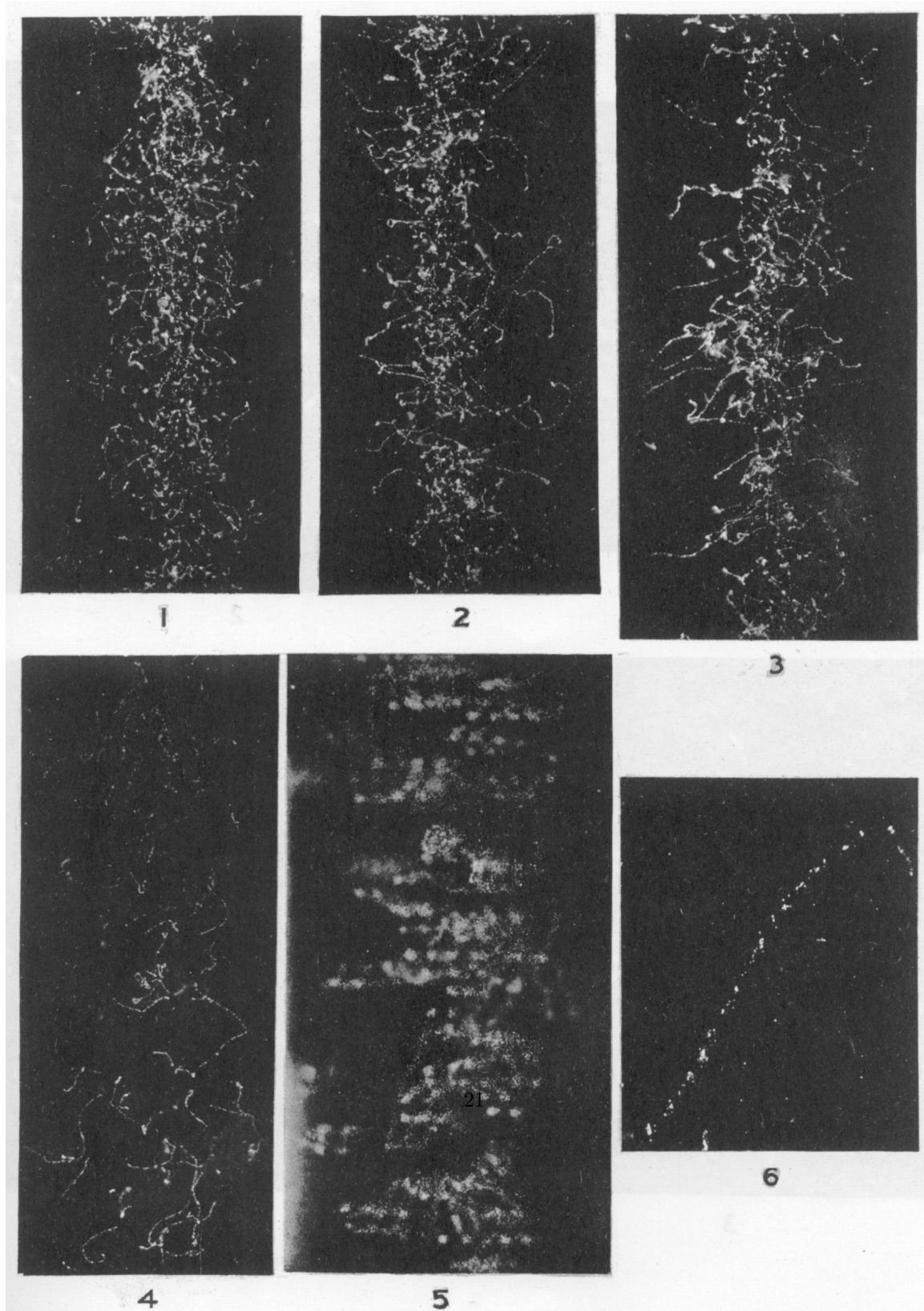


Figure 7:

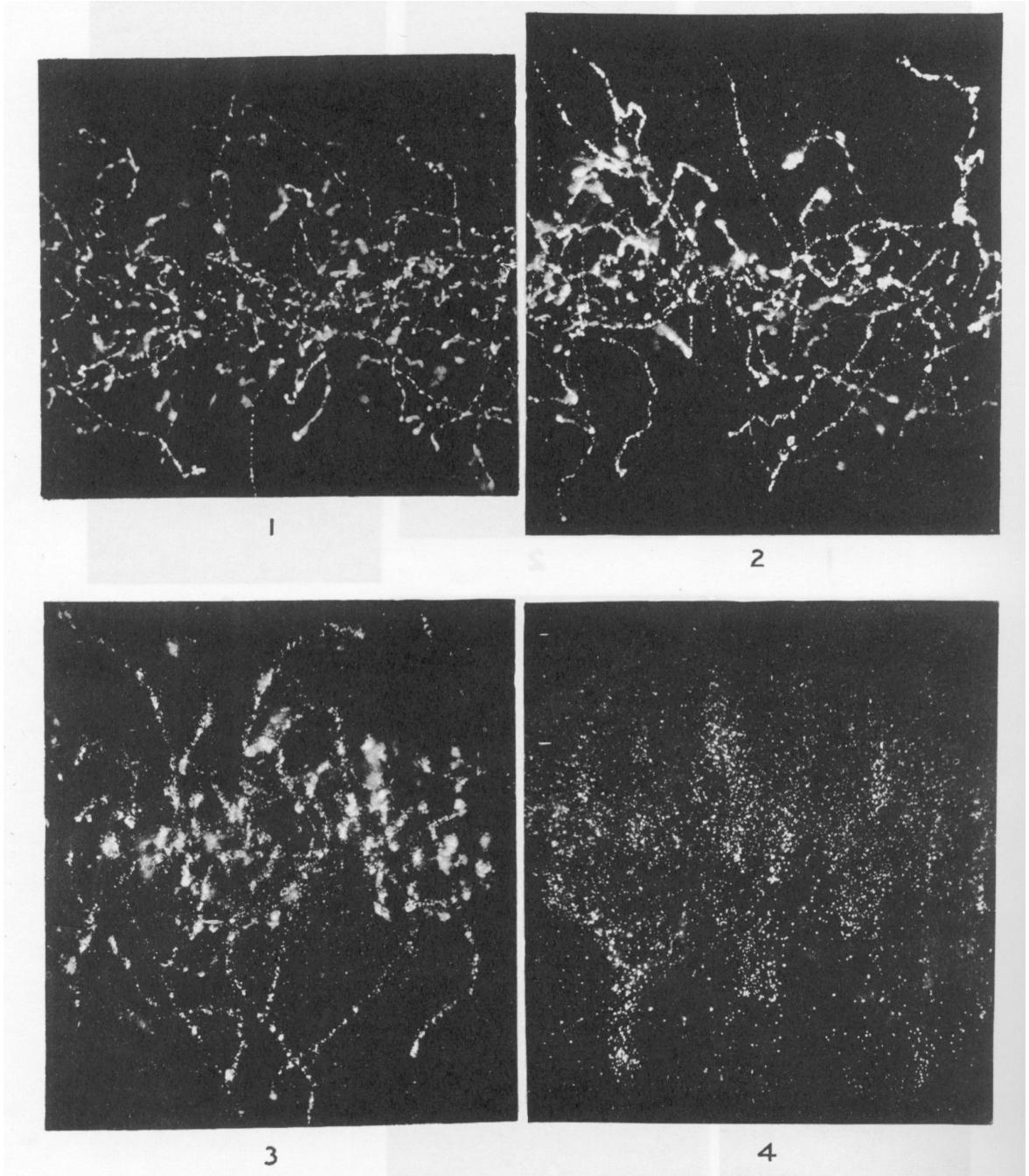


Figure 8:  
22

## Penetrating Radiation in Seven Free Balloon Flights.

V.F. Hess

(Received 1912)

Last year I had the opportunity to investigate the penetrating radiation during two balloon flights. I reported on the first of these to the scientific meeting of Karlsruhe. In both flights no essential variation in the radiation could be observed up to heights of 1,100 meters.

Gockel, also, in two balloon flights, was unable to detect the expected diminution in intensity of the radiation with height. From this it was concluded that besides the  $\gamma$  radiation from the radioactive element in the earth's crust there must be another source of penetrating radiation.

Two Wulf radiation apparatuses of 3 mm. wall thickness were used for the observations.

Apparatus No. 1 had an ionization cell with a value of 2039 cc and a capacity of 1.597 cm. Apparatus No. 2 had a volume of 2970 cc and a capacity of 1.097 cm.

A charge loss corresponding to a decrease of 1 volt per hour thus represented an intensity of ionization of  $q = 1.56$  ions per cc per second in Apparatus No 1 and  $q = 0.7355$  ions per cc per second in Apparatus No. 2.

Whereas all observers of the penetrating radiation on the top of towers have always confirmed a decrease of the penetrating radiation, Gockel and I in balloon flights could not detect such a decrease with certainty. In order to obtain reliable mean values it was necessary to carry out observation in long-lasting flights at modest heights. Parallel observations with a third

thin-wall apparatus were accordingly undertaken to determine if the soft rays behave like the intrinsic radiation.

Further attention had also to be given to the fluctuations of the radiation. Pocini, making parallel observations with two Wulf radiation apparatuses, detected during single hour reading intervals undoubtedly simultaneous fluctuations of the rate of discharge over land as well as over the sea. The cause of the fluctuations therefore clearly lay outside the apparatus and in the radiation itself. It was therefore very important to determine whether this kind of simultaneous fluctuation can also be observed in balloon flights. Since such observations can be carried out with the least difficulty in long-lasting balloon flights at the same height, I have made the major part of the observations during night flights.

The last and most important objective of this investigation was the measurement of the radiation at the greatest possible height. In the six flights originating in Vienna the low-load capacity of the gases used as well as the poor meteorological conditions prevented us from achieving this objective. In a hydrogen flight from Aussing a.d. Elbe succeeded in carrying out measurements to a height of 5350 meters.

Before each flight, control observations were made for several hours with the three sets of instruments. In this procedure the instruments were attached to the balloon cabin in the same manner as during the flight.

In [Table 41-1]  $q_1$ ,  $q_2$ ,  $q_3$  are the reading in ions per cc second of the penetrating radiation made with the instruments 1, 2, and 3 respectively.

**Table 41 - 1**

Trip (26–27 April 1912).

Balloon: "Excelsior."

Pilot: Hauptmann W. Hoffory.

Observer: V.F. Hess.

NN	Time	Mean abso- lute m	Height rela- tive m	Ap- para		Ap- para		Apparatus 3
				tus 1	$q_1$	tus 2	$q_2$	$q_3$
1	16 <sup>h</sup> 40 – 17 <sup>h</sup> 40	156	0	15.6	11.5	—	—	from the
2	17 <sup>h</sup> 40 – 18 <sup>h</sup> 40	156	0	18.7	11.8	21.0	21.0	take-off
3	18 <sup>h</sup> 40 – 21 <sup>h</sup> –	156	0	17.8	11.6	19.5	19.5	point in
4	21 <sup>h</sup> 30 – 22 <sup>h</sup> 30	156	0	17.8	11.3	20.0	20.0	the Klub-
5	23 <sup>h</sup> 26 – 0 <sup>h</sup> 26	300	140	14.4	9.6	19.4	19.8	platz.
6	0 <sup>h</sup> 26 – 1 <sup>h</sup> 26	350	0	16.2	9.9	17.4	17.9	Vienna
7	1 <sup>h</sup> 26 – 2 <sup>h</sup> 26	300	140	14.4	10.1	17.7	18.1	
8	2 <sup>h</sup> 26 – 3 <sup>h</sup> 32	330	160	15.0	9.6	18.2	18.7	
9	3 <sup>h</sup> 32 – 4 <sup>h</sup> 32	320	150	14.4	9.8	18.5	19.0	
10	4 <sup>h</sup> 32 – 5 <sup>h</sup> 35	300	70	17.2	13.2	20.6	21.0	
11	5 <sup>h</sup> 35 – 6 <sup>h</sup> 35	540	240	17.8	11.8	19.6	20.8	
12	6 <sup>h</sup> 35 – 7 <sup>h</sup> 35	1050	800	17.6	10.0	18.1	20.3	
13	7 <sup>h</sup> 35 – 8 <sup>h</sup> 35	1400	1200	12.2	8.8	17.3	20.3	
14	8 <sup>h</sup> 35 – 9 <sup>h</sup> 35	1800	1600	17.5	10.9	17.3	21.3	

The mean height of the balloon during any particular observing interval (as a rule, about an hour) was determined graphically from the barometric trace. A mean value for the height was then obtained from altitude above sea level of the particular spot under the balloon.

[Table 41-1] shows first of all that for small heights above the ground the radiation is really weaker than at the ground itself. If we compute mean values we obtain [Table 41-2].

**Table 41 - 2**

	App. No. 1	App. No. 2	App. No. 3	
Before ascent	$q_1 = 17.5$	$q_2 = 11.55$	$q_3 = 20.2$	ions/cc/second
140 to 190 meters above ground	$q_1 = 14.7$	$q_2 = 9.8$	$q_3 = 18.7$	ions/cc/second

The ion count differences are 2.6, 1.8, and 1.5. The mean difference is thus about 2 ions. This decrease of about 2 ions in the radiation is clearly due to the absorption by the air of  $\gamma$  rays of the radioactive material in the earth's crust. The above-mentioned difference of 2 ions represents about three quarters of the total ionizing power arising from the  $\gamma$  rays of the radioactive material in the earth's crust. The total  $\gamma$  radiation from the earth's crust must thus give rise to about 3 ions per cc per second in the zinc container.

(For the small height of 160 meters we may disregard the possible increase of radiation coming from above).

[The data for several balloon flights are omitted – Editors.]

In order to get an over-all picture of the variation of the penetrating radiations with heights as given by the mean values, I have arranged all the 88 balloon observations that I have made of radiation intensities in vertical steps. Since in this procedure each mean value for a given height is computed from several individual values which were obtained under different conditions and which may be influenced by the temporal fluctuations already discussed, we must not expect at this point to obtain a very exact picture of the variation of the radiation with increasing height. . .

[A discussion of some minor variations of the ion count near the earth's surface is omitted – Editors.]

We see that the  $\gamma$  radiation from the surface of the earth and the air layers close to the earth accounts for the excitation in the zinc containers of about 3 ions per cc per second.

At heights of more than 2,000 meters there is a marked increase in the radiation. It reaches 4 ions from 3,000 to 4,000 meters and 16 to 18 ions from 4,000 to 5,200 meters in two counters. The increase is even stronger in the thin-walled counter No. 3.

What is the source of this penetrating radiation which is observed simul-

taneously in the three counter? . . .

[Here Hess gives various reasons and cogent arguments for dismissing the earth's radioactivity as the source – Editors.]

*The discoveries revealed by the observations here given are best explained by assuming that radiation of great penetrating power enters our atmosphere from the outside and engenders ionization even in counter lying deep in the atmosphere. The intensity of this radiation appears to vary hourly.* Since I found no diminution of this radiation for balloon flights during an eclipse or at night time we can hardly consider the sun as its source.



*On the Constitution of Atoms and Molecules*

N. Bohr,  
*Dr. phil. Copenhagen*  
(Received July 1913)

*Introduction*

In order to explain the results of experiments on scattering of  $\alpha$  rays by matter Prof. Rutherford<sup>1</sup> has given a theory of the structure of atoms. According to this theory, the atom consist of a positively charged nucleus surrounded by a system of electrons kept together by attractive forces from the nucleus; the total negative charge of the electrons is equal to the positive charge of the nucleus. Further, the nucleus is assumed to be the seat of the essential part of the mass of the atom, and to have linear dimensions exceedingly small compared with the linear dimensions of the whole atom. The number of electrons in an atom is deduced to be approximately equal to half the atomic weight. Great interest is to be attributed to this atom-model; for, as Rutherford has shown, the assumption of the existence of nuclei, as those in question, seems to be necessary in order to account for the results of the experiments on large angle scattering of the  $\alpha$  rays.<sup>2</sup>

In an attempt to explain some of the properties of matter on the basis of this atom-model we meet, however, with difficulties of a serious nature arising from the apparent instability of the system of electrons: difficulties purposely avoid in atom-models previously considered, for instance, in the one proposed by Sir. J.J. Thomson<sup>3</sup> According to the theory of the latter the atom consist of a sphere of uniform positive electrification, inside which the electrons move in circular orbits.

<sup>1</sup>E. Rutherford, Phil. Mag. XXI. p. 669 (1911)

<sup>2</sup>See also Geiger and Marsden, Phil. Mag. April 1913.

<sup>3</sup>J.J. Thomson, Phil. Mag. VII. p. 237 (1904).

The principal difference between the atom-models proposed by Thomson and Rutherford consist in the circumstance that the forces acting on the electrons in the atom-model of Thomson allow of certain configurations and motion of the electrons for which the system is in a stable equilibrium; such configurations, however, apparently do not exist for the second atom-model. The nature of the difference in question will perhaps be most clearly seen by noticing that among the quantities characterizing the first atom a quantity appears – the radius of the positive sphere – of dimensions of a length and of the same order of magnitude as the linear extension of the atom, while such a length does not appear among the quantities characterizing the second atom, viz. the charges and masses of the electrons and the positive nucleus; nor can it be determined solely by help of the latter quantities.

The way of considering a problem of this kind has, however, undergone essential alterations in recent years owing to the development of the theory of the energy radiation, and the direct affirmation of the new assumptions introduced in this theory, found by experiments on very different phenomena such as specific heats, photoelectric effect, Röntgen-rays, &c. The result of the discussion of these questions seems to be a general acknowledgment of the inadequacy of the classical electrodynamics in describing the behaviour of system of atomic size.<sup>4</sup> Whatever the alteration in the laws of motion of the electrons may be, it seems necessary to introduce in the laws in question a quantity foreign to the classical electrodynamics, i.e., Planck's constant, or as it often is called the elementary quantum of action. By the introduction of this quantity the question of the stable configuration of the electrons in the atoms is essentially changed, as this constant is of such dimensions and magnitude that it, together with the mass and charge of the particles, can determine a length of the order of magnitude required.

This paper is an attempt to show that the application of the above ideas to Rutherford's atom-model affords a basis for a theory of the constitution of atoms. It will further be shown that from this theory we are led to a theory of the constitution of molecules.

In the present first part of the paper the mechanism of the binding of electrons by a positive nucleus is discussed in relation to Planck's theory. It will be shown that it is possible from the point of view taken to account in a simple way for the law of the line spectrum of hydrogen. Further, reasons are given for a principal hypothesis on which the considerations contained in the following parts are based.

---

<sup>4</sup>See f. inst., "Theorie du ravalement et les quanta." Rapports de la renion a Bruxelles, Nov. 1911, Paris, 1912.

I wish here to express my thanks to Prof. Rutherford for his kind and encouraging interest in this work.

## Part I. – Binding of Electrons by Positive Nuclei.

### § 1. General Considerations

The inadequacy of the classical electrodynamics in accounting for the properties of atoms from an atom-model as Rutherford's, will appear very clearly if we consider a simple system consisting of a positively charged nucleus of very small dimensions and an electron describing closed orbits around it. For simplicity, let us assume that the mass of the electron is negligibly small in comparison with that of the nucleus, and further, that the velocity of the electron is small compared with that of light.

Let us at first assume that there is no energy radiation. In this case the electron will describe stationary elliptical orbits. The frequency of revolution  $\omega$  and the major-axis of the orbit  $2a$  will depend on the amount of energy  $W$  which must be transferred to the system in order to remove the electron to an infinitely great distance apart from the nucleus. Denoting the charge of the electron and of the nucleus by  $-e$  and  $E$  respectively and the mass of the electron by  $m$ , we thus get

$$\omega = \frac{\sqrt{2}}{\pi} \cdot \frac{W^{3/2}}{eE\sqrt{m}}, \quad 2a = \frac{eE}{W}. \quad (1)$$

Further, it can easily be shown that the mean value of the kinetic energy of the electron taken for a whole revolution is equal to  $W$ . We see that if the value of  $W$  is not given, there will be no values of  $\omega$  and  $a$  characteristic for the system in question.

Let us now, however, take the effect of the energy radiation into account, calculated in the ordinary way from the acceleration of the electron. In this case the electron will no longer describe stationary orbits.  $W$  will continuously increase, and the electron will approach the nucleus describing orbits of smaller and smaller dimensions, and with greater and greater frequency; the electron on the average gaining in kinetic energy at the same time as the whole system loses energy. This process will go on until the dimensions of

the orbit are the same order of magnitude as the dimensions of the electron or those of the nucleus. A simple calculation shows that the energy radiated out during the process considered will be enormously great compared with that radiated out by ordinary molecular processes.

It is obvious that the behaviour of such a system will be very different from that of an atomic system occurring in nature. In the first place, the actual atoms in their permanent state have absolutely fixed dimensions and frequencies. Further, if we consider any process, the result seems always to be that after a certain amount of energy characteristic for the systems in question is radiated out, the system will again settle down in a stable state of equilibrium, in which the distance apart of the particles are of the same order of magnitude as before the process.

Now the essential point in Planck's theory of radiation is that the energy radiation from an atomic system does not take place in the continuous way assumed in the ordinary electrodynamics, but that it, on the contrary, takes place in distinctly separated emissions, the amount of energy radiated out from an atomic vibrator of frequency  $\nu$  in a single emission being equal to  $\tau h\nu$ , where  $\tau$  is an entire number, and  $h$  is a universal constant.<sup>5</sup>

Returning to the simple case of an electron and a positive nucleus considered above, let us assume that the electron at the beginning of the interaction with the nucleus was at a great distance apart from the nucleus, and had no sensible velocity relative to the latter. Let us further assume that the electron after interaction has taken place has settled down in a stationary orbit around the nucleus. We shall, for reasons referred to later, assume that the orbit in question is circular: this assumption will, however, make no alteration in the calculations for system containing only a single electron.

Let us now assume that, during the binding of the electron, a homogeneous radiation is emitted of a frequency  $\nu$ , equal to half the frequency of revolution of the electron in its final orbit; then from Planck's theory, we might expect that the amount of energy emitted by the process considered is equal to  $\tau h\nu$ , where  $h$  is Planck's constant an entire number. If we assume that the radiation emitted is homogeneous, the second assumption concerning the frequency of the radiation suggests itself, since the frequency of revolution of the electron at the beginning of the emission is 0. The question, however, of the rigorous validity of both assumptions, and also of the application made of Planck's theory, will be more closely discussed in § 3.

---

<sup>5</sup>See f. inst., M. Planck, Ann. d. Phys. XXXI. p. 758 (1910); XXXVII. p. 612 (1912); Verh. Phys. Ges. 1911, p. 138.

Putting

$$W = \tau h \frac{\omega}{2}, \quad (2)$$

we get by help of the formula (1)

$$W = \frac{2\pi^2 me^2 E^2}{\tau^2 h^2}, \quad \omega = \frac{4\pi^2 me^2 E^2}{\tau^3 h^3}, \quad 2a = \frac{\tau^2 h^2}{2\pi^2 me E}. \quad (3)$$

If in these expressions we give  $\tau$  different values, we get a series of values for  $W$ ,  $\omega$ , and  $a$  corresponding to a series of configurations of the system. According to the above considerations, we are led to assume that these configurations will correspond to states of the system in which there is no radiation of energy; states which consequently will be stationary as long as the system is not disturbed from outside. We see that the value of  $W$  is greatest if  $\tau$  has its smallest value 1. This case will therefore correspond to the most stable of the system, i.e., will correspond to the binding of the electron for the breaking up of which the greatest amount of energy is required.

Putting in the above expressions  $\tau = 1$  and  $E = e$ , and introducing the experimental values

$$e = 4.7 \cdot 10^{-10}, \quad \frac{e}{m} = 5.31 \cdot 10^{17}, \quad h = 6.5 \cdot 10^{-27},$$

we get

$$2a = 1.1 \cdot 10^{-8} \text{ cm}, \quad \omega = 6.2 \cdot 10^{15} \frac{1}{\text{sec}}, \quad \frac{W}{e} = 13 \text{ volt}.$$

We see that these values are of the same order of magnitude as the linear dimensions of the atoms, the optical frequencies, and the ionization- potentials.

The general importance of Planck's theory for the discussion of the behaviour of atomic system was originally pointed out by Einstein.<sup>6</sup> The considerations of Einstein have been developed and applied on a number of different phenomena, especially by Stark, Nernst, and Sommerfield. The agreement as to the order of magnitude between values observed for the frequencies and dimensions of the atoms, and values for these quantities calculated by considerations similar to those given above, has been the subject of much discussion. It was first pointed out by Haas,<sup>7</sup> in an attempt to

---

<sup>6</sup>A. Einstein, Ann. d.Phys. XVII. p. 132 (1905); XX. p. 199 (1906); XXII. p. 180 (1907).

<sup>7</sup>A.E. Haas, Jahrb. d. Rad. u.El. VII. p. 261 (1910). See further, A.Schidlof, Ann. d. Phys. XXXV. p. 90 (1911); E. Wertheimer, Phys. Zietschr. XII. p. 409 (1911), Verh. deutsch. Phys. Ges. 1912, p. 431; F.A. Lindermann, Verh.deutsch.Phys.Ges. 1911, pp. 482, 1107; F. Haber, Verh. deutsch. Phys. Ges. 1911, p. 1117.

explain the meaning and the value of Planck's constant on the basis of J.J. Thomson's atom-model, by help of the linear dimensions and frequency of an hydrogen atom. Systems of the kind considered in this paper, in which the forces between the particles vary inversely as the square of the distance, are discussed in relation to Planck's theory by J.W. Nicholson.<sup>8</sup> In a series of papers this author has shown that it seems to be possible to account for lines of hitherto unknown origin in the spectra of the stellar nebulae and that of the solar corona, by assuming the presence in these bodies of certain hypothetical elements of exactly indicated constitution. The atoms of these elements are supposed to consist simply of a ring of a few electrons surrounding a positive nucleus of negligibly small dimensions. The ratios between the frequencies corresponding to the lines in question are compared with the ratios between the frequencies corresponding to different modes of vibration of the ring of electrons. Nicholson has obtained a relation to Planck's theory showing that the ratios between the wave-lenth of different sets of lines of the coronal spectrum can be accounted for with great accuracy by assuming that the ratio between the energy of the system and the frequency of rotation of the ring is equal to an entire multiple of Planck's constant. The quantity Nicholson refers to as the energy is equal to twice the quantity which we have denoted above by W. In the latest paper cited Nicholson has found it necessary to give the theory a more complicated form, still, however, representing the ratio of energy to frequency by a simple function of whole numbers.

The excellent agreement between the calculated and observed values of the ratios between the wave-length in question seems a strong argument in favour of the validity of the foundation of Nicholson's calculations. Serious objections, however, may be raised against the theory. These objections are intimately connected with the problem of the homogeneity of the radiation emitted. In Nicholson's calculations the frequency of lines in a line-spectrum is identified with the frequency of vibration of a mechanical system in a distinctly indicated state of equilibrium. As a relation from Planck's theory is used, we might expect that the radiation is sent out in quanta; but systems like those considered, in which the frequency is a function of the energy, cannot emit a finite amount of a homogeneous radiation; for, as soon as the emission of radiation is started, the energy and also the frequency of the system are altered. Further, according to the calculation of Nicholson, the systems are unstable for some modes of vibration. Apart from such

---

<sup>8</sup>J.W. Nicholson, Month. Not. Roy. Astr. Soc. LXXII. pp. 49, 139, 677, 693, 729 (1912).

objections – which may be only formal (see p. 23)?????? – it must be remarked, that the theory in the form given does not seem to be able to account for the well-known laws of Balmer and Rydberg connecting the frequencies of the lines in the line-spectra of the ordinary elements.

It will now be attempted to show that the difficulties in question disappear if we consider the problems from the point of view taken in this paper. Before proceeding it may be useful to restate briefly the ideas characterizing the calculations on p. 5. The principal assumptions used are:

- (1) That the dynamical equilibrium of the systems in the stationary states can be discussed by help of the ordinary mechanics, while the passing of the systems between different stationary states cannot be treated on that basis.
- (2) That the latter is followed by the emission of a *homogeneous* radiation, for which the relation between the frequency and the amount of energy emitted is the one given by Planck's theory.

The first assumption seems to present itself; for it is known that the ordinary mechanism cannot have an absolute validity, but will only hold in calculations of certain mean values of the motion of the electrons. On the other hand, in the calculations of the dynamical equilibrium in a stationary state in which there is no relative displacement of the particles, we need not distinguish between the actual motions and their mean values. The second assumption is in obvious constant to the ordinary ideas of electrodynamics, but appears to be necessary in order to account for experimental facts.

In the calculations on page 5 we have further made use of the more special assumptions, viz., that the different stationary states correspond to the emission of a different number of Planck's energy-quanta, and that the frequency of the radiation emitted during the passing of the system from a state in which no energy is yet radiated out to one of the stationary states, is equal to half the frequency of revolution of the electron in the latter state. We can, however (see § 3), also arrive at the expressions (3) for the stationary states by using assumptions of somewhat different from. We shall, therefore, postpone the discussion of the spacial assumptions, and first show how by the help of the above principal assumptions, and of the expressions (3) for the stationary states, we can account for the line-spectrum of hydrogen.

## § 2. Emission of Line-spectra

*Spectrum of Hydrogen.* – General evidence indicates that an atom of hydrogen consist simply of a single electron rotating round a positive nucleus of charge  $e$ .<sup>9</sup> The reformation of a hydrogen atom, when the electron has been removed to great distances away from the nucleus – e.g. by the effect of electrical discharge in a vacuum tube – will accordingly correspond to the binding of an electron by a positive nucleus considered on p. 5. If in (3) we put  $E = e$ , we get for the total amount of energy radiated out by the formation of one of the stationary states,

$$W_r = \frac{2\pi^2 me^4}{\tau^2 h^2}.$$

The amount of energy emitted by the passing of the system from a state corresponding to  $\tau = \tau_1$  to one corresponding to  $\tau = \tau_2$ , is consequently

$$W_{r_2} - W_{r_1} = \frac{2\pi^2 me^4}{h^2} \cdot \left( \frac{1}{\tau_2^2} - \frac{1}{\tau_1^2} \right).$$

If now we suppose that the radiation in question is homogeneous, and that the amount of energy emitted is equal to  $h\nu$ , where  $\nu$  is the frequency of the radiation, we get

$$W_{r_2} - W_{r_1} = h\nu$$

and from this

$$\nu = \frac{2\pi^2 me^4}{h^3} \cdot \left( \frac{1}{\tau_2^2} - \frac{1}{\tau_1^2} \right). \quad (4)$$

We see that this expression accounts for the law connecting the lines in the spectrum of hydrogen. If we put  $\tau_2 = 2$  and let  $\tau_1$  vary, we get the ordinary Balmer series. If we put  $\tau_3 = 3$ , we get the series in the ultra-red observed by Paschen<sup>10</sup> and previously suspected by Ritz. If we put  $\tau_2 = 1$  and  $\tau = 4, 5, \dots$ , we get series respectively in the extreme ultraviolet and the extreme ultra-red, which are not observed, but the existence of which may be expected.

---

<sup>9</sup>See f. inst. N. Bohr, Phil. Mag. XXV. p. 24 (1913). The conclusion drawn in the paper cited is strongly supported by the fact that hydrogen, in the experiments on positive rays of Sir. J.J. Thomson, is the only element which never occurs with a positive charge corresponding to the loss of more than one electron (comp. Phil. Mag. XXIV. p. 672 (1912)).

<sup>10</sup>F. Paschen, Ann. d. Phys. XXVII. p. 565 (1908).

The agreement in question is quantitative as well as qualitative. Putting

$$e = 4.7 \cdot 10^{-10}, \quad \frac{e}{m} = 5.31 \cdot 10^{17} \quad \text{and} \quad h = 6.5 \cdot 10^{-27},$$

we get

$$\frac{2\pi^2 me^4}{h^3} = 3.1 \cdot 10^{15}.$$

The observed value for the factor outside the bracket in the formula (4) is

$$3.290 \cdot 10^{15}.$$

We agreement between the theoretical and observed values is inside the uncertainty due to experimental errors in the constants entering in the expression for the theoretical value. We shall in § 3 return to consider the possible importance of the agreement in question.

It may be remarked that the fact, that it has not been possibly to observe more than 12 lines of the Balmer series in experiments with vacuum tubes, while 33 lines are observed in the spectra of some celestial bodies, is just what we should expect from the above theory. According to the equation (3) the diameter of the orbit of the electron in the different stationary states is proportional to  $\tau^2$ . For  $\tau = 12$  the diameter is equal to  $1.6 \cdot 10^{-6}$  cm, or equal to mean distance between the molecules in a gas at a pressure of about 7 mm mercury; for  $\tau = 33$  the diameter is equal to  $1.2 \cdot 10^{-5}$  cm, corresponding to the mean distance of the molecules at a pressure of about 0.02 mm mercury. According to the theory the necessary condition for the appearance of a great number of lines is therefore a very small density of the gas; for simultaneously to obtain an intensity sufficient for observation the space filled with the gas must be very great. If the theory is right, we may therefore never expect to be able in experiments with vacuum tubes to observe the lines corresponding to high numbers of the Balmer series of the emission spectrum of hydrogen; it might, however, be possible to observe the lines by investigation of the absorption spectrum of this gas. (see § 4).

It will be observed that we in the above way do not obtain other series of lines, generally ascribed to hydrogen; for instance, the series first observed by Pickering<sup>11</sup> in the spectrum of the star  $\zeta$  Puppis, and the set of series recently found by Fowler<sup>12</sup> by experiments with vacuum tubes containing a mixture of hydrogen and helium. We shall, however, see that, by help of the above theory, we can account naturally for these series of lines if we ascribe them to helium.

---

<sup>11</sup>E.C. Pickering, *Astrophys. J.* IV. p. 369 (1896); v. p. 92 (1897).

<sup>12</sup>A. Fowler, *Mouth. Not. Roy. Astr. Soc.* LXXIII. Dec. 1912.

A neutral atom of the latter element consists, according to Rutherford's theory, of a positive nucleus of charge  $2e$  and two electrons. Now considering the binding of a single electron by a helium nucleus, we get putting  $E = 2e$  in the expressions (3) on page 5, and proceeding in exactly the same way as above,

$$\nu = \frac{8\pi^2 me^4}{h^3} \cdot \left( \frac{1}{\tau_2^2} - \frac{1}{\tau_1^2} \right) = \frac{2\pi^2 me^4}{h^3} \cdot \left( \frac{1}{(\frac{\tau_2}{2})^2} - \frac{1}{(\frac{\tau_1}{2})^2} \right).$$

If we in this formula put  $\tau_1 = 1$  or  $\tau_2 = 2$ , we get series of lines in the extreme ultra-violet. If we put  $\tau_2 = 3$ , and let  $\tau_1$  vary, we get a series which includes 2 of the series observed by Folwer, and denoted by him as the first and second principal series of the hydrogen spectrum. If we put  $\tau_2 = 4$ , we get the series observed by Pickering in the spectrum of  $\zeta$  Puppis. Every second of the lines in this series is identical with a line in the Balmer series of the hydrogen spectrum; the presence of hydrogen in the star in question may therefore account for the fact that these lines are of a greater intensity than the rest of the lines in the series. The series is also observed in the experiments of Fowler, and denoted in his paper as the Sharp series of the hydrogen spectrum. If we finally in the above formula put  $\tau_2 = 5, 6, \dots$ , we get series, the strong lines of which are to be expected in the ultra-red.

The reason why the spectrum considered is not observed in ordinary helium tubes may be that in such tubes the ionization of helium is not so complete in the star considered or in the experiments of Fowler, where a strong discharge was sent through a mixture of hydrogen and helium. The condition for the appearance of the spectrum is, according to the above theory, that helium atoms are present in a state in which they have lost both their electrons. Now we must assume that the amount of energy to be used in removing the second electron from a helium atom is much greater than that to be used in removing the first. Further, it is known from experiments on positive rays, that hydrogen atoms can acquire a negative charge; therefore the presence of hydrogen in the experiments of Fowler may effect that more electrons are removed from some of the helium atoms than would be the case if only helium were present.

*Spectra of other substances.* — in case of systems containing more electrons we must — in conformity with the result of experiments — expect more complicated laws for the line-spectra than those considered. I shall try to show that the point of view taken above allows, at any rate, a certain understanding of the laws observed. According to Rydberg's theory — with the

generalization given by Ritz<sup>13</sup> – the frequency corresponding to the lines of the spectrum of an element can be expressed by

$$\nu = F_\tau(\tau_1) - F_s(\tau_2),$$

where  $\tau_1$  and  $\tau_2$  are entire numbers, and  $F_1, F_2, F_3, \dots$  are functions of  $\tau$  which approximately are equal to  $\frac{K}{(\tau+a_1)^2}, \frac{K}{(\tau+a_2)^2}, \dots$   $K$  is a universal constant, equal to the factor outside the bracket in the formula (4) for the spectrum of hydrogen. The different series appear if we put  $\tau_1$  or  $\tau_2$  equal to a fixed number and let the other vary.

The circumstance that the frequency can be written as a difference between two functions of entire numbers suggests an origin of the lines in the spectra in question similar to the one we have assumed for hydrogen; i.e. that the lines correspond to a radiation emitted during the passing of the system between two different stationary states. For system containing more than one electron the detailed discussion may be very complicated, as there will be many different configurations of the electrons which can be taken into consideration as stationary states. This may account for the difference sets of series in the line spectra emitted from the substances in question. Here I shall only try to show how, by help of the theory, it can be simple explained that the constant  $K$  entering in Rydberg's formula is the same for all substances. Let us assume that the spectrum in question corresponds to the radiation emitted during the binding of an electron; and let us further assume that the system including the electron considered is neutral. The force on the electron, when at a great distance apart the nucleus and the electrons previously bound, will be very nearly the same as the above case of the binding of an electron by a hydrogen nucleus. The energy corresponding to one of the stationary states will therefore for  $\tau$  great be very nearly equal to that given by the expression (3) on p. 5, if we put  $E = e$ . For  $\tau$  great we consequently get

$$\lim[\tau^2 \cdot F_1(\tau)] = \lim[\tau^2 \cdot F_2(\tau)] = \dots = \frac{2\pi^2 me^4}{h^3},$$

in conformity with Rydberg's theory.

---

<sup>13</sup>W. Ritz, Phys. Zeitschr. IX. p. 521 (1908).

### § 3. General Considerations Continued

We shall now return to the discussion (see p. 7) of the special assumptions used in deducing the expression (3) on p. 5 for the stationary states of a system consisting of an electron rotating round a nucleus.

For one, we have assumed that the different stationary states correspond to an emission of a different number of energy-qyanta. Considering systems in which the frequency is a function of the energy, this assumption, however, may be regarded as improbable; for as soon as one quantum is sent out the frequency is altered. We shall now see that we can leave the assumption used and still retain the equation (2) on p. 5, and thereby the formal analogy with Planck's theory.

Firstly, it will be observed that it has not been necessary, in order to account for the law of the spectra by help of the expressions (3) for the stationary states, to assume that in any case a radiation is sent out corresponding to more than a single energy-quantum,  $h\nu$ . Further information on the frequency of the radiation may be obtained by comparing calculations of the energy radiation in the region of slow vibrations based on the above assumptions with calculations based on the ordinary mechanics. As is known, calculations on the latter basis are in agreement with experiments on the energy radiation in the named region.

Let us assume that the ratio between the total amount of energy emitted and the frequency of revolution of the electron for the different stationary states is given by the equation  $W = f(\tau) \cdot h\omega$ , instead of by the equation (2). Proceeding in the same way as above, we get in this case instead of (3)

$$W = \frac{\pi^2 meE^2}{2h^2 f^2(\tau)}, \quad \omega = \frac{\pi^2 me^2 E^2}{2h^3 f^3(\tau)}.$$

Assuming as above that the amount of energy emitted during the passing of the system from a state corresponding to  $\tau = \tau_1$  to one for which  $\tau = \tau_2$  is equal to  $h\nu$ , we get instead of (4)

$$\nu = \frac{\pi^2 me^2 E^2}{2h^3} \cdot \left( \frac{1}{f^2(\tau_2)} - \frac{1}{f^2(\tau_1)} \right).$$

We see that in order to get an expression of the same form as the Balmer series we must put  $f(\tau) = c\tau$ .

In order to determine  $c$  let us now consider the passing of the system between two successive stationary states corresponding to  $\tau = N$  and  $\tau =$

$N - 1$ ; introducing  $f(\tau) = c\tau$ , we get for the frequency of the radiation emitted

$$\nu = \frac{\pi^2 me^2 E^2}{2c^2 h^3} \cdot \frac{2N - 1}{N^2(N - 1)^2}.$$

For the frequency of revolution of the electron before and after the emission we have

$$\omega_N = \frac{\pi^2 me^2 e^2}{2c^3 h^3 N^3} \quad \text{and} \quad \omega_{N-1} = \frac{\pi^2 me^2 E^2}{2c^3 h^3 (N - 1)^3}.$$

If  $N$  is great the ratio between the frequency before and after the emission will be very near equal to 1; and according to the ordinary electrodynamics we should therefore expect that the ratio between the frequency of radiation and the frequency of revolution also very nearly equal to 1. This condition will only be satisfied if  $c = 1/2$ . Putting  $f(\tau) = \tau/2$ , we, however, again arrive at the equation (2) and consequently at the expression (3) for the stationary states.

If we consider the passing of the system between two states corresponding to  $\tau = N$  and  $\tau = N - n$ , where  $n$  is small compared with  $N$ , we get with the same approximation as above, putting  $f(\tau) = \tau/2$ ,

$$\nu = n\omega.$$

The possibility of an emission of a radiation of such a frequency may also be interpreted from analogy with the ordinary electrodynamics, as an electron rotating round a nucleus in an elliptical orbit will emit a radiation which according to Fourier's theorem can be resolved into homogeneous components, the frequency of which are  $n\omega$ , if  $\omega$  is the frequency of revolution of the electron.

We are thus led to assume that the interpretation of the equation (2) is not that the different stationary states correspond to an emission of different numbers of energy-quanta, but that the frequency of the energy emitted during the passing of the system from a state in which no energy is yet radiated out to one of the different stationary states, is equal to different multiples of  $\omega/2$ , where  $\omega$  is the frequency of revolution of the electron in the state considered. From this assumption we get exactly the same expressions as before for the stationary states, and from these by help of the principal assumptions on p. 7 the same expression for the law of the hydrogen spectrum. Consequently we may regard our preliminary considerations on p. 5 only as a simple form of representing the results of the theory.

Before we leave the discussion of this question, we shall for a moment return to the question of the significance of the agreement between the observed and calculated values of the constant entering in the expressions (4) for the Balmer series of the hydrogen spectrum. From the above consideration it will follow that, taking the starting-point in the form of the law of the hydrogen spectrum and assuming that the different lines correspond to a homogeneous radiation emitted during the passing between different, stationary states, we shall arrive at exactly the same expression for the constant in question as that given by (4), if we only assume (1) that the radiation is sent out in quanta  $h\nu$ , and (2) that the frequency of the radiation emitted during the passing of the system between successive stationary states will coincide with the frequency of revolution of the electron in the region of slow vibrations.

As all the assumptions used in this latter way of representing the theory are of what we may call a qualitative character, we are justified in expecting — if the whole way of considering is a sound one — an absolute agreement between the values calculated and observed for the constant in question, and not only an approximate agreement. The formula (40) may therefore be of value in the discussion of the results of experimental determinations of the constants  $e$ ,  $m$ , and  $h$ .

While there obviously can be no question of a mechanical foundation of the calculations given in this paper, it is, however, possible to give a very simple interpretation of the result of the calculation on p. 5 by help of symbols taken from the ordinary mechanics. Denoting the angular momentum of the electron round the nucleus by  $M$ , we have immediately for a circular orbit  $\pi M = T/\omega$ , where  $\omega$  is the frequency of revolution and  $T$  the kinetic energy of the electron; for a circular orbit we further have  $T = W$  (see p. 3) and from (2), p. 5, we consequently get

$$M = \tau M_0,$$

where

$$M_0 = \frac{h}{2\pi} = 1.04 \cdot 10^{-27}.$$

If we therefore assume that the orbit of the electron in the stationary states is circular, the result of the calculation on p. 5 can be expressed by the simple condition: that the angular momentum of the electron round the nucleus in a stationary state of the system is equal to an entire multiple of a universal value, independent of the charge on the nucleus. The possible importance of the angular momentum in the discussion of atomic systems

in relation to Planck's theory is emphasized by Nicholson.<sup>14</sup>

The great number of different stationary states we do not observe expect by investigation of the emission and absorption of radiation. In most of the other physical phenomena, however, we only observe the atoms of the matter in a single distinct state, i.e., the state of the atoms at low temperature. From the preceding considerations we are immediately led to the assumption that the "permanent" state is the one among the stationary states during the formation of which the greatest amount of energy is emitted. According to the equation (3) on p. 5, this state is the one which corresponds to  $\tau = 1$ .

#### *§ 4. Absorption of Radiation*

In order to account for Kirchhoff's law it is necessary to introduce assumptions on the mechanism of absorption of radiation which correspond to those we have used considering the emission. Thus we must assume that a system consisting of a nucleus and an electron rotating round it under certain circumstances can absorb a radiation of a frequency equal to the frequency of the homogenous radiation emitted during the passing of the system between different stationary states. Let us consider the radiation emitted during the passing of the system between two stationary states  $A_1$  and  $A_2$  corresponding to values for  $\tau$  equal to  $\tau_1$  and  $\tau_2$ ,  $\tau_1 > \tau_2$ . As the necessary condition of the radiation in question was the presence of systems in the state  $A_1$ , we must assume that the necessary condition for an absorption of the radiation is the presence of systems in the state  $A_2$ .

These considerations seems to be in conformity with experiments on absorption in gases. In hydrogen gas at ordinary conditions for instance there is no absorption of a radiation of a frequency corresponding to the line-spectrum of this gas; such an absorption is only observed in hydrogen gas in a luminous state. This is what we should expect according to the above. We have on p. 9 assumed that the radiation in question was emitted during the passing of the systems between stationary states corresponding to  $\tau \geq 2$ . The state of the atoms in hydrogen gas at ordinary conditions should, however, correspond to  $\tau = 1$ ; furthermore, hydrogen atoms at ordinary conditions combine into molecules, i.e., into system in which the electrons have frequencies different from those in the atoms (see Part III.) From the circumstance that certain substances in a non-luminous state, as,

---

<sup>14</sup>J.W. Nicholson, loc. cit. p. 679.

for instance, sodium vapour, absorb radiation corresponding to lines in the line-spectra of the substances, we may, on the other hand, conclude that the lines in question are emitted during the passing of the system between two states, one of which is the permanent state.

How much the above considerations differ from an interpretation based on the ordinary electrodynamic of perhaps most early shown by the fact that we have been forced to assume that a system of electrons will absorb a radiation of a frequency different from the frequency of vibration of the electrons calculated in the ordinary way. It may in this connexion be of interest to mention a generalization of the considerations to which we are led by experiments on the photo-electric effect and which may be able to throw some light on the problem in question. Let us consider state of the system in which the electron is free, i.e., in which the electron possesses kinetic energy sufficient to remove to infinite distances from the nucleus. If we assume that the motion of the electron is governed by the ordinary mechanics and that there is no (sensible) energy radiation, the total energy of the system – as in the above considered stationary states – will be constant. Further, there will be perfect continuity between the two kinds of states, as the difference between frequency and dimensions of the system in successive stationary states will diminish without limit if  $\tau$  increases. In the following considerations we shall for the sake of brevity refer to the two kinds of states in question as “mechanical” states; by this notation only emphasizing the assumption that the motion of the electron in both cases can be assumed for by the ordinary mechanics.

Tracing the analogy between the two kinds of mechanical states, we might now expect the possibility of an absorption of radiation, not only corresponding to the passing of the system between two different stationary states, but also corresponding to the passing between one of the stationary states and a state in which the electron is free; and as above, we might expect that the frequency of this radiation was determined by the equation  $E = h\nu$ , where  $E$  is the difference between the total energy of the system in the two states. As it will be seen, such an absorption of radiation is just what is observed in experiments on ionization by ultra-violet light and by Röntgen rays. Obviously, we get in this way the same expression for the kinetic energy of an electron ejected from an atom by photo-electron effect as that deduced by Einstein<sup>15</sup> i.e.,  $T = h\nu - W$ , where  $T$  is the kinetic energy of the electron ejected, and  $W$  the total amount of energy emitted during the original binding of the electron.

---

<sup>15</sup>A. Einstein, Ann. d. Phys. XVII. p. 146 (1905).

The above considerations may further account for the result of some experiments of R.W. Wood<sup>16</sup> on absorption of light by sodium vapour. In these experiments, an absorption corresponding to a very great number of lines in the principal series of the sodium spectrum is observed, and in addition a continuous absorption which begins at the head of the series and extends to the extreme ultra-violet. This is exactly what we should expect according to the analogy in question, and, as we shall see, a closer consideration of the above experiments allows us to trace the analogy still further. As mentioned on p. 9 the radii of the orbits of the electrons will for stationary states, corresponding to high values for  $\tau$  be very great compared with ordinary atomic dimensions. This circumstance was used as an explanation of the non-appearance in experiments with vacuum-tubes of lines corresponding to the higher numbers in the Balmer series of the hydrogen spectrum. This is also in conformity with experiments on the emission spectrum of sodium; in the principal series of the emission spectrum of this substance rather few lines are observed. Now in Wood's experiments the pressure was not very low, the states corresponding to high values for  $\tau$  could therefore not appear; yet in the absorption spectrum about 50 lines were detected. In the experiments in question we consequently observe an absorption of radiation which is not accompanied by a complete transition between two different stationary states. According to the present theory we must assume that this absorption is followed by an emission of energy during which the systems pass back to the original stationary state. If there are no collisions between the different systems this energy will be emitted as a radiation of the same frequency as that absorbed, and there will be no true absorption but only a scattering of the original radiation; a true absorption will not occur unless the energy in question is transformed by collisions into kinetic energy of free particles. In analogy we may now from the above experiments conclude that a bound electron – also in cases in which there is no ionization – will have an absorbing (scattering) influence on a homogeneous radiation, as soon as the frequency of the radiation is greater than  $W/h$ , where  $W$  is the total amount of energy emitted during the binding of the electron. This would be highly in favour of a theory of absorption as the one sketched above, as there can in such a case be no question of a coincidence of the frequency of the radiation and a characteristic frequency of vibration of the electron. It will further be seen that the assumption, that there will be an absorption (scattering) of any radiation corresponding to a transition between two different mechanical states, is in perfect analogy with the

---

<sup>16</sup>R.W. Wood, Physical Optics, p. 513 (1911).

assumption generally used that a free electron will have an absorbing (scattering) influence on light of any frequency. Corresponding considerations will hold for the emission of radiation.

In analogy to the assumption used in this paper that the emission of line-spectra is due to the re-formation of atoms after one or more of the lightly bound electrons are removed, we may assume that the homogeneous Röntgen radiation is emitted during the setting down of the systems after one of the firmly bound electrons escapes, e.g. by impact of cathode particles.<sup>17</sup> In the next part in this paper, dealing with the constitution of atoms, we shall consider the question more closely and try to show that a calculation based on this assumption is in quantitative agreement with the results of experiments: here we shall only mention briefly a problem with which we meet in such a calculation.

Experiments on the phenomena of X-rays suggest that not only the emission and absorption of radiation cannot be treated by the help of the ordinary electrodynamics, but not even the result of a collision between two electrons of which the one is bound in an atom. This is perhaps most easily shown by some very instructive calculations on the energy of  $\beta$ -particles emitted from radioactive substances recently published by Rutherford.<sup>18</sup> These calculations strongly suggest that an electron of great velocity in passing through an atom and colliding with the electrons bound will lose energy in distinct finite quanta. As is immediately seen, this is very different from what we might expect if the result of the collisions was governed by the usual mechanical laws. The failure of the classical mechanics in such a problem might also be expected beforehand from the absence of anything like equipartition of kinetic energy between free electrons and electrons bound in atoms. From the point of view of the "mechanical" states we see, however, that the following assumption – which is in accord with the above analogy – might be able to account for the result of Rutherford's calculation and for the absence of equipartition of kinetic energy; two colliding electrons, bound or free, will, after the collision as well as before, be in mechanical states. Obviously, the introduction of such an assumption would not make any alteration necessary in the classical treatment of a collision between two free particles. But, considering a collision between a free and a bound electron, it would follow that the bound electron by the collision could not acquire a less amount of energy than the difference in energy corresponding to successive stationary states, and consequently that the free electron which collides with it could

---

<sup>17</sup> Compare J.J. Thomson, Phil. Mag. XXIII. p. 456 (1912).

<sup>18</sup> E. Rutherford, Phil. Mag. XXIV. pp. 453 & 893 (1912).

not lose a less amount.

The preliminary and hypothetical character of the above considerations needs not to be emphasized. The intention, however, has been to show that the sketched generalization of the theory of the stationary states possibly may afford a simple basis of representing a number of experimental facts which cannot be explained by help of the ordinary electrodynamics, and that assumptions used do not seem to be inconsistent with experiments on phenomena for which a satisfactory explanation has been given by the classical dynamics and the wave theory of light.

### *§ 5. The permanent State of an Atomic System*

We shall now return to the main object of this paper – the discussion of the “permanent” state of a system consisting of nuclei and bound electrons. For a system consisting of a nucleus and an electron rotating round it, this state is, according to the above, determined by the condition that the angular momentum of the electron round the nucleus is equal to  $h/2\pi$ .

On the theory of this paper the only neutral atom which contains a single electron is the hydrogen atom. The permanent state of this atom should correspond to the values of  $a$  and  $\omega$  calculated on p. 5. Unfortunately, however, we know very little of the behaviour of hydrogen atoms on account of the small dissociation of hydrogen molecules at ordinary temperatures. In order to get a closer comparison with experiments, it is necessary to consider more complicated systems.

Considering systems in which more electrons are bound by a positive nucleus, a configuration of the electrons which presents itself as a permanent state is in which the electrons are arranged in a ring round the nucleus. In the discussion of this problem on the basis of the ordinary electrodynamics, we meet – apart from the question of the energy radiation – with new difficulties due to the question of the stability of the ring. Disregarding for a moment this latter difficulty, we shall first consider the dimensions and frequency of the systems in relation to Planck’s theory of radiation.

Let us consider a ring consisting of  $n$  electrons rotating round a nucleus of charge  $E$ , the electrons being arranged at equal angular intervals the circumference of a circle of radius  $a$ .

The total potential energy of the system consisting of the electrons and

the nucleus is

$$P = -\frac{ne}{a} \cdot (E - es_n),$$

where

$$s_n = \frac{1}{4} \sum_{s=1}^{s=n-1} \operatorname{cosec} \frac{s\pi}{n}.$$

For the radial force exerted on an electron by the nucleus and the other electrons we get

$$F = -\frac{1}{n} \cdot \frac{dP}{da} = -\frac{e}{a^2} \cdot (E - es_n).$$

Denoting the kinetic energy of an electron by  $T$  and neglecting the electromagnetic forces due to the motion of the electrons (see Part II), we get, putting the centrifugal force on an electron equal to the radial force,

$$\frac{2T}{a} = \frac{e}{a^2} \cdot (E - es_n),$$

or

$$T = \frac{e}{2a} \cdot (E - es_n).$$

From this we get for the frequency of revolution

$$\omega = \frac{1}{2\pi} \cdot \sqrt{\frac{e(E - es_n)}{ma^3}}.$$

The total amount of energy  $W$  necessary transferred to the system in order to remove the electrons to infinite distances apart from the nucleus and from each other is

$$W = -P - nT = \frac{ne}{2a} \cdot (E - es_n) = nT,$$

equal to the total kinetic energy of the electrons.

We see that the only difference in the above formula and those holding for the motion of a single electron in a circular orbit round a nucleus is the exchange of  $E$  for  $E - es_n$ . It is also immediately seen that corresponding to the motion of an electron in an elliptical orbit round a nucleus, there will be a motion of the  $n$  electrons in which each rotates in an elliptical orbit with the nucleus in the focus, and the  $n$  electrons at any moment are situated at equal angular intervals on a circle with the nucleus as the centre. The major axis and frequency of the orbit of the single electrons will for this motion be given by the expressions (1) on p. 3 if we replace  $E$  by  $E - es_n$  and  $W$  by  $W/n$ . Let us now suppose that the system of  $n$  electrons rotating in a ring round a nucleus is formed in a way analogous to the one assumed for

a single electron rotating round a nucleus. It will thus be assumed that the electrons, before the binding by the nucleus, were at a great distance apart from the latter and possessed no sensible velocities, and also that during the binding a homogeneous radiation is emitted. As in the case of a single electron, we have here that the total amount of energy emitted during the formation of the system is equal to the final kinetic energy of the electrons. If we now suppose that during the formation of the system the electrons at any moment are situated at equal angular intervals on the circumference of a circle with the nucleus in the centre, from analogy with the considerations, on p. 5 we are here led to assume the existence of a series of stationary configurations in which the kinetic energy per electron is equal to  $\tau h\omega/2$ , where  $\tau$  is an entire number,  $h$  Planck's constant, and  $\omega$  the frequency of revolution. The configuration in which the greatest amount of energy is emitted is, as before, the one in which  $\tau = 1$ . This configuration we shall assume to be the permanent state of the system if the electrons in this state are arranged in a single ring. As for the case of a single 3electron we get that the angular momentum of each of the electrons is equal to  $h/2\pi$ . It may be remarked that instead of considering the single electrons we might have considered the ring as an entity. This would, however, lead to the same result, for in this case the frequency of revolution  $\omega$  will be replaced by the frequency  $n\omega$  of the radiation from the whole ring calculated from ordinary electrodynamics, and  $T$  by the total kinetic energy  $nT$ .

There may be many other stationary states corresponding to other ways of forming the system. The assumption of the existence of such states seems necessary in order to account for the line-spectra of systems containing more than one electron (p. 11); it is also suggested by the theory of Nicholson mentioned on p. 6, to which we shall return in a moment. The consideration of the spectra, however, gives, as far as I can see, no indication of the existence of stationary states in which all the electrons are arranged in a ring and which correspond to greater values for the total energy emitted than the one we above have assumed to be the permanent state. Further, there may be stationary configurations of a system of  $n$  electrons and a nucleus of charge  $E$  in which all the electrons are not arranged in a single ring. The question, however, of the existence of such stationary configurations is not essential for our determination of the permanent state, as long as we assume that the electrons in this state of the system are arranged in a single ring. Systems corresponding to more complicated configurations will be discussed on p. 24.?????

Using the relation  $T = h\omega/2$  we get, by help of the above expressions for  $T$  and  $\omega$ , values for  $a$  and  $\omega$  corresponding to the permanent state of the

system which only differ from those given by the equations (3) on p. 5, by exchange of  $E$  for  $E - es_n$ .

The question of stability of a ring of electrons rotating round a positive charge is discussed in great detail by Sir. J.J. Thomson<sup>19</sup> An adaption of Thomson's analysis for the case here considered of a ring rotating round a nucleus of negligibly small linear dimensions is given by Nicholson.<sup>20</sup> The investigation of the problem in question naturally divides in two parts: one concerning the stability for displacements of the electrons on the plane of the ring; one concerning displacements perpendicular to this plane. As Nicholson's calculations show, the answer to the question of stability differs very much in the two cases in question. While the ring for the latter displacements in general is stable if the number of electrons is not great; the ring is in no case considered by Nicholson stable for displacement of the first kind.

According, however, to the point of view taken in this paper, the question of stability for displacements of the electrons in the plane of the ring is most intimately connected with the question of the mechanism of the binding of the electrons, and like the latter cannot be treated on the basis of the ordinary dynamics. The hypothesis of which we shall make use in the following is that the stability of a ring of electrons rotating round a nucleus is secured through the above condition of the universal constancy of the angular momentum, together with the further condition that the configuration of the particles is the one by the formation of which the greatest of energy is emitted. As will be shown, this hypothesis is, concerning the question of stability for a displacement of the electrons perpendicular to the plane of the ring, equivalent to that used in ordinary mechanical calculations.

Returning to the theory of Nicholson on the origin of lines observed in the spectrum of the solar corona, we shall now see that the difficulties mentioned on p. 7 may be only formal. In the first place, from the point of view considered above the objection as to the instability of the systems for displacements of the electrons in the plane of the ring may not be valid. Further, the objection as to emission of the radiation in quanta will not have reference to the calculations in question, if we assume that in the coronal spectrum we are not dealing with a true emission but only with a scattering of radiation. This assumption seems probable if we consider the conditions in the celestial body in question: for on account comparatively few collisions to disturb the stationary states and to cause a true emission of light corresponding to the transition between different stationary states; on the

---

<sup>19</sup>Loc. cit.

<sup>20</sup>Loc. cit.

other hand there will in the solar corona be intense illumination of light of all frequencies which may excite the natural vibrations of the systems in the different stationary states. If the above assumption is correct, we immediately understand the entirely different from for the laws connecting the lines discussed by Nicholson and those connecting the ordinary line-spectra considered in this paper.

Proceeding to consider systems of more complicated constitution, we shall make use of the following theorem, which can be very simply proved; – “In every system consisting of electrons and positive nuclei, in which the nuclei are at rest and the electrons move in circular orbits with a velocity small compared with the velocity of light, the kinetic energy will be numerically equal to half the principal energy.”

By help of this theorem we get – as in the previous cases of a single electron or of a ring rotating round a nucleus – that the total amount of energy emitted, by the formation of the systems from a configuration in which the distances apart of the particles are infinitely great and in which the particles have no velocities relative to each other, is equal to the kinetic energy of the electrons in the final configuration.

In analogy with the case of a single ring we are here led to assume that corresponding to any configuration of equilibrium a series of geometrically similar, stationary configuration of the system will exist in which the kinetic energy of every electron is equal to the frequency of revolution multiplied by  $\tau/2h$  where  $\tau$  is an entire number and  $h$  Planck's constant. In any such series of stationary configurations the one corresponding to the greatest amount of energy emitted will be the one in which  $\tau$  for every electron is equal to 1. Considering that the ratio of kinetic energy to frequency for a particle rotating in a circular orbit is equal to  $\pi$  times the angular momentum round the center of the orbit, we are therefore led to the following simple generalization of the hypotheses mentioned on pp. 15 and 22. ??????

*“In any molecular system consisting of positive nuclei and electrons in which the nuclei are at rest relative to each other and the electrons move in circular orbits, the angular momentum of every electron round the centre of its orbit will in the permanent state of the system be equal to  $h/2\pi$ , where  $h$  is Planck's constant.”<sup>21</sup>*

In analogy with the considerations on p. 23, we shall assume that a

<sup>21</sup>In the considerations leading to this hypothesis we have assumed that the velocity of the electrons is small compared with the velocity of light. The limits of the validity of this assumption will be discussed in Part II.

configuration satisfying this condition is stable if the total energy of the system is less than in any neighbouring configuration satisfying the same condition of the angular momentum of the electrons.

As mentioned in the introduction, the above hypothesis will be used in a following communication as a basis for a theory of the constitution of atoms and molecules. It will be shown that it leads to results which seem to be in conformity with experiments on a number of different phenomena.

The foundation of the hypothesis has been sought entirely in its relation with Planck's theory of radiation; by help of considerations given later it will be attempted to throw some further light on the formation of it from another point of view.

*April 5, 1913*

*On the Constitution of Atoms and Molecules*

N. Bohr,  
*Dr. phil. Copenhagen*  
(Received July 1913)

*Part II. – Systems containing only a  
Single Nucleus*

1

*§ 1 General Assumptions*

Following the theory of Rutherford, we shall assume that the atoms of the elements consist of a positively charged nucleus surrounded by a cluster of electrons. The nucleus is the seat of the essential part of the mass of the atom, and has linear dimensions exceedingly small compared with the distance apart of the electrons in the surrounding cluster.

As in the previous paper, we shall assume that the cluster of electrons is formed by the successive binding by the nucleus of electrons initially nearly at rest, energy at the same time being radiated away. This will go on until, when the total negative charge on the bound electrons is numerically equal to the positive charge on the nucleus, the system will be neutral and no longer able to exert sensible forces on electrons at distances from the nucleus great in comparison with the dimensions of the orbits of the bound electrons. We may regard the formation of helium from  $\alpha$  rays as an observed example of

---

<sup>1</sup>Part I was published in Phil. Mag. XXVI. p. 1 (1913).

a process of this kind, an  $\alpha$  particle on this view being identical with the nucleus of a helium atom.

On account of the small dimensions of the nucleus, its internal structure will not be of sensible influence on the constitution of the cluster of electrons, and consequently will have no effect on the ordinary physical and chemical properties of the atom. The latter properties on this theory will depend entirely on the total charge and mass of the nucleus; the internal structure of the nucleus will be of influence only on the phenomena of radioactivity.

From the result of experiments on large-angle scattering of  $\alpha$ -rays, Rutherford<sup>2</sup> found an electric charge on the nucleus corresponding per atom to a number of electrons approximately equal to half the atomic weight. This result seems to be in agreement with the number of electrons per atom calculated from experiments on scattering of Röntgen radiation.<sup>3</sup> The total experimental evidence supports the hypothesis<sup>4</sup> that the actual number of electrons in a neutral atom with a few exceptions is equal to the number which indicated the position of the corresponding element in the series of elements arranged in order of increasing atomic weight. For example on this view, the atom of oxygen which is the eighth element of the series has eight electrons and a nucleus carrying eight unit charges.

We shall assume that the electrons are arranged at equal angular intervals in coaxial rings rotating round the nucleus. In order to determine the frequency and dimensions of the rings we shall use the main hypothesis of the first paper, viz.; that in the permanent state of an atom the angular momentum of every electron round the centre of its orbit is equal to the universal value  $h/2\pi$ , where  $h$  is Planck's constant. We shall take as a condition of stability, that the total energy of the system in the configuration in question is less than in any neighbouring configuration satisfying the same condition of the angular momentum of the electrons.

If the charge on the nucleus and the number of electrons in the different rings is known, the condition in regard to the angular momentum of the electrons will, as shown in § 2, completely determine the configuration of the system. i.e., the frequency of revolution and the linear dimensions of the rings. Corresponding to different distributions of the electrons in the rings, however, there will, in general, be more than one configuration which will satisfy the condition of the angular momentum together with the condition of stability.

<sup>2</sup>Comp. also Geiger and Marsden, Phil. Mag. XXV. p. 604 (1913).

<sup>3</sup>Comp. C.G. Barkla, Phil. Mag. XXI. p. 648 (1911).

<sup>4</sup>Comp. A.v.d. Broek, Phys. Zeitschr. XIV. p. 32 (1913).

In § 3 and § 4 it will be shown that, on the general view of the formation of the atoms, we are led to indications of the arrangement of the electrons in the rings which are consistent with those suggested by the chemical properties of the corresponding element.

In § 5 will be shown that it is possible from the theory to calculate the momentum velocity of cathode rays necessary to produce the characteristic Röntgen radiation from the element, and that this is in approximate agreement with the experimental values.

In § 6 the phenomena of radioactivity will be briefly considered in relation of the theory.

## § 2 Configuration and Stability of the System

Let us consider an electron of charge  $e$  and mass  $m$  which moves in a circular orbit of radius  $a$  with a velocity  $v$  small compared with the velocity of light. Let us denote the radial force acting on the electrons by  $e^2/a^2F$ ;  $F$  will in general be dependent on  $a$ . The condition of dynamical equilibrium gives

$$\frac{mv^2}{a} = \frac{e^2}{a^2}F.$$

Introducing the condition of universal constancy of the angular momentum of the electron, we have

$$mva = \frac{h}{2\pi}.$$

From these two conditions we now get

$$a = \frac{h^2}{4\pi^2 e^2 m} \cdot F^{-1} \quad \text{and} \quad v = \frac{2\pi e^2}{h} \cdot F; \quad (1)$$

and for the frequency of revolution  $w$  consequently

$$\omega = \frac{4\pi^2 e^2 m}{h^2} \cdot F^2. \quad (2)$$

If  $F$  is known, the dimensions and frequency of the corresponding orbit are simply determined by (1) and (2). For a ring of  $n$  electrons rotating round a nucleus of charge  $ne$  we have (comp. Part I., p. 20)????

$$F = N - s_n, \quad \text{where} \quad s_n = \frac{1}{4} \cdot \sum_{s=1}^{s=n-1} \operatorname{cosec} \frac{s\pi}{n}.$$

The values for  $s_n$  from  $n = 1$  to  $n = 16$  are given in the table 1.

For systems consisting of nuclei and electrons in which the first are at rest and the latter move in circular orbits with a velocity small compared with the velocity of light, we have shown (see part I., p. 21)???? that the total kinetic energy of the electrons is equal to the total amount of energy emitted during the formation of the system from an original configuration in which all the particles are at rest and at infinite distances from each other. Denoting this amount of energy by  $W$ , we consequently get

$$W = \sum \frac{m}{2} v^2 = \frac{2\pi^2 e^4 m}{h^2} \sum F^2. \quad (3)$$

Putting in (1), (2), and (3)  $e = 4.7 \cdot 10^{-10}$ ,  $\frac{e}{m} = 5.31 \cdot 10^{-17}$ , and  $h = 6.5 \cdot 10^{-27}$  we get

$$\begin{aligned} a &= 0.55 \cdot 10^{-8} F^{-1}, & v &= 2.1 \cdot 10^8 F, \\ \omega &= 6.2 \cdot 10^{15} F^2, & W &= 2.0 \cdot 10^{-11} \sum F^2. \end{aligned} \quad (4)$$

In neglecting the magnetic forces due to the motion of the electrons we have in Part I. assumed that the velocities of the particles are small compared with the velocity of light. The above calculations show that for this to hold,  $F$  must be small compared with 150. As will be seen, the latter condition will be satisfied for all the electrons in the atoms of elements of low atomic weight and for a greater part of the electrons contained in the atoms of the other elements.

If the velocity of the electrons is not small compared with the velocity of light, the constancy of the angular momentum no longer involved a constant ratio between the energy and the frequency of revolution. Without introducing new assumptions, we cannot therefore in this case determine the configuration of the systems on the basis of the consideration in Part I. Considerations given later suggest, however, that the constancy of the angular momentum is the principal condition. applying this condition for velocities not small compared with the velocity of light, we get the same expression for  $v$  as that given by (1), while the quantity  $m$  in the expressions for  $a$  and  $\omega$  is replaced by  $m/\sqrt{1 - v^2/c^2}$ , and in the expression for  $W$  by

$$m \cdot 2 \frac{c^2}{v^2} \cdot \left( 1 - \sqrt{1 - \frac{v^2}{c^2}} \right).$$

As stated in Part I., a calculation based on the ordinary mechanics given the result, that a ring of electrons rotating round a positive nucleus in general is unstable for displacement of the electrons in the plane of the ring. In order

to escape from this difficulty, we have assumed that the ordinary principles of mechanics cannot be used in the discussion of the problem in question, any more than in the discussion of the connected problem of the mechanism of binding of electrons. We have also assumed that the stability for such displacement is secured through the introduction of the hypothesis of the universal constancy of the angular momentum of the electrons.

As is easily shown, the latter assumption is included in the condition of stability in § 1. Consider a ring of electrons rotation round a nucleus, and assume that the system is in dynamical equilibrium and that the radius of the ring is  $a_0$ , the  $v_0$ , the total kinetic energy  $T_0$ , and the potential energy  $P_0$ . As shown in Part i. (p. 21) we have  $P_0 = -2T_0$ . Next consider a configuration of the system in which the electrons, under influence of extraneous forces, rotate with the same angular momentum round the nucleus in a ring of radius  $a = \alpha a_0$ . In this case we have  $P = \frac{1}{\alpha} P_0$ , and on account of the uniformity of the angular momentum  $v = 1/\alpha \cdot v_0$  and  $T = 1/\alpha^2 \cdot T_0$ . Using the relation  $P_0 = -2T_0$ , we get

$$P + T = \frac{1}{\alpha} \cdot P_0 + \frac{1}{\alpha^2} T_0 = P_0 + T_0 + T_0 \cdot \left(1 - \frac{1}{\alpha}\right)^2.$$

We see that the total energy of the new configuration is greater than in the original. according to the condition of stability in § 1 the system is consequently stable for the displacement considered. In this connexion, it may be remarked that in Part I. we have assumed that the frequency of radiation emitted or absorbed by the systems cannot be determined from the frequencies of vibration of the electrons in the plane of the orbits, calculated by help of the ordinary mechanics. We have, on the contrary, assumed that the frequency of the radiation is determined by the condition  $h\nu = E$ , where  $\nu$  is the frequency,  $h$  Planck's constant, and  $E$  the difference in energy corresponding to two different "stationary" states of the system.

In considering the stability of a ring of electrons rotating round a nucleus for displacements of the electrons perpendicular to the plane of the ring, imagine a configuration of the system in which the electrons are displaced by  $\delta z_1, \delta z_2, \dots, \delta z_n$  respectively, and suppose that the electrons, under influence of extraneous forces, rotate in circular orbits parallel to the original plane with the same radial and the same angular momentum round the axis of the system as before. The kinetic energy is unaltered by the displacement, and neglecting powers of the quantities  $\delta z_1, \dots, \delta z_n$  higher than the second, the increase of the potential energy of the system is given by

$$\frac{1}{2} \cdot \frac{e^2}{a^3} \cdot N \sum (\delta z)^2 - \frac{1}{32} \cdot \frac{e^2}{a^2} \cdot \sum \sum |\cosec^3 \frac{\pi(r-s)}{n}| (\delta z_r - \delta z_s)^2,$$

where  $a$  is the radius of the ring,  $Ne$  the charge on the nucleus, and  $n$  the number of electrons. According to the condition of stability in § 1 the system is stable for the displacement considered, if the above expression is positive for arbitrary values of  $\delta z_1, \dots, \delta z_n$ . By a simple calculation it can be shown that the latter condition is equivalent to the condition

$$N > p_{n,0} - p_{n,m}, \quad (5)$$

where  $m$  denotes the whole number (smaller than  $n$ ) for which

$$p_{n,k} = \frac{1}{8} \sum_{s=1}^{s=n-1} \cos 2k \cdot \frac{s\pi}{n} \operatorname{cosec}^3 \frac{s\pi}{n}$$

has its smallest value. This condition is identical with the condition of stability for displacements of the electrons perpendicular to the plane of the ring, deduced by help of ordinary mechanical considerations.<sup>5</sup>

A suggestive illustration is obtained by imagining that the displacements considered are produced by the effect of extraneous forces acting on the electrons in a direction parallel to the axis of the ring. If the displacements are produced infinitely slowly the motion of the electrons will at any moment be parallel to the original plane of the ring, and the angular momentum of each of the electrons round the centre of its orbit will obviously be equal to its original value; the increase in the potential energy of the system will be equal to the work done by the extraneous forces during the displacements we are led to assume that the ordinary mechanics can be used in calculating the vibrations of the electrons perpendicular to the plane of the ring – contrary to the ease of vibrations in the plane of the ring. This assumption is supposed by the apparent agreement with observations obtained by Nicholson in his theory of the origin of lines in the spectra of the solar corona and stellar nebulae (see Part I. pp. 6 & 23).????? In addition it will be shown later that the assumption seems to be in agreement with experiments on dispersion.

The following table gives the values of  $s_n$  and  $P_{n,0} - P_{n,m}$  from  $n = 1$  to  $n = 16$ .

Table 1.

---

<sup>5</sup>Comp. J.W. Nicholson, Month. Not. Roy. Astr. Soc. 72. p. 52 (1912).

$n$	$s_n$	$p_{n,0} - p_{n,m}$	$n$	$s_n$	$p_{n,0} - p_{n,m}$
1	0	0	9	3.328	13.14
2	0.25	0.25	10	3.863	18.13
3	0.577	0.58	11	4.416	23.60
4	0.957	1.41	12	4.984	30.80
5	1.377	2.43	13	5.565	38.57
6	1.828	4.25	14	6.159	48.38
7	2.305	6.35	15	6.764	58.83
8	2.805	9.56	16	7.379	71.65

We see from the table that the number of electrons which can rotate in a single ring round a nucleus of charge  $Ne$  increases only very slowly for increasing  $N$ ; for  $N = 20$  the maximum value is  $n = 10$ ; for  $N = 13$ ; for  $N = 60$ ,  $n = 15$ . We see, further, that a ring of  $n$  electrons cannot rotate in a single ring round a nucleus of charge  $ne$  unless  $n < 8$ .

In the above we have suppose that the electrons move under the influence of a stationary radial force and that their orbits are exactly circular. The first condition will not be satisfied if we consider a system containing several rings of electrons which rotate with different frequencies. If, however, the distance between the rings is not small in comparison with their radii, if the ratio between their frequency is not near to unity, the deviation from circular orbits may be very small and the motion of the electrons to a close approximation may be identical with that obtained on the assumption that the charge on the electrons is uniformly distributed along the circumference of the rings. If the ratio between the radii of the rings is not near to unity, the conditions of stability on this assumption may also be considered sufficient.

We have assumed in § 1 that the electrons in the atoms rotate in coaxial rings. The calculation indicated that only in the case of systems containing a great number of electrons will the planes of the rings separate; in the case of systems containing a moderate number of electrons, all the rings will be situated in a single plane through the nucleus. For the sake of brevity, we shall therefore here only consider the latter case.

Let us consider an electric charge  $E$  uniformly distributed along the circumference of a circle of radius  $a$ .

At a point distant  $z$  from the plane of the ring, and at a distance  $r$  from

the axis of the ring, the electrostatic potential is given by

$$U = \frac{1}{\pi} \cdot E \int_0^\pi \frac{d\vartheta}{\sqrt{a^2 + r^2 + z^2 - 2ar \cos \vartheta}}.$$

Putting in this expression  $z = 0$  and  $\frac{r}{a} = \tan^2 \alpha$ , and using the notation

$$K(\alpha) = \int_0^{\pi/2} \frac{d\vartheta}{\sqrt{1 - \sin^2 \alpha \cos^2 \vartheta}},$$

we get for the radial force exerted on an electron in a point in the plane of the ring

$$e \frac{\partial U}{\partial r} = \frac{Ee}{r^2} Q(\alpha),$$

where

$$Q(\alpha) = \frac{1}{\pi} \sin^4 \alpha (K(2\alpha) - \cot \alpha \cdot K'(2\alpha)).$$

The corresponding force perpendicular to the plane of the ring at a distance  $r$  from the center of the ring and at a small distance  $\delta z$  from its plane is given by

$$e \frac{\partial U}{\partial z} = \frac{Ee \delta z}{r^3} R(\alpha),$$

where

$$R(\alpha) = \frac{2}{\pi} \sin^6 \alpha [K(2\alpha) + \tan(2\alpha) \cdot K'(2\alpha)].$$

A short table of the functions  $Q(\alpha)$  and  $R(\alpha)$  is given on p. 485.???

Next consider a system consisting of a number of concentric rings of electrons which rotate in the same plane round a nucleus of charge  $Ne$ . Let the radial of the rings be  $a_1, a_2, \dots$ , and the number of electrons on the different rings  $n_1, n_2, \dots$

Putting  $a_r/a_s = \tan^2(\alpha_{r,s})$  we get for the radial force acting on an electron in the  $r$ th ring  $e^2/a_r^2 F_r$  where

$$F_r = N - s - \sum n_s Q(\alpha_{r,s}).$$

the summation is to be taken over all the rings except the one considered.

If we know the distribution of the electrons in the different rings, from the relation (1) on p. 478.???? we can, by help of the above, determine  $a_1, a_2, \dots$ . The calculation can be made by successive approximations, starting from a set of values for the  $\alpha$ 's, and from them calculating the  $F$ 's, and then

redetermining the  $\alpha_s$  by the relation (1) which gives  $F_s/F_r = a_r/a_s = \tan^2(\alpha_{r,s})$ , and so on.

As in the case of a single ring it is supposed that the systems are stable for displacements of the electrons in the plane of their orbits. In a calculation such as that on p. 480,???? the interaction of the rings ought strictly to be taken into account. This interaction will involve that the quantities  $F$  are not constant, as for a single ring rotating round a nucleus, but will vary with the radii of the rings; the variation in  $F$ , however, if the ratio between the radii of the rings is not very near to unity, will be too small to be of influence on the result of the calculation.

Considering the stability of the systems for a displacement of the electrons perpendicular to the plane of the rings, it is necessary to distinguish between displacements in which the centres of gravity of the electrons in the single rings are unaltered, and displacements in which all the electrons inside the same ring are displaced in the same direction. The condition of stability for the first kind of displacements is given by the condition (5) on p. 481,???? if for every ring we replace  $N$  by a quantity  $G_r$  determined by the condition that  $e^2/a_r^3 G_r \delta z$  is equal to the component perpendicular to the plane of the ring of the force – due to the nucleus and the electrons in the other rings – acting on one of the electrons if it has received a small displacement  $\delta z$ . Using the same notation as above, we get

$$G_r = N - \sum n_s R(\alpha_{r,s}).$$

If all the electrons in one of the rings are displaced in the same direction by help of extraneous forces, the displacement will produce corresponding displacements of the electrons in the other rings; and this interaction will be of influence on the stability. For example, consider a system of  $m$  concentric rings rotating in a plane round a nucleus of charge  $Ne$ , and let us assume that the electrons in the different rings are displaced perpendicular to the plane by  $\delta z_1, \delta z_2, \dots, \delta z_m$  respectively. With the above notation the increase in the potential energy of the system is given by

$$\frac{1}{2} \cdot N \sum n_r \frac{e^2}{a_n^3} (\delta z_n)^2 - \frac{1}{4} \cdot \sum \sum n_r n_s \frac{e^2}{a_r^3} R(\alpha_{r,s}) (\delta z_r - \delta z_s)^2.$$

The condition of stability is that this expression is positive for arbitrary values  $\delta z_1, \dots, \delta z_m$ . This condition can be worked out simply in the usual way. It is not of sensible influence compared with the condition of stability for the displacements considered above, except in cases where the system contains several rings of few electrons.

The following Table. containing the values of  $Q(\alpha)$  and  $R(\alpha)$  for every fifth degree from  $\alpha = 20^\circ$  to  $\alpha = 70^\circ$ , gives an estimate of the order of magnitude of these functions: —

Table 2.

$\alpha$	$\tan^2\alpha$	$Q(\alpha)$	$R(\alpha)$
20	0.132	0.001	0.002
25	0.217	0.005	0.011
30	0.333	0.021	0.048
35	0.490	0.080	0.217
40	0.704	0.373	1.549
45	1.000	-	-
50	1.420	1.708	4.438
55	2.040	1.233	1.839
60	3.000	1.093	1.301
65	4.599	1.037	1.115
70	7.548	1.013	1.041

$\tan^2\alpha$  indicated the ratio between the radii of the rings ( $\tan^2(a_{r,s}) = \frac{a_r}{a_s}$ ). The values of  $Q(\alpha)$  show that unless the ratio of the radii of the rings is nearly unity the effect of outer rings on the dimensions of inner rings is very small, and that the corresponding effect of inner rings on outer is to neutralize approximately the effect of a part of the charge on the nucleus corresponding to the number of electrons on the ring. The values of  $R(\alpha)$  show that the effect of outer rings on the stability of inner — though greater than the effect on the dimensions — is small, but that unless the ratio between the radii is very great, the effect of inner rings on the stability of outer is considerably greater than to neutralize a corresponding part of the charge of the nucleus.

The maximum number of electrons which the innermost ring can contain being unstable is approximately equal to that calculated on p. 482 for a single ring rotating round a nucleus. For the outer rings, however, we get considerably smaller numbers than those determined by the condition (5) if

we replace  $Ne$  by the total charge on the nucleus and on the electrons of inner rings.

If system of rings rotating round a nucleus in a single plane is stable for small displacements of the electrons perpendicular to this plane, there will in general be no stable configurations of the rings, satisfying the condition of the constancy of the angular momentum of the electrons, in which all the rings are not situated in the plane. An exception occurs in the special case of two rings containing equal numbers of electrons; in this case there may be a stable configuration in which the two rings have equal radii and rotate in parallel planes at equal distances from the nucleus, the electrons in the one ring being situated just opposite the intervals between the electrons in the other ring. The latter configuration, however, is unstable if the configuration in which all the electrons in the two rings are arranged in a single ring is stable.

### *§ 3 Constitution of Atoms containing very few Electrons*

As stated in § 1, the condition of the universal constancy of the angular momentum of the electrons, together with the condition of stability, is in most cases not sufficient to determine completely the constitution of the system. On the general view of formation of atoms, however, and by making use of the knowledge of the properties of the corresponding elements, it will be attempted, in this section and the next, to obtain indications of what configurations of the electrons may be expected to occur in the atoms. In these considerations we shall assume that the number of electrons in the atom is equal to the number which indicates the position of the corresponding element in the series of elements arranged in order of increasing atomic weight.

Exceptions to this rule will be supposed to occur only at such places in the series where deviation from the periodic law of the chemical properties of the elements are observed. In order to show clearly the principles used we shall first consider with some detail those atoms containing very few electrons.

Forsake of brevity we shall, by the symbol  $N(n_1, n_2 \dots)$ , refer to a plane system of rings of electrons rotating round a nucleus of charge  $Ne$ , satisfying the condition of the angular momentum of the electrons with the approximation used in § 2.  $n_1, n_2 \dots$  are the numbers of electrons in the rings,

starting from inside. By  $a_1, a_2, \dots$  and  $\omega_1, \omega_2, \dots$  we shall denote the radii and frequency of the rings taken in the same order. The total amount of energy  $W$  emitted by the formation of the system shall simply be denoted by  $W[N(n_1, n_2, \dots)]$ .

$$N = 1 \quad \text{Hydrogen.}$$

In Part I. we have considered the binding of an electron by a positive nucleus of charge  $e$ , and have shown that it is possible to account for the Balmer spectrum of hydrogen on the assumption of the existence of a series of stationary states in which the angular momentum of the electron round the nucleus is equal to entire multiplies of the value  $h/2\pi$ , where  $h$  is Planck's constant. The formula found for the frequencies of the spectrum was

$$\nu = \frac{2\pi^2 e^4 m}{h^3} \cdot \left( \frac{1}{\tau_2^2} - \frac{1}{\tau_1^2} \right),$$

where  $\tau_1$  and  $\tau_2$  are entire numbers. Introducing the values for  $e$ ,  $m$ , and  $h$  used on p. 479, we get for the factor before the bracket  $3.1 \cdot 10^{15}$ ; <sup>6</sup> the value observed for the constant in the Balmer spectrum is  $3.290 \cdot 10^{15}$ .

For the permanent state of a neutral hydrogen atom we get from the formula (1) and (2) in § 2, putting  $F = 1$ ,

$$\begin{aligned} \mathbf{1}(1) : \quad \alpha &= \frac{h^2}{4\pi e^2 m} = 0.55 \cdot 10^{-8}, \quad \omega = \frac{4\pi^2 e^4 m}{h^3} = 6.2 \cdot 10^{15}, \\ W &= \frac{2\pi^2 e^4 m}{h^2} = 2.0 \cdot 10^{-11}. \end{aligned}$$

These values are of the order of magnitude to be expected. For  $W/e$  we get 0.043, which corresponds to 13 volts; the value for the ionizing potential of a hydrogen atom, calculated by Sir J.J. Thomson from experiments on positive rays, is 11 volt.<sup>7</sup> No other definite data, however are available for hydrogen atoms. For sake of brevity, we shall in the following denote the values for  $a, \omega$  and  $W$  corresponding to the configuration 1(1) by  $a_0, \omega_0$ , and  $W_0$ .

---

<sup>6</sup>This value is that calculated in the first part of the paper. Using the values  $e = 4.78 \cdot 10^{-10}$  (see R.A. Millikan, Brit. Assoc. Rep. 1912, p. 410),  $e/m = 5.31 \cdot 10^{17}$  (see P. Gmelin, Ann. d. Phys. XXVIII. p. 1086 (1909) and A.H. Bucherer, Ann. d. Phys. XXXVII p. 597 (1912)), and  $e/h = 7.27 \cdot 10^{16}$  calculated by Planck's theory from the experiments of E. Warbung G. Leithäuser, E. Hupka, and C. Müller, Ann.d.Phys. XL. p. 611 (1913)) we get  $2\pi^2 e^4 m/h^3 = 3.26 \cdot 10^{15}$  in very close agreement with observations.

<sup>7</sup>J.J. Thomson, Phil. Mag. XXIV. p. 218 (1912).

At distance from the nucleus, great in comparison with  $a_0$ , the system **1** (1) will not exert sensible forces on free electrons. Since, however, the configuration:

$$\mathbf{1}(2) \quad a = 1.33a_0, \quad \omega = 0.563\omega_0, \quad W = 1.13W_0.$$

corresponds to a greater value for  $W$  than the configuration **1**(1), we may expect that a hydrogen atom under certain conditions can acquire a negative charge. This is in agreement with experiments on positive rays. Since  $W[1(3)]$  is only 0.54, a hydrogen atom cannot be expected to be able to acquire a double negative charge.

$$N = 2$$

*Helium.*

As shown in Part I., using the same assumptions as for hydrogen, we must expect that during the binding of an electron by a nucleus of charge  $2e$ , a spectrum is emitted, expressed by

$$\nu = \frac{2\pi^2 me^4}{h^3} \cdot \left( \frac{1}{(\frac{\tau_2}{2})^2} - \frac{1}{(\frac{\tau_1}{2})^2} \right).$$

This spectrum includes the spectrum observed by Pickering in the star *xi* Puppis and the spectra recently observed by Folwer in experiments with vacuum tubes filled with a mixture of hydrogen and helium. These spectra are generally ascribed to hydrogen.

For the permanent state of a positively charge helium atom, we get

$$\mathbf{2}(1) \quad a = \frac{1}{2}a_0, \quad \omega = 4\omega_0, \quad W = 4W_0.$$

At distances from the nucleus great compared with the radius of the bound electron, the system **2**(1) will, to a close approximation, act an an electron as a simple nucleus of charge  $e$ . For a system consisting of two electrons and a nucleus of charge  $2e$ , we may therefore assume the existence of a series of stationary states in which the electron most lightly bound moves approximately in the same way as the electron in the stationary states of a hydrogen atom. Such an assumption has already been used in Part I. in an attempt to explain the appearance of Rydberg's constant in the formula for the line-spectrum of any element. We can, however, hardly assume the existence of a stable configuration in which the two electrons have the same angular momentum round the nucleus and move in different orbits, the one outside the other. In such a configuration the electrons would be so near to

each other that the deviations from circular orbits would be very great. For the permanent state of a neutral helium atom, we shall therefore adopt the configuration

$$\mathbf{2}(2) \quad a = 0.571a_0, \quad \omega = 3.06\omega_0, \quad W = 6.13W_0.$$

Since

$$W[\mathbf{2}(2)] - W[\mathbf{2}(1)] = 2.13W_0,$$

we see that both electrons in a neutral helium atom are more firmly bound than the electron in a hydrogen atom. Using the values on p. 488,???? we get

$$2.13 \cdot \frac{W_0}{e} = 27, \quad 2.13 \cdot \frac{W_0}{h} = 6.6 \cdot 10^{15} \text{ 1/sec.}$$

these values are of the same order of magnitude as the value observed for the ionization potential in helium, 20.5 volt,<sup>8</sup> and the value for the frequency of the ultra-violet absorption in helium determined by experiments on dispersion  $5.9 \cdot 10^{15} \text{ 1/sec.}^9$

The frequency in question may be regarded as corresponding to vibrations in the plane of the ring (see p. 480).???? The frequency of vibration of the whole ring perpendicular to the plane, calculated in the ordinary way (see p. 482), is given by  $\nu = 3.27\omega_0$ . The fact that the latter frequency is great compared with that observed might explain that the number of electrons in a helium atom, calculated by help of Drude's theory from the experiments on dispersion, is only about two-thirds of the number to be expected. (Using  $\frac{e}{m} = 5.31 \cdot 10^{17}$  the value calculated is 1.2.)

For a configuration of a helium nucleus and three electrons, we get

$$\mathbf{2}(3) \quad a = 0.703a_0, \quad \omega = 2.02\omega_0, \quad W = 6.07W_0.$$

Since  $W$  for this configuration is smaller than for the configuration  $\mathbf{2}(2)$ , the theory indicates that a helium atom cannot acquire a negative charge. This is in agreement with experimental evidence, which shows that helium atoms have no "affinite" for free electrons.<sup>10</sup>

<sup>8</sup>J.Franck u. G. Hertz, Verh. d. Deutsch. Phys. Ges. XV. p. 34 (1913).

<sup>9</sup>C. and M. Cuthbertson, Proc. Roy. Soc. A. LXXXIV. p. 13 (1910). In a previous paper (Phil. Mag. Jan. 1913) the author took the values for the refractive index in helium, given by M. and C. Cuthbertson, as corresponding to atmosphere pressure; these values, however, refer to double atmosphere pressure. Consequently the value there given for the number of electrons in a helium atom calculated from Drude's theory has to be divided by 2.)

<sup>10</sup>See J. Franck, Verh. d. Deutsch. Phys. Ges. XII. p. 613 (1910).

In a later paper it will be shown that the theory offers a simple explanation of the marked tendency of hydrogen and helium atoms to combine into molecules.

$N = 3$

*Lithium.*

In analogy with the cases of hydrogen and helium we must expect that during the binding of an electron by a nucleus of charge  $3e$ , a spectrum is emitted, given by

$$\nu = \frac{2\pi^2 me^4}{h^3} \cdot \left( \frac{1}{(\frac{\tau_2}{3})^2} - \frac{1}{(\frac{\tau_1}{3})^2} \right).$$

On account of the great energy to be spent in removing all the electrons bound in a lithium atom (see below) the spectrum considered can only be expected to be observed in extraordinary cases.

In a recent note Nicholson<sup>11</sup> has drawn attention to the fact that in the spectra of certain stars, which show the Pickering spectrum with special brightness, some lines occur the frequencies of which to a close approximation can be expressed by the formula

$$\nu = K \cdot \left( \frac{1}{4} - \frac{1}{(m \pm 1/3)^2} \right).$$

where  $K$  is the same constant as in the Balmer spectrum of hydrogen. From analogy with the Balmer- and Pickering-spectra, Nicholson has suggested that the lines in question are due to hydrogen.

It is seen that the lines discussed by Nicholson are given by the above formula if we put  $\tau_2 = 6$ . The lines in question correspond to  $\tau_1 = 10, 13$  and  $14$ ; if we for  $\tau_2 = 6$  put  $\tau_1 = 9, 12$  and  $15$ , we get lines coinciding with lines of the ordinary Balmer-spectrum of hydrogen. If we in the above formula put  $\tau = 1, 2$ , and  $3$ , we get series of lines in the ultra-violet. If we put  $\tau_2 = 4$  we get only a single line in visible spectrum, viz.: for  $\tau_1 = 5$  which gives  $\nu = 6.662 \cdot 10^{14}$ , or a wave-length  $\lambda = 4.503 \cdot 10^{-8}$  cm closely coinciding with the wave-length  $4.504 \cdot 10^{-8}$  cm of one of the lines of unknown origin in the table quoted by Nicholson. In this table, however, no lines occur corresponding to  $\tau_2 = 5$ .

For the permanent state of a lithium atom with two positive charges we get a configuration

---


$$\mathbf{3}(1) \quad a = \frac{1}{3}a_0, \quad \omega = 9\omega_0, \quad W = 9W_0.$$

<sup>11</sup>J.W. Nicholson, Month. Not. Roy. Astr. Soc. LXXIII. 382 (1913).

The probability of a permanent configuration in which two electrons move in different orbits around each other must for lithium be considered still less probable than for helium, as the ratio between the radii of the orbits would be still nearer to unity. For a lithium atom with a single positive charge we shall, therefore, adopt the configuration:

$$\mathbf{3}(2) \quad a = 0.364a_0, \quad \omega = 7.56\omega_0, \quad W = 15.13W_0.$$

Since  $W[\mathbf{3}(2)] - W[\mathbf{3}(1)] = 6.13W_0$  we see that the first two electrons in a lithium atom very strongly bound compared with the electron in a hydrogen atom; they are still more rigidly bound than the electrons in a helium atom.

From a consideration of the chemical properties we should expect the following configuration for the electrons in a neutral lithium atom:

$$\begin{aligned} \mathbf{3}(2,1) \quad a_1 &= 0.362a_0, \quad \omega_1 = 7.65\omega_0, \\ &\qquad \qquad \qquad W = 16.02W_0 \\ a_2 &= 1.182a_0, \quad \omega_2 = 0.716\omega_0, \end{aligned}$$

This configuration may be considered as highly probable also from a dynamical point view. The deviation of the outermost electron from a circular orbit will be very small, partly on account of the great values of the ratio between the radii, and of the ratio between the frequencies of the orbits of the inner and outer electrons, partly also on account of the symmetrical arrangement of the inner electrons. accordingly, it appears probable that the three electrons will not arrange themselves in a single ring and from the system:

$$\mathbf{3}(3) \quad a = 0.413a_0, \quad \omega = 5.87\omega_0, \quad W = 17.61W_0,$$

although  $W$  for this configuration is greater than for  $\mathbf{3}(2,1)$ .

Since  $W[\mathbf{3}(2,1)] - W[\mathbf{3}(3)] = 0.89W_0$ , we see that the outer electron in the configuration  $\mathbf{3}(2,1)$  is bound even more lightly than the electron in a hydrogen atom. the difference in the firmness of the binding corresponds to a difference of 1.4 volts in the ionization potential. A marked difference between the electron in hydrogen and the outermost electron in lithium lies also in the greater tendency of the latter electron top leave the plane of this orbits. The quantity  $G$  considered in § 2, which gives a kind of measure for the stability for displacements perpendicular to this plane, is thus for the outer electron in lithium only 0.55, while for hydrogen it is 1. This may have a bearing on the explanation of the apparent tendency of lithium atoms to take a positive charge in chemical combinations with other elements.

For a possible negatively charged lithium atom we may expect the configuration:

$$\begin{aligned} \mathbf{3}(2,2) \quad a_1 &= 0.362a_0, \quad \omega_1 = 7.64\omega_0, \\ &\qquad\qquad\qquad W = 16.16W_0 \\ a_2 &= 1.516a_0, \quad \omega_2 = 0.436\omega_0, \end{aligned}$$

it should be remarked that we have no detailed knowledge of the properties in the atomic state, either for lithium or hydrogen, or for most of the electrons considered below.

$$N = 4$$

*Beryllium.*

For reasons analogous to those considered for helium and lithium we may for the formation of a neutral beryllium atom assume the following states:

$$\begin{aligned} \mathbf{4}(1) \quad a &= 0.25a_0, \quad \omega = 16\omega_0, \quad W = 16W_0, \\ \mathbf{4}(2) \quad a &= 0.267a_0, \quad \omega = 14.06\omega_0, \quad W = 28.13W_0, \\ \mathbf{4}(2,1) \quad a_1 &= 0.263a_0, \quad \omega_1 = 14.46\omega_0, \\ &\qquad\qquad\qquad W = 31.65W_0, \\ a_2 &= 0.605a_0, \quad \omega_2 = 2.74\omega_0, \\ \mathbf{4}(2,2) \quad a_1 &= 0.262a_0, \quad \omega_1 = 14.60\omega_0, \\ &\qquad\qquad\qquad W = 33.61W_0, \\ a_2 &= 0.673a_0, \quad \omega_2 = 2.21\omega_0, \end{aligned}$$

although the configurations:

$$\begin{aligned} \mathbf{4}(3) \quad a &= 0.292a_0, \quad \omega = 11.71\omega_0, \quad W = 35.14W_0, \\ \mathbf{4}(4) \quad a &= 0.329a_0, \quad \omega = 9.26\omega_0, \quad W = 37.04W_0, \end{aligned}$$

correspond to less values for the total energy than the configuration  $\mathbf{4}(2,1)$  and  $\mathbf{4}(2,2)$ .

From analogy we get further for the configuration of a possible negatively charged atom,

$$\begin{aligned} \mathbf{4}(2,3) \quad a_1 &= 0.263a_0, \quad \omega_1 = 14.51\omega_0, \\ &\qquad\qquad\qquad W = 33.66W_0 \\ a_2 &= 0.803a_0, \quad \omega_2 = 1.55\omega_0, \end{aligned}$$

Comparing the outer ring of the atom considered with the ring of a helium atom, we see that the presence of the inner ring of two electrons in the beryllium atom markedly changes the properties of the outer ring;

partly because the outer electrons in the configuration adopted for a neutral beryllium atom are more lightly bound than the electrons in a helium atom, and partly because the quantity  $G$ , which for helium is equal to 2, for the outer ring in the configuration  $4(2,2)$  is only equal 1.12.

Since  $W[4(2,3)] - W[4(2,2)] = 0.05W_0$ , the beryllium atom will further have a definite, although very small affinity for free electrons.

#### *§ 4 Atoms containing greater numbers of electrons*

From the examples discussed in the former section it will appear that the problem of the arrangement of the electrons in the atoms is intimately connected with the question of the confluence of two rings of electrons rotating round a nucleus outside each other, and satisfying the condition of the universal constancy of the angular momentum. apart from the necessary conditions of stability for displacements of the electrons perpendicular to the plane of the orbits, the present theory gives very little information on this problem. It seems, however, possible by the help of simple considerations to throw some light on the question.

Let us consider two rings rotating round a nucleus in a single plane, the one outside the other. Let us assume that the electrons in the one ring act upon the electrons in the other as if the electric charge were uniformly distributed along the circumference of the ring, and that the ring with this approximation satisfy the condition of the angular momentum of the electrons and stability for displacements perpendicular to their plane.

Now suppose that, by help of suitable imaginary extraneous forces acting parallel to the axis of the rings, we pull the inner ring slowly to one side. During this process, on account of the repulsion from the inner ring, the outer will move to the opposite side of the original plane of the rings. During the displacements of the rings angular momentum of the electrons round the axis of the system will remain constant, and the diameter of the inner ring will increase while that of the outer will diminish. At the beginning of the displacement the magnitude of the extraneous forces to be applied to the original inner ring will increase but thereafter decrease, and at a certain distance between the plane of the rings the system will be in a configuration of equilibrium. This equilibrium, however, will not be stable. If we let the rings slowly return they will either reach their original position, or they arrive at a position in which the ring, which originally was the outer, is now

the inner, and *vise versa*.

If the charge of the electrons were uniformly distributed along the circumference of the rings, we could by the process considered at most obtain an interchange of the rings, but obviously not a junction of them. Taking, however, the discrete distribution of the electrons into account, it can be shown that in the special case when the number of electrons on the two rings are equal, and when the rings rotate in the same direction, the rings will unite by the process, provided that the final configuration is stable. In this case the radii and the frequency of the rings will be equal in the unstable configuration of equilibrium mentioned above. In reaching this configuration the electrons in the one ring will further be situated just opposite the intervals between the electrons in the outer, since such an arrangement will correspond to the smallest total energy. If now we let the rings return to their original plane, the electrons in the one ring will pass into the intervals between the electrons in the other, and from a single ring. Obviously the ring thus formed will satisfy the same condition of the angular momentum of the electrons as the original rings.

If the two rings contain unequal numbers of electrons the system will during a process such as that considered behave very differently, and, contrary to the former case, we cannot expect that the rings will flow together, if by help of extraneous forces acting parallel to the axis of the system they are displaced slowly from their original plane. It may in this connexion be noticed that the characteristic for the displacements considered is not the special assumption about the extraneous forces, but only invariance of the angular momentum of the electrons round the centre of the rings; displacements of this kind take in the present theory a similar position to arbitrary displacements in the ordinary mechanics.

The above considerations may be taken as an indication that there is greater tendency for the confluence of two rings when each contains the same number of electrons. Considering the successive binding of electrons by a positive nucleus, we conclude from this that, unless the charge on the nucleus is very great, rings of electrons will only join together if they contain equal numbers of electrons; and that accordingly the numbers of electrons on inner rings will only be 2, 4, 8, .... If the charge of the nucleus is very great the rings of electrons first bound, if few in number, will be very close together, and we must expect that the configuration will be very unstable, and that a gradual interchange of electrons between the rings will be greatly facilitated.

This assumption in regard to the number of electrons in the rings is strongly supported by the fact that the chemical properties of the elements

of low atomic weight vary with a period of 8. Further, it follows that the number of electrons on the outermost ring will always be odd or even, according as the total number of electrons in the atom is odd or even. This has a suggestive relation to the fact that the valency of an element of low atomic weight always is odd or even according as the number of the element in the periodic series is odd or even.

For the atoms of the elements considered in the former section we have assumed that the two electrons first bound are arranged in a single ring, and, further, that the two next electrons are arranged in another ring. If  $N \geq 4$  the configuration  $N(4)$  will correspond to a smaller value for the total energy than the configuration  $N(2,2)$ . The greater the value of  $N$  the closer will the ratio between the radii of the rings in the configuration  $N(2,2)$  approach unity, and the greater will be the energy emitted by an eventual confluence of the rings. The particular member of the series of the elements for which the four innermost electrons will be arranged for the first time in a single ring cannot be determined from the theory. From a consideration of the chemical properties we can hardly expect that it will have taken place before boron ( $N = 5$ ) or carbon ( $N = 6$ ), on account of the observed trivalence and tetrivalence respectively of these elements; on the other hand, the periodic system of the elements strongly suggests that already in neon ( $N = 10$ ) an inner ring of eight electrons will occur. Unless  $N > 14$  the configuration  $N(4,4)$  corresponds to smaller value for the total energy than the configuration  $N(8)$ ; already for  $N \geq 10$  the latter configuration, however, will be stable for displacements of the electrons perpendicular to the plane of their orbits. A ring of 16 electrons will not be stable unless  $N$  is very great; but in such a case the simple considerations mentioned do not apply.

The confluence of two rings of equal number of electrons, which rotate round a nucleus of charge  $Ne$  outside a ring of  $n$  electrons already bound, must be expected to take place more easily than the confluence of two similar rings rotating round a nucleus of charge  $(N - n) \cdot e$ ; for the stability of the rings for a displacement perpendicular to their plane will (see § 2) be smaller in the first than the latter case. This tendency for stability to decrease for displacements perpendicular to the plane of the ring will be especially marked for the outer rings of electrons of a neutral atom. In the latter case we must expect the confluence of rings to be greatly facilitated and in certain cases it may even happen that the number of electrons in the outer ring may be greater than in the next, and that the outer ring may show deviations from the assumption of 1, 2, 4, 8 electrons in the rings, e.g. the configurations **5**(2,3) and **6**(2,4) instead of the configuration **5**(2,2,1) and **6**(2,2,2). We shall here not discuss further the intricate question of the

arrangement of the electrons in the outer ring. In the scheme given below the number of electrons in this rings is arbitrary put equal to the normal valency of the corresponding element; i.e. for electronegative and electropositive elements respectively the number of hydrogen atoms and twice the number of oxygen atoms with which one atom of the element combines.

Such an arrangement of the outer electrons is suggested by considerations of atomic volumes. As is well known, the atomic volume of the elements is a periodic function of the atomic weights. If arranged in the usual way according to the periodic system, the elements inside the same column have approximately the same atomic volume, while this volume changes considerably from one column to another, being greatest for columns corresponding to the smallest valency 1 and smallest for the greatest valency 4. An approximate estimate of the radius of the outer ring of a neutral atom can be obtained by assuming that the total forces due to the nucleus and the inner electrons is equal to that from a nucleus of charge  $ne$ , where  $n$  is the number of electrons in the ring. Putting  $F = n - s_n$  in the equation (1) on p. 478, ?????? and denoted the value of  $a$  for  $n = 1$  by  $a_0$ , we get for  $n = 2$ ,  $a = 0.41a_0$ ; and for  $n = 4$ ,  $a = 0.33a_0$ . According the arrangement chosen for the electrons will involve a variation in the dimensions of the outer ring similar to the variation in the atomic volumes of the corresponding elements. It must, however, be borne in mind that the experimental determinations of atomic volumes in most cases are deduced from consideration of molecules rather than atoms.

From the above we are led to the following possible scheme for the arrangement of the electrons in light atoms: –

<b>1(1)</b>	<b>9(4,4,1)</b>	<b>17(8,4,4,1)</b>
<b>2(2)</b>	<b>10(8,2)</b>	<b>18(8,8,2)</b>
<b>3(2,1)</b>	<b>11(8,2,1)</b>	<b>19(8,8,2,1)</b>
<b>4(2,2)</b>	<b>12(8,2,2)</b>	<b>20(8,8,2,2)</b>
<b>5(2,3)</b>	<b>13(8,2,3)</b>	<b>21(8,8,2,3)</b>
<b>6(2,4)</b>	<b>14(8,2,4)</b>	<b>22(8,8,2,4)</b>
<b>7(4,3)</b>	<b>15(8,4,3)</b>	<b>23(8,8,4,3)</b>
<b>8(4,2,2)</b>	<b>16(8,4,2,2)</b>	<b>24(8,8,4,2,2)</b>

Without any fuller discussion it seems not unlikely that this constitution of the atoms will correspond to properties of the elements similar with those observed.

In the first place there will be a marked periodicity with a period of 8. Further, the binding of the outer electrons in every horizontal series of the

above scheme will become weaker with increasing number of electrons per atom, corresponding to the observed increase of the electropositive character for an increase of atomic weight of the elements in every single group of the periodic system. A corresponding agreement holds for the variation of the atomic volumes.

In the case of atoms of higher atomic weight the simple assumptions used do not apply. A few indications, however, are suggested from consideration of the variations in the chemical properties of the elements. At the end of the 3rd period of 8 elements we meet with the iron-group. This group takes a particular position in the system of the elements, since it is the first time that elements of neighbouring atomic weight show similar chemical properties. This circumstance indicates that the configurations of the electrons in the elements of this group differ only in the arrangement of the inner electrons. The fact that the period in the chemical properties of the elements after the iron-group is no longer 8, but 18, suggests that elements of higher atomic weight contain a recurrent configuration of 18 electrons in the innermost rings. The deviation from 2, 4, 8, 16 may be due to a gradual interchange of electrons between the rings, such as is indicated on p. 495. Since a ring of 18 electrons will not be stable the electrons may be arranged in two parallel rings (see p. 486). ??????? Such a configuration of the inner electrons will act upon the outer electrons in very nearly the same way as nucleus of charge  $(N - 18) \cdot e$ . It might therefore be possible that with increase of  $N$  another configuration of the same type will be formed outside the first, such as is suggested by the presence of a second period of 18 elements.

On the same lines, the presence of the group of the rare earths indicates that for still greater values of  $N$  another gradual alteration of the innermost rings will take place. Since, however, for elements of higher atomic weight than those of this group, the laws connection the vibration of the chemical properties with the atomic weight are similar to these between the elements of low atomic weight, we may conclude that the configuration of the innermost electrons will be again repeated. The theory, however, is not sufficiently complete to give a definite answer to such problems.

## § 5 *Characteristic Röntgen Radiation*

According to the theory of emission of radiation given in Part I., the ordinary line-spectrum of an element is emitted during the reformation of an atom

when one or more of the electrons in the other rings are removed. In analogy it may be supposed that the characteristic Röntgen radiation is sent out during the setting down of the system if electrons in inner rings are removed by some agency, e.g. by impact of cathode particles. This view of the origin of the characteristic Röntgen radiation has been proposed by Sir. J.J. Thomson.

Without any special assumption in regard to the constitution of the radiation, we can from this view determine the minimum velocity of the cathode rays necessary to produce the characteristic Röntgen radiation of a spacial type by calculating the energy necessary to remove one of the electrons from the different rings. Even if we know the numbers of electrons in the rings, a rigorous calculation of this momentum energy might still be complicated, and the result largely dependent on the assumptions used; for, as mentioned in Part I., p. 19, ?????????? the calculation cannot be performed entirely on the basis of the ordinary mechanics. We can, however, obtain very simply an approximate comparison with experiments if we consider the innermost ring and as a first approximation neglect the repulsion from the electrons in comparison with the attraction of the nucleus. Let us consider a simple system consisting of a bound electron rotating in a circular orbit round a positive nucleus of charge  $Ne$ . From the expressions (1) on p. 478 ??????? we get for the velocity of the electron, putting  $F = N$ ,

$$v = \frac{2\pi e^2}{h} N = 2.1 \cdot 10^8 \cdot N.$$

The total energy to be transferred to the system in order to remove the electron to an infinite distance from the nucleus is equal to the kinetic energy of the bound electron. If, therefore, the electron is removed to a great distance from the nucleus by impact of another rapidly moving electron, the smallest kinetic energy possessed by the latter when at a great distance from the nucleus must necessarily be equal to the kinetic energy of the bound electron before the collision. The velocity of the free electron therefore must be at least equal to  $e$ .

According to Whiddington's experiments<sup>12</sup> the velocity of cathode rays just able to produce the characteristic Röntgen radiation of the so-called  $K$ -type—the hardest type of radiation observed—from an element of atomic weight  $A$  is for elements from Al to Se approximately equal to  $A \cot 10^8$  cm/sec. As seen this is equal to the above calculated value for  $r$ , if we put  $N = A/2$ .

---

<sup>12</sup>R. Whiddington, Proc. Roy. Soc. A. LXXXV. p. 323 (1911).

Since we have obtained approximate agreement with experiment by ascribing the characteristic Röntgen radiation of the *K*-type to the innermost ring, it is to be expected that no harder type of characteristic radiation will exist. This is strongly indicated by observations of the penetrating power of  $\gamma$  rays.<sup>13</sup>

It is worthy of remark that the theory gives not only nearly the right value for the energy required to remove an electron from the outer ring, but also the energy required to remove an electron from the innermost ring. The approximate agreement between the calculated and experimental values is all the more striking it is recalled that the energies required in the two cases for an element of atomic weight 70 differ by a ratio of 1000.

In connexion with this it should be emphasized that the remarkable homogeneity of the characteristic Röntgen radiation – indicated by experiments on absorption of the rays, as well as by the interference observed in recent experiments on diffraction of Röntgen rays in crystals – is in agreement with the main assumption used in part I. (see p. 7) in considering the emission of line-spectra, viz. that the radiation emitted during the passing of the systems between different stationary states is homogeneous.

Putting in (4)  $F = N$ , we get for the diameter of the innermost ring approximately  $2a = 1/N \cdot 10^{-8}$  cm. For  $N = 100$  this gives  $2a = 10^{-10}$  cm, a value which is very small in comparison with ordinary atomic dimensions but still very great compared with the dimensions to be expected for the nucleus. according to Rutherford's calculation the dimensions of the latter are of the same order of magnitude as  $10^{-12}$  cm.

## § 6 *Radioactive Phenomena*

According to the present theory the cluster of electrons surrounding the nucleus is formed with emission of energy, and the configuration is determined by the condition that the energy emitted is a maximum. The stability involved by these assumptions seems to be in agreement with the general properties of matter. It is, however, in striking opposition to the phenomena of radioactivity, and according to the theory the origin of the latter phenomena may therefore be sought elsewhere than in the electronic distribution round the nucleus.

---

<sup>13</sup>Comp. E. Rutherford, Phil. Mag. XXIV. p. 453 (1912).

A necessary consequence of Rutherford's theory of the structure of atoms is that the  $\alpha$ -particles have their origin in the nucleus. On the present theory it seems also necessary that the nucleus is the seat of the expulsion of the high-speed  $\beta$ -particles. In the first place, the spontaneous expulsion of a  $\beta$ -particle from the cluster of electrons surrounding the nucleus would be something quite foreign to the assumed properties of the system. further, the expulsion of an  $\alpha$ -particle can hardly be expected to produce a lasting effect on the stability of the cluster of electrons. The effect of the expulsion will be of two different kinds. Partly the particle may collide with the bound electrons during its passing through the atom. This effect will be analogous to that produced by bombardment of atoms of other substances by  $\alpha$ -rays and cannot be expected to give rise to a subsequent expulsion of  $\beta$ -rays. Partly the expulsion of the particle will involve an alteration in the configuration of the bound electrons, since the charge remaining on the nucleus is different from the original. In order to consider the latter effect let us regard a single ring of electrons rotating round a nucleus of charge  $Ne$ , and let us assume that an  $\alpha$ -particle is expelled from the nucleus in a direction perpendicular to the plane of the ring. The expulsion of the particle will obviously not produce any alteration in the angular momentum of the electrons; and if the velocity of the  $\alpha$ -particle is small compared with the velocity of the electrons – as it will be if we consider inner rings of an atom of high atomic weight – the ring during the expulsion will expand continuously, and after the expulsion will take the position claimed by the theory for a stable ring rotating round a nucleus of charge  $(N - 2) \cdot e$ . The consideration of this simple case strongly indicates that the expulsion of an  $\alpha$ -particle will not have a lasting effect on the stability of the internal rings of electrons in the residual atom.

The question of the origin of  $\beta$ -particles may also be considered from another point of view, based on a consideration of the chemical and physical properties of the radioactive substances. As is well known, several of these substances have very similar chemical properties and have hitherto resisted every attempt to separate them by chemical means. There is also some evidence that the substances in question show the same line-spectrum.<sup>14</sup> It has been suggested by several writers that the substances are different only in radio-active properties and atomic weight but identical in all other physical and chemical respects. according to the theory, this would mean that the charge on the nucleus, as well as the configuration of the surrounding electrons, was identical in some of the elements, the only difference being

---

<sup>14</sup>see A.S. Russel and R. Rossi, Proc. Roy. Soc. A. LXXXVII. p. 478 (1912).

the mass and the internal condition of the nucleus. From the considerations of § 4 this assumption is already strongly suggested by the fact that the number of radioactive substances is greater than the number of places at our disposal in the periodic system. If, however, the assumption is right, the fact that two apparently identical elements emit  $\beta$ -particles of different velocities, shows that the  $\beta$ -rays as well as the  $\alpha$ -rays have their origin in the nucleus.

This view of the origin of  $\alpha$ - and  $\beta$ -particles explains very simply the way in which the change in the chemical properties of the radioactive substances is connected with the nature of the particles emitted. The results of experiments are expressed in the two rules:<sup>15</sup>

1. Whenever an  $\alpha$ -particle is expelled the group in the periodic system to which the resultant product belongs is two units less than that to which the parent body belongs.
2. Whenever a  $\beta$ -particle is expelled the group of the resultant body is 1 unit greater than that of the parent.

As will be seen this is exactly what is to be expected according to the considerations of § 4.

In escaping from the nucleus, the  $\beta$ -rays may be expected to collide with the bound electrons in the inner rings. This will give rise to an emission of a characteristic radiation of the same type as the characteristic Röntgen radiation emitted from elements of lower atomic weight by impact of cathode-rays. The assumption that the emission of  $\gamma$ -rays is due to collisions of  $\beta$ -rays with bound electrons is proposed by Rutherford<sup>16</sup> in order to account for the numerous groups of homogeneous  $\beta$ -rays expelled from certain radioactive substances.

In the present paper it has been attempted to show that the application of Planck's theory of radiation to Rutherford's atom-model through the introduction of the hypothesis of the universal constancy of the angular momentum of the bound electrons, leads to results which seem to be in agreement with experiments.

In a later paper the theory will be applied to systems containing more than one nucleus.

---

<sup>15</sup>See A.S. Russell, Chem. News, CVII. p. 49 (1913); G.v. Hevesy, Phys. Zeitschr. XIV. p. 49 (1913); K. Fajaus, Phys. Zeitschr. XIV. pp. 131 & 136 (1913); Verh. d. deutsch. Phys. Ges. XV. p. 240 (1913); F. Soddy, Chem. News, CVII. p. 97 (1913).

<sup>16</sup>E. Rutherford, Phil. Mag. XXIV. pp. 453&893 (1912).

**The Structure of the Atom**

E. Rutherford,  
University of Manchester  
(Received february 1914)

The present paper and the accompanying paper by Mr.C. Darwin deal with certain points in connexion with the “nucleus” theory of the atom which were purposely omitted in my first communication on that subject (Phil. Mag. May 1911). I brief account is given of the later investigations which have been made to test the theory and of the deductions which can be drawn from them. At the same time a brief statement is given of recent observations on the passage of  $\alpha$  particles through hydrogen, which throw important light on the dimensions of the nucleus.

In my previous paper (*loc. cit.*) I pointed out the importance of the study of the passage of the high speed  $\alpha$  and  $\beta$  particles through matter as a means of throwing light on the internal structure of the atom. attention was drawn to the remarkable fact, first observed by Geiger and Marsden<sup>1</sup> that a small fraction of the swift  $\alpha$  particles from radioactive substances were able to be deflected through an angle of more than  $90^\circ$  as the results of an encounter with a single atom. It was shown that the type of atom devised by Lord Kelvin and worked out in great detail by J.J. Thomson was unable to produce such large deflexions unless the diameter of the positive sphere was exceedingly small. In order to account for this large angle scattering of  $\alpha$  particles, I supposed that the atom consisted of a positively charged nucleus of small dimensions in which practically all the mass of the atom was concentrated. The nucleus was supposed to be surrounded by a distribution of electrons to make the atom electrically neutral, and extending to distance from the nucleus comparable with the ordinary accepted radius of the atom. Some of the swift  $\alpha$  particles passed through the atoms in their path and entered the intense electric field in the neighbourhood of the nucleus and

<sup>1</sup>Proc. Roy. Soc. A. LXXXII. p. 495 (1909).

were deflected from their rectilinear path. In order to suffer a deflexion of more than a few degrees, the  $\alpha$  particle has to pass very close to the nucleus, and it was assumed that the field of force in this region was not appreciably affected by the external electronic distribution. Supposing that the forces between the nucleus and the  $\alpha$  particle are repulsive and follow the law of inverse squares, the  $\alpha$  particle described a hyperbolic orbit round the nucleus and its deflexion can be simply calculated.

It was deduced from this theory that the number of  $\alpha$  particles falling normally on unit area of a surface and making an angle  $\phi$  with the direction of the incident rays is proportional to

- (1)  $\text{cosec}^4 \phi/2$  or  $1/\phi^4$  if  $\phi$  be small;
- (2) the number of atoms per unit volume of the scattering material;
- (3) thickness of scattering material  $t$  provided this is small;
- (4) square of the nucleus charge  $Ne$ ;
- (5) and is inversely proportional to  $(mu^2)^2$ , where  $m$  is the mass of the  $\alpha$  particle and  $u$  its velocity.

From the data of scattering on  $\alpha$  particles previously given by Geiger<sup>2</sup>, it was deduced that the value of the nucleus charge was equal to about half the atomic weight multiplied by the electronic charge. Experiments were begun by Geiger and Marsden<sup>3</sup> to test whether the laws of single scattering of  $\alpha$  particles were in agreement with the theory. The general experimental method employed by them consisted in allowing a narrow pencil of  $\alpha$  particles to fall normally on a thin film of matter, and observing by the scintillation method the number scattering through different angles. This was a very difficult and laborious piece of work involving the counting of many thousands of particles. They found that their results were in very close accord with the theory. When the thickness of the scattering film was very small, the amount of scattering was directly proportional to the thickness and varied inversely as the fourth power of the velocity of the incident  $\alpha$  particles. A special study was made of the number of  $\alpha$  particles scattered through angles varying between  $5^\circ$  and  $150^\circ$ . Although over this range the number decreased in the ratio 200.000 to 1, the relation between number and angle agreed with the theory within the limit of experimental errors. They found that the scattering of different atoms of matter was approximately

<sup>2</sup>Proc. Roy. Soc. A. LXXXIII. p. 492 (1910).

<sup>3</sup>Geiger and Marsden, Phil. Mag. XXV. p. 604 (1913).

proportional to the square of the atomic weight, showing that the charge on the nucleus was nearly proportional to the atomic weight. By determining the number of  $\alpha$  particles scattered from thin films of gold, they concluded that the nucleus charge was equal to about the atomic weight multiplied by the electronic charge. On account of the difficulties of this experiment, the actual number could not be considered correct within more than 20 per cent.

The experimental results of Geiger and Marsden were thus in complete accord with the predictions of the theory, and indicated the essential correctness of this hypothesis of the structure of the atom.

In determining the magnitude of single scattering, I assumed in my previous paper, for simplicity of calculation, that the atom was at rest during an encounter with an  $\alpha$  particle. In an accompanying paper, Mr.C. Darwin has worked out the relations to be expected when account is taken of the motion of the recoiling atom. He has shown that no sensible error has been introduced in this way even for atoms of such low atomic weight as carbon. Mr. Darwin has also worked out the scattering to be expected if the law of force is not that of the inverse square, and has shown that it is not in accord with experiment either with regard to the variation of scattering with angle or with the variation of scattering with velocity. The general evidence certainly indicates that the law of force between the  $\alpha$  particle and the nucleus is that of the inverse square.

It is of interest to note that C.T.R. Wilson<sup>4</sup> by photographing the trails of the  $\alpha$  particle, later showed that the  $\alpha$  particle occasionally suffers a sudden deflexion through a large angle. This affords convincing evidence of the correctness of the view that large deflexions do occasionally occur as a result of an encounter with a single atom.

On the theory outlined, the large deflexions of the  $\alpha$  particle are supposed to be due to its passage close to the nucleus where the field is very intense and to be not appreciably affected by its passage through the external distribution of electrons. This assumption seems to be legitimate when we remember that the mass and energy of the  $\alpha$  particle are very large compared with that of an electron even moving with a velocity comparable with that of light. Simple considerations show that the deflexions which an  $\alpha$  particle would experience even in passing through the complex electronic distribution of a heavy atom like gold, must be small compared with the large deflexions actually observed. In fact, the passage of swift  $\alpha$  particles through matter affords the most definite and straightforward method

---

<sup>4</sup>C.T.R. Wilson, Proc. Roy. Soc. A. LXXXVII. p. 277 (1912).

of throwing light on the gross structure of the atom, for the  $\alpha$  particle is able to penetrate the atom without serious disturbance from the electronic distribution, and thus is only affected by the intense field associated with the nucleus of the atom.

This independence of the large angle scattering on the external distribution of electrons is only true for charged particles whose kinetic energy is very large. It is not to be expected that it will hold for particles moving at very much lower speeds and with much less energy – such, for example, as the ordinary cathode particles or the recoil atoms from active matter. In such cases it is probable that the external electronic distribution plays a far more prominent part in governing the scattering than in the case under consideration.

### *Scattering of $\beta$ Particles*

It is to be anticipated on the nucleus theory that swift  $\beta$  particles should suffer deflexions through large angles in their passage close to the nucleus. There seems to be no doubt that such large deflexions are actually produced, and I showed in my previous paper that the result of scattering of  $\beta$  particles found by Crowther<sup>5</sup> could be generally explained on the nucleus theory of atomic structure. It should be borne in mind, however, that there are several important points of distinction between the effects to be expected for an  $\alpha$  particle and a  $\beta$  particle. Since the force between the nucleus and  $\beta$  particle is attractive, the  $\beta$  particle increases rapidly in speed in approaching the nucleus. On the ordinary electrodynamics, this entails a loss of energy by radiation, and also an increase of the apparent mass of the electron. Darwin<sup>6</sup> has worked out mathematically the result of these effects on the orbit of the electron, and has shown that, under certain conditions, the  $\beta$  particle does not escape from the atom but describes a spiral orbit ultimately falling into the nucleus. This result is of great interest, for it may offer an explanation of the disappearance of swift  $\beta$  particles in their passage through matter. In addition, it must be borne in mind that the swiftest  $\beta$  particle expelled from radium C possesses only about one-third of the energy of the corresponding  $\alpha$  particle, while the average energy of the  $\beta$  particle is less than one-sixth of that of the  $\alpha$  particle. It is thus to be

---

<sup>5</sup>Crowther, Proc. Roy. Soc. A. LXXXIV. p. 226 (1910).

<sup>6</sup>Darwin, Phil. Mag. XXV. p. 210 (1913).

anticipated that the large angle scattering of a  $\beta$  particle by the nucleus will take place in regions where the  $\alpha$  particle will only suffer a small deflexion—regions for which the application of the simple theory may not have been accurately tested. For these reasons, it is of great importance to determine the laws of large angle scattering of  $\beta$  particles of different speeds in passing through matter, as it should throw light on a number of important points connected with atomic structure. Experiments are at present in progress in the laboratory to examine the scattering of such swift  $\beta$  particles in detail.

It is obvious that a  $\beta$  particle in passing close to an electron will occasionally suffer a large deflexion. The problem is mathematically similar to that for a close encounter of an  $\alpha$  particle with a helium atom of the same mass, which is discussed by Mr. Darwin in the accompanying paper. Such large deflexions due to electronic encounter, however, should be relatively small in number compared with those due to the nucleus of a heavy atom.

### *Scattering in Hydrogen*

Special interest attaches to the effects to be expected when  $\alpha$  particles pass through light gases like hydrogen and helium. In a previous paper by Mr. Nuttall and the author<sup>7</sup>, it has been shown that the scattering of  $\alpha$  particles in hydrogen and helium is in good agreement with the view that the hydrogen nucleus has one positive charge, while the  $\alpha$  particle, or helium, has too. Mr. Darwin has worked out in detail the simple scattering to be anticipated when  $\alpha$  particles pass through hydrogen and helium. It is only necessary here to refer to the fact that on the nucleus theory a small number of hydrogen atoms should acquire, as the result of close encounters with  $\alpha$  particles, velocities about 1.6 times that of the velocity of the  $\alpha$  particle itself. On account of the fact that the hydrogen atom carries one positive charge while the  $\alpha$  particle carries two, it can be calculated that the some of the hydrogen atoms should have a range in hydrogen of nearly four times that of the  $\alpha$  particle which sets them in motion.

Mr. Marsden has kindly made experiments for me to test whether the presence of such hydrogen atoms can be detected. A detail account of his experiments will appear later, but it suffices to mention here that undoubtedly evidence has been obtained by him that some of the hydrogen atoms are set in such swift motion that they are able to produce a visible scintillation on a zinc sulphide screen and are able to travel through hydrogen a distance

---

<sup>7</sup>Rutherford and Nuttall, Phil. Mag. XXVI. p. 702 ( 1913).

three or four times greater than the colliding  $\alpha$  particle. The general method employed was to place a thin  $\alpha$ -ray tube containing about 100 millicuries of purified emanation in a tube filled with hydrogen. The scintillations due to the  $\alpha$  particle from the tube disappeared in air after traversing a distance of about 5 cm. When the air was displaced by hydrogen, the great majority of the scintillations disappeared at about 20 cm. from the source, which corresponds to the range of the  $\alpha$  particle in hydrogen. A small number of scintillations, however, persisted in hydrogen up to a distance of about 90 cm. The scintillations were of less intensity than those due to the ordinary  $\alpha$  particle. The number of scintillations observed is of the order of magnitude to be anticipated on the theory of single scattering, supposing that the nucleus in hydrogen and helium has such small dimensions, and that they behave like point charges for distance up to  $10^{-13}$  cm.

There appears to be no doubt that the scintillations observed beyond 20 cm. are due to charged hydrogen atoms which are set in swift motion by a close encounter with an  $\alpha$  particle. Experiments are at present in progress by Mr. Marsden to determine the number of hydrogen atoms set in motion, and the variation of the number with the scattering angle.

It does not possible to explain the appearance of such swift hydrogen atoms unless it be supposed that the forces of repulsion between the  $\alpha$  particle and the hydrogen atom are exceedingly intense. Such intense forces can only arise if the positive nuclei have exceedingly small dimensions, so that a close approach between them is possible.

### *Dimensions and Constitution of the Nucleus*

In my previous paper I showed that the nucleus must have exceedingly small dimensions, and calculated that in the case of gold its radius was not greater than  $3 \times 10^{-12}$  cm. In order to account for the velocity given to hydrogen atoms by the collision with  $\alpha$  particles, it can be simply calculated (see Darwin) that the center of nuclei of helium and hydrogen must approach within a distance of  $1.7 \times 10^{-13}$  cm. of each other. Supposing for simplicity the nuclei to have dimensions and to be spherical in shape, it is clear that the sum of the radii of the hydrogen and helium nuclei is not greater than  $1.7 \times 10^{-13}$  cm. This is an exceedingly small quantity, even *smaller* than the ordinarily accepted value of the diameter of the electron, viz.  $2 \times 10^{-13}$  cm. It is obvious that the method we have considered gives a maximum estimate of the dimensions of the nuclei, and it is not improbable that the

hydrogen nucleus itself may have still smaller dimensions. This rises the question whether the hydrogen nucleus is so small that its mass may be accounted for in the same way as the mass of the negative electron.

It is well known from the experiments of Sir J.J. Thomson and others, that no positively charged carrier has been observed of mass less than of the hydrogen atom. The exceedingly small dimensions found for the hydrogen nucleus add weight to the suggestion that the hydrogen nucleus is the *positive electron*, and that its mass is entirely electromagnetic in origin. According to the electromagnetic theory, the electrical mass of a charged body, supposed spherical, is  $\frac{2}{3} \frac{e^2}{a}$  where  $e$  is the charge and  $a$  the radius. The hydrogen nucleus consequently must have a radius about  $1/1830$  of the electron if its mass is to be explained in this way. There is no experimental evidence at present contrary to such an assumption.

The helium nucleus has a mass nearly four times that of hydrogen. If one supposes that the positive electron, i.e. the hydrogen atom, is a unit of which all atoms are composed, it is to be anticipated that the helium atom contains, four positive electrons and two negative.

It is well known that a helium atom is expelled in many cases in the transformation of radioactive matter, but no evidence has so far been obtained of the expulsion of a hydrogen atom. In conjunction with Mr. Robinson, I have examined whether any other charged atoms are expelled from radioactive matter except helium atoms, and the recoil atoms which accompany the expulsion  $\alpha$  particles. The examination showed that if such particles are expelled, their number is certainly less than 1 in 10,000 of the number of helium atoms. It thus follows that the helium nucleus is a very stable configuration which survives the intense disturbances resulting in its expulsion with high velocity from the radioactive atom, and is one of the units, of which possibly the great majority of the atoms are composed. The radioactive evidence indicates that the atomic weight of successive products decreases by four units consequent on the expulsion of an  $\alpha$  particle, and it has often been pointed out that the atomic weight of many of the permanent atoms differ by about four units.

It will be seen later that the resultant positive charge on the nucleus determines the main physical and chemical properties of the atom. The mass of the atom is, however, dependent on the number and arrangement of the positive and negative electrons constituting the atom. Since the experimental evidence indicates that the nucleus has very small dimensions, the constituent positive and negative electrons must be very closely packed together. As Lorentz has pointed out, the electrical mass of a system of charged particles, if close together, will depend not only on the number of

these particles, but on the way their fields interact. For the dimensions of the positive and negative electrons considered, the packing must be very close in order to produce an appreciable alteration in the mass due to this cause. This may, for example, be the explanation of the fact that the helium atom has not quite four times the mass of the hydrogen atom. Until, however, the nucleus theory has been more definitely tested, it would appear premature to discuss the possible structure of the nucleus itself. The general theory would indicate that the nucleus of a heavy atom is an exceedingly complicated system, although its dimensions are very minute.

An important question arises whether the atomic nuclei, which all carry a positive charge, contain negative electrons. This question has been discussed by Bohr<sup>8</sup>, who concluded from the radioactive evidence that the high speed  $\beta$  particles have their origin in the nucleus. The general radioactive evidence certainly supports such a conclusion. It is well known that the radioactive transformations which are accompanied by the expulsion of high speed  $\beta$  particles are, like the  $\alpha$  ray changes, unaffected by wide ranges of temperature or by physical and chemical conditions. On the nucleus theory, there can be no doubt that the  $\alpha$  particle has its origin in the nucleus and gains a great part, if not all, of its energy of motion in escaping from the atom. It seems reasonable, therefore, to suppose that a  $\beta$  ray transformation also originates from the expulsion of a negative electron from the nucleus. It is well known that the energy expelled in the form of  $\beta$  and  $\gamma$  rays during the transformation of radium C<sup>9</sup> is about one-quarter of the energy of the expelled  $\alpha$  particle. It does not seem easy to explain this large emission of energy by supposing it to have its origin in the electronic distribution. It seems more likely that a very high speed electron is liberated from the nucleus, and in its escape from the atom sets the electronic distribution in violent vibration, giving rise to intense  $\gamma$  rays and also to secondary  $\beta$  particles. The general evidence certainly indicates that many of the high speed electrons from radioactive matter are liberated from the electronic distribution in consequence of the disturbance due to the primary electron escape from the nucleus.

### *Charge on the Nucleus*

We have seen that from an examination of the scattering of  $\alpha$  particles

---

<sup>8</sup>Bohr, Phil. Mag. XXVI. p. 476 (1913).

<sup>9</sup>See Rutherford and Robinson, Phil. Mag. XXV. p. 301 ( 1913).

by matter, it has been found that the positive charge on the nucleus is approximately equal to  $\frac{1}{2}Ae$ , when A is the atomic weight and e the unit charge. This is equivalent to the statement that the number of electrons in the external distribution is about half the atomic weight in terms of hydrogen. It is of interest to note that this is the value deduced by Barkla<sup>10</sup> from entirely different evidence, viz. the scattering of X rays in their passage through matter. This is founded on the theory of scattering given by Sir J.J. Thomson, which supposes that each electron in an atom scatters as an independent unit. It seems improbable that the electrons within the nucleus would contribute to this scattering, for they are packed together with positive nuclei and must be held in equilibrium by forces of a different order of magnitude from those which bind the external electrons.

It is obvious from the consideration of the cases of hydrogen and helium, where hydrogen has one electron and helium two, that the number of electrons cannot be exactly half the atomic weight in all cases. This has led to an interesting suggestion by van Broek<sup>11</sup> that the number of units of charge on the nucleus, and consequently the number of external electrons, may be equal to the number of the elements when arranged in order of increasing atomic weight. On this view, the nucleus charges of hydrogen, helium, and carbon are 1, 2, 6 respectively, and so on far the other elements, provided there is no gap due to a missing element. This view has been taken by Bohr in his theory of the constitution of simple atoms and molecules.

Recently strong evidence of two distinct kinds has been brought in support of such a contention. Soddy<sup>12</sup> has pointed out that the recent generalisation of the relation between the chemical properties of the elements and the radiations can be interpreted by supposing that the atom loses two positive charges by the expulsion of an  $\alpha$  particle, and one negative by the expulsion of a high speed electron. From a consideration of the series of products of the three main radioactive branches of uranium, thorium, and actinium, it follows that some of the radioactive elements may be arranged so that the nucleus charge decreases by one unit as we pass from one element to another. It would thus appear that van den Broek's suggestion probably holds for some if not all of the heavy radioactive elements. Recently Moseley<sup>13</sup> has supplied very valuable evidence that this rule also holds for a number of the lighter elements. By examination of the wave-length of the characteristic X ray emitted by twelve elements varying in atomic weight

<sup>10</sup>Barkla, Phil. Mag. XXI. p. 648 (1911)

<sup>11</sup>van der Broek, Phys. Zeit. XIV. p. 32 (1913).

<sup>12</sup>Soddy, Jahr. d. Rad. X. p. 188 (1913).

<sup>13</sup>Moseley, Phil. Mag. XXVI. p. 1024 (1913).

between calcium (40) and zinc (65.4), he has shown that the variation of wave-length can be simply explained by supposing that the charge on the nucleus increases from element to element by exactly one unit. This holds true for cobalt and nickel, although it has long been known that they occupy an anomalous relative position in the periodic classification of the electrons according to atomic weights.

There appears to be no reason why this new and powerful method of analysis, depending on an examination of the frequency of the characteristic X ray spectra of the elements, should not be extended to a large number of elements, so that further definite data on the point may be expected in the near future.

It is clear on the nucleus theory that the physical and chemical properties of the ordinary elements are for the most part dependent entirely on the charge of the nucleus, for the latter determines the number and distribution of the external electrons on which the chemical and physical properties must mainly depend. As Bohr has pointed out, the properties of gravitation and radioactivity, which are entirely uninfluenced by chemical or physical agencies, must be ascribed mainly if not entirely to the nucleus, while the ordinary physical and chemical properties are determined by the number and distribution of the external electrons. On this view, the nucleus charge is a fundamental constant of the atom, while the atomic mass of an atom may be a complicated function of the arrangement of the units which make up the nucleus.

It should be borne in mind that there is no inherent impossibility on the nucleus theory that atoms may differ considerably in atomic weight and yet have the same nucleus charge. This is most simply illustrated by radioactive evidence. In the following table the atomic weight and nucleus charge are given for a few of the successive elements arising from the transformation of uranium. The actual nucleus charge of uranium is unknown, but for simplicity it is assumed to be 100.

successive Elements	Ur <sub>1</sub> →	Ur X <sub>1</sub> →	Ur X <sub>2</sub> →	U r <sub>2</sub> →	Io →	Ra
Atomic weights	238.5	234.5	234.5	234.5	230.5	226.5
Charge on nucleus	100	98	99	100	98	96

Following the recent theories, it is supposed that the emission of an  $\alpha$  particle lowers the nucleus charge by two units, while the emission of a  $\beta$  particle rises it by one unit. It is seen that Ur<sub>1</sub> and Ur<sub>2</sub> have the same

nucleus charge although they differ in atomic weight by four units.

If the nucleus is supposed to be composed of a mixture of hydrogen nuclei with one charge and of helium nuclei with two charges, it is *a priori* conceivable that a number of atoms may exist with the same nucleus charge but of different atomic masses. The radioactive evidence certainly supports such a view, but probably only a few of such possible atoms would be stable enough to survive for a measurable time.

Bohr<sup>14</sup> has drawn attention to the difficulties of constructing atoms on the "nucleus" theory, and has shown that the stable positions of the external electrons cannot be deduced from the classical mechanics. By the introduction of a conception connected with Planck's quantum, he has shown that on certain assumptions it is possible to construct simple atoms and molecules out of positive and negative nuclei, e.g. by hydrogen atom and molecule and the helium atom, which behave in many respects like the actual atoms or molecules. While there may be much difference of opinion as to the validity and of the underlying physical meaning of the assumptions made by Bohr, there can be no doubt that the theories of Bohr are of great interest and importance to all physicists as the first definite attempt to construct simple atoms and molecules and to explain their spectra.

University of Manchester,

February 1914.

---

<sup>14</sup>Bohr, Phil. Mag. XXVI. pp. 476, 857 (1913)



ON THE GRAVITATIONAL FIELD OF A MASS POINT  
ACCORDING TO EINSTEIN'S THEORY †

BY K. SCHWARZSCHILD

(Communicated January 13th, 1916 [see above p. 42].)

TRANSLATION‡ AND FOREWORD BY  
S. Antoci\* and A. Loinger\*\*

**Foreword.** This fundamental memoir contains the ORIGINAL form of the solution of Schwarzschild's problem. It is regular in the whole space-time, with the only exception of the origin of the spatial co-ordinates; consequently, it leaves no room for the science fiction of the black holes. (In the centuries of the decline of the Roman Empire people said: "Graecum est, non legitur" ...).

§1. In his work on the motion of the perihelion of Mercury (see Sitzungsberichte of November 18th, 1915) Mr. Einstein has posed the following problem:

Let a point move according to the prescription:

$$\left\{ \begin{array}{l} \delta \int ds = 0, \\ \text{where} \\ ds = \sqrt{\sum g_{\mu\nu} dx_\mu dx_\nu} \quad \mu, \nu = 1, 2, 3, 4, \end{array} \right. \quad (1)$$

where the  $g_{\mu\nu}$  stand for functions of the variables  $x$ , and in the variation the variables  $x$  must be kept fixed at the beginning and at the end of the path of integration. In short, the point shall move along a geodesic line in the manifold characterised by the line element  $ds$ .

The execution of the variation yields the equations of motion of the point:

$$\frac{d^2 x_\alpha}{ds^2} = \sum_{\mu, \nu} \Gamma_{\mu\nu}^\alpha \frac{dx_\mu}{ds} \frac{dx_\nu}{ds}, \quad \alpha, \beta = 1, 2, 3, 4, \quad (2)$$

where

$$\Gamma_{\mu\nu}^\alpha = -\frac{1}{2} \sum_\beta g^{\alpha\beta} \left( \frac{\partial g_{\mu\beta}}{\partial x_\nu} + \frac{\partial g_{\nu\beta}}{\partial x_\mu} - \frac{\partial g_{\mu\nu}}{\partial x_\beta} \right), \quad (3)$$

and the  $g^{\alpha\beta}$  stand for the normalised minors associated to  $g_{\alpha\beta}$  in the determinant  $|g_{\mu\nu}|$ .

According to Einstein's theory, this is the motion of a massless point in the gravitational field of a mass at the point  $x_1 = x_2 = x_3 = 0$ , if the "components of the gravitational field"  $\Gamma$  fulfil everywhere, with the exception of the point  $x_1 = x_2 = x_3 = 0$ , the "field equations"

† Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin, Phys.-Math. Klasse 1916, 189-196.

‡ The valuable advice of D.-E. Liebscher is gratefully acknowledged.

\* Dipartimento di Fisica "A. Volta", Università di Pavia, Via Bassi 6 - 27100 Pavia (Italy).

\*\* Dipartimento di Fisica, Università di Milano, Via Celoria 16 - 20133 Milano (Italy).

$$\sum_{\alpha} \frac{\partial \Gamma_{\mu\nu}^{\alpha}}{\partial x_{\alpha}} + \sum_{\alpha\beta} \Gamma_{\mu\beta}^{\alpha} \Gamma_{\nu\alpha}^{\beta} = 0, \quad (4)$$

and if also the “equation of the determinant”

$$|g_{\mu\nu}| = -1 \quad (5)$$

is satisfied.

The field equations together with the equation of the determinant have the fundamental property that they preserve their form under the substitution of other arbitrary variables in lieu of  $x_1, x_2, x_3, x_4$ , as long as the determinant of the substitution is equal to 1.

Let  $x_1, x_2, x_3$  stand for rectangular co-ordinates,  $x_4$  for the time; furthermore, the mass at the origin shall not change with time, and the motion at infinity shall be rectilinear and uniform. Then, according to Mr. Einstein's list, *loc. cit.* p. 833, the following conditions must be fulfilled too:

1. All the components are independent of the time  $x_4$ .
2. The equations  $g_{\rho 4} = g_{4\rho} = 0$  hold exactly for  $\rho = 1, 2, 3$ .
3. The solution is spatially symmetric with respect to the origin of the co-ordinate system in the sense that one finds again the same solution when  $x_1, x_2, x_3$  are subjected to an orthogonal transformation (rotation).
4. The  $g_{\mu\nu}$  vanish at infinity, with the exception of the following four limits different from zero:

$$g_{44} = 1, \quad g_{11} = g_{22} = g_{33} = -1.$$

*The problem is to find out a line element with coefficients such that the field equations, the equation of the determinant and these four requirements are satisfied.*

§2. Mr. Einstein showed that this problem, in first approximation, leads to Newton's law and that the second approximation correctly reproduces the known anomaly in the motion of the perihelion of Mercury. The following calculation yields the exact solution of the problem. It is always pleasant to avail of exact solutions of simple form. More importantly, the calculation proves also the uniqueness of the solution, about which Mr. Einstein's treatment still left doubt, and which could have been proved only with great difficulty, in the way shown below, through such an approximation method. The following lines therefore let Mr. Einstein's result shine with increased clearness.

§3. If one calls  $t$  the time,  $x, y, z$ , the rectangular co-ordinates, the most general line element that satisfies the conditions 1-3 is clearly the following:

$$ds^2 = Fdt^2 - G(dx^2 + dy^2 + dz^2) - H(xdx + ydy + zdz)^2$$

where  $F, G, H$  are functions of  $r = \sqrt{x^2 + y^2 + z^2}$ .

The condition (4) requires: for  $r = \infty : F = G = 1, H = 0$ .

When one goes over to polar co-ordinates according to  $x = r \sin \vartheta \cos \phi, y = r \sin \vartheta \sin \phi, z = r \cos \vartheta$ , the same line element reads:

$$\begin{aligned} ds^2 &= Fdt^2 - G(dr^2 + r^2 d\vartheta^2 + r^2 \sin^2 \vartheta d\phi^2) - Hr^2 dr^2 \\ &= Fdt^2 - (G + Hr^2)dr^2 - Gr^2(d\vartheta^2 + \sin^2 \vartheta d\phi^2). \end{aligned} \quad (6)$$

Now the volume element in polar co-ordinates is equal to  $r^2 \sin \vartheta dr d\vartheta d\phi$ , the functional determinant  $r^2 \sin \vartheta$  of the old with respect to the new coordinates is different from 1; then the field equations

would not remain in unaltered form if one would calculate with these polar co-ordinates, and one would have to perform a cumbersome transformation. However there is an easy trick to circumvent this difficulty. One puts:

$$x_1 = \frac{r^3}{3}, \quad x_2 = -\cos \vartheta, \quad x_3 = \phi. \quad (7)$$

Then we have for the volume element:  $r^2 dr \sin \vartheta d\vartheta d\phi = dx_1 dx_2 dx_3$ . The new variables are then *polar co-ordinates with the determinant 1*. They have the evident advantages of polar co-ordinates for the treatment of the problem, and at the same time, when one includes also  $t = x_4$ , the field equations and the determinant equation remain in unaltered form.

In the new polar co-ordinates the line element reads:

$$ds^2 = F dx_4^2 - \left( \frac{G}{r^4} + \frac{H}{r^2} \right) dx_1^2 - Gr^2 \left[ \frac{dx_2^2}{1-x_2^2} + dx_3^2(1-x_2^2) \right], \quad (8)$$

for which we write:

$$ds^2 = f_4 dx_4^2 - f_1 dx_1^2 - f_2 \frac{dx_2^2}{1-x_2^2} - f_3 dx_3^2(1-x_2^2). \quad (9)$$

Then  $f_1, f_2, f_3, f_4$  are three functions of  $x_1$  which have to fulfil the following conditions:

1. For  $x_1 = \infty$ :  $f_1 = 1/r^4 = (3x_1)^{-4/3}$ ,  $f_2 = f_3 = r^2 = (3x_1)^{2/3}$ ,  $f_4 = 1$ .
2. The equation of the determinant:  $f_1 \cdot f_2 \cdot f_3 \cdot f_4 = 1$ .
3. The field equations.
4. Continuity of the  $f$ , except for  $x_1 = 0$ .

§4. In order to formulate the field equations one must first form the components of the gravitational field corresponding to the line element (9). This happens in the simplest way when one builds the differential equations of the geodesic line by direct execution of the variation, and reads out the components from these. The differential equations of the geodesic line for the line element (9) result from the variation immediately in the form:

$$\begin{aligned} 0 &= f_1 \frac{d^2 x_1}{ds^2} + \frac{1}{2} \frac{\partial f_4}{\partial x_1} \left( \frac{dx_4}{ds} \right)^2 + \frac{1}{2} \frac{\partial f_1}{\partial x_1} \left( \frac{dx_1}{ds} \right)^2 - \frac{1}{2} \frac{\partial f_2}{\partial x_1} \left[ \frac{1}{1-x_2^2} \left( \frac{dx_2}{ds} \right)^2 + (1-x_2^2) \left( \frac{dx_3}{ds} \right)^2 \right] \\ 0 &= \frac{f_2}{1-x_2^2} \frac{d^2 x_2}{ds^2} + \frac{\partial f_2}{\partial x_1} \frac{1}{1-x_2^2} \frac{dx_1}{ds} \frac{dx_2}{ds} + \frac{f_2 x_2}{(1-x_2^2)^2} \left( \frac{dx_2}{ds} \right)^2 + f_2 x_2 \left( \frac{dx_3}{ds} \right)^2 \\ 0 &= f_2 (1-x_2^2) \frac{d^2 x_3}{ds^2} + \frac{\partial f_2}{\partial x_1} (1-x_2^2) \frac{dx_1}{ds} \frac{dx_3}{ds} - 2 f_2 x_2 \frac{dx_2}{ds} \frac{dx_3}{ds} \\ 0 &= f_4 \frac{d^2 x_4}{ds^2} + \frac{\partial f_4}{\partial x_1} \frac{dx_1}{ds} \frac{dx_4}{ds}. \end{aligned}$$

The comparison with (2) gives the components of the gravitational field:

$$\begin{aligned}
\Gamma_{11}^1 &= -\frac{1}{2} \frac{1}{f_1} \frac{\partial f_1}{\partial x_1}, & \Gamma_{22}^1 &= +\frac{1}{2} \frac{1}{f_1} \frac{\partial f_2}{\partial x_1} \frac{1}{1-x_2^2}, \\
\Gamma_{33}^1 &= +\frac{1}{2} \frac{1}{f_1} \frac{\partial f_2}{\partial x_1} (1-x_2^2), \\
\Gamma_{44}^1 &= -\frac{1}{2} \frac{1}{f_1} \frac{\partial f_4}{\partial x_1}, \\
\Gamma_{21}^2 &= -\frac{1}{2} \frac{1}{f_2} \frac{\partial f_2}{\partial x_1}, & \Gamma_{22}^2 &= -\frac{x_2}{1-x_2^2}, & \Gamma_{33}^2 &= -x_2(1-x_2^2), \\
\Gamma_{31}^3 &= -\frac{1}{2} \frac{1}{f_2} \frac{\partial f_2}{\partial x_1}, & \Gamma_{32}^3 &= +\frac{x_2}{1-x_2^2}, \\
\Gamma_{41}^4 &= -\frac{1}{2} \frac{1}{f_4} \frac{\partial f_4}{\partial x_1}
\end{aligned}$$

(the other ones are zero).

Due to the rotational symmetry around the origin it is sufficient to write the field equations only for the equator ( $x_2 = 0$ ); therefore, since they will be differentiated only once, in the previous expressions it is possible to set everywhere since the beginning  $1 - x_2^2$  equal 1. The calculation of the field equations then gives

$$\begin{aligned}
a) \quad & \frac{\partial}{\partial x_1} \left( \frac{1}{f_1} \frac{\partial f_1}{\partial x_1} \right) = \frac{1}{2} \left( \frac{1}{f_1} \frac{\partial f_1}{\partial x_1} \right)^2 + \left( \frac{1}{f_2} \frac{\partial f_2}{\partial x_1} \right)^2 + \frac{1}{2} \left( \frac{1}{f_4} \frac{\partial f_4}{\partial x_1} \right)^2, \\
b) \quad & \frac{\partial}{\partial x_1} \left( \frac{1}{f_1} \frac{\partial f_2}{\partial x_1} \right) = 2 + \frac{1}{f_1 f_2} \left( \frac{\partial f_2}{\partial x_1} \right)^2, \\
c) \quad & \frac{\partial}{\partial x_1} \left( \frac{1}{f_1} \frac{\partial f_4}{\partial x_1} \right) = \frac{1}{f_1 f_4} \left( \frac{\partial f_4}{\partial x_1} \right)^2.
\end{aligned}$$

Besides these three equations the functions  $f_1, f_2, f_3$  must fulfil also the equation of the determinant

$$d) \quad f_1 f_2^2 f_4 = 1, \text{ i. e. } \frac{1}{f_1} \frac{\partial f_1}{\partial x_1} + \frac{2}{f_2} \frac{\partial f_2}{\partial x_1} + \frac{1}{f_4} \frac{\partial f_4}{\partial x_1} = 0.$$

For now I neglect (b) and determine the three functions  $f_1, f_2, f_4$  from (a), (c), and (d). (c) can be transposed into the form

$$c') \quad \frac{\partial}{\partial x_1} \left( \frac{1}{f_4} \frac{\partial f_4}{\partial x_1} \right) = \frac{1}{f_1 f_4} \frac{\partial f_1}{\partial x_1} \frac{\partial f_4}{\partial x_1}.$$

This can be directly integrated and gives

$$c'') \quad = \frac{1}{f_4} \frac{\partial f_4}{\partial x_1} = \alpha f_1, \quad (\alpha \text{ integration constant})$$

The addition of (a) and (c') gives

$$\frac{\partial}{\partial x_1} \left( \frac{1}{f_1} \frac{\partial f_1}{\partial x_1} + \frac{1}{f_4} \frac{\partial f_4}{\partial x_1} \right) = \left( \frac{1}{f_2} \frac{\partial f_2}{\partial x_1} \right)^2 + \frac{1}{2} \left( \frac{1}{f_1} \frac{\partial f_1}{\partial x_1} + \frac{1}{f_4} \frac{\partial f_4}{\partial x_1} \right)^2.$$

By taking (d) into account it follows

$$-2 \frac{\partial}{\partial x_1} \left( \frac{1}{f_2} \frac{\partial f_2}{\partial x_1} \right) = 3 \left( \frac{1}{f_2} \frac{\partial f_2}{\partial x_1} \right)^2.$$

By integrating

$$\frac{1}{\frac{1}{f_2} \frac{\partial f_2}{\partial x_1}} = \frac{3}{2} x_1 + \frac{\rho}{2} \quad (\rho \text{ integration constant})$$

or

$$\frac{1}{f_2} \frac{\partial f_2}{\partial x_1} = \frac{2}{3x_1 + \rho}.$$

By integrating once more,

$$f_2 = \lambda (3x_1 + \rho)^{2/3}. \quad (\lambda \text{ integration constant})$$

The condition at infinity requires:  $\lambda = 1$ . Then

$$f_2 = (3x_1 + \rho)^{2/3}. \quad (10)$$

Hence it results further from (c'') and (d)

$$\frac{\partial f_4}{\partial x_1} = \alpha f_1 f_4 = \frac{\alpha}{f_2^2} = \frac{\alpha}{(3x_1 + \rho)^{4/3}}.$$

By integrating while taking into account the condition at infinity

$$f_4 = 1 - \alpha (3x_1 + \rho)^{-1/3}. \quad (11)$$

Hence from (d)

$$f_1 = \frac{(3x_1 + \rho)^{-4/3}}{1 - \alpha (3x_1 + \rho)^{-1/3}}. \quad (12)$$

As can be easily verified, the equation (b) is automatically fulfilled by the expressions that we found for  $f_1$  and  $f_2$ .

Therefore all the conditions are satisfied apart from the *condition of continuity*.  $f_1$  will be discontinuous when

$$1 = \alpha (3x_1 + \rho)^{-1/3}, \quad 3x_1 = \alpha^3 - \rho.$$

In order that this discontinuity coincides with the origin, it must be

$$\rho = \alpha^3. \quad (13)$$

Therefore the condition of continuity relates in this way the two integration constants  $\rho$  and  $\alpha$ .

The complete solution of our problem reads now:

$$f_1 = \frac{1}{R^4} \frac{1}{1 - \alpha/R}, \quad f_2 = f_3 = R^2, \quad f_4 = 1 - \alpha/R,$$

where the auxiliary quantity

$$R = (3x_1 + \rho)^{1/3} = (r^3 + \alpha^3)^{1/3}$$

has been introduced.

*When one introduces these values of the functions  $f$  in the expression (9) of the line element and goes back to the usual polar co-ordinates one gets the line element that forms the exact solution of Einstein's problem:*

$$ds^2 = (1 - \alpha/R)dt^2 - \frac{dR^2}{1 - \alpha/R} - R^2(d\vartheta^2 + \sin^2 \vartheta d\phi^2), \quad R = (r^3 + \alpha^3)^{1/3}. \quad (14)$$

The latter contains only the constant  $\alpha$  that depends on the value of the mass at the origin.

§5. *The uniqueness of the solution* resulted spontaneously through the present calculation. From what follows we can see that it would have been difficult to ascertain the uniqueness from an approximation procedure in the manner of Mr. Einstein. Without the continuity condition it would have resulted:

$$f_1 = \frac{(3x_1 + \rho)^{-4/3}}{1 - \alpha(3x_1 + \rho)^{-1/3}} = \frac{(r^3 + \rho)^{-4/3}}{1 - \alpha(r^3 + \rho)^{-1/3}}.$$

When  $\alpha$  and  $\rho$  are small, the series expansion up to quantities of second order gives:

$$f_1 = \frac{1}{r^4} \left[ 1 + \frac{\alpha}{r} - 4/3 \frac{\rho}{r^3} \right].$$

This expression, together with the corresponding expansions of  $f_2$ ,  $f_3$ ,  $f_4$ , satisfies up to the same accuracy all the conditions of the problem. Within this approximation the condition of continuity does not introduce anything new, since discontinuities occur spontaneously only in the origin. Then the two constants  $\alpha$  and  $\rho$  appear to remain arbitrary, hence the problem would be physically undetermined. The exact solution teaches that in reality, by extending the approximations, the discontinuity does not occur at the origin, but at  $r = (\alpha^3 - \rho)^{1/3}$ , and that one must set just  $\rho = \alpha^3$  for the discontinuity to go in the origin. With the approximation in powers of  $\alpha$  and  $\rho$  one should survey very closely the law of the coefficients in order to recognise the necessity of this link between  $\alpha$  and  $\rho$ .

§6. Finally, one has still to derive the *motion of a point in the gravitational field*, the geodesic line corresponding to the line element (14). From the three facts, that the line element is homogeneous in the differentials and that its coefficients do not depend on  $t$  and on  $\phi$ , with the variation we get immediately three intermediate integrals. If one also restricts himself to the motion in the equatorial plane ( $\vartheta = 90^\circ$ ,  $d\vartheta = 0$ ) these intermediate integrals read:

$$(1 - \alpha/R) \left( \frac{dt}{ds} \right)^2 - \frac{1}{1 - \alpha/R} \left( \frac{dR}{ds} \right)^2 - R^2 \left( \frac{d\phi}{ds} \right)^2 = \text{const.} = h, \quad (15)$$

$$R^2 \frac{d\phi}{ds} = \text{const.} = c, \quad (16)$$

$$(1 - \alpha/R) \frac{dt}{ds} = \text{const.} = 1 \quad (\text{determination of the time unit}). \quad (17)$$

From here it follows

$$\left( \frac{dR}{d\phi} \right)^2 + R^2(1 - \alpha/R) = \frac{R^4}{c^2} [1 - h(1 - \alpha/R)]$$

or with  $1/R = x$

$$\left(\frac{dx}{d\phi}\right)^2 = \frac{1-h}{c^2} + \frac{h\alpha}{c^2}x - x^2 + \alpha x^3. \quad (18)$$

If one introduces the notations:  $c^2/h = B$ ,  $(1-h)/h = 2A$ , this is identical to Mr. Einstein's equation (11), *loc. cit.* and gives the observed anomaly of the perihelion of Mercury.

Actually Mr. Einstein's approximation for the orbit goes into the exact solution when one substitutes for  $r$  the quantity

$$R = (r^3 + \alpha^3)^{1/3} = r \left(1 + \frac{\alpha^3}{r^3}\right)^{1/3}.$$

Since  $\alpha/r$  is nearly equal to twice the square of the velocity of the planet (with the velocity of light as unit), for Mercury the parenthesis differs from 1 only for quantities of the order  $10^{-12}$ . Therefore  $r$  is virtually identical to  $R$  and Mr. Einstein's approximation is adequate to the strongest requirements of the practice.

Finally, the exact form of the third Kepler's law for circular orbits will be derived. Owing to (16) and (17), when one sets  $x = 1/R$ , for the angular velocity  $n = d\phi/dt$  it holds

$$n = cx^2(1 - \alpha x).$$

For circular orbits both  $dx/d\phi$  and  $d^2x/d\phi^2$  must vanish. Due to (18) this gives:

$$0 = \frac{1-h}{c^2} + \frac{h\alpha}{c^2}x - x^2 + \alpha x^3, \quad 0 = \frac{h\alpha}{c^2} - 2x + 3\alpha x^2.$$

The elimination of  $h$  from these two equations yields

$$\alpha = 2c^2x(1 - \alpha x)^2.$$

Hence it follows

$$n^2 = \frac{\alpha}{2}x^3 = \frac{\alpha}{2R^3} = \frac{\alpha}{2(r^3 + \alpha^3)}.$$

The deviation of this formula from the third Kepler's law is totally negligible down to the surface of the Sun. For an ideal mass point, however, it follows that the angular velocity does not, as with Newton's law, grow without limit when the radius of the orbit gets smaller and smaller, but it approaches a determined limit

$$n_0 = \frac{1}{\alpha\sqrt{2}}.$$

(For a point with the solar mass the limit frequency will be around  $10^4$  per second). This circumstance could be of interest, if analogous laws would rule the molecular forces.



## The Foundation of the General Theory of Relativity

by A. Einstein

[This first page was missing in the existing translation.]

The theory which is presented in the following pages conceivably constitutes the farthest-reaching generalization of a theory which, today, is generally called the [1] “theory of relativity”; I will call the latter one—in order to distinguish it from the [2] first named—the “special theory of relativity,” which I assume to be known. The generalization of the theory of relativity has been facilitated considerably by [3] Minkowski, a mathematician who was the first one to recognize the formal equivalence of space coordinates and the time coordinate, and utilized this in the construction of the theory. The mathematical tools that are necessary for general [4] relativity were readily available in the “absolute differential calculus,” which is based upon the research on non-Euclidean manifolds by Gauss, Riemann, and Christoffel, and which has been systematized by Ricci and Levi-Civita and has already been [5] applied to problems of theoretical physics. In section B of the present paper I developed all the necessary mathematical tools—which cannot be assumed to be known to every physicist—and I tried to do it in as simple and transparent a manner as possible, so that a special study of the mathematical literature is not required for [6] the understanding of the present paper. Finally, I want to acknowledge gratefully my friend, the mathematician Grossmann, whose help not only saved me the effort of studying the pertinent mathematical literature, but who also helped me in my search for the field equations of gravitation.

[The balance of this translation is reprinted from H. A. Lorentz et al., *The Principle of Relativity*, trans. W. Perrett and G. B. Jeffery (Methuen, 1923; Dover rpt., 1952).]

## THE FOUNDATION OF THE GENERAL THEORY OF RELATIVITY

By A. EINSTEIN

### A. FUNDAMENTAL CONSIDERATIONS ON THE POSTULATE OF RELATIVITY

#### § I. 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

[7]

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 experiences.

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  $K'$ .

[8]

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

[9]

\* 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.

[10] 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  $\pi$ . With a measuring-rod at rest relatively to  $K'$ , the quotient would be greater than  $\pi$ . 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-covariant with respect to any substitutions whatever (generally co-covariant).*

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-covariance, 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

[11]

there is a corresponding system of values of the variables  $x_1 \dots x_4$ . To two coincident point-events there corresponds one system of values of the variables  $x_1 \dots x_4$ , i.e. coincidence is characterized by the identity of the co-ordinates. If, in place of the variables  $x_1 \dots x_4$ , 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

$$ds^2 = -dX_1^2 - dX_2^2 - dX_3^2 + dX_4^2 \quad . \quad . \quad . \quad (1)$$

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  $dX_1 \dots dX_4$  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  $dx_1 \dots dx_4$  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  $dX_\nu$  will allow themselves to be represented here by definite linear homogeneous expressions of the  $dx_\sigma$  :—

$$dX_\nu = \sum_\sigma a_{\nu\sigma} dx_\sigma \quad . \quad . \quad . \quad . \quad (2)$$

Inserting these expressions in (1), we obtain

$$ds^2 = \sum_{\sigma\tau} g_{\sigma\tau} dx_\sigma dx_\tau, \quad . \quad . \quad . \quad . \quad (3)$$

where the  $g_{\sigma\tau}$  will be functions of the  $x_\sigma$ . 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  $\sigma$  and  $\tau$ , so that the sum consists of  $4 \times 4$  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

$$\begin{matrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & +1 \end{matrix} \quad \dots \quad . \quad . \quad . \quad (4)$$

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 stand-point 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  $g_{\tau\sigma}$  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  $g_{\sigma\tau}$ . 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  $x_1 \dots x_4$ , 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"  $dx_\nu$ , for which the law of transformation is expressed by the equation

$$dx'_\sigma = \sum_\nu \frac{\partial x'_\sigma}{\partial x_\nu} dx_\nu . . . . . \quad (5)$$

The  $dx'_\sigma$  are expressed as linear and homogeneous functions of the  $dx_\nu$ . 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  $A^\nu$ , and which is transformed by the same law

$$A'^\sigma = \sum_\nu \frac{\partial x'^\sigma}{\partial x_\nu} A^\nu, . . . . . \quad (5a)$$

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  $A^\sigma$  and  $B^\sigma$  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  $A_\nu$ , the components of a covariant four-vector, if for any arbitrary choice of the contravariant four-vector  $B^\nu$

$$\sum_{\nu} A_\nu B^\nu = \text{Invariant} \quad . \quad . \quad . \quad (6)$$

The law of transformation of a covariant four-vector follows from this definition. For if we replace  $B^\nu$  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

$$\sum_{\sigma} B'^{\sigma} \sum_{\nu} \frac{\partial x_\nu}{\partial x'_\sigma} A_\nu = \sum_{\sigma} B'^{\sigma} A'_\sigma.$$

Since this equation is true for arbitrary values of the  $B'^{\sigma}$ , 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  $\nu$  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.

[13]

### § 6. Tensors of the Second and Higher Ranks

*Contravariant Tensors.*—If we form all the sixteen products  $A^{\mu\nu}$  of the components  $A^\mu$  and  $B^\nu$  of two contravariant four-vectors

$$A^{\mu\nu} = A^\mu B^\nu \quad . \quad (8)$$

then by (8) and (5a)  $A^{\mu\nu}$  satisfies the law of transformation

$$A'^{\sigma\tau} = \frac{\partial x'_\sigma}{\partial x_\mu} \frac{\partial x'_\tau}{\partial x_\nu} A^{\mu\nu} \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

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  $A_\mu$  and  $B_\nu$ ,

$$A_{\mu\nu} = A_\mu B_\nu, \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

the law of transformation for these is

$$A'_{\sigma\tau} = \frac{\partial x_\mu}{\partial x'_\sigma} \frac{\partial x_\nu}{\partial x'_\tau} A_{\mu\nu} \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

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  $\mu$ , and contravariant with respect to the index  $\nu$ . Its law of transformation is

$$A_{\sigma}^{\tau} = \frac{\partial x'^{\tau}}{\partial x_{\nu}} \frac{\partial x_{\mu}}{\partial x'^{\sigma}} A_{\mu}^{\nu} \quad . \quad . \quad . \quad . \quad (13)$$

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  $\mu, \nu$ ,

$$A^{\mu\nu} = A^{\nu\mu}, \quad . \quad . \quad . \quad . \quad . \quad (14)$$

or respectively,

$$A_{\mu\nu} = A_{\nu\mu}. \quad . \quad . \quad . \quad . \quad . \quad (14a)$$

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  $\mu$  and  $\nu$ , 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

$$A^{\mu\nu} = - A^{\nu\mu}, \quad \dots \quad \dots \quad \dots \quad (15)$$

or respectively,

$$A_{\mu\nu} = - A_{\nu\mu}, \quad \dots \quad \dots \quad \dots \quad (15a)$$

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,

$$\begin{aligned} T_{\mu\nu\sigma} &= A_{\mu\nu}B_\sigma, \\ T^{\mu\nu\sigma\tau} &= A^{\mu\nu}B^{\sigma\tau}, \\ T_{\mu\nu}^{\sigma\tau} &= A_{\mu\nu}B^{\sigma\tau}. \end{aligned}$$

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  $A_{\mu\nu}^{\sigma\tau}$ , we obtain the mixed tensor of the second rank,

$$A_\nu^\tau = A_{\mu\nu}^{\mu\tau} \quad (= \sum_\mu A_{\mu\nu}^{\mu\tau}),$$

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  $B^\sigma$  we form by outer multiplication the mixed tensor

$$D_{\mu\nu}^\sigma = A_{\mu\nu}B^\sigma.$$

On contraction with respect to the indices  $\nu$  and  $\sigma$ , 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  $B^\sigma$ . 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  $\mu$  and  $\tau$ , and “inner” with respect to the indices  $\nu$  and  $\sigma$ .

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  $B^\mu$  and  $C^\nu$  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  $B^\mu$  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  $B^\nu$  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  $C^\mu$  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  $B^\nu$  and  $C^\mu$ . 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  $dx_\mu$  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

$$g_{\mu\sigma}g^{\nu\sigma} = \delta_{\mu}^{\nu} . . . . . \quad (16)$$

where the symbol  $\delta_{\mu}^{\nu}$  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  $dx_{\mu}$  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  $\sigma$  and  $\tau$ , it follows from the results of the preceding paragraph that  $g^{\sigma\tau}$  is a contravariant tensor.

It further follows from (16) that  $\delta_{\mu}^{\nu}$  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

$$| g_{\mu\nu} | \times | g^{\mu\nu} | = 1 . . . . . \quad (17)$$

*The Volume Scalar.*—We seek first the law of transfor-

mation of the determinant  $g = |g_{\mu\nu}|$ . In accordance with (11)

$$g' = \left| \frac{\partial x_\mu}{\partial x'_\sigma} \frac{\partial x_\nu}{\partial x'_\tau} g_{\mu\nu} \right|.$$

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  $dX_1 \dots dX_4$ . 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

$$\left| \frac{\partial x'^\sigma}{\partial x_\mu} \right| = 1 \quad . \quad . \quad . \quad . \quad . \quad (19)$$

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 covariant in face of all substitutions for which the determinant is unity?" but our question is "What are the generally covariant 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,

$$\begin{aligned} A^\mu &= g^{\mu\sigma} A_\sigma, \\ A &= g_{\mu\nu} A^{\mu\nu}. \end{aligned}$$

The following forms may be specially noted:—

$$\begin{aligned} A^{\mu\nu} &= g^{\mu\alpha} g^{\nu\beta} A_{\alpha\beta}, \\ A_{\mu\nu} &= g_{\mu\alpha} g_{\nu\beta} A^{\alpha\beta} \end{aligned}$$

(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  $\lambda$  be a function of the co-ordinates  $x_\nu$ , 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  $x_\nu$  as functions of  $\lambda$ . Let the symbol  $\delta$  indicate the transition from a point of the required geodetic to the point corresponding to the same  $\lambda$  on a neighbouring line. Then for (20) we may substitute

$$\left. \begin{aligned} \int_{\lambda_1}^{\lambda_2} \delta w d\lambda &= 0 \\ w^2 &= g_{\mu\nu} \frac{dx_\mu}{d\lambda} \frac{dx_\nu}{d\lambda} \end{aligned} \right\} \quad . \quad . \quad . \quad . \quad (20a)$$

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

$$[14] \quad \kappa_\sigma = \frac{d}{d\lambda} \left\{ \frac{g_{\mu\nu}}{w} \frac{dx_\mu}{d\lambda} \right\} - \frac{1}{2w} \frac{\partial g_{\mu\nu}}{\partial x_\sigma} \frac{dx_\mu}{d\lambda} \frac{dx_\nu}{d\lambda} \quad . \quad (20b)$$

Since the values of  $\delta x_\sigma$  are arbitrary, it follows from this that

$$\kappa_\sigma = 0 \quad . \quad . \quad . \quad . \quad . \quad (20c)$$

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  $\lambda$ . Then  $w = 1$ , and in place of (20c) we obtain

$$[15] \quad g_{\mu\nu} \frac{d^2 x_\mu}{ds^2} + \frac{\partial g_{\mu\nu}}{\partial x_\sigma} \frac{dx_\sigma}{ds} \frac{dx_\mu}{ds} - \frac{1}{2} \frac{\partial g_{\mu\nu}}{\partial x_\sigma} \frac{dx_\mu}{ds} \frac{dx_\nu}{ds} = 0$$

or, by a mere change of notation,

$$g_{\alpha\sigma} \frac{d^2 x_\alpha}{ds^2} + [\mu\nu, \sigma] \frac{dx_\mu}{ds} \frac{dx_\nu}{ds} = 0 \quad . \quad . \quad . \quad (20d)$$

where, following Christoffel, we have written

$$[\mu\nu, \sigma] = \frac{1}{2} \left( \frac{\partial g_{\mu\sigma}}{\partial x_\nu} + \frac{\partial g_{\nu\sigma}}{\partial x_\mu} - \frac{\partial g_{\mu\nu}}{\partial x_\sigma} \right) \quad . \quad . \quad . \quad (21)$$

Finally, if we multiply (20d) by  $g^{\sigma\tau}$  (outer multiplication with respect to  $\tau$ , inner with respect to  $\sigma$ ), we obtain the equations of the geodetic line in the form

$$\frac{d^2 x_\tau}{ds^2} + \{\mu\nu, \tau\} \frac{dx_\mu}{ds} \frac{dx_\nu}{ds} = 0 \quad . \quad . \quad . \quad (22)$$

where, following Christoffel, we have set

$$\{\mu\nu, \tau\} = g^{\tau\alpha} [\mu\nu, \alpha] \quad . \quad . \quad . \quad (23)$$

### § 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,  $\phi$  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

$$\psi = \frac{\partial\phi}{\partial x_\mu} \frac{dx_\mu}{ds}$$

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  $dx_\mu$ . 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  $\phi$ .

According to our rule, the differential quotient

$$\chi = \frac{d\psi}{ds}$$

taken on a curve, is similarly an invariant. Inserting the value of  $\psi$ , 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  $d^2x_\nu/ds^2$  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  $\mu$  and  $\nu$ , it follows that the expression in brackets is symmetrical in  $\mu$  and  $\nu$ . Since a geodetic line can be drawn in any direction from a point of the continuum, and therefore  $dx_\mu/ds$  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  $A_\mu$ . In the first place we can readily show that the operation leads to a tensor, even if the vector  $A_\mu$  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  $\psi$  and  $\phi$  are scalars. The sum of four such terms

$$S_\mu = \psi^{(1)} \frac{\phi^{(1)}}{\partial x_\mu} + \dots + \psi^{(4)} \frac{\phi^{(4)}}{\partial x_\mu},$$

is also a covariant vector, if  $\psi^{(1)}, \phi^{(1)}, \dots, \psi^{(4)}, \phi^{(4)}$  are scalars. But it is clear that any covariant vector can be represented in the form  $S_\mu$ . For, if  $A_\mu$  is a vector whose components are any given functions of the  $x_\nu$ , 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  $S_\mu$  shall be equal to  $A_\mu$ .

Therefore, in order to demonstrate that  $A_{\mu\nu}$  is a tensor if any covariant vector is inserted on the right-hand side for  $A_\mu$ , we only need show that this is so for the vector  $S_\mu$ . 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  $\psi$ ,

$$\psi \frac{\partial^2 \phi}{\partial x_\mu \partial x_\nu} - \{\mu\nu, \tau\} \psi \frac{\partial \phi}{\partial x_\tau}$$

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  $A_\mu$ .

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

$$\begin{array}{cccc} A_{11} & A_{12} & A_{13} & A_{14} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0. \end{array}$$

By the addition of four tensors of this type, we obtain the tensor  $A_{\mu\nu}$  with any assigned 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  $B_\nu$ , and of the second by  $A_\mu$ , we obtain in each case a tensor of the third rank. By adding these, we have the tensor of the third rank

$$A_{\mu\nu\sigma} = \frac{\partial A_{\mu\nu}}{\partial x_\sigma} - \{\sigma\mu, \tau\}A_{\tau\nu} - \{\sigma\nu, \tau\}A_{\mu\tau}. \quad . \quad (27)$$

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.

### § II. 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

$$dg = g^{\mu\nu}gdg_{\mu\nu} = -g_{\mu\nu}gdg^{\mu\nu} \quad . \quad . \quad (28)$$

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  $g_{\mu\sigma}g^{\nu\sigma} = \delta_\mu^\nu$ , it follows on differentiation that

$$\left. \begin{aligned} g_{\mu\sigma} \frac{\partial g^{\nu\sigma}}{\partial x_\lambda} &= -g^{\nu\sigma} \frac{\partial g_{\mu\sigma}}{\partial x_\lambda} \\ g_{\mu\sigma} \frac{\partial g^{\nu\sigma}}{\partial x_\lambda} &= -g^{\nu\sigma} \frac{\partial g_{\mu\sigma}}{\partial x_\lambda} \end{aligned} \right\} \quad . . . . \quad (30)$$

From these, by mixed multiplication by  $g^{\sigma\tau}$  and  $g_{\nu\lambda}$  respectively, and a change of notation for the indices, we have

$$\left. \begin{aligned} dg^{\mu\nu} &= -g^{\mu a} g^{\nu b} \frac{\partial g_{ab}}{\partial x_\sigma} \\ \frac{\partial g^{\mu\nu}}{\partial x_\sigma} &= -g^{\mu a} g^{\nu b} \frac{\partial g_{ab}}{\partial x_\sigma} \end{aligned} \right\} \quad . . . . \quad (31)$$

and

$$\left. \begin{aligned} dg_{\mu\nu} &= -g_{\mu a} g_{\nu b} \frac{\partial g^{ab}}{\partial x_\sigma} \\ \frac{\partial g_{\mu\nu}}{\partial x_\sigma} &= -g_{\mu a} g_{\nu b} \frac{\partial g^{ab}}{\partial x_\sigma} \end{aligned} \right\} \quad . . . . \quad (32)$$

The relation (31) admits of a transformation, of which we also have frequently to make use. From (21)

$$\frac{\partial g_{ab}}{\partial x_\sigma} = [a\sigma, b] + [\beta\sigma, a] \quad . . . . \quad (33)$$

Inserting this in the second formula of (31), we obtain, in view of (23)

$$\frac{\partial g^{\mu\nu}}{\partial x_\sigma} = -g^{\mu\tau} \{ \tau\sigma, \nu \} - g^{\nu\tau} \{ \tau\sigma, \mu \} \quad . . . . \quad (34)$$

Substituting the right-hand side of (34) in (29), we have

$$\frac{1}{\sqrt{-g}} \frac{\partial \sqrt{-g}}{\partial x_\sigma} = \{ \mu\sigma, \mu \} \quad . . . . \quad (29a)$$

*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

$$\frac{\partial}{\partial x_\nu} (g^{\mu\nu} A_\mu) - A_\mu \frac{\partial g^{\mu\nu}}{\partial x_\nu} - \frac{1}{2} g^{\tau a} \left( \frac{\partial g_{\mu a}}{\partial x_\nu} + \frac{\partial g_{\nu a}}{\partial x_\mu} - \frac{\partial g_{\mu\nu}}{\partial x_a} \right) g^{\mu\nu} A_\tau. \quad [18]$$

- [19] In accordance with (31) and (29), the last term of this expression may be written

$$[20] \quad \frac{1}{2} \frac{\partial g^{\tau\nu}}{\partial x_\nu} A_\tau + \frac{1}{2} \frac{\partial g^{\tau\mu}}{\partial x_\mu} A_\tau + \frac{1}{\sqrt{-g}} \frac{\partial \sqrt{-g}}{\partial x_\alpha} g^{\mu\nu} A_\tau.$$

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  $A^\nu$  like  $A_\mu$  is an arbitrary vector, we finally obtain

$$\Phi = \frac{1}{\sqrt{-g}} \frac{\partial}{\partial x_\nu} (\sqrt{-g} A^\nu). \quad . \quad . \quad . \quad (35)$$

This scalar is the *divergence* of the contravariant vector  $A^\nu$ .

*The "Curl" of a Covariant Vector.*—The second term in (26) is symmetrical in the indices  $\mu$  and  $\nu$ . 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  $A_\sigma^{\alpha\beta}$  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

$$A_{\sigma}^{\alpha\beta} = \frac{\partial A^{\alpha\beta}}{\partial x_{\sigma}} + \{\sigma\gamma, \alpha\}A^{\gamma\beta} + \{\sigma\gamma, \beta\}A^{\alpha\gamma}. \quad . \quad (38)$$

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 :—

$$A_{\mu\sigma}^{\alpha} = \frac{\partial A_{\mu}^{\alpha}}{\partial x_{\sigma}} - \{\sigma\mu, \tau\}A_{\tau}^{\alpha} + \{\sigma\tau, \alpha\}A_{\mu}^{\tau}. \quad . \quad (39)$$

On contracting (38) with respect to the indices  $\beta$  and  $\sigma$  (inner multiplication by  $\delta_{\beta}^{\sigma}$ ), 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  $\beta$  and  $\gamma$ , 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

$$A^{\alpha} = \frac{1}{\sqrt{-g}} \frac{\partial(\sqrt{-g}A^{\alpha\beta})}{\partial x_{\beta}}. \quad . \quad . \quad . \quad (40)$$

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  $\alpha$  and  $\sigma$ , and taking (29a) into consideration, we obtain

$$\sqrt{-g}A_{\mu}^{\alpha} = \frac{\partial(\sqrt{-g}A_{\mu}^{\alpha})}{\partial x_{\sigma}} - \{\sigma\mu, \tau\}\sqrt{-g}A_{\tau}^{\alpha}. \quad . \quad (41)$$

If we introduce the contravariant tensor  $A^{\rho\sigma} = g^{\rho\tau}A_{\tau}^{\sigma}$  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

$$-\frac{1}{2}\sqrt{-g}\frac{\partial g_{\rho\sigma}}{\partial x_\mu}A^{\rho\sigma}.$$

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

$$\frac{1}{2}\sqrt{-g}\frac{\partial g^{\rho\sigma}}{\partial x_\mu}A_{\rho\sigma}.$$

In the case of symmetry in question, (41) may therefore be replaced by the two forms

$$\sqrt{-g}A_\mu = \frac{\partial(\sqrt{-g}A_\mu^\sigma)}{\partial x_\sigma} - \frac{1}{2}\frac{\partial g_{\rho\sigma}}{\partial x_\mu}\sqrt{-g}A^{\rho\sigma}. \quad (41a)$$

$$\sqrt{-g}A_\mu = \frac{\partial(\sqrt{-g}A_\mu^\sigma)}{\partial x_\sigma} + \frac{1}{2}\frac{\partial g^{\rho\sigma}}{\partial x_\mu}\sqrt{-g}A_{\rho\sigma}. \quad (41b)$$

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  $A_\mu$ . Then (with a somewhat different naming of the indices) we get the tensor of the third rank

$$A_{\mu\sigma\tau} = \frac{\partial^2 A_\mu}{\partial x_\sigma \partial x_\tau} - \{\mu\sigma, \rho\}\frac{\partial A_\rho}{\partial x_\tau} - \{\mu\tau, \rho\}\frac{\partial A_\rho}{\partial x_\sigma} - \{\sigma\tau, \rho\}\frac{\partial A_\rho}{\partial x_\mu} \\ + \left[ -\frac{\partial}{\partial x_\tau}\{\mu\sigma, \rho\} + \{\mu\tau, a\}\{\sigma a, \rho\} + \{\sigma\tau, a\}\{\sigma a, \rho\} \right] A_\rho.$$

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  $\sigma$  and  $\tau$ . The same holds good for the sum of the second and third terms. Thus we obtain

$$A_{\mu\sigma\tau} - A_{\mu\tau\sigma} = B_{\mu\sigma\tau}^\rho A_\rho \quad . \quad . \quad . \quad (42)$$

where

$$\begin{aligned} B_{\mu\sigma\tau}^\rho = & - \frac{\partial}{\partial x_\tau} \{ \mu\sigma, \rho \} + \frac{\partial}{\partial x_\sigma} \{ \mu\tau, \rho \} - \{ \mu\sigma, \alpha \} \{ \alpha\tau, \rho \} \\ & + \{ \mu\tau, \alpha \} \{ \alpha\sigma, \rho \} \quad (43) \end{aligned}$$

The essential feature of the result is that on the right side of (42) the  $A_\rho$  occur alone, without their derivatives. From the tensor character of  $A_{\mu\sigma\tau} - A_{\mu\tau\sigma}$  in conjunction with the fact that  $A_\rho$  is an arbitrary vector, it follows, by reason of § 7, that  $B_{\mu\sigma\tau}^\rho$  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  $B_{\mu\sigma\tau}^\rho$  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  $B_{\mu\sigma\tau}^\rho$  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  $\tau$  and  $\rho$  we obtain the covariant tensor of second rank

[21]

\* The mathematicians have proved that this is also a *sufficient* condition.

$$\left. \begin{aligned} G_{\mu\nu} &= B_{\mu\nu\rho}^\rho = R_{\mu\nu} + S_{\mu\nu} \\ \text{where} \\ R_{\mu\nu} &= - \frac{\partial}{\partial x_a} \{ \mu\nu, a \} + \{ \mu a, \beta \} \{ \nu\beta, a \} \\ S_{\mu\nu} &= \frac{\partial^2 \log \sqrt{-g}}{\partial x_\mu \partial x_\nu} - \{ \mu\nu, a \} \frac{\partial \log \sqrt{-g}}{\partial x_a} \end{aligned} \right\} \quad (44)$$

*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 diffi-

culty 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

$$\Gamma_{\mu\nu}^{\tau} = - \{ \mu\nu, \tau \} \quad . \quad . \quad . \quad (45)$$

the equation of the motion of the point with respect to  $K_1$ , becomes

$$\frac{d^2x_{\tau}}{ds^2} = \Gamma_{\mu\nu}^{\tau} \frac{dx_{\mu}}{ds} \frac{dx_{\nu}}{ds} \quad . \quad . \quad . \quad (46)$$

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  $\Gamma_{\mu\nu}^{\tau}$  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  $E_{\mu\sigma\tau}^{\rho} = 0$  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  $B_{\mu\nu\tau}^{\rho}$ , 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  $B_{\mu\nu\tau}^{\rho}$  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  $B_{\mu\nu\tau}^{\rho}$ , 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  $B_{\mu\nu\tau}^{\rho}$ . 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

$$\left. \begin{aligned} \frac{\partial \Gamma_{\mu\nu}^a}{\partial x_a} + \Gamma_{\mu\beta}^a \Gamma_{\nu a}^{\beta} &= 0 \\ \sqrt{-g} &= 1 \end{aligned} \right\} \quad . . . . \quad (47)$$

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  $\lambda$  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 :—

$$\left. \begin{array}{l} \delta \int H d\tau = 0 \\ H = g^{\mu\nu} \Gamma_{\mu\beta}^a \Gamma_{\nu a}^\beta \\ \sqrt{-g} = 1 \end{array} \right\} \quad . \quad . \quad . \quad (47a)$$

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

$$\begin{aligned} \delta H &= \Gamma_{\mu\beta}^a \Gamma_{\nu a}^\beta \delta g^{\mu\nu} + 2g^{\mu\nu} \Gamma_{\mu\beta}^a \delta \Gamma_{\nu a}^\beta \\ &= -\Gamma_{\mu\beta}^a \Gamma_{\nu a}^\beta \delta g^{\mu\nu} + 2\Gamma_{\mu\beta}^a \delta(g^{\mu\nu} \Gamma_{\nu a}^\beta). \end{aligned}$$

But

$$\delta(g^{\mu\nu} \Gamma_{\nu a}^\beta) = -\frac{1}{2}\delta \left[ g^{\mu\nu} g^{\beta\lambda} \left( \frac{\partial g_{\nu\lambda}}{\partial x_a} + \frac{\partial g_{a\lambda}}{\partial x_\nu} - \frac{\partial g_{a\nu}}{\partial x_\lambda} \right) \right]. \quad [22]$$

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  $\mu$  and  $\beta$ . They cancel each other in the expression for  $\delta H$ , because they are multiplied by the

quantity  $\Gamma_{\mu\beta}^{\alpha}$ , which is symmetrical with respect to the indices  $\mu$  and  $\beta$ . Thus there remains only the first term in round brackets to be considered, so that, taking (31) into account, we obtain

$$\delta H = - \Gamma_{\mu\beta}^{\alpha} \Gamma_{\nu a}^{\beta} \delta g^{\mu\nu} + \Gamma_{\mu\beta}^{\alpha} \delta g_a^{\mu\beta}.$$

Thus

$$\left. \begin{aligned} \frac{\partial H}{\partial g^{\mu\nu}} &= - \Gamma_{\mu\beta}^{\alpha} \Gamma_{\nu a}^{\beta} \\ \frac{\partial H}{\partial g_a^{\mu\nu}} &= \Gamma_{\mu\nu}^{\sigma} \end{aligned} \right\} . . . . . \quad (48)$$

Carrying out the variation in (47a), we get in the first place

$$\frac{\partial}{\partial x_a} \left( \frac{\partial H}{\partial g_a^{\mu\nu}} \right) - \frac{\partial H}{\partial g^{\mu\nu}} = 0, \quad . . . . . \quad (47b)$$

which, on account of (48), agrees with (47), as was to be proved.

If we multiply (47b) by  $g_a^{\mu\nu}$ , then because

$$\frac{\partial g_a^{\mu\nu}}{\partial x_a} = \frac{\partial g_a^{\mu\nu}}{\partial x_\sigma}$$

and, consequently,

$$g_\sigma^{\mu\nu} \frac{\partial}{\partial x_a} \left( \frac{\partial H}{\partial g_a^{\mu\nu}} \right) = \frac{\partial}{\partial x_a} \left( g_\sigma^{\mu\nu} \frac{\partial H}{\partial g_a^{\mu\nu}} \right) - \frac{\partial H}{\partial g_a^{\mu\nu}} \frac{\partial g_a^{\mu\nu}}{\partial x_\sigma},$$

we obtain the equation

$$\frac{\partial}{\partial x_a} \left( g_\sigma^{\mu\nu} \frac{\partial H}{\partial g_a^{\mu\nu}} \right) - \frac{\partial H}{\partial x_\sigma} = 0$$

or \*

$$\left. \begin{aligned} \frac{\partial t_\sigma^\alpha}{\partial x_a} &= 0 \\ - 2\kappa t_\sigma^\alpha &= g_\sigma^{\mu\nu} \frac{\partial H}{\partial g_a^{\mu\nu}} - \delta_\sigma^\alpha H \end{aligned} \right\} . . . . . \quad (49)$$

where, on account of (48), the second equation of (47), and (34)

$$\kappa t_\sigma^\alpha = \frac{1}{2} \delta_\sigma^\alpha g^{\mu\nu} \Gamma_{\mu\beta}^\lambda \Gamma_{\nu\lambda}^\beta - g^{\mu\nu} \Gamma_{\mu\beta}^\alpha \Gamma_{\nu\sigma}^\beta \quad . . . . . \quad (50)$$

\* The reason for the introduction of the factor  $-2\kappa$  will be apparent later.

It is to be noticed that  $t_\sigma^a$  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

$$\frac{d}{dx_4} \int t_\sigma^a dV = \int (lt_\sigma^1 + mt_\sigma^2 + nt_\sigma^3) dS. \quad . \quad (49a)$$

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  $t_\sigma^a$  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

$$g^{\nu\sigma} \frac{\partial \Gamma_{\mu\nu}^a}{\partial x_a} = \frac{\partial}{\partial x_a} (g^{\nu\sigma} \Gamma_{\mu\nu}^a) - \frac{\partial g^{\nu\sigma}}{\partial x_a} \Gamma_{\mu\nu}^a,$$

which quantity, by reason of (34), is equal to

$$\frac{\partial}{\partial x_a} (g^{\nu\sigma} \Gamma_{\mu\nu}^a) = g^{\nu\beta} \Gamma_{a\beta}^\sigma \Gamma_{\mu\nu}^a - g^{\sigma\beta} \Gamma_{\beta a}^\nu \Gamma_{\mu\nu}^a,$$

or (with different symbols for the summation indices)

$$\frac{\partial}{\partial x_a} (g^{\sigma\beta} \Gamma_{\mu\beta}^a) = g^{\nu\delta} \Gamma_{\gamma\beta}^\sigma \Gamma_{\delta\mu}^a - g^{\nu\sigma} \Gamma_{\mu\beta}^a \Gamma_{\nu\alpha}^\beta.$$

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

$$\kappa(t_\mu^\sigma - \frac{1}{2}\delta_\mu^\sigma t),$$

where  $t = t_\alpha^a$ . Thus instead of equations (47) we obtain

$$\left. \begin{aligned} \frac{\partial}{\partial x_a} (g^{\sigma\beta} \Gamma_{\mu\beta}^a) &= -\kappa(t_\mu^\sigma - \frac{1}{2}\delta_\mu^\sigma t) \\ \sqrt{-g} &= 1 \end{aligned} \right\} \quad . \quad . \quad . \quad (51)$$

### § 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  $\rho$  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  $T_\sigma^\alpha$ , which, like the energy-components  $t_\sigma$  [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  $\rho$  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  $t_\mu^\sigma + T_\mu^\sigma$  of the energy-components of matter and of gravitational field. Thus instead of (51) we obtain the tensor equation

$$\frac{\partial}{\partial x_\alpha} (g^{\sigma\beta} T_{\mu\beta}^\alpha) = -\kappa [(t_\mu^\sigma + T_\mu^\sigma) - \frac{1}{2}\delta_\mu^\sigma(t + T)], \quad \left. \begin{array}{l} \\ \sqrt{-g} = 1 \end{array} \right\}. \quad (52)$$

where we have set  $T = T_\mu^\mu$  (Laue's scalar). These are the

\*  $g_{\alpha\tau} T_\sigma^\alpha = T_{\sigma\tau}$  and  $g^{\sigma\beta} T_\sigma^\alpha = T^{\alpha\beta}$  are to be symmetrical tensors.

required general field equations of gravitation in mixed form. Working back from these, we have in place of (47)

$$\left. \begin{aligned} \frac{\partial}{\partial x_\alpha} \Gamma_{\mu\nu}^\alpha + \Gamma_{\mu\beta}^\alpha \Gamma_{\nu\alpha}^\beta &= -\kappa(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T), \\ \sqrt{-g} &= 1 \end{aligned} \right\} . \quad (53)$$

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  $\mu$  and  $\sigma$ , and after multiplying the resulting equation by  $\frac{1}{2}\delta_\mu^\sigma$ , subtract it from equation (52). This gives

$$\frac{\partial}{\partial x_\alpha} (g^{\sigma\beta} T_{\mu\beta}^\alpha - \frac{1}{2}\delta_\mu^\sigma g^{\lambda\beta} T_{\lambda\beta}^\alpha) = -\kappa(t_\mu^\sigma + T_\mu^\sigma). \quad (52a)$$

On this equation we perform the operation  $\partial/\partial x_\sigma$ . We have

$$\frac{\partial^2}{\partial x_\alpha \partial x_\sigma} \left( g^{\sigma\beta} T_{\beta\mu}^\alpha \right) = -\frac{1}{2} \frac{\partial^2}{\partial x_\alpha \partial x_\sigma} \left[ g^{\sigma\beta} g^{\alpha\lambda} \left( \frac{\partial g_{\mu\lambda}}{\partial x_\beta} + \frac{\partial g_{\beta\lambda}}{\partial x_\mu} - \frac{\partial g_{\mu\beta}}{\partial x_\lambda} \right) \right].$$

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  $\alpha$  and  $\sigma$  on the one hand, and  $\beta$  and  $\lambda$  on the other. The second term may be re-modelled by (31), so that we have

$$\frac{\partial^2}{\partial x_\alpha \partial x_\sigma} \left( g^{\sigma\beta} T_{\beta\mu}^\alpha \right) = \frac{1}{2} \frac{\partial^3 g^{\alpha\beta}}{\partial x_\alpha \partial x_\beta \partial x_\mu} . \quad (54)$$

The second term on the left-hand side of (52a) yields in the

first place

$$-\frac{1}{2} \frac{\partial^2}{\partial x_\alpha \partial x_\mu} (g^{\lambda\beta} T_{\lambda\beta}^\alpha)$$

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

$$\frac{\partial^2}{\partial x_\alpha \partial x_\sigma} \left( g^{\rho\beta} T_{\mu\beta} - \frac{1}{2} \delta_\mu^\sigma g^{\lambda\beta} T_{\lambda\beta}^\alpha \right) \equiv 0 \quad . \quad . \quad . \quad (55)$$

From (55) and (52a), it follows that

$$\frac{\partial(t_\mu^\sigma + T_\mu^\sigma)}{\partial x_\sigma} = 0. \quad . \quad . \quad . \quad (56)$$

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  $t^\sigma$  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

$$\frac{\partial t_\sigma^\alpha}{\partial x_\alpha} + \frac{1}{2} \frac{\partial g^{\mu\nu}}{\partial x_\sigma} T_{\mu\nu} = 0,$$

or, in view of (56),

$$\frac{\partial T_{\sigma}^a}{\partial x_a} + \frac{1}{2} \frac{\partial g^{\mu\nu}}{\partial x_{\sigma}} T_{\mu\nu} = 0 \quad . \quad . \quad . \quad (57)$$

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

$$\frac{\partial T_{\sigma}^a}{\partial x_a} = - \Gamma_{a\sigma}^{\beta} T_{\beta}^a \quad . \quad . \quad . \quad (57a)$$

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 Göttingen, 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  $\rho$  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

$$T^{\alpha\beta} = -g^{\alpha\beta}p + \rho \frac{dx_\alpha}{ds} \frac{dx_\beta}{ds} . . . . . \quad (58)$$

be the contravariant energy-tensor of the fluid. To it belongs the covariant tensor

$$T_{\mu\nu} = -g_{\mu\nu}p + g_{\mu a}g_{\mu\beta} \frac{dx_a}{ds} \frac{dx_\beta}{ds} \rho, . . . . \quad (58a)$$

as well as the mixed tensor \*

$$T_\sigma^\alpha = -\delta_\sigma^\alpha p + g_{\sigma\beta} \frac{dx_\beta}{ds} \frac{dx_\alpha}{ds} \rho . . . . \quad (58b)$$

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  $\rho$ . Thus  $\rho$  is not constant for an incompressible fluid.

with the given equation between  $p$  and  $\rho$ , 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

[28]

Let  $\phi_\nu$  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

$$F_{\rho\sigma} = \frac{\partial \phi_\rho}{\partial x_\sigma} - \frac{\partial \phi_\sigma}{\partial x_\rho} . . . . . \quad (59)$$

It follows from (59) that the system of equations

$$\frac{\partial F_{\rho\sigma}}{\partial x_\tau} + \frac{\partial F_{\sigma\tau}}{\partial x_\rho} + \frac{\partial F_{\tau\rho}}{\partial x_\sigma} = 0 . . . . . \quad (60)$$

[29]

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.

$$\left. \begin{aligned} \frac{\partial F_{23}}{\partial x_4} + \frac{\partial F_{34}}{\partial x_2} + \frac{\partial F_{42}}{\partial x_3} &= 0 \\ \frac{\partial F_{34}}{\partial x_1} + \frac{\partial F_{41}}{\partial x_3} + \frac{\partial F_{13}}{\partial x_4} &= 0 \\ \frac{\partial F_{41}}{\partial x_2} + \frac{\partial F_{12}}{\partial x_4} + \frac{\partial F_{24}}{\partial x_1} &= 0 \\ \frac{\partial F_{12}}{\partial x_3} + \frac{\partial F_{23}}{\partial x_1} + \frac{\partial F_{31}}{\partial x_2} &= 0 \end{aligned} \right\} \quad . . . . \quad (60a)$$

This system corresponds to the second of Maxwell's systems of equations. We recognize this at once by setting

$$\left. \begin{aligned} F_{23} &= H_x, \quad F_{14} = E_x \\ F_{31} &= H_y, \quad F_{24} = E_y \\ F_{12} &= H_z, \quad F_{34} = E_z \end{aligned} \right\} \quad . . . . \quad (61)$$

Then in place of (60a) we may set, in the usual notation of three-dimensional vector analysis,

$$\left. \begin{aligned} - \frac{\partial H}{\partial t} &= \text{curl } E \\ \text{div } H &= 0 \end{aligned} \right\} \quad . . . . \quad (60b)$$

We obtain Maxwell's first system by generalizing the form given by Minkowski. We introduce the contravariant six-vector associated with  $F^{\alpha\beta}$

$$[30] \quad F^{\mu\nu} = g^{\mu\alpha} g^{\nu\beta} F_{\alpha\beta} \quad . . . . \quad (62)$$

and also the contravariant vector  $J^\mu$  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 coordinates) :—

$$\frac{\partial}{\partial x_\nu} F^{\mu\nu} = J^\mu \quad . . . . \quad (63)$$

Let

$$\left. \begin{aligned} F^{23} &= H'_x, \quad F^{14} = -E'_x \\ F^{31} &= H'_y, \quad F^{24} = -E'_y \\ F^{12} &= H'_z, \quad F^{34} = -E'_z \end{aligned} \right\} \quad . . . . \quad (64)$$

which quantities are equal to the quantities  $H_x \dots E_z$  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)

$$\left. \begin{aligned} \frac{\partial E'}{\partial t} + j &= \text{curl } H' \\ \text{div } E' &= \rho \end{aligned} \right\} \quad . \quad . \quad . \quad . \quad (63a)$$

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

$$\kappa_\sigma = F_{\sigma\mu} J^\mu \quad . \quad . \quad . \quad . \quad . \quad (65)$$

By (61) its components, written in the three-dimensional manner, are

$$\left. \begin{aligned} \kappa_1 &= \rho E_x + [j \cdot H]^x \\ \vdots &\quad \vdots \quad \vdots \quad \vdots \\ \kappa_4 &= - (j \cdot E) \end{aligned} \right\} \quad . \quad . \quad . \quad . \quad . \quad (65a)$$

$\kappa_\sigma$  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  $\kappa_\sigma$  will vanish.

To obtain the energy-components  $T_\sigma^\nu$  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

$$\kappa_\sigma = F_{\sigma\mu} \frac{\partial F^{\mu\nu}}{\partial x_\nu} = \frac{\partial}{\partial x_\nu} (F_{\sigma\mu} F^{\mu\nu}) - F^{\mu\rho} \frac{\partial F_{\sigma\mu}}{\partial x_\nu}.$$

The second term of the right-hand side, by reason of (60), permits the transformation

$$F^{\mu\nu} \frac{\partial F_{\sigma\mu}}{\partial x_\nu} = - \frac{1}{2} F^{\mu\nu} \frac{\partial F_{\mu\nu}}{\partial x_\sigma} = - \frac{1}{2} g^{\mu\alpha} g^{\nu\beta} F_{\alpha\beta} \frac{\partial F_{\mu\nu}}{\partial x_\sigma},$$

which latter expression may, for reasons of symmetry, also be written in the form

$$-\frac{1}{4} \left[ g^{\mu\alpha} g^{\nu\beta} F_{\alpha\beta} \frac{\partial F_{\mu\nu}}{\partial x_\sigma} + g^{\mu\alpha} g^{\nu\beta} \frac{\partial F_{\alpha\beta}}{\partial x_\sigma} F_{\mu\nu} \right].$$

But for this we may set

$$-\frac{1}{4} \frac{\partial}{\partial x_\sigma} (g^{\mu\alpha} g^{\nu\beta} F_{\alpha\beta} F_{\mu\nu}) + \frac{1}{4} F_{\alpha\beta} F_{\mu\nu} \frac{\partial}{\partial x_\sigma} (g^{\mu\alpha} g^{\nu\beta}).$$

The first of these terms is written more briefly

$$-\frac{1}{4} \frac{\partial}{\partial x_\sigma} (F^{\mu\nu} F_{\mu\nu});$$

the second, after the differentiation is carried out, and after some reduction, results in

$$-\frac{1}{2} F^{\mu\tau} F_{\mu\nu} g^{\nu\rho} \frac{\partial g_{\sigma\tau}}{\partial x_\sigma}.$$

[31]

Taking all three terms together we obtain the relation

$$\kappa_\sigma = \frac{\partial T_\sigma^\nu}{\partial x_\nu} - \frac{1}{2} g^{\tau\mu} \frac{\partial g_{\mu\nu}}{\partial x_\sigma} T_\tau^\nu . . . . . \quad (66)$$

where

$$T_\sigma^\nu = -F_{\sigma\alpha} F^{\nu\alpha} + \frac{1}{4} \delta_\sigma^\nu F_{\alpha\beta} F^{\alpha\beta}.$$

Equation (66), if  $\kappa_\sigma$  vanishes, is, on account of (30), equivalent to (57) or (57a) respectively. Therefore the  $T_\sigma^\nu$  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

[32]

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  $\Gamma_{\mu\nu}^{\tau}$  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

$$\frac{d^2x_{\tau}}{dt^2} = \Gamma_{44}^{\tau}$$

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

$$\frac{d^2x_\tau}{dt^2} = -\frac{1}{2}\frac{\partial g_{44}}{\partial x_\tau} \quad (\tau = 1, 2, 3) \quad . \quad . \quad . \quad (67)$$

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,

$$-\frac{1}{2}\left(\frac{\partial^2 g_{44}}{\partial x_1^2} + \frac{\partial^2 g_{44}}{\partial x_2^2} + \frac{\partial^2 g_{44}}{\partial x_3^2}\right) = -\frac{1}{2}\nabla^2 g_{44}.$$

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  $6.7 \times 10^{-8}$ , usually called the constant of gravitation. By comparison we obtain

$$\kappa = \frac{8\pi K}{c^2} = 1.87 \times 10^{-27} \quad . \quad . \quad (69)$$

### § 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

$$\left. \begin{aligned} g_{\rho\sigma} &= -\delta_{\rho\sigma} - \alpha \frac{x_\rho x_\sigma}{r^3} (\rho, \sigma = 1, 2, 3) \\ g_{\rho 4} &= g_{4\rho} = 0 \quad (\rho = 1, 2, 3) \\ g_{44} &= 1 - \frac{\alpha}{r} \end{aligned} \right\} \quad . \quad (70)$$

[33] 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)

$$\alpha = \frac{\kappa M}{4\pi}, \quad . \quad . \quad . \quad . \quad . \quad (70a)$$

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  $dx_\nu$  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,

$$dx = 1 - \frac{a}{2r} \quad . \quad . \quad . \quad . \quad . \quad (71)$$

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

$$ds^2 = -1; \quad dx_1 = dx_3 = dx_4 = 0; \quad x_1 = r, \quad x_2 = x_3 = 0.$$

The result is

$$-1 = g_{22}dx_2^2 = -dx_2^2 \quad . \quad . \quad . \quad . \quad (71a)$$

With the tangential position, therefore, the gravitational field of the point of mass 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

$$1 = g_{44}dx_4^2;$$

$$dx_4 = \frac{1}{\sqrt{g_{44}}} = \frac{1}{\sqrt{(1 + (g_{44} - 1))}} = 1 - \frac{1}{2}(g_{44} - 1)$$

or

$$dx_4 = 1 + \frac{\kappa}{8\pi} \int \rho \frac{d\tau}{r} \quad . . . . \quad (72)$$

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

$$- dx_1^2 - dx_2^2 - dx_3^2 + dx_4^2 = 0$$

and therefore by the general theory of relativity by the equation

$$ds^2 = g_{\mu\nu} dx_\mu dx_\nu = 0 \quad . . . . \quad (73)$$

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

$$\sqrt{\left(\frac{dx_1}{dx_4}\right)^2 + \left(\frac{dx_2}{dx_4}\right)^2 + \left(\frac{dx_3}{dx_4}\right)^2} = \gamma$$

[35] 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  $(\gamma, n)$ , has the curvature  $- dy/dn$ .

We examine the curvature undergone by a ray of light passing by a mass  $M$  at the distance  $\Delta$ . If we choose the system of co-ordinates in agreement with the accompanying diagram, the total bending of the ray (calculated positively if

[34] \* 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

$$B = \frac{2a}{\Delta} = \frac{\kappa M}{2\pi\Delta}. \quad . . . \quad (74)$$

[36]

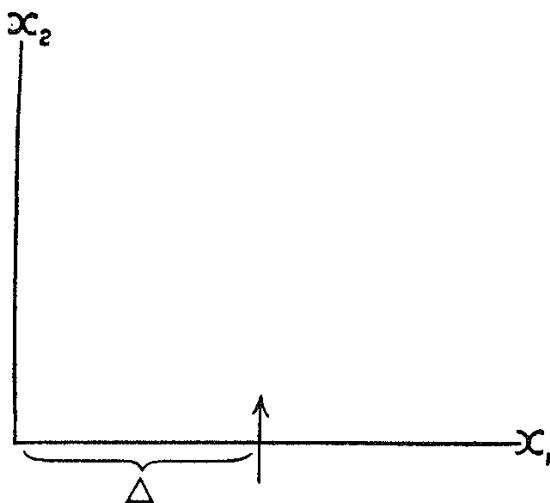


FIG. 8.

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

$$\epsilon = 24\pi^3 \frac{a^3}{T^2 c^2 (1 - e^2)} \quad . . . \quad (75)$$

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.

[37]

\* 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.

---



# Invariant Variation Problems

Emmy Noether

M. A. Tavel's English translation of "Invariante Variationsprobleme," *Nachr. d. König. Gesellsch. d. Wiss. zu Göttingen, Math.-phys. Klasse*, 235–257 (1918), which originally appeared in *Transport Theory and Statistical Physics*, **1** (3), 183–207 (1971).<sup>0</sup>

## Abstract

The problems in variation here concerned are such as to admit a continuous group (in Lie's sense); the conclusions that emerge from the corresponding differential equations find their most general expression in the theorems formulated in Section 1 and proved in following sections. Concerning these differential equations that arise from problems of variation, far more precise statements can be made than about arbitrary differential equations admitting of a group, which are the subject of Lie's researches. What is to follow, therefore, represents a combination of the methods of the formal calculus of variations with those of Lie's group theory. For special groups and problems in variation, this combination of methods is not new; I may cite Hamel and Herglotz for special finite groups, Lorentz and his pupils (for instance Fokker), Weyl and Klein for special infinite groups.<sup>1</sup> Especially Klein's second Note and the present developments have been mutually influenced by each other, in which regard I may refer to the concluding remarks of Klein's Note.

## § 1. Preliminary Remarks and Formulation of Theorems

All functions occurring in the sequel are to be assumed analytic, or at least continuous and continuously differentiable a definite number of times, and unique in the interval considered.

By a "group of transformation," familiarly, is meant a system of transformations such that for each transformation, there exists an inverse contained in the system, and such that the composition of any two transformations of the system in turn belongs to the system. The group will be called a finite continuous group  $\mathfrak{G}_\rho$  if its transformations are contained in a most general (transformation) depending analytically on  $\rho$  essential parameters  $\epsilon$  (i.e., the  $\rho$  parameters are not to be representable as  $\rho$  function of fewer parameters). Correspondingly, an infinite continuous group  $\mathfrak{G}_{\infty\rho}$  is understood to be a group whose most general transformations depend on  $\rho$  essential arbitrary functions  $p(x)$  and their derivatives analytically, or at least in a continuous and finite-fold continuously differentiable manner. The group depending on infinitely many parameters but not on arbitrary functions stands as an intermediate term between the two. Finally, a group depending both on arbitrary functions and on parameters is called a mixed group.<sup>2</sup>

Let  $x_1, \dots, x_n$  be independent variables and  $u_1(x), \dots, u_\mu(x)$  functions depending upon them. If the  $x$ 's and  $u$ 's are subjected to the transformations of a group, then, by hypothesis of invertibility

---

<sup>0</sup>This paper is reproduced by Frank Y. Wang ([fwang@lagcc.cuny.edu](mailto:fwang@lagcc.cuny.edu)) with LATEX.

<sup>1</sup>Hamel, Math. Ann. **59** and Z. f. Math. u. Phys. **50**. Herglotz, Ann. d. Phys. (4) **36**, esp. § 9, p. 511. Fokker, Verslag d. Amsterdamer Akad. Jan. 27, 1917. For further bibliography, compare Klein's second Note, Göttinger Nachrichten, July 19, 1918. The recently published work by Kneser (Math. Zschr. 2) deals with the setting up of invariants by a similar method.

<sup>2</sup>Lie, in "Grundlagen für die Theorie der Unendlichen kontinuierlichen Transformationsgruppen" (Foundations of the theory of infinite continuous groups of transformations), Ber. d. K. Sachs. Ges. d. Wissenschaft 1981 (cited as Grundlagen), defines the infinite continuous group as a group of transformations which are given by the most general solutions of a system of partial differential equations, provided these solutions do not depend only on a finite number of parameters. One of the above-mentioned types differing from the finite group will thus be thereby obtained; whereas conversely the limiting case of infinitely many parameters need not necessarily satisfy a system of differential equations.

of the transformations, there must again be exactly  $n$  independent variables  $y_1, \dots, y_n$  among the transformed quantities; let the other depending upon them be designated by  $v_1(y), \dots, v_\mu(y)$ . In the transformations, the derivatives of the  $u$ 's with respect to the  $x$ 's, namely  $\frac{\partial u}{\partial x}, \frac{\partial^2 u}{\partial x^2}, \dots$  may also occur.<sup>3</sup> A function is called an invariant of the group if there subsists a relationship

$$P\left(x, u, \frac{\partial u}{\partial x}, \frac{\partial^2 u}{\partial x^2}, \dots\right) = P\left(y, v, \frac{\partial v}{\partial y}, \frac{\partial^2 v}{\partial y^2}, \dots\right).$$

In particular, then, an integral  $I$  will be an invariant of the group if there subsists a relationship

$$I = \int \dots \int f\left(x, u, \frac{\partial u}{\partial x}, \frac{\partial^2 u}{\partial x^2}, \dots\right) dx = \int \dots \int f\left(y, v, \frac{\partial v}{\partial y}, \frac{\partial^2 v}{\partial y^2}, \dots\right) dy \quad ^4 \quad (1)$$

integrated over an *arbitrary* real  $x$ -interval and the corresponding  $y$ -interval.<sup>5</sup>

Then, for an arbitrary, not necessarily invariant integral  $I$ ,  $I$  form the first variation  $\delta I$  and transform it by partial integration according to the rules of the calculus of variations. As we know, provided  $\delta u$  with all derivatives that occur is assumed to vanish at the boundary, while otherwise arbitrary,

$$\delta I = \int \dots \int \delta f dx = \int \dots \int \left( \sum \psi_i \left( x, u, \frac{\partial u}{\partial x}, \dots \right) \delta u_i \right) dx, \quad (2)$$

where  $\psi$  stands for the Lagrange expressions, i.e., the left-hand sides of the Lagrange equations of the corresponding variation problem  $\delta I = 0$ . To this integral relationship there corresponds an integral-free identity in  $\delta u$  and its derivatives, generated by writing in the boundary terms as well. As the partial integration shows, these boundary terms are integrals over divergences, i.e., over expressions

$$\text{Div } A = \frac{\partial A_1}{\partial x_1} + \dots + \frac{\partial A_n}{\partial x_n},$$

where  $A$  is linear in  $\delta u$  and its derivatives. Hence

$$\sum \psi_i \delta u_i = \delta f + \text{Div } A. \quad (3)$$

If in particular  $f$  contains only first derivatives of the  $u$ 's, then in the case of the single integral the identity (3) is identical with what Heun calls the “central equation of Lagrange”

$$\sum \psi_i \delta u_i = \delta f - \frac{d}{dx} \left( \sum \frac{\partial f}{\partial u'_i} \delta u_i \right), \quad \left( u'_i = \frac{du_i}{dx} \right), \quad (4)$$

whereas for the  $n$ -fold integral, (3) goes over into

$$\sum \psi_i \delta u_i = \delta f - \frac{\partial}{\partial x_1} \left( \sum \frac{\partial f}{\partial \frac{\partial u_i}{\partial x_1}} \delta u_i \right) - \dots - \frac{\partial}{\partial x_n} \left( \sum \frac{\partial f}{\partial \frac{\partial u_i}{\partial x_n}} \delta u_i \right). \quad (5)$$

<sup>3</sup>I suppress the subscripts, insofar as feasible, even in summations; thus,  $\frac{\partial^2 u}{\partial x^2}$  for  $\frac{\partial^2 u_\alpha}{\partial x_\beta \partial x_\gamma}$ , etc.

<sup>4</sup>By way of abbreviation, I write  $dx, dy$  for  $dx_1 \dots dx_n, dy_1 \dots dy_n$ .

<sup>5</sup>All arguments  $x, u, \epsilon, p(x)$  occurring in the transformations are to be assumed real, whereas the coefficients may be complex. But since the final results are concerned with *identities* in the  $x$ 's,  $u$ 's, parameters and arbitrary functions, these hold also for complex values, provided only that all functions that occur are assumed analytic. A large portion of the results, incidentally, can be justified without integrals, so that here the restriction to reals is not necessary even to the arguments. On the other hand, the developments at the close of Section 2 and beginning of Section 5 do not appear to be feasible without integrals.

For the single integral and  $\kappa$  derivatives of the  $u$ 's, (3) is given by

$$\begin{aligned} \sum \psi_i \delta u_i &= \delta f - \\ &- \frac{d}{dx} \left\{ \sum \left( \binom{1}{1} \frac{\partial f}{\partial u_i^{(1)}} \delta u_i + \binom{2}{1} \frac{\partial f}{\partial u_i^{(2)}} \delta u_i^{(1)} + \dots + \binom{\kappa}{1} \frac{\partial f}{\partial u_i^{(\kappa)}} \delta u_i^{(\kappa-1)} \right) \right\} + \\ &+ \frac{d^2}{dx^2} \left\{ \sum \left( \binom{2}{2} \frac{\partial f}{\partial u_i^{(2)}} \delta u_i + \binom{3}{2} \frac{\partial f}{\partial u_i^{(3)}} \delta u_i^{(1)} + \dots + \binom{\kappa}{2} \frac{\partial f}{\partial u_i^{(\kappa)}} \delta u_i^{(\kappa-2)} \right) \right\} + \dots \\ &+ (-1)^\kappa \frac{d^\kappa}{dx^\kappa} \left\{ \sum \binom{\kappa}{\kappa} \frac{\partial f}{\partial u_i^{(\kappa)}} \delta u_i \right\} \quad (6) \end{aligned}$$

and a corresponding identity holds for the  $n$ -fold integral; in particular,  $A$  contains  $\delta u$  as far as the  $(\kappa - 1)$ st derivative. The fact that (4), (5), and (6) actually define the Lagrange expressing  $\psi_i$  follows from the fact that the combinations of the right-hand sides eliminate all higher derivatives of the  $\delta u$ 's, while on the other hand the relation (2), to which the partial integration leads *uniquely*, is satisfied.

Now in the following we shall be concerned with these two theorems:

- I. If the integral  $I$  is invariant with respect to a  $\mathfrak{G}_\rho$ , then  $\rho$  linearly independent combinations of the Lagrange expressions become divergences — and from this, conversely, invariance of  $I$  with respect to a  $\mathfrak{G}_\rho$  will follow. The theorem holds good even in the limiting case of infinitely many parameters.
- II. If the integral  $I$  is invariant with respect to a  $\mathfrak{G}_{\infty\rho}$  in which the arbitrary functions occur up to the  $\sigma$ -th derivative, then there subsist  $\rho$  identity relationships between the Lagrange expressions and their derivatives up to the  $\sigma$ -th order. In this case also, the converse holds.<sup>6</sup>

For mixed groups, the statements of both theorems hold; that is, both dependencies and divergence relations independent thereof occur.

Passing over from these identities to the corresponding variation problem, i.e., putting  $\psi = 0$ ,<sup>7</sup> Theorem I in the one-dimensional case — where the divergence goes over into a total differential — asserts the existence of  $\rho$  first integrals, between which, however, non-linear dependencies may subsist;<sup>8</sup> in the multidimensional case, the divergence equations often referred to of late as “laws of conservation” are obtained; Theorem II states that  $\rho$  of the Lagrange equations are a consequence of the rest.

The simplest example of Theorem II — without converse — is afforded by the Weierstrass parametric representation; here the integral, with homogeneity of first order, is as we know invariant if the independent variable  $x$  is replaced by an arbitrary function of  $x$  that leaves  $u$  unchanged ( $y = p(x)$ ;  $v_i(y) = u_i(x)$ ). Thus one arbitrary function occurs, but without derivatives, and to this corresponds the known linear relationship among the Lagrange expressions themselves  $\sum \psi_i \frac{du_i}{dx} = 0$ . Another example is presented by the “general theory of relativity” of the physicists; there we have the group of *all* transformations  $y_i = p_i(x)$  of the  $x$ 's, while the  $u$ 's (designated as  $g_{\mu\nu}$  and  $q$ ) are subjected to the transformations thereby induced for the coefficients of a quadratic and linear differential form — transformations which contain the first derivatives of the arbitrary function  $p(x)$ . To this correspond the familiar  $n$  dependencies between the Lagrange expressions and their first derivatives.<sup>9</sup>

If in particular we specialize the group by allowing no derivatives of the  $u(x)$ 's in the transformations, and moreover let the transformed independent quantities depend only on the  $x$ 's, not on the  $u$ 's, then (as is shown in Section 5) the invariance of  $I$  entails the relative invariance of  $\sum \psi_i \delta u_i$ ,<sup>10</sup>

<sup>6</sup>For certain trivial exceptions, compare Section 2, Note 13.

<sup>7</sup>Somewhat more generally, we may alternatively put  $\psi_i = T_i$ ; cf. Section 3, Note 15.

<sup>8</sup>Cf. close of Section 3.

<sup>9</sup>Cf. e.g., Klein's presentation.

<sup>10</sup>That is,  $\sum \psi_i \delta u_i$  acquires a factor upon transformation.

and likewise of the divergences occurring in Theorem I, once the parameters are subjected to suitable transformations. For Theorem II, similarly, we get relative invariance of the left-hand sides of the dependencies as associated with the aid of the arbitrary functions; and as a consequence of this, another function whose divergence vanishes identically and admits of the group — mediating, in the physicists' theory of relativity, the connection between dependencies and the law of conservation of energy.<sup>11</sup> Theorem II, finally, in terms of group theory, furnishes the proof of a related Hilbertian assertion about the failure of laws of conservation of energy proper in "general relativity." With these supplementary remarks, Theorem I comprises all theorems on first integrals known to mechanics etc., while Theorem II may be described as the utmost possible generalization of the "general theory of relativity" in group theory.

## § 2. Divergence Relationships and Dependencies

Let  $\mathfrak{G}$  be a — finite or infinite — continuous group; then it is always possible to arrange for the zero values of the parameters  $\epsilon$ , or of the arbitrary function  $p(x)$ , to correspond to the identity transformation.<sup>12</sup> The most general transformation will therefore be of the form

$$y_i = A_i \left( x, u, \frac{\partial u}{\partial x}, \dots \right) = x_i + \Delta x_i + \dots$$

$$v_i(y) = B_i \left( x, u, \frac{\partial u}{\partial x}, \dots \right) = u_i + \Delta u_i + \dots$$

where  $\Delta x_i$ ,  $\Delta u_i$  stand for the terms of lowest dimension in  $\epsilon$ , or  $p(x)$  and its derivatives; in which, in fact, they will be assumed linear. As will afterwards, this is no restriction of generality.

Now let the integral  $I$  be an invariant with respect to  $\mathfrak{G}$ , satisfying, that is, the relationship (1). Then in particular,  $I$  will also be invariant with respect to the infinitesimal transformation

$$y_i = x_i + \Delta x_i; \quad v_i(y) = u_i + \Delta u_i;$$

contained in  $\mathfrak{G}$ , and for this relation (1) goes over into

$$0 = \Delta I = \int \dots \int f \left( y, v(y), \frac{\partial v}{\partial y}, \dots \right) dy - \int \dots \int f \left( x, u(x), \frac{\partial u}{\partial x}, \dots \right) dx, \quad (7)$$

where the first integral is to be extended over the  $x + \Delta x$  interval corresponding to the  $x$ -interval. But this integration may alternatively be transformed into an integration over the  $x$ -interval, but virtue of the transformation, valid for infinitesimal  $\Delta x$ ,

$$\int \dots \int f \left( y, v(y), \frac{\partial v}{\partial y}, \dots \right) dy = \int \dots \int f \left( x, u(x), \frac{\partial u}{\partial x}, \dots \right) dx + \int \dots \int \text{Div}(f \cdot \Delta x) dx. \quad (8)$$

So if in place of the infinitesimal transformation  $\Delta u$ , we introduce the variation

$$\bar{\delta} u_i = v_i(x) - u_i(x) = \Delta u_i - \sum \frac{\partial u_i}{\partial x_\lambda} \Delta x_\lambda, \quad (9)$$

then (7) and (8) go over into

$$0 = \int \dots \int \{ \bar{\delta} f + \text{Div}(f \cdot \Delta x) \} dx. \quad (10)$$

The right-hand side is the familiar formula for simultaneous variation of the dependent and independent variables. Since the relation (10) is satisfied for integration over any arbitrary interval, the

---

<sup>11</sup>Compare Klein's second note.

<sup>12</sup>Cf. e.g., Lie, *Grundlagen*, p. 331. Where arbitrary functions are concerned, the special values  $a^\sigma$  of the parameters are to be replaced by fixed functions  $p^\sigma$ ,  $\frac{\partial p^\sigma}{\partial x}$ ,  $\dots$ ; and correspondingly, the values  $a^\sigma + \epsilon$  by  $p + p(x)$ ,  $\frac{\partial p^\sigma}{\partial x} + \frac{\partial p}{\partial x}$ , etc.

integrand must vanish identically; therefore Lie's differential equations for the invariance of  $I$  goes over into the relation

$$\bar{\delta}f + \text{Div}(f \cdot \Delta x) = 0. \quad (11)$$

If in this, by (3),  $\bar{\delta}f$  is expressed in terms of the Lagrange expressions, we get

$$\sum \psi_i \bar{\delta}u_i = \text{Div } B \quad (B = A - f \cdot \Delta x), \quad (12)$$

and this relationship, therefore, for every invariant integral  $I$ , represents an identity in all arguments that occur; it is the required form of Lie's differential equations for  $I$ .<sup>13</sup>

Now for the present let  $\mathfrak{G}$  be taken to be a finite continuous group  $\mathfrak{G}_\rho$ ; since by hypothesis  $\Delta u$  and  $\Delta x$  are linear in the parameters  $\epsilon_1, \dots, \epsilon_\rho$ , hence by (9) the same holds for  $\bar{\delta}u$  and its derivatives; therefore  $A$  and  $B$  are linear in the  $\epsilon$ 's. So if I let

$$B = B^{(1)}\epsilon_1 + \dots + B^{(\rho)}\epsilon_\rho; \quad \bar{\delta}u = \bar{\delta}u^{(1)}\epsilon_1 + \dots + \bar{\delta}u^{(\rho)}\epsilon_\rho,$$

where, that is  $\bar{\delta}u^{(1)}, \dots$  are functions of  $x, u, \frac{\partial u}{\partial x}, \dots$ , the required divergence relationships follow from (12):

$$\sum \psi_i \bar{\delta}u_i^{(1)} = \text{Div } B^{(1)}; \quad \dots \quad \sum \psi_i \bar{\delta}u_i^{(\rho)} = \text{Div } B^{(\rho)}. \quad (13)$$

Thus  $\rho$  linearly independent combinations of the Lagrange expressions become divergences; the linear independence follows from the fact that by (9),  $\bar{\delta}u = 0, \Delta x = 0$  would entail  $\Delta u = 0, \Delta x = 0$ , or in other words a dependency between the infinitesimal transformations. But by hypothesis, none such is satisfied for any value of the parameters, since otherwise the  $\mathfrak{G}_\rho$  regenerated by integration from the infinitesimal transformations would depend on fewer than  $\rho$  essential parameters. But the further possibility  $\bar{\delta}u = 0, \text{Div}(f \cdot \Delta x) = 0$  was excluded. These conclusions hold good even in the limiting case of infinitely many parameters.

Now let  $\mathfrak{G}$  be an infinite continuous group  $\mathfrak{G}_{\infty\rho}$ ; then  $\bar{\delta}u$  and its derivatives, and hence  $B$  also, will again be linear in the arbitrary functions of  $p(x)$  and their derivatives;<sup>14</sup> independently of (12), further, by substitution of the values of  $\bar{\delta}u$ , let

$$\sum \psi_i \bar{\delta}u_i = \sum_{\lambda, i} \psi_i \left\{ a_i^{(\lambda)}(x, u, \dots) p^{(\lambda)}(x) + b_i^{(\lambda)}(x, u, \dots) \frac{\partial p^{(\lambda)}}{\partial x} + \dots + c_i^{(\lambda)}(x, u, \dots) \frac{\partial^\sigma p^{(\lambda)}}{\partial x^\sigma} \right\}.$$

Now, by the identity

$$\varphi(x, u, \dots) \frac{\partial^\tau p(x)}{\partial x^\tau} = (-1)^\tau \cdot \frac{\partial^\tau \varphi}{\partial x^\tau} \cdot p(x) \quad \text{mod Divergences}$$

and analogously to the partial integration formula, the derivatives of  $p$  can be replaced by  $p$  itself and by divergences that will be linear in  $p$  and its derivatives; hence we get

$$\sum \psi_i \bar{\delta}u_i = \sum_\lambda \left\{ (a_i^{(\lambda)} \psi_i) - \frac{\partial}{\partial x} (b_i^{(\lambda)} \psi_i) + \dots + (-1)^\sigma \frac{\partial^\sigma}{\partial x^\sigma} (c_i^{(\lambda)} \psi_i) \right\} p^{(\lambda)} + \text{Div } \Gamma \quad (14)$$

and in conjunction with (12)

$$\sum \left\{ (a_i^{(\lambda)} \psi_i) - \frac{\partial}{\partial x} (b_i^{(\lambda)} \psi_i) + \dots + (-1)^\sigma \frac{\partial^\sigma}{\partial x^\sigma} (c_i^{(\lambda)} \psi_i) \right\} p^{(\lambda)} = \text{Div}(B - \Gamma). \quad (15)$$

I now form the  $n$ -fold integral over (15), extended over any interval; and choose the  $p(x)$ 's such that they, with all derivatives occurring in  $(B - \Gamma)$ , will vanish at the boundary. Since the integral

<sup>13</sup>(12) goes over into  $0 = 0$  for the trivial case — which can occur only if  $\Delta x, \Delta u$  depend also on derivatives of the  $u$ 's — when  $\text{Div}(f \cdot \Delta x) = 0, \bar{\delta}u = 0$ ; thus these infinitesimal transformations are always to be eliminated from the groups, and only the number of remaining parameters, or arbitrary functions, is to be counted in the formulation of the theorems. Whether the remaining infinitesimal transformations still form a group must be left moot.

<sup>14</sup>That it signifies no restriction to assume the  $p$ 's free from  $u, \frac{\partial u}{\partial x}$ , is shown by the converse.

over a divergence reduces to a boundary integral, then, the integral over the left side of (15) will also vanish for  $p(x)$ 's which are arbitrary except that they and sufficiently many of their derivatives vanish at the boundary; and thence follows, by known inferences, the vanishing of the integrand for every  $p(x)$ , or in other words the  $\rho$  relationships:

$$\sum \left\{ (a_i^{(\lambda)} \psi_i) - \frac{\partial}{\partial x} (b_i^{(\lambda)} \psi_i) + \dots + (-1)^\sigma \frac{\partial^\sigma}{\partial x^\sigma} (c_i^{(\lambda)} \psi_i) \right\} = 0 \quad (\lambda = 1, 2, \dots, \rho). \quad (16)$$

These are the required dependencies between the Lagrange expressions and their derivatives for invariance of  $I$  with respect to  $\mathfrak{G}_{\infty\rho}$ ; the linear independence is proved as above, since the converse leads back to (12), and since we can again argue back from the infinitesimal transformations to the finite ones, as will be explained more fully in Section 4. In the case of a  $\mathfrak{G}_{\infty\rho}$ , that is to say, even in the infinitesimal transformations there always occur  $\rho$  arbitrary transformations. Equation (15) and (16) further entail  $\text{Div}(B - \Gamma) = 0$ .

If, as corresponds to a “mixed group,”  $\Delta x$  and  $\Delta u$  are taken linear in the  $\epsilon$ 's and  $p(x)$ 's, then we see, by equating first the  $p(x)$ 's and then the  $\epsilon$ 's to zero, that both divergence relationships (13) and dependencies (16) hold.

### § 3. Converse in Case of Finite Group

To prove the converse, we are first to run through essentially the foregoing arguments in reverse order. From the fact of (13), by multiplication by the  $\epsilon$ 's and adding, the fact of (12) follows; and thence, by virtue of the identity (3), a relationship  $\bar{\delta}f + \text{Div}(A - B) = 0$ . So if we put  $\Delta x = \frac{1}{f}(A - B)$ , we have thereby arrived at (11); whence finally, by integration, there follows (7),  $\Delta I = 0$ , or in other words the invariance of  $I$  with respect to the infinitesimal transformation determined by  $\Delta x$ ,  $\Delta u$ , where  $\Delta u$ 's by virtue of (9) are determined from  $\Delta x$  and  $\bar{\delta}u$ , and  $\Delta x$  and  $\Delta u$  become linear in the parameters. But  $\Delta I = 0$  entails, in known manner, the invariance of  $I$  with respect to the finite transformations generated by integration of the simultaneous system

$$\frac{dx}{dt} = \Delta x_i; \quad \frac{du_i}{dt} = \Delta u_i; \quad (x_i = y, \quad u_i = v_i, \quad \text{for } t = 0). \quad (17)$$

These finite transformations contain  $\rho$  parameters  $a_1 \dots a_\rho$ , namely the combinations  $t\epsilon_1, \dots, t\epsilon_\rho$ . From the assumption that there are  $\rho$  and only  $\rho$  linearly independent divergence relationships (13), it follows further that the finite transformations, once they do not contain the derivatives  $\frac{\partial u}{\partial x}$ , always form a group. For in the contrary case, at least one infinitesimal transformation generated by Lie's bracketing process would fail to be a linear combination of the other  $\rho$ ; and since  $I$  admits of this transformation also, there would be more than  $\rho$  linearly independent divergence relationships; or else that infinitesimal transformation would be of the special form where  $\bar{\delta}u = 0$ ,  $\text{Div}(f \cdot \Delta x) = 0$ , but in that case  $\Delta x$  or  $\Delta u$ , contrary to hypothesis, would depend on derivatives. Whether this case can arise when derivatives occur in  $\Delta x$  or  $\Delta u$  must be left moot; in that case, the  $\Delta x$  determined above must be augmented by all functions  $\Delta x$  for which  $\text{Div}(f \cdot \Delta x) = 0$  to restore the group property, but by agreement the parameters thereby adjoined are not to count. This completes the proof of the converse.

From this conversion, it follows further that  $\Delta x$  and  $\Delta u$  can actually be assumed linear in the parameters. For if  $\Delta u$  and  $\Delta x$  were forms of higher degree in  $\epsilon$ , then by the linear independence of the power products of the  $\epsilon$ 's, quite analogous relations (13) would follow, only in greater number, from which, by the converse, invariance of  $I$  follows with respect to a group whose infinitesimal transformations contain the parameters *linearly*. If this group is to contain exactly  $\rho$  parameters, then linear dependencies must subsist between the divergence relationships originally obtained through the terms of higher degree in  $\epsilon$ .

Let us add the remark that in the case where  $\Delta x$  and  $\Delta u$  also contain derivatives of the  $u$ 's, the finite transformations may depend on infinitely many derivatives of the  $u$ 's; for in that case the

integration of (17), in the determination of  $\frac{d^2x_i}{dt^2}$ ,  $\frac{d^2u_i}{dt^2}$  leads to  $\Delta \left( \frac{\partial u}{\partial x_\kappa} \right) = \frac{\partial \Delta u}{\partial x_\kappa} - \sum_\lambda \frac{\partial u}{\partial x_\lambda} \frac{\partial \Delta x_\lambda}{\partial x_\kappa}$ , so that the number of derivatives in general increases at each step. By way of example, say,

$$f = \frac{1}{2}u'^2; \quad \psi = -u''; \quad \psi \cdot x = \frac{d}{dx}(u - u'x); \quad \bar{\delta}u = x \cdot \epsilon;$$

$$\Delta x = \frac{-2u}{u'^2}\epsilon; \quad \Delta u = \left( x - \frac{2u}{u'} \right) \cdot \epsilon$$

Since the Lagrange expressions of a divergence vanish identically, the converse shows, finally, the following: if  $I$  admits of a  $\mathfrak{G}_\rho$ , then any integral that differs from  $I$  only by a boundary integral, i.e., by an integral over a divergence, likewise admits of a  $\mathfrak{G}_\rho$  having the same  $\bar{\delta}u$ 's whose infinitesimal transformations will in general contain derivatives of the  $u$ 's. Thus for instance, corresponding to the above example,  $f^* = \frac{1}{2} \left\{ u'^2 - \frac{d}{dx} \left( \frac{u^2}{x} \right) \right\}$  admits of the infinitesimal transformation  $\Delta u = x\epsilon$ ,  $\Delta x = 0$ ; while derivatives of the  $u$ 's occur in the infinitesimal transformations corresponding to  $f$ .

Passing over to the variations problem, i.e., putting  $\psi_i = 0$ ,<sup>15</sup> (13) goes over into the equation  $\text{Div } B^{(1)} = 0, \dots, \text{Div } B^{(\rho)} = 0$ , often referred to as “laws of conservation.” In the one-dimensional case, it follows from this that  $B^{(1)} = \text{const.}$ ,  $B^{(\rho)} = \text{const.}$ ; and here the  $B$ 's contain at most  $(2\kappa-1)$ st derivatives of the  $u$ 's (by (6)), provided  $\Delta u$  and  $\Delta x$  contain no higher derivatives than the  $\kappa$ -th one occurring in  $f$ . Since  $2\kappa$ -th derivatives in general occur in  $\psi$ ,<sup>16</sup> therefore, we have the existence of  $\rho$  first integrals. That there may be non-linear dependencies among these is again shown by the above  $f$ . To the linearly independent  $\Delta u = \epsilon_1$ ,  $\Delta x = \epsilon_2$  there correspond the linearly independent relations  $u'' = \frac{d}{dx}u'$ ;  $u'' \cdot u' = \frac{1}{2}\frac{d}{dx}(u')^2$ ; whereas between the first integrals  $u' = \text{const.}$ ,  $u'^2 = \text{const.}$  a non-linear dependency exists. This relates to the elementary case where  $\Delta u$ ,  $\Delta x$  contain no derivatives of the  $u$ 's.<sup>17</sup>

## § 4. Converse in Case of Infinite Group

First let us show that the assumption of linearity of  $\Delta x$  and  $\Delta u$  constitutes no restriction, a conclusion which follows, even without the converse, from the fact that  $\mathfrak{G}_{\infty\rho}$  formally depends on  $\rho$  and only  $\rho$  arbitrary functions. For it turns out that in the non-linear case, upon composition of the transformations, whereby the terms of lowest order are added together, the number of arbitrary functions would increase. In fact, say, let

$$y = A \left( x, u, \frac{\partial u}{\partial x}, \dots; p \right) = x + \sum a(x, u, \dots) p^\nu + b(x, u, \dots) p^{\nu-1} \frac{\partial p}{\partial x}$$

$$+ c p^{\nu-2} \left( \frac{\partial p}{\partial x} \right)^2 + \dots + d \left( \frac{\partial p}{\partial x} \right)^\nu + \dots \quad (p^\nu = (p^{(1)})^{\nu_1} \dots (p^{(\rho)})^{\nu_\rho});$$

and correspondingly  $v = B \left( x, u, \frac{\partial u}{\partial x}, \dots; p \right)$ ; then by composition with  $z = A \left( y, v, \frac{\partial v}{\partial y}, \dots; q \right)$ , for the terms of lowest order, we get

$$z = x + \sum a(p^\nu + q^\nu) + b \left\{ p^{\nu-1} \frac{\partial p}{\partial x} + q^{\nu-1} \frac{\partial q}{\partial x} \right\} + c \left\{ p^{\nu-2} \left( \frac{\partial p}{\partial x} \right)^2 + q^{\nu-2} \left( \frac{\partial q}{\partial x} \right)^2 \right\} + \dots$$

<sup>15</sup> $\psi_i = 0$ , or, somewhat more generally,  $\psi_i = T_i$ , where  $T_i$  are newly adjoined functions, are referred to in physics as “field equations.” In the case  $\psi_i = T_i$ , the identities (13) goes over into equations  $\text{Div } B^{(\lambda)} = \sum T_i \delta u_i^{(\lambda)}$ , likewise known in physics as laws of conservation.

<sup>16</sup>Provided  $f$  is non-linear in the  $\kappa$ -th derivatives.

<sup>17</sup>Otherwise we also have  $u'^\lambda = \text{const.}$  for every  $\lambda$ , corresponding to

$$u'' \cdot (u')^{\lambda-1} = \frac{1}{\lambda} \frac{d}{dx} (u')^\lambda.$$

Here, if any coefficient different from  $a$  and  $b$  is different from zero, in other words, if a term  $p^{\nu-\sigma} \left( \frac{\partial p}{\partial x} \right)^\sigma + q^{\nu-\sigma} \left( \frac{\partial q}{\partial x} \right)^\sigma$  actually occurs for  $\sigma > 1$  it cannot be written as a differential quotient of a single function or power product of one; the number of arbitrary functions, contrary to hypothesis, has thus increased. If all coefficients different from  $a$  and  $b$  vanish, then, according to the values of the exponents  $\nu_1, \dots, \nu_\rho$ , the second term will become the differential quotient of the first (as always, for example, for a  $\mathfrak{G}_{\infty 1}$ ), so that linearity does actually result; or else the number of arbitrary functions must again increase. The infinitesimal transformations, then, owing to the linearity of the  $p(x)$ 's, satisfy a system of linear partial differential equations; and since the group property is satisfied, they constitute an “infinite group of infinitesimal transformations” accord to Lie's definition (Grundlagen, § 10).

Now the converse is arrived at similarly to the case of the finite group. The fact that the dependencies (16) hold leads, through multiplication by  $p^{(\lambda)}$  and addition, by virtue of the identity transformation (14), to  $\sum \psi_i \bar{\delta} u_i = \text{Div } \Gamma$  and thence, as in Section 3 follows the determination of  $\Delta x$  and  $\Delta u$  and the invariance of  $I$  with respect to these infinitesimal transformations, if they contain no derivatives  $\frac{\partial u}{\partial x}, \dots$ , certainly form a group, follows, as in Section 3, from the fact that otherwise, by composition more arbitrary functions would occur, whereas by assumption there are to be only  $\rho$  dependencies (16); hence they form an “infinite group of infinitesimal transformations.” But such a one consists (Grundlagen, Theorem VII, p. 391) of the most general infinitesimal transformations of a certain “infinite group  $\mathfrak{G}$  of finite transformations,” in Lie's sense, thereby defined. Each such finite transformation is generated from infinitesimal ones (Grundlagen, § 7),<sup>18</sup> and so arises through integration of the simultaneous system

$$\frac{dx_i}{dt} = \Delta x_i; \quad \frac{du_i}{dt} = \Delta u_i; \quad (x_i = y_i, u_i = v, \text{ for } t = 0),$$

where, however, it may be necessary further to assume the arbitrary  $p(x)$ 's dependent on  $t$ . Thus  $\mathfrak{G}$  does actually depend on  $\rho$  arbitrary functions; if in particular it suffices to assume  $p(x)$  free from  $t$ , then this dependency becomes analytic in the arbitrary function  $q(x) = t \cdot p(x)$ .<sup>19</sup> If derivatives  $\frac{\partial u}{\partial x}, \dots$ , occurs, it may be necessary also to adjoin infinitesimal transformation(s)  $\bar{\delta} u = 0$ ,  $\text{Div}(f \cdot \Delta x) = 0$  before drawing the same conclusions.

In terms of an example of Lie's (Grundlagen, § 7), let us add mention of a fairly general case in which it is possible to break through to explicit formulas, which at the same time show that the derivatives of the arbitrary functions up to the  $\sigma$ -th order to occur; where, in other words, the converse is complete. I refer to such groups of infinitesimal transformations of the  $u$ 's thereby “induced” corresponds; i.e., such transformations of the  $u$ 's for which  $\Delta u$ , and consequently  $u$ , depend only on the arbitrary functions occurring in  $\Delta x$ ; assuming further that the derivatives  $\frac{\partial u}{\partial x}, \dots$  do not occur in  $\Delta u$ . That is we have

$$\Delta x_i = p^{(i)}(x); \quad \Delta u_i = \sum_{\lambda=1}^n \left\{ a^{(\lambda)}(x, u) p^{(\lambda)} + b^{(\lambda)} \frac{\partial p^{(\lambda)}}{\partial x} + \dots + c^{(\lambda)} \frac{\partial^\sigma p^{(\lambda)}}{\partial x^\sigma} \right\}.$$

Since the infinitesimal transformation  $\Delta x = p(x)$  generates every transformation  $\Delta x = y + g(y)$  with arbitrary  $g(y)$ , we can in particular determine  $p(x)$  to depend on  $t$  in such a matter as to generate the single-member group

$$x_i = y_i + t \cdot g_i(y), \tag{18}$$

<sup>18</sup>Hence it follows in particular that the group  $\mathfrak{G}$  generated from the infinitesimal transformations  $\Delta x, \Delta u$  of a  $\mathfrak{G}_{\infty\rho}$  reduces back to  $\mathfrak{G}_{\infty\rho}$ . For  $\mathfrak{G}_{\infty\rho}$  contains no infinitesimal transformations distinct from  $\Delta x, \Delta u$  dependent on arbitrary functions, and cannot contain any independent of them but depending on parameters, as otherwise it would be a mixed group. But according to the above, the infinitesimal transformations determine the finite ones.

<sup>19</sup>The question whether perhaps this latter case always occurs was raised in a different formulation by Lie (Grundlagen, § 7 and § 13 at end).

which goes over into the identity for  $t = 0$  and into the required  $x = y + g(y)$  for  $t = 1$ . For by differentiation of (18), it follows that

$$\frac{dx_i}{dt} = g_i(y) = p^{(i)}(x, t), \quad (19)$$

where  $p(x, t)$  is determined from  $g(y)$  by inversion of (18); and conversely, (18) is generated from (19) by virtue of the auxiliary condition  $x_i = y_i$  for  $t = 0$ , by which the integral is uniquely determined. By means of (18), the  $x$ 's can be replaced in  $\Delta u$  by the “constants of integration”  $y$  and by  $t$ ; the  $g(y)$ 's occurring just up to the  $\sigma$ -th derivative, the  $\frac{\partial y}{\partial x}$ 's being expressed in terms of  $\frac{\partial x}{\partial y}$  in  $\frac{\partial p}{\partial x} = \sum \frac{\partial g}{\partial y_\kappa} \frac{\partial y_\kappa}{\partial x}$ , and  $\frac{\partial^\sigma p}{\partial x^\sigma}$  being in general replaced by its value in  $\frac{\partial g}{\partial y}, \dots, \frac{\partial x}{\partial y}, \dots, \frac{\partial^\sigma x}{\partial y^\sigma}$ . For the determination of the  $u$ 's we thus obtain the system of equations

$$\frac{du_i}{dt} = F_i \left( g(y), \frac{\partial g}{\partial y}, \dots, \frac{\partial^\sigma g}{\partial y^\sigma}, u, t \right) \quad (u_i = v_i \text{ for } t = 0)$$

in which only  $t$  and  $u$  are variables, while the  $g(y), \dots$  pertain to the field of coefficients, so that integration yields

$$u_i = v_i + B_i \left( v, g(y), \frac{\partial g}{\partial y}, \dots, \frac{\partial^\sigma g}{\partial y^\sigma}, t \right)_{t=1},$$

or transformations depending on exactly  $\sigma$  derivatives of the arbitrary functions. The identity is contained in this, by (18), for  $g(y) = 0$ ; and the group property follows from the fact that the method specified affords every transformation  $x = y + g(y)$ , whereby the induced transformation of the  $u$ 's is uniquely determined, and the group  $\mathfrak{G}$  accordingly exhausted.

From the converse it follows incidentally that it constitutes no restriction to assume the arbitrary functions to be dependent only on the  $x$ 's, not on the  $u$ ,  $\frac{\partial u}{\partial x}, \dots$ . For in the latter event, the identity transformation (14), and hence also (15), would involve not only the  $p^{(\lambda)}$ 's but also  $\frac{\partial p^{(\lambda)}}{\partial u}$ ,  $\frac{\partial p^{(\lambda)}}{\partial \frac{\partial u}{\partial x}}, \dots$ . Now if we assume the  $p^{(\lambda)}$ 's to be successively of the zeroth, first,  $\dots$  degree in  $u$ ,  $\frac{\partial u}{\partial x}, \dots$ , with arbitrary functions of  $x$  as coefficients, then we again obtain dependencies (16), only in greater number; which, however, according to the above converse, through conjunction with arbitrary functions dependent on  $x$  only, reduce to the previous case. In the same way it is shown that mixed groups correspond to simultaneous occurrence of dependencies and of divergence relationships independent of them.<sup>20</sup>

<sup>20</sup>As in Section 3, it here again follows from the converse that besides  $I$ , every integral  $I^*$  different from it by an integral over a divergence likewise admits of an infinite group, with the same  $\bar{\delta}u$ 's, though  $\Delta x$  and  $\Delta u$  will in general involve derivatives of the  $u$ 's. Such an integral  $I^*$  was introduced by Einstein in the general theory of relativity to obtain a simpler version of laws of conservation of energy;  $I$  specify the infinitesimal transformations that this  $I^*$  admits of, adhering precisely in nomenclature to Klein's second Note. The integral  $I = \int \dots \int K d\omega = \int \dots \int \mathfrak{K} dS$  admits of the group of all transformations of the  $\omega$ 's and those induced thereby for the  $g_{\mu\nu}$ 's; to this correspond the dependencies (Klein's (30))

$$\sum \mathfrak{K}_{\mu\nu} g_{\tau}^{\mu\nu} + 2 \sum \frac{\partial g^{\mu\nu} \mathfrak{K}_{\mu\tau}}{\partial \omega^\sigma} = 0.$$

Now  $I^* = \int \dots \int \mathfrak{K}^* dS$ , where  $\mathfrak{K}^* = \mathfrak{K} + \text{Div}$ , and consequently  $\mathfrak{K}_{\mu\nu}^* = \mathfrak{K}_{\mu\nu}$ , where  $\mathfrak{K}_{\mu\nu}^*$ ,  $\mathfrak{K}_{\mu\nu}$  stand in each instance for the Lagrange expressions. The dependencies specified are therefore such for  $\mathfrak{K}_{\mu\nu}^*$  also; and after multiplication by  $p^\tau$  and addition, we obtain, applying the transformations of product differentiation in reverse,

$$\sum \mathfrak{K}_{\mu\nu} p^{\mu\nu} + 2 \text{Div} \left( \sum g^{\mu\sigma} \mathfrak{K}_{\mu\tau} p^\tau \right) = 0;$$

$$\delta \mathfrak{K}^* + \text{Div} \left( \sum 2g^{\mu\sigma} \mathfrak{K}_{\mu\tau} p^\tau - \frac{\partial \mathfrak{K}^*}{\partial g_\sigma^{\mu\nu}} p^{\mu\nu} \right) = 0.$$

Comparing this with Lie's differential equation  $\delta \mathfrak{K}^* + \text{Div}(\mathfrak{K}^* \Delta \omega) = 0$ ,

$$\Delta \omega^\sigma = \frac{1}{\mathfrak{K}^*} \cdot \left( \sum 2g^{\mu\sigma} \mathfrak{K}_{\mu\tau} p^\tau - \frac{\partial \mathfrak{K}^*}{\partial g_\sigma^{\mu\nu}} p^{\mu\nu} \right); \quad \Delta g^{\mu\nu} = p^{\mu\nu} + \sum g_\sigma^{\mu\nu} \Delta \omega^\sigma$$

## § 5. Invariance of the Several Constituents of the Relations

If we specialize the group  $\mathfrak{G}$  to be the simplest case usually considered by allowing no derivatives of the  $u$ 's in the transformations, and in that the transformed independent variables depend only on the  $x$ 's not on the  $u$ 's, we can infer invariance of the several constituents in formulas. To begin with, by known arguments, we get invariance of  $\int \dots \int (\sum \psi_i \delta u_i) dx$ ; relative invariance, that is of  $\sum \psi_i \delta u_i$ ,<sup>21</sup> meaning by  $\delta$  any variation. For we have in the first place

$$\delta I = \int \dots \int \delta f \left( x, u, \frac{\partial u}{\partial x}, \dots \right) dx = \int \dots \int \delta f \left( y, v, \frac{\partial v}{\partial y}, \dots \right) dy,$$

and in the second place, for  $\delta u, \delta \frac{\partial u}{\partial x}, \dots$  vanishing at the boundary, according to which  $\delta v, \delta \frac{\partial v}{\partial y}, \dots$  vanishing at the boundary also owing to the linear homogeneous transformation of the  $\delta u, \delta \frac{\partial u}{\partial x}, \dots$ ,

$$\int \dots \int \delta f \left( x, u, \frac{\partial u}{\partial x}, \dots \right) dx = \int \dots \int (\sum \psi_i(u, \dots) \delta u_i) dx;$$

$$\int \dots \int \delta f \left( y, v, \frac{\partial v}{\partial y}, \dots \right) dy = \int \dots \int (\sum \psi_i(v, \dots) \delta v_i) dy,$$

and consequently, for  $\delta u, \delta \frac{\partial u}{\partial x}, \dots$  vanishing at the boundary,

$$\begin{aligned} \int \dots \int (\sum \psi_i(u, \dots) \delta u_i) dx &= \int \dots \int (\sum \psi_i(v, \dots) \delta v_i) dy \\ &= \int \dots \int (\sum \psi_i(v, \dots) \delta v_i) \left| \frac{\partial y_i}{\partial x_\kappa} \right| dx. \end{aligned}$$

If in the third integral  $y, v, \delta v$  are expressed in terms of  $x, u, \delta u$ , and the third is equated to the first, we thus have a relationship

$$\int \dots \int (\sum \chi_i(u, \dots) \delta u_i) dx = 0$$

for  $\delta u$  vanishing at the boundary but otherwise arbitrary, and thence follows, familiarly, the vanishing of the integrand for any  $\delta u$  whatever; the relation

$$\sum \psi_i(u, \dots) \delta u_i = \left| \frac{\partial y_i}{\partial x_\kappa} \right| (\sum \psi_i(v, \dots) \delta v_i),$$

identical in  $\delta u$ , therefore holds, asserting the relative invariance of  $\sum \psi_i \delta u_i$  and consequently the invariance of  $\int \dots \int (\sum \psi_i \delta u_i) dx$ .<sup>22</sup>

follow as infinitesimal transformations of which  $I^*$  admits. These infinitesimal transformations, then depend on the first and second derivatives of the  $g^{\mu\nu}$ 's, and contain the arbitrary  $p$ 's as far as the first derivative.

<sup>21</sup>That is,  $\sum \psi_i \delta u_i$  takes on a factor upon transformation, and this always used to be termed relative invariance in the algebraic theory of invariance.

<sup>22</sup>These conclusions fail if  $y$  depends also on the  $u$ 's, since in that case  $\delta f \left( y, v, \frac{\partial v}{\partial y}, \dots \right)$  also contains terms  $\sum \frac{\partial f}{\partial y} \delta y$ , so that the divergence transformation does not lead to the Lagrange expressions; and similarly if derivatives of the  $u$ 's are admitted; for in that case the  $\delta v$ 's become linear combinations of  $\delta u, \delta \frac{\partial u}{\partial x}, \dots$ , and so lead only after another divergence transformation to an identity  $\int \dots \int (\sum \chi_i(u, \dots) \delta u) dx = 0$ , so that again the Lagrange expressions do not appear on the right.

The question whether it is possible to argue from the invariance of  $\int \dots \int (\sum \psi_i \delta u_i) dx$  back to the subsistence of divergence relationships is synonymous, according to the converse, with the question whether one can thence infer the invariance of  $I$  with respect to a group leading not necessarily to the same  $\Delta u, \Delta x$ , but to the same  $\delta u$ 's. In the special case of the single integral and only first derivatives in  $f$ , it is possible for the finite group to argue from the invariance of the Lagrange expressions to the existence of first integrals (c.f. e.g., Engel, Gött. Nachr. 1916, p. 270).

To apply this to the divergence relationships and dependencies derived, we must first demonstrate that the  $\bar{\delta}u$  derived from the  $\Delta u$ ,  $\Delta x$ 's does in fact satisfy the laws of transformation for the variation  $\delta u$ , provided only that the parameters, or arbitrary functions, in  $\bar{\delta}v$  are so determined as corresponds to the similar group of infinitesimal transformations in  $y, v$ ; if  $\mathfrak{T}_q$  designates the transformation that carries  $x, u$  over into  $y, v$ , and  $\mathfrak{T}_p$  and infinitesimal one in  $x, u$ , then the one similar thereto in  $y, v$  is given by  $\mathfrak{T} = \mathfrak{T}_q \mathfrak{T}_p \mathfrak{T}_q^{-1}$ , where the parameters, or arbitrary functions  $r$ , are thus determined from  $p$  and  $q$ . In formulas, this is expressed as follows:

$$\begin{aligned}\mathfrak{T}_p : \xi &= x + \Delta x(x, p); & u^* &= u + \Delta u(x, u, p); \\ \mathfrak{T}_q : y &= A(x, q); & v &= B(x, u, q); \\ \mathfrak{T}_q \mathfrak{T}_p : \eta &= A(x + \Delta x(x, p), q); & v^* &= B(x + \Delta x(p), u + \Delta u(p), q).\end{aligned}$$

But this generates  $\mathfrak{T}_r = \mathfrak{T}_q \mathfrak{T}_p \mathfrak{T}_q^{-1}$ , or

$$\eta = y + \Delta y(r); \quad v^* = v + \Delta v(r),$$

if by the inverse  $\mathfrak{T}_q$  we regard the  $x$ 's as functions of the  $y$ 's and consider only the infinitesimal terms; so we have the identity

$$\begin{aligned}\eta &= y + \Delta y(r) = y + \sum \frac{\partial A(x, q)}{\partial x} \Delta x(p); \\ v^* &= v + \Delta v(r) = v + \sum \frac{\partial B(x, u, q)}{\partial x} \Delta x(p) + \sum \frac{\partial B(x, u, q)}{\partial u} \Delta u(p).\end{aligned}\tag{20}$$

Replacing  $\xi = x + \Delta x$  by  $\xi - \Delta \xi$  in this so that  $\xi$  goes back into  $x$ , and  $\Delta x$  vanishes, by the first equation (20) too will go back over into  $y = \eta - \Delta \eta$ ; if by this substitution  $\Delta u(p)$  goes over into  $\bar{\delta}u(p)$ , then  $\Delta v(r)$  will go over into  $\bar{\delta}v(r)$  as well, and the second formula (20) gives

$$\begin{aligned}v + \bar{\delta}v(y, v, \dots r) &= v + \sum \frac{\partial B(x, u, q)}{\partial u} \bar{\delta}u(p), \\ \bar{\delta}v(y, v, \dots r) &= \sum \frac{\partial B}{\partial u} \bar{\delta}u_\kappa(x, u, p),\end{aligned}$$

so that the transformation formulas for variations are actually satisfied provided  $\bar{\delta}v$  is assumed to depend only on the parameters or arbitrary functions  $r$ .<sup>23</sup>

So in particular, the relative invariance of  $\sum \psi_i \bar{\delta}u_i$  follows: hence also, by (12), since the divergence relationships are satisfied in  $y, v$  as well, the relative invariance of  $\text{Div } B$ ; and further, by (14) and (13), the relative invariance of  $\text{Div } \Gamma$  and of the left-hand sides of the dependencies as conjoined with the  $p^{(\lambda)}$ 's, where the arbitrary  $p(x)$ 's (or the parameters) are to be replaced by the  $r$ 's everywhere in the transformed formulas. This leads as well to the relative invariance of  $\text{Div}(B - \Gamma)$ , or of a divergence of a not identically vanishing system of functions  $B - \Gamma$  whose divergence vanishes identically.

From the relative invariance of  $\text{Div } B$ , we can draw additional inference of invariance of the first integral in the one-dimensional case and for finite group. The parametric transformation corresponding to the infinitesimal transformation becomes, by (20), linear and homogeneous, and owing to the invertibility of all transformations, the  $\epsilon$ 's also will be linear and homogeneous in the transformed parameters  $\epsilon^*$ . This invertibility is certainly preserved if we put  $\psi = 0$ , since no derivatives of the  $u$ 's occur in (20). Through equating the coefficients of the  $\epsilon^*$ 's in

$$\text{Div } B(x, u, \dots \epsilon) = \frac{dy}{dx} \cdot \text{Div } B(y, v, \dots \epsilon^*)$$

<sup>23</sup>It turns out again that  $y$  must be taken independent of  $u$  in order for the conclusions to hold. As an example, consider the  $\delta g^{\mu\nu}$  and  $\delta q_\rho$  given by Klein, which satisfy the transformations for variations provided the  $p$ 's are subjected to a vector transformation.

the  $\frac{d}{dy}B^{(\lambda)}(y, v, \dots)$ 's therefore also become linear homogeneous functions of  $\frac{d}{dx}B^{(\lambda)}(x, u, \dots)$ 's so that  $\frac{d}{dx}B^{(\lambda)}(x, u, \dots) = 0$  or  $B^{(\lambda)}(x, u) = \text{const.}$  duly entails  $\frac{d}{dy}B^{(\lambda)}(y, v, \dots) = 0$  or  $B^{(\lambda)}(y, v) = \text{const.}$  as well. So the  $\rho$  first integrals corresponding to a  $\mathfrak{G}_\rho$  in each instance admit of the group, with result that the further integration is simplified. The simplest example of this is that  $f$  is free of  $x$  or of a  $u$ , which corresponds to the infinitesimal transformation  $\Delta x = \epsilon$ ,  $\Delta u = 0$ , or  $\Delta x = 0$ ,  $\Delta u = \epsilon$ . We shall have  $\bar{\delta}u = -\epsilon \frac{du}{dx}$  or  $\epsilon$  respectively, and since  $B$  is derived from  $f$  and  $\bar{\delta}u$  by differentiation and rational combination, it is free accordingly of  $x$  or  $u$  respectively, and admits of the corresponding groups.<sup>24</sup>

## § 6. A Hilbertian Assertion

From the foregoing, finally, we also obtain the proof of a Hilbertian assertion about the connection of the failure of laws of conservation of energy proper with “general relativity” (Klein’s first Note, Göttinger Nachr. 1917, Reply 1st paragraph), and that in a generalized group theory version.

Let the integral  $I$  admit of a  $\mathfrak{G}_{\infty\rho}$ , and let  $\mathfrak{G}_\rho$  be any finite group generated by specializing the arbitrary functions, and hence a subgroup of  $\mathfrak{G}_{\infty\rho}$ . Then to the infinite group  $\mathfrak{G}_{\infty\rho}$  there correspond dependencies (16), and to the finite one  $\mathfrak{G}_\sigma$ , divergence relationships (13); and conversely from the subsistence of any divergence relationships, the invariance of  $I$  follows, with respect to some finite group which will be identical with  $\mathfrak{G}_\sigma$  if and only if the  $\bar{\delta}u$ 's are linear combinations of those obtained from  $\mathfrak{G}_\sigma$ . Thus the invariance with respect to  $\mathfrak{G}_\sigma$  cannot lead to any divergence relationships different from (13). But since the subsistence of (16) entails the invariance of  $I$  with respect to the infinitesimal transformations,  $\Delta u$ ,  $\Delta x$  of  $\mathfrak{G}_{\infty\rho}$  for *any*  $p(x)$ , it entails in particular nothing less than invariance with respect to the infinitesimal transformations of  $\mathfrak{G}_\sigma$  arising therefrom by specialization and consequently with respect to  $\mathfrak{G}_\sigma$ . Thus the divergence relationships  $\sum \psi_i \bar{\delta}u_i^{(\lambda)} = \text{Div } B^{(\lambda)}$  must be consequences of the dependencies (16), which latter may alternatively be written  $\sum \psi_i a_i^{(\lambda)} = \text{Div } \chi^{(\lambda)}$  where the  $\chi^{(\lambda)}$ 's are linear combinations of the Lagrange expressions and their derivatives. Since the  $\psi$ 's occur linearly in both (13) and (16), the divergence relations must thus in particular be *linear* combinations of the dependencies (16); Accordingly,  $\text{Div } B^{(\lambda)} = \text{Div} \left( \sum \alpha \cdot \chi^{(\kappa)} \right)$ ; and the  $B^{(\lambda)}$ 's themselves are thus linearly composed of the  $\chi$ 's, i.e., the Lagrange expressions and their derivatives, and of functions whose divergence vanishes identically, say like the  $B - \Gamma$ 's encountered at the close of Section 2, for which  $\text{Div}(B - \Gamma) = 0$ , and where the divergence at the same time has the invariant property. I shall refer to divergence relationships in which the  $B^{(\lambda)}$ 's can be composed from the Lagrange expressions and their derivatives in the specified manner as “improper,” and to all other as “proper.”

If conversely the divergence relations are linear combinations of the dependencies (16), and so “improper,” invariance with respect to  $\mathfrak{G}_\sigma$  follows from that with respect to  $\mathfrak{G}_{\infty\rho}$ ;  $\mathfrak{G}_\sigma$  becomes a subgroup of  $\mathfrak{G}_{\infty\rho}$ . The divergence relationships corresponding to an infinite group  $\mathfrak{G}_\sigma$  will thus be improper if and only if  $\mathfrak{G}_\sigma$  is a subgroup of an infinite group invariant with respect to  $I$ .

By specialization of the groups, this yields the original Hilbertian assertion. Let “displacement group” be understood to mean the finite

$$y_i = x_i + \epsilon_i; \quad v_i(y) = u_i(x);$$

---

<sup>24</sup>In the cases where mere invariance of  $\int (\sum \psi_i \delta u_i) dx$  entails the existence of first integrals, these do not admit of the entire group  $\mathfrak{G}_\rho$ ; for example,  $\int (u'' \delta u) dx$  admits of the infinitesimal transformation  $\Delta x = \epsilon_2$ ,  $\Delta u = \epsilon_1 + x\epsilon_3$ ; whereas the first integral  $u - u'x = \text{const.}$ , corresponding to  $\Delta x = 0$ ,  $\Delta u = x\epsilon_3$ , does not admit of the other two infinitesimal transformations, since it explicitly contains both  $u$  and  $x$ . To this first integral, there happen to correspond infinitesimal transformations for  $f$  that contain derivatives. So we see that invariance  $\int \dots \int (\sum \psi_i \delta u_i) dx$  is at all events a weaker condition than invariance of  $I$ , and this should be noted as to a question raised in a previous remark.

that is

$$\Delta x_i = \epsilon_i, \quad \Delta u_i = 0, \quad \bar{\delta} u_i = - \sum_{\lambda} \frac{\partial u_i}{\partial x_{\lambda}} \epsilon_{\lambda}.$$

Invariance with respect to the displacement group asserts, as we know, that in

$I = \int \dots \int f \left( x, u, \frac{\partial u}{\partial x}, \dots \right) dx$ , the  $x$ 's do not occur explicitly in  $f$ . The associated  $n$  divergence relationships

$$\sum \psi_i \frac{\partial u_i}{\partial x_{\lambda}} = \operatorname{Div} B^{(\lambda)} \quad (\lambda = 1, 2, \dots, n)$$

will be referred to as “energy relationships,” since the laws of conservation  $\operatorname{Div} B^{(\lambda)} = 0$  corresponding to the variation problem answer to “laws of conservation of energy,” and the  $B^{(\lambda)}$ 's to the “energy components.” So then we have: If  $I$  admits of the displacement group, then the energy relationships become improper if and only if  $I$  is invariant with respect to an infinite group containing the displacement group as subgroup.<sup>25</sup>

An example of such infinite groups is presented by the group of *all* transformations of the  $x$ 's and such of the induced transformations of the  $u(x)$ 's in which only *derivatives* of the arbitrary functions  $p(x)$  occur; the displacement group is generated by the specialization  $p^{(i)}(x) = \epsilon_i$ ; but it must remain undecided whether this — and the groups generated by change of  $I$  by a boundary integral — suffices to give the most general of these groups. Induced transformations of the specified kind arise, say, when the  $u$ 's are subjected to the coefficient transformations of a “total differential form,” i.e., a form  $\sum a d^{\lambda} x_i + \sum b d^{\lambda-1} x_i dx_{\kappa} + \dots$  containing higher differentials besides the  $dx$ 's; more special induced transformations, in which the  $p(x)$ 's occur in first derivative only, are furnished by the coefficient transformations of ordinary differential forms  $\sum c dx_{i_1} \dots dx_{i_{\lambda}}$ , and only these have ordinarily been considered.

Another group of the specified kind — one which, owing to the occurrence of the logarithmic term, cannot be coefficient transformation — is, say, the following:

$$y = x + p(x); \quad v_i = u_i + \ln(1 + p'(x)) = u_i + \ln \frac{dy}{dx};$$

$$\Delta x = p(x); \quad \Delta u_i = p'(x);^{26} \quad \bar{\delta} u_i = p'(x) - u'_i p(x).$$

The dependencies (16) here become

$$\sum_i \left( \psi_i u'_i + \frac{d\psi_i}{dx} \right) = 0,$$

and the improper energy relationships

$$\sum \left( \psi_i u'_i + \frac{d(\psi_i + \text{const.})}{dx} \right) = 0.$$

A simple invariant integral of the group is

$$I = \int \frac{e^{-2u_1}}{u'_1 - u'_2} dx.$$

The most general  $I$  is determined by integration of Lie's differential equation

$$\bar{\delta} f + \frac{d}{dx}(f \cdot \Delta x) = 0,$$

<sup>25</sup>The laws of conservation of energy of classical mechanics as well as those of the old “theory of relativity” (where  $\sum dx^2$  goes over into itself) are “proper,” since no infinite groups occur.

<sup>26</sup>From these infinitesimal transformations, the finite ones are calculated backwards by the method given in Section 4 at end.

which by substitution of their values for  $\Delta x$  and  $\bar{\delta}u$ , provided  $f$  is assumed to depend on only first derivatives of the  $u$ 's, goes over into

$$\frac{\partial f}{\partial x} p(x) + \left\{ \sum \frac{\partial f}{\partial u_i} - \frac{\partial f}{\partial u'_i} u'_i + f \right\} p'(x) + \left\{ \sum \frac{\partial f}{\partial u''_i} \right\} p''(x) = 0$$

(identically in  $p(x)$ ,  $p'(x)$ ,  $p''(x)$ ). This system of equations has solutions for as few as two functions  $u(x)$  actually containing the derivatives, namely

$$f = (u'_1 - u'_2) \Phi \left( u_1 - u_2, \frac{e^{-u_1}}{u'_1 - u'_2} \right),$$

where  $\Phi$  stands for an arbitrary function of the specified arguments.

Hilbert enunciates his assertion to the effect that the failure of proper laws of conservation of energy is a characteristic feature of the “general theory of relativity.” In order for this assertion to hold good literally, therefore, the term “general relativity” should be taken in a broader sense than usual, and extended also to the forgoing groups depending on  $n$  arbitrary functions.<sup>27</sup>

---

<sup>27</sup>This again confirms the correctness of a comment of Klein's that the term “relativity” current in physics is replaceable by “invariance relative to a group.” (“Über die geometrischen Grundlagen der Lorentzgruppe,” Jhrber. d. Deutsch. Math. Vereinig. **19** (1910), p. 287, reprinted in the Phys. Zeitschrift.)

ON THE QUANTUM THEORY OF LINE-SPECTRA.

DEDICATED TO THE MEMORY OF MY VENERATED TEACHER  
PROFESSOR C. CHRISTIANSEN  
October 9, 1843 November 28, 1917

N. Bohr,  
*Dr. phil. Copenhagen*  
(Received 1918)

*Introduction*

In an attempt to develop certain outlines of a theory of line-spectra based on a suitable application of the fundamental ideas introduced by Planck in his theory of temperature-radiation to the theory of the nucleus atom of Sir Ernest Rutherford, the writer has shown that it is possible in this way to obtain a sample interpretation of some of the main laws governing the line-spectra of the elements, and especially to obtain a deduction of the well known Balmer formula for the hydrogen spectrum<sup>1</sup>. The theory in the form given allowed of a detailed discussion only in the case of periodic systems, and obviously was not able to account in detail for the characteristic difference between the hydrogen spectrum and the spectra of other elements, or for the characteristic effects on the hydrogen spectrum of external electric

---

<sup>1</sup>N.Bohr, Phil. Mag. **26** (1913) 1, 476, 857; **27** (1914) 506; **29** (1915) 332; **30** (1915) 394.

and magnetic fields. Recently, however, a way out of this difficulty has been opened by Sommerfeld<sup>2</sup> who, by introducing a suitable

Editor's note. This paper is the Introduction and Part I (pp. 1–36) of the paper *On the Quantum Theory of Line-spectra*, which was first published (1918–1922) in the *Memoires de l'Academie Royale des Sciences et des Letters de Danemark*, Copenhague, Section des Sciences, 8me serie t. IV, n° 1, fasc. 1–3 (D. Kgl. Danske Vidensk. Selsk. Skrifter, Naturvidensk. og Mathem. Afd. 8. Række, IV.1, 1–3). The Introduction was signed 'Copenhagen, November 1917'. The titles of Parts II and III, which are not reproduced here, are respectively *on the hydrogen spectrum* and *On the spectra elements of higher atomic number*.

generalisation of the theory to a simple type of non-periodic motions and by taking the small variation of the mass of the electron with its velocity into account, obtained an explanation of the line-structure of the hydrogen lines which was found to be in brilliant conformity with the measurements. Already in his first paper on this subject, Sommerfeld pointed out that his theory evidently offered a clue to the interpretation of the more intricate structure of the spectra of other elements. Briefly afterwards Epstein<sup>3</sup> and Schwarzschild<sup>4</sup> independent of each other, by adapting Sommerfeld's ideas to the treatment of a more extended class of non-periodic systems obtained a detailed explanation of the characteristic effect of an electric field on the hydrogen spectrum discovered by Stark. Subsequently Sommerfeld<sup>5</sup> himself and Debye<sup>6</sup> have on the same lines indicated an interpretation of the effect of a magnetic field on the hydrogen spectrum which, although no complete explanation of the observations was obtained, undoubtedly represents an important step towards a detailed understanding of this phenomenon.

In spite of the great progress involved in these investigations many difficulties of fundamental nature remained unsolved, not only as regards the limited applicability of the methods used in calculating the frequencies of the spectrum of a given system, but especially as regards the question of the

<sup>2</sup>A. Sommerfeld, Ber. Akad. München. 1915, pp. 425, 459; 1916, p. 131; 1917, p. 83. Ann. d. Phys. **51** (1916) 1.

<sup>3</sup>P. Epstein, Phys. Zeitschr. **17** (1916) 148; Ann. d. Phys. **50** (1916) 489; **51** (1916) 168.

<sup>4</sup>K. Schwarzschild, Ber. Akad. Berlin (1916) 548.

<sup>5</sup>A. Sommerfeld, Phys. Zeitschr. **17** (1916) 491.

<sup>6</sup>P. Debye, Nachr. K. Ges. d. Wiss. Göttingen, 1916; Phys. Zeitschr. **17** (1916) 507.

polarization and intensity of the emitted spectral lines. These difficulties are intimately connected with the radical departure from the ordinary ideas of mechanics and electrodynamics involved in the main principles of the quantum theory, and with the fact that it has not been possible hitherto to replace these ideas by others forming an equally consistent and developed structure. Also in this respect, however, great progress has recently been obtained by the work of Einstein<sup>7</sup> and Ehrenfest<sup>8</sup>. On this state of the theory it might therefore be of interest to make an attempt to discuss the different applications from a uniform point of view, and especially to consider the underlying assumptions in their relations to ordinary mechanics and electrodynamics. Such an attempt has been made in the present paper, and it will be shown that it seems possible to throw some light on the outstanding difficulties by trying to trace the analogy between the quantum theory and the ordinary theory of radiation as closely as possible.

The paper is divided into four parts.

- |          |   |
|----------|---|
| Part I   | contains a brief discussion of the general principles of the theory and deals   |
| Part II  | with the application of the general theory to periodic systems of one degree  |
| Part III | of freedom and to the class of non-periodic systems referred to above.  |
| Part IV  | contains a detailed discussion of the theory of the hydrogen spectrum in<br>order to illustrate the general considerations.<br>contains a discussion of the questions arising in connection with the expla-<br>nation of the spectra of other elements.<br>contains a general discussion of the theory of the constitution of atoms and<br>molecules based on the application of the quantum theory to the nucleus<br>atom. |

*Copenhagen, November 1917.*

---

<sup>7</sup>A. Einstein, Verh. d. D. phys. Ges. **18** (1916) 318; Phys. Zeitschr. **18** (1917) 121.

<sup>8</sup>P. Ehrenfest, Proc. Acad. Amsterdam **16** (1914) 591; Phys. Zeitschr. **15** (1914) 657;  
Ann. d. Phys. **51** (1916) 327; Phil. Mag. **33** (1917) 500.

## PART I. ON THE GENERAL THEORY

### 1. General principles

The quantum theory of line-spectra rests upon the following fundamental assumptions:

- I. That an atomic system can, and can only, exist permanently in a certain series of states corresponding to a discontinuous series of values for its energy, and that consequently any change of the energy of the system, including emission and absorption of electromagnetic radiation, must take place a complete transition between two such states. These states will be denoted as the stationary states of the system.
- II. That the radiation absorbed or emitted during a transition between two stationary states is ‘unifrequentic’ and possesses a frequency  $\nu$ , given by the relation

$$E' = E'' = h\nu, \quad (1)$$

Where  $h$  is Planck’s constant and where  $E'$  and  $E''$  are the values of the energy in the two states under consideration.

As pointed out by the writer in the papers referred to in the introduction, these assumptions offer an immediate interpretation of the fundamental principle of combination of spectral lines deduced from the measurements of the frequencies of the series spectra of the elements. according to the laws discovered by Balmer, Rydberg and Ritz, the frequencies of the lines of the series spectrum of an element can be expressed by a formula of the type:

$$\nu = f_{\tau''}(n'') - f_{\tau'}(n'), \quad (2)$$

where  $n'$  and  $n''$  are whole numbers and  $f_{\tau}(n)$  is one among a set of functions of  $n$ , characteristic for the element under consideration. On the above assumptions this formula may obviously be interpreted by assuming that the stationary states of an atom of an element form a set of series, and that the energy in the  $n$ th state of the  $\tau$ th series, omitting an arbitrary constant, is given by

$$E_{\tau}(n) = -hf_{\tau}(n). \quad (3)$$

We thus see that the values for the energy in the stationary states of an atom may be obtained directly from the measurements of the spectrum by

means of relation (1). In order, however, to obtain a theoretical connection between these values and the experimental evidence about the constitution of the atom obtained from other sources, it is necessary to introduce further assumptions about the laws which govern the stationary states of a given atomic system and the transitions between these states.

Now on the basis of a vast amount of experimental evidence, we are forced to assume that an atom or molecule consists of a number of electrified particles in motion, and, since the above fundamental assumptions imply that no emission of radiation takes place in the stationary states, we must consequently assume that *the ordinary laws of electrodynamics cannot be applied* to these states without radical alterations. In many cases, however, the effect of that part of the electrodynamical forces which is connected with the emission of radiation will at any moment be very small in comparison with the effect of the simple electrostatic attractions or repulsions of the charged particles corresponding to Coulomb's law. Even if the theory of radiation must be completely altered, it is therefore a natural assumption that it is possible in such cases to obtain a close approximation in the description of the motion in the stationary states, by retaining only the latter forces. In the following we shall therefore, as in all the papers mentioned in the introduction, for the present *calculate the motions of the particles in the stationary states as the motions of mass-points according to ordinary mechanics* including the modifications claimed by the theory of relativity, and we shall later in the discussion of the special applications come back to the question of the degree of approximation which may be obtained in this way.

If next we consider a transition between two stationary states, it is obvious at once from the essential discontinuity, involved in the assumptions I and II, that in general it is impossible even approximately to describe this phenomenon by means of ordinary mechanics or to calculate the frequency of the radiation absorbed or emitted by such a process by means of ordinary electrodynamics. On the other hand, from the fact that it has been possible by means of ordinary mechanics and electrodynamics to account for the phenomenon of temperature-radiation in the limiting region of slow vibrations, we may expect that any theory capable of describing this phenomenon in accordance with observations will form some sort of natural generalisation of the ordinary theory of radiation. Now the theory of temperature-radiation in the form originally given by Planck confessedly lacked internal consistency, since, in the deduction of his radiation formula, assumptions of similar character as I and II were used in connection with assumptions which were in obvious contrast to them. Quite recently, how-

ever, Einstein<sup>9</sup> has succeeded, on the basis of the assumptions I and II, to give a consistent and instructive deduction of Planck's formula by introducing certain supplementary assumptions about the *probability of transition of a system between two stationary states* and about the manner in which this probability depends on the density of radiation of the corresponding frequency in the surrounding space, suggested from analogy with the ordinary theory of radiation. Einstein compares the emission or absorption of radiation of frequency  $\nu$  corresponding to a transition between two stationary states with the emission or absorption to be expected on ordinary electrodynamics for a system consisting of a particle executing harmonic vibrations of this frequency. In analogy with the fact that on the latter theory such a system will without external excitation emit a radiation of frequency  $\nu$ , Einstein assumes in the first place that on the quantum theory there will be a certain probability  $A_{n''}^{n'} dt$  that the system in the stationary state of greater energy, characterised by the letter  $n'$ , in the time interval  $dt$  will start *spontaneously* to pass to the stationary state of smaller energy, characterised by the letter  $n''$ . Moreover, on ordinary electrodynamics the harmonic vibrator will, in addition to the above mentioned independent emission, in the presence of a radiation of frequency  $\nu$  in the surrounding space, and dependent on the accidental phase-difference between this radiation and the vibrator, emit or absorb radiation-energy. In analogy with this, Einstein assumes secondly that in the presence of a radiation in the surrounding space, the system will on the quantum theory, in addition to the above mentioned probability of spontaneous transition from the state  $n'$  to the state  $n''$ , possess a certain probability, depending on this radiation, of passing in the time  $dt$  from the state  $n'$  to the state  $n''$ , as well as from the state  $n''$  to the state  $n'$ . These latter probabilities are assumed to be proportional to the intensity of the surrounding radiation and are denoted by  $\varrho_\nu B_{n''}^{n'} dt$  and  $\varrho_\nu B_{n'}^{n''} dt$  respectively, where  $\varrho_\nu d\nu$  denotes the amount of radiation in unit volume of the surrounding space distributed on frequencies between  $\nu$  and  $\nu + d\nu$ , while  $B_{n''}^{n'}$  and  $B_{n'}^{n''}$  are constants which, like  $A_{n''}^{n'}$ , depend only on the stationary states under consideration. Einstein does not introduce any detailed assumption as to the values of these constants, no more than to the conditions by which the different stationary states of a given system are determined or to the 'a-priory probability' of these states on which their relative occurrence in a distribution of statistical equilibrium depends. He shows, however, how it is possible from the above general assumptions of Boltzmann's principle on the relation between entropy and

---

<sup>9</sup>A. Einstein, loc. cit. [ [paper 1 and 1d]

probability and Wien's well known displacement-law, to deduce a formula for the temperature radiation which apart from an undetermined constant factor coincides with Planck's, if we only assume that the frequency corresponding to the transition between the two states is determined by (1). It will therefore be seen that by reversing the line of argument, Einstein's theory may be considered as a very direct support of the latter relation.

In the following discussion of the application of the quantum theory to determine the line-spectrum of a given system, it will, just as in the theory of temperature-radiation, not be necessary to introduce detailed assumptions as to the mechanism of transition between two stationary states. We shall show, however, that the conditions which will be used to determine the values of the energy in the stationary states are of such a type that the frequencies calculated by (1), in the limit where the motions in successive stationary states comparatively differ very little from each other, will tend to coincide with the frequencies to be expected on the ordinary theory of radiation from the motion of the system in the stationary states. In order to obtain the necessary relation to the ordinary theory of radiation in the limit of slow vibrations, we are therefore led directly to certain conclusions about the probability of transition between two stationary states in this limit. This leads again to certain general considerations about connection between the probability of a transition between any two stationary states and the motion of the system in these states, which will be shown to throw light on the question of the polarisation and intensity of the different lines of the spectrum of a given system.

In the above considerations we have by an atomic system tacitly understood a number of electrified particles which move in a field of force which, with approximation mentioned, possesses a potential depending only on the position of the particles. This may more accurately be denoted as a system under constant external conditions, and the question next arises about the variation in the stationary states which may be expected to take place during a variation of the external conditions, e.g. when exposing the atomic system to some variable external field to force. Now, in general, we must obviously assume that this variation cannot be calculated by ordinary mechanics, no more than the transition between two different stationary states corresponding to constant external conditions. If, however, the variation of the external conditions is very slow, we may from the necessary stability of the stationary states expect that the motion of the system at any given moment during the variation will differ only very little from the motion in a stationary state corresponding to the instantaneous external conditions. If now, moreover, the variation is performed at a constant or very slowly changing rate, the

forces to which the particles of the system will be exposed will not differ at any moment from those to which they would be exposed if we imagine that the external forces arise from a number of slowly moving additional particles which together with the original system form a system in a stationary state. From this point of view it seems therefore natural to assume that, with the approximation mentioned, the motion of an atomic system in the stationary states can be calculated direct application of ordinary mechanics, not only under constant external conditions, but in general also during a slow and uniform variation of these conditions. This assumption, which may be denoted as the principle of the '*mechanical transformability*' of the stationary states, has been introduced in the quantum theory by Ehrenfest<sup>10</sup> and is, as it will be seen in the following sections, of great importance in the discussion of the conditions to be used to fix the stationary states of an atomic system among the continuous multitude of mechanically possible motions. In this connection it may be pointed out that the principle of the mechanical transformability of the stationary states allows us to overcome a fundamental difficulty which at sight would seem to be involved in the definition of the energy difference between two stationary states which enters in relation (1). In fact we have assumed that the direct transition between two such states cannot be described by ordinary mechanics, while on the other hand we possess no means of defining an energy difference between two states if there exists no possibility for a continuous mechanical connection between them. It is clear, however, that such a connection is just afforded by Ehrenfest's principle which allows us to transform mechanically the stationary states of a given system into those of another, because for the latter system we may take one in which the forces which act on the particles are very small and where we may assume that the values of the energy in all the stationary states will tend to coincide.

As regards the problem of the statistical distribution of the different stationary states between a great number of atomic systems of the same kind in temperature equilibrium, the number of systems present in the different states may be deduced in the well known way from Boltzmann's fundamental relation between entropy and probability, if we know the values of the energy in these states and the *a-priori probability* to be ascribed to each state in the calculation of the probability of the whole distribution. In contrast

---

<sup>10</sup>P. Ehrenfest, loc.cit. In these papers the principle in question is called the adiabatic hypothesis in accordance with the line of argumentation followed by Ehrenfest in which considerations of thermodynamical problems play an important part. From the point of view taken in the present paper, however, the above notation might in a more direct way indicate the content of the principle and the limits of its applicability.

considerations of ordinary statistical mechanics we possess on the quantum theory no direct of determining these a-priori probabilities, because we have no detailed information about the mechanism of transition between the different stationary states. If the a-priori probabilities are known for the states of a given atomic system, however, they may be deduced for any other system which can be formed from this by a continuous transformation without passing through one of the singular systems referred to below. In fact, in examining the necessary conditions for the explanation of the second law of thermodynamics Ehrenfest<sup>11</sup> has deduced a certain general condition as regards the variation of the a-priori probability corresponding to a small change of the external conditions from which it follows, that the a-priori probability of a given stationary state of an atomic system must remain unaltered during a continuous transformation, except in special cases in which the values of the energy in some of the stationary states will tend to coincide during the transformation. In this result we possess, as we shall see, a rational basis for the determination of the a-priori probability of the different stationary states of a given atomic system.

## 2. System of one degree of freedom

As the simplest illustration of the principles discussed in the former section we shall begin by considering systems of a single degree of freedom, in which case it has been possible to establish a general theory of stationary states. This is due to the fact, that *the motion will be simply periodic*, provided the distance between the parts of the system will not increase infinitely with the time, a case which for obvious reasons cannot represent a stationary state in the sense defined above. On account of this, the discussion of the mechanical transformability of the stationary states can, as pointed out by Ehrenfest<sup>12</sup> for systems of one degree of freedom be based on a mechanical theorem about periodic systems due to Boltzmann and originally applied by this author in a discussion of the bearing of mechanics on the explanation of the laws of thermodynamics. For the sake of the considerations in the following sections it will be convenient here to give the proof in a form which differs slightly from that given by Ehrenfest, and which takes also regard to

<sup>11</sup>P. Ehresfest, Phys. Zeitschr. **15** (1914) 660. The above interpretation of this relation is not stated explicitly by Ehrenfest, but it presents itself directly if the quantum theory is taken in the form corresponding to the fundamental assumption I.

<sup>12</sup>P. Ehrenfest, loc.cit. Proc. Acad. Amsterdam (1914) 591.

the modifications in the ordinary laws of mechanics claimed by the theory of relativity.

Consider for the sake of generality a conservative mechanical system of  $s$  degrees of freedom, the motion of which is governed by Hamilton's equations:

$$\frac{dp_k}{dt} = -\frac{\partial E}{\partial q_k}, \quad \frac{dq_k}{dt} = \frac{\partial E}{\partial p_k}, \quad (k = 1, \dots, s) \quad (4)$$

where  $E$  is the total energy considered as a function of the generalised positional coordinates  $q_1, \dots, q_s$  and the corresponding canonically conjugated momenta  $p_1, \dots, p_s$ . If the velocities are so small that the variation in the mass of the particles due to their velocities can be neglected, the  $p$ 's are defined in the usual way by

$$p_k = \frac{\partial T}{\partial \dot{q}_k}, \quad (k = 1, \dots, s)$$

where  $T$  is the kinetic energy of the system considered as a function of the generalised velocities  $\dot{q}_1, \dots, \dot{q}_s$  ( $\dot{q}_k = dq_k/dt$ ) and of  $q_1, \dots, q_s$ . If the relativity modifications are taken into account the  $p$ 's are defined by a similar set of expressions in which the kinetic energy is replaced by  $T' = \sum m_0 c^2 \cdot (1 - \sqrt{1 - v^2/c^2})$ , where the summation is to be extended over all the particles of the system, and  $v$  is the velocity of one of the particles and  $m_0$  its mass for zero velocity, while  $c$  is the velocity of light.

Let us now assume that the system performs a periodic motion with the period  $\sigma$ , and let us form the expression

$$I = \int_0^\sigma \sum_1^s p_k \dot{q}_k \, dt, \quad (5)$$

which is easily seen to be independent of the special choice of coordinates  $q_1, \dots, q_s$  used to describe the motion of the system. In fact, if the variation of the mass with the velocity is neglected we get

$$I = 2 \int_0^\sigma T \, dt,$$

and if the relativity modifications are included, we get a quite analogous in which the kinetic energy is replaced by  $T'' = \sum \frac{1}{2} m_0 v^2 / \sqrt{1 - v^2/c^2}$ .

Consider next some new periodic motion of the system formed by a small variation of the first motion, but which may need the presence of external

forces in order to be a mechanically possible motion. For the variation in  $I$  we get then

$$\delta I = \int_0^\sigma \sum_1^s (\dot{q}_k \cdot \delta \cdot p_k + p_k \cdot \delta \dot{q}_k) dt + \left| \sum_1^s p_k \cdot \dot{q}_k \delta t \right|_0^\sigma,$$

where the last term refers to the variation of the limit of the integral due to the variation in the period  $\sigma$ . By partial integration of the second term in the bracket under integral we get next

$$\delta I = \int_0^\sigma \sum_1^s (\dot{q}_k \cdot \delta p_k - p_k \cdot \delta q_k) dt + \left| \sum_1^s p_k \cdot (\dot{q}_k \cdot \delta t + \delta q_k) \right|_0^\sigma,$$

where the last term is seen to be zero, because the term in the bracket as well as  $p_k$  will be the same in both limits, since the varied motion as well as the original motion is assumed to be periodic. By means of equations (4) we get therefore

$$\delta I = \int_0^\sigma \sum_1^s \left( \frac{\partial E}{\partial p_k} \cdot \delta \cdot p_k + \frac{\partial E}{\partial q_k} \cdot \delta \cdot q_k \right) dt = \int_0^\sigma \delta E dt. \quad (6)$$

Let us now assume that the small variation of the motion is produced by a small external field established at a uniform rate during a time interval  $\vartheta$ , long compared with  $\sigma$ , so that the comparative increase during a period is very small. In this case  $\delta E$  is at any moment equal to the total work done by the external forces on the particles of the system since the beginning of the establishment of the field. Let this moment be  $t = -\vartheta$  and let the potential of the external field at  $t \geq 0$  be given by  $\Omega$ , expressed as a function of the  $q$ 's. at any given moment  $t > 0$  we have then

$$\delta E = - \int_{-\vartheta}^0 \frac{\vartheta + t}{\vartheta} \cdot \sum_1^s \frac{\partial \Omega}{\partial q_k} \cdot \dot{q}_k \cdot dt - \int_0^t \sum_1^s \frac{\partial \Omega}{\partial q_k} \cdot \dot{q}_k dt,$$

which gives by partial integration

$$\delta E = \frac{1}{\vartheta} \int_{-\vartheta}^0 \Omega dt - \Omega_t,$$

where the values for the  $q$ 's to be introduced in  $\Omega$  in the first term are those corresponding to the motion under the influence of the increasing

external field, and the values to be introduced in the second term are those corresponding to the configuration at the time  $t$ . neglecting small quantities of the same order as the square of the external force, however, we may in this expression for  $\delta E$  instead of the values for the  $q$ 's corresponding to the perturbed motion take those corresponding to the original motion of the system. With this approximation the first term is equal to the mean value of the second taken over a period  $\sigma$ , and we have consequently

$$\int_0^\sigma \delta E \, dt = 0. \quad (7)$$

From (6) and (7) it follows that  $I$  will remain constant during the slow establishment of the small external field, if the motion corresponding to a constant value of the field is periodic. If next the external field corresponding to  $\Omega$  is considered as an inherent part of the system, it will be seen in the same way that  $I$  will remain unaltered during the establishment of a new small external field, and so on. Consequently *I will be invariant for any finite transformation of the system which is sufficiently slowly performed*, provided the motion at any moment during the process is periodic and the effect of the variation is calculated on ordinary mechanics.

Before we proceed to the applications of this result we shall mention a simple consequence of (6) for systems for which every orbit is periodic independent of the initial conditions. In that case we may for the varied motion take an undisturbed motion of the system corresponding to slightly different initial conditions. This gives  $\delta E$  constant, and from (6) we get therefore

$$\delta E = \omega \cdot \delta \cdot I, \quad (8)$$

where  $\omega = 1/\sigma$  is the frequency of the motion. This equation forms a simple relation between the variations in  $E$  and  $I$  for periodic systems, which will be often used in the following.

Returning now to systems of one degree of freedom, we shall take our starting point from Planck's original theory of a *linear harmonic vibrator*. according to this theory the stationary states of a system, consisting of a particle executing linear harmonic vibrations with a constant frequency  $\omega_0$  independent of the energy, are given by the well known relation

$$E = n \cdot h \cdot \omega_0, \quad (9)$$

where  $n$  is a positive entire number,  $h$  Planck's constant, and  $E$  the total energy which is supposed to be zero if the particle is at rest.

From (8) it follows at once, that (9) is equivalent to

$$I = \int_0^\sigma p \dot{q} dt = \int p dq = nh, \quad (10)$$

where the latter integral is to be taken over a complete oscillation of  $q$  between its limits. On the principle of the mechanical transformability of the stationary states we shall therefore assume, following Ehrenfest, that (10) holds not only for a Planck's vibrator but for *any periodic system of one degree of freedom* which can be formed in a continuous manner from a linear harmonic vibrator by a gradual variation of the field of force in which the particle moves. This condition is immediately seen to be fulfilled by all such systems in which the motion is of oscillating type i.e. where the moving particle during a period passes twice through any point of its orbit once in each direction. If, however, we confine ourselves to systems of one degree of freedom, it will be seen that system in which the motion is of rotating type, i. e. where the particle during a period passes only through every point of its orbit, cannot be formed in a continuous manner from a linear harmonic vibrator without passing through singular states in which the period becomes infinite long and the result becomes ambiguous . We shall not here enter more closely on this difficulty which has been pointed out by Ehrenfest, because it disappears when we consider systems of several degrees of freedom, where we shall see that a simple generalisation of holds for any system for which every motion is periodic.

As regards the application of (9) to statistical problems it was assumed in Planck's theory that the different states of the vibrator corresponding to different values of  $n$  are *a-priori equally probable*, and this assumption was strongly supported by the agreement obtained on this basis with the measurements of the specific heat of solids at low temperatures. Now it follows from the considerations of Ehrenfest, mentioned in the former section, that the a-priori probability of a given stationary state is not changed by a continuous transformation, and we shall therefore expect that for any system of one degree of freedom the different corresponding to different entire values of  $n$  in (10) are a-priori equally probable.

As pointed out by Planck in connection with the application of (9), it is simply seen that statistical considerations, based on the assumption of equal probability for the different states given by (10), will show the necessary relation to considerations of ordinary statistical mechanics in the limit where the latter theory has been found to give results in agreement with experiments. Let the configuration and motion of a mechanical system be characterised

by  $s$  independent variables  $q_1, \dots, q_s$  and corresponding momenta  $p_1, \dots, p_s$ , and let the state of the system be represent in a  $2s$ -dimensional phase-space by a point with coordinates  $q_1, \dots, q_s, p_1, \dots, p_s$ . Then, according to ordinary statistical mechanics, the probability for this point to lie within a small element in the phase-space is independent of the position and shape of this element and simply proportional to its volume, defined in the usual way by

$$\delta W = \int dq_1 \dots dq_s dp_1 \dots dp_s. \quad (11)$$

In the quantum theory, however, these considerations cannot be directly applied, since the point representing the state of a system cannot be displaced continuously in the  $2s$ -dimensional phase-space, but can lie only on certain surfaces of lower dimensions in this space. For systems of one degree of freedom the phase-space is a two dimensional surface, and the points representing the states of some system given by (10) will be situated on closed curves on this surface. Now, in general, the motion will differ considerably for any two states corresponding to successive entire values of  $n$  in (10), and a simple general connection between the quantum theory and ordinary statistical mechanics is therefore out of question. In the limit, however, where  $n$  is large, the motions in successive states will only differ very little from each other, and it would therefore make little difference whether the points representing the systems are distributed continuously on the phase-surface or situated only on the curves corresponding to (10), provided the number of systems which in the first case are situated between two such curves is equal to the number which in the second case lies on one of these curves. But it will be seen that this condition is just fulfilled in consequence of the above hypothesis of equal a-priori probability of the different stationary states, because the element of phase-surface limited by two successive curves corresponding to (10) is equal to

$$\delta W = \int dp dq = [\int p dq]_n - [\int p dq]_{n-1} = I_n - I_{n-1} = h, \quad (12)$$

so that on ordinary statistical mechanics the probabilities for the point to lie within any two such elements is the same. We see consequently that the hypothesis of equal probability of the different states given by (10) gives the same result as ordinary statistical mechanics in all such applications in which the states of the great majority of the systems correspond to large values of  $n$ . Considerations of this kind have led Debye<sup>13</sup> to point out that

---

<sup>13</sup>P. Debye, Wolfskehl-Vortrag, Göttingen (1913).

condition (10) might have a general validity for systems of one degree of freedom, already before Ehresfest, on the basis of his theory of the mechanical transformability of the stationary states, had shown that this condition forms the only rational generalisation of Planck's condition (9).

We shall now discuss the relation between the theory of *spectra of atomic systems of one degree of freedom*, based on (1) and (10), and the ordinary theory of radiation, and we shall see that this relation in several respects shows a close analogy to the relation, just considered, between the statistical applications of (10) and considerations based on ordinary statistical mechanics. Since the values for the frequency  $\omega$  in two states corresponding to different values of  $n$  in (10) in general are different, we see at once that we cannot expect a simple connection between the frequency by (1) of the radiation corresponding to a transition between two stationary states and the motions of the system in these states, except in the limit where  $n$  is very large, and where the ratio between the frequencies of the motion in successive stationary states differs very little from unity. Consider now a transition between the state corresponding to  $n = n'$  and the state corresponding to  $n = n''$ , and let us assume that  $n'$  and  $n''$  are large numbers and that  $n' - n''$  is small compared with  $n'$  and  $n''$ . In that case we may in (8) for  $\delta E$  put  $E' - E''$  and for  $\delta I$  put  $I' - I''$ , we get therefore from (1) and (10) for the frequency of the radiation emitted or absorbed during the transition between the two states

$$\nu = \frac{1}{h} \cdot (E' - E'') = \frac{\omega}{h} \cdot (I' - I'') = (n' - n'') \cdot \omega. \quad (13)$$

Now in a stationary state of a periodic system the displacement of the particles in any given direction may always be expressed by means of a Fourier-series as a sum of harmonic vibrations:

$$\xi = \sum C_\tau \cos 2\pi \cdot (\tau \cdot \omega \cdot t + c_\tau), \quad (14)$$

where the  $C$ 's and  $c$ 's are constants and the summation is to be extended over all positive entire values of  $\tau$ . On the ordinary theory of radiation we should therefore expect the system to emit a spectrum consisting of a series of lines of frequencies equal to  $\tau\omega$ , but as it is seen, this is just equal to the series of frequencies which we obtain from (13) by introducing different values for  $n' - n''$ . As far as the frequencies are concerned we see therefore that in the limit where  $n$  is large there exists a close relation between the ordinary theory of radiation and the theory spectra based on (1) and (10). It may be noticed, however, that, while on the first theory radiations of the different frequencies  $\tau\omega$  corresponding to different values of  $\tau$  are emitted

or absorbed at the same time, these frequencies will on the present theory, based on the fundamental assumption I and II, be connected with entirely different processes of emission or absorption corresponding to the transition of the system from a given state to different neighbouring stationary states.

In order to obtain the necessary connection, mentioned in the former section, to the ordinary theory of radiation in the limit of slow vibrations, we must further claim that a relation, as that just proved for the frequencies, will, in the limit of large  $n$ , hold also for the intensities of the different lines in the spectrum. Since now on ordinary electrodynamics the intensities of the radiations corresponding to different values of  $\tau$  are directly determined from the coefficients  $C_\tau$  in (14), we must therefore expect that for large values of  $n$  these coefficients will on the quantum theory determine the *probability of spontaneous transition* from a given stationary state for which  $n = n'$  to a neighbouring state for which  $n = n'' = n' - \tau$ . Now, this connection between the amplitudes of the different harmonic vibrations into which the motion can be resolved, characterised by different values of  $\tau$ , and the probabilities of transition from a given stationary state to the different neighbouring stationary states, characterized by different values of  $n' - n''$ , may clearly be expected to be of a general nature. Although, of course, we cannot without a detailed theory of the mechanism of transition obtain an exact calculation of the latter probabilities, unless  $n$  is large, we may expect that also for small values of  $n$  the amplitude of the harmonic vibrations corresponding to a given value of  $\tau$  will in some way give a measure for the probability of a transition between two states for which  $n' - n''$  is equal to  $\tau$ . Thus in general there will be a certain probability of an atomic system in a stationary state to pass spontaneously to any other state of smaller energy, but if for all motions of a given system the coefficients  $C$  in (14) are zero for certain values of  $\tau$ , we are led to expect that no transition will be possible, for which  $n' - n''$  is equal to one of these values.

A simple illustration of these considerations is offered by the linear harmonic vibrator mentioned above in connection with Planck's theory. Since in this case  $C_\tau$  is equal to zero for any  $\tau$  different from 1, we shall expect that for this system only such transitions are possible in which  $n$  alters by one unit. From (1) and (9) we obtain therefore the simple result that the frequency of any radiation emitted or absorbed by a linear harmonic is equal to the constant frequency  $\omega_0$ . This result seems to be supported by observations on the absorption-spectra of diatomic gases, showing that certain strong absorption-lines, which according to general evidence may be ascribed to vibrations of the two atoms in the molecule relative to each other, are not accompanied by lines of the same order of intensity and correspond-

ing to entire multipla of the frequency, such as it should be expected from (1) if the system had any considerable tendency to pass between non-successive states. In this connection it may be noted that the fact, that in the absorption spectra of some diatomic gases faint lines occur corresponding to the double frequency of the main lines<sup>14</sup> obtains a natural explanation by assuming that for finite amplitudes the vibrations are not exactly harmonic and that therefore the molecules possess a small probability of passing also between non-successivestates.

### 3. Conditionally periodic systems

If we consider systems of several degrees of freedom the motion will be periodic only in singular cases and the general conditions which determine the stationary states cannot therefore be derived by means of the same simple kind of considerations as in the former section. As mentioned in the introduction, however, Sommerfeld and others have recently succeeded, by means of a suitable generalisation of (10), to obtain conditions for an important class of systems of several degrees of freedom, which, in connection with (1), have been found to give results in convincing agreement with experimental results about line-spectra. Subsequently these conditions have been proved by Ehrenfest and especially Burgers<sup>15</sup> to be invariant for slow mechanical transformations.

To the generalisation under consideration we are naturally led; if we first consider such systems for which the motions corresponding to the different degrees of freedom are dynamically independent of each other. This occurs if the expression for the total energy  $E$  in Hamilton's equations (4) for a system of  $s$  degrees of freedom can be written as a sum  $E_1 + \dots + E_s$ , where  $E_k$  contains  $q_k$  and  $p_k$  only. An illustration of a system of this kind is presented by a particle moving in a field of force in which the force-components normal to three mutually perpendicular fixed planes are functions of the distances from these planes respectively. Since in such a case the motion corresponding to each degree of freedom in general will be periodic, just as for a system of one degree of freedom, we may obviously expect that the condition (10) is

---

<sup>14</sup>See E.C. Kemble, Phys. Rev. **8** (1916) 701.

<sup>15</sup>J.M. Burgers, Versl. Akad. Amsterdam **25** (1917) 849, 918, 1055; Ann. d. Phys. **52** (1917) 195; Phil. Mag. **33** (1917) 514.

here replaced by a set of  $s$  conditions:

$$I_k = \int p_k \cdot dq_k = n_k \cdot h, \quad (k = 1, \dots, s) \quad (15)$$

where the integrals are taken over a complete period of the different  $q$ 's respectively, and where  $n_1, \dots, n_s$  are entire numbers. It will be seen at once that these conditions are invariant for any slow transformation of the system for which the independency of the motions corresponding to the different coordinates is maintained.

A more general class of systems for which a similar analogy with systems of a single degree of freedom exists and where conditions of the same type as (15) present themselves is obtained in the case where, although the motions corresponding to the different degrees of freedom are not independent of each other, it is possible nevertheless by a suitable choice of coordinates to express each of the momenta  $p_k$  as a function of  $q_k$  only. A simple of this kind consists of a particle moving in a plane orbit in a central field of force. Taking the length of the radius-vector from the centre of the field to the particle as  $q_1$ , and the angular distance of this radius-vector from the centre of the field to the particle as  $q_1$ , and the angular distance of this radius-vector from a fixed line in the plane of the orbit as  $q_2$ , we get at once from (4), since  $E$  does not constant  $q_2$ , the well known result that during the motion the angular momentum  $p_2$  is constant and that the radial motion, given by the variation of  $p_1$  and  $q_1$  with the time, will be exactly the same as for a system of one degree of freedom. In his fundamental application of the quantum theory to the spectrum of a *non-periodic system* Sommerfeld assumed therefore that the stationary states of the above system are given by two conditions of the form:

$$I_1 = \int p_1 \cdot dq_1 = n_1 \cdot h, \quad I_2 = \int p_2 \cdot dq_2 = n_2 \cdot h. \quad (16)$$

While the first integral obviously must be taken over a period of the radial motion, there might at first sight seem to be a difficulty in fixing the limits of integration of  $q_2$ . This disappears, however, if we notice that an integral of the type under consideration will not be altered by a change of coordinates in which  $q$  is replaced by some function of this variable. In fact, if instead of the angular distance of the radius-vector we take for  $q_2$  some continuous periodic function of this angle with period  $2\pi$ , every point in the plane of the orbit will correspond to one set of coordinates only and the relation between  $p$  and  $q$  will be exactly of the same type as for a periodic system of one degree of freedom for which the motion is of oscillating type.

It follows therefore that the integration in the second of the conditions (16) has to be taken over a complete revolution of the radius–vector, and that consequently this condition is equivalent with the simple condition that the angular momentum of the particle round the centre of the field is equal to an entire multiplum of  $h/2\pi$ . As pointed out by Ehrenfest, the conditions (16) are invariant for such special transformations of the system for which the central symmetry is maintained. This follows immediately from the fact that the angular momentum in transformations of this type remains invariant, and that the equations of motion for the radial coordinate as long as  $p_2$  remains constant are the same as for a system of one degree of freedom. On the basis of (16), Sommerfeld has, as mentioned in the introduction, obtained a brilliant explanation of the fine structure of the lines in the hydrogen spectrum, due to the change of the mass of the electron with its velocity.<sup>16</sup> To this theory we shall come back in Part II. As pointed by Epstein<sup>17</sup> and Schwarzschild<sup>18</sup> the central systems considered by Sommerfeld form a special case of a more general class of systems for which conditions of the same type as (15) may be applied. These are the socalled *conditionally periodic systems*, to which we are led if the equations of motion are discussed by means of the Hamilton–Jacobi partial differential equation.<sup>19</sup> In the expression for the total energy  $E$  as a function of the  $q$ 's and the  $p$ 's, let the latter quantities be replaced by the partial differential coefficients of some function  $S$  with respect to the corresponding  $q$ 's respectively, and consider the partial differential equation:

$$E \cdot \left( q_1, \dots, q_s, \frac{\partial S}{\partial q_1}, \dots, \frac{\partial S}{\partial q_s} \right) = \alpha_1, \quad (17)$$

---

<sup>16</sup>In this connection it may be remarked that conditions of the type as (16) were proposed independently by W. Wilson [Phil. Mag. **29** (1915) 795 and **31** (1916) 156] but by him applied only to the simple Keplerian motion described by the electron in the hydrogen atom if the relativity modifications are neglected. Due to the singular position of periodic systems in the quantum theory of systems of several degrees of freedom this application, however, involves, as it will appear from the following discussion, an ambiguity which deprives the result of an immediate physical interpretation. Conditions analogous to (16) have also been established by Planck in his interesting theory of the physical structure of the phase space of systems of several degrees of freedom [Verh. d. D. Phys. Ges. **17** (1915) 407 and 438; Ann. d. Phys. **50** (1916) 385]. This theory, which has no direct relation to the problem of line-spectra discussion in the present paper, rests upon a profound analysis of the geometrical problem of dividing the multiple-dimensional space corresponding to a system of several degrees of freedom into ‘cells’ in a way analogous to the division of the phase surface of a system of one degree of freedom by the curves given by (10).

<sup>17</sup>P. Epstein, loc. cit.

<sup>18</sup>K. Schwarzschild, loc. cit.

<sup>19</sup>See f. inst. C.V.L. Charlier, Die Mechanik des Himmels, Bd. I, Abt. 2.

obtained by putting this expression equal to an arbitrary constant  $\alpha_1$ .

If then

$$S = F \cdot (q_1, \dots, q_s, \alpha_1, \dots, \alpha_s) + C,$$

where  $\alpha_1, \dots, \alpha_s$  and  $C$  are arbitrary constants like  $\alpha_1$ , is a total integral of (17), we get, as shown by Hamilton and Jacobi, the general solution of the equations of motion (4) by putting

$$\frac{\partial S}{\partial \alpha_1} = t + \beta_1, \quad \frac{\partial S}{\partial \alpha_k} = \beta_k, \quad (k = 2, \dots, s) \quad (18)$$

and

$$\frac{\partial S}{\partial q_k} = p_k, \quad (k = 1, \dots, s) \quad (19)$$

where  $t$  is the time and  $\beta_1, \dots, \beta_s$  a new set of arbitrary constants. By means of (18) the  $q$ 's are given as functions of the time  $t$  and the  $2s$  constants  $\alpha_1, \dots, \alpha_s, \beta_1, \dots, \beta_s$  which may be determined for instance from the values of the  $q$ 's and  $\dot{q}$ 's at a given moment.

Now the class of systems, referred to, is that for which, for a suitable choice of orthogonal coordinates, it is possible to find a total integral of (17) of the form

$$S = \sum_1^s S_k \cdot (q_k, \alpha_1, \dots, \alpha_s), \quad (20)$$

where  $S_k$  is a function of the  $s$  constants  $\alpha_1, \dots, \alpha_s$  and of  $q_k$  only. In this case, in which the equation (17) allows of which is called ‘separation of variables’, we get from (19) that every  $p$  is a function of the  $\alpha$ 's and of the corresponding  $q$  only. If during the motion the coordinates do not become infinite in the course of time or converge to fixed limits, every  $q$  will, just as for systems of one degree of freedom, oscillate between two fixed values, different for the different  $q$ 's and depending on the  $\alpha$ 's. Like in the case of a system of one degree of freedom,  $p_k$  will become zero and change its sign whenever  $q_k$  passes through one of these limits. Apart from special cases, the system will during the motion never pass twice through a configuration corresponding to the same set of values for the  $q$ 's and  $p$ 's, but it will in the course of time pass within any given, however small, distance from any configuration corresponding to a given set of values  $q_1, \dots, q_s$ , representing a point within a certain closed  $s$ -dimensional extension limits by  $s$  pairs of  $(s-1)$ -dimensional surface corresponding to constant values of the  $q$ 's equal to the above mentioned limits of oscillation. A motion of this kind is called ‘conditionally periodic’. It will be seen that the character of the motion will depend only on the  $\alpha$ 's and not on the  $\beta$ 's, which latter constants serve

only to fix the exact configuration of the system at a given moment, when the  $\alpha$ 's are known. For special systems it may occur that the orbit will not cover the above mentioned  $s$ -dimensional extension everywhere dense, but will, for all values of the  $\alpha$ 's, be confined to an extension of less dimensions. Such a case we will refer to in the following as a case of 'degeneration'.

Since for a conditionally periodic system which allows of separation in the variables  $q_1, \dots, q_s$  the  $p$ 's are functions of the corresponding  $q$ 's only, we may, just as in the case of independent degrees of freedom or in the case of quasiperiodic motion in a certain field, from a set of expressions of the type

$$I_k = \int p_k \cdot (q_k, \alpha_1, \dots, \alpha_s) \cdot dq_k, \quad (k = 1, \dots, s) \quad (21)$$

where the integration is taken over a complete oscillation of  $q_k$ . As, in general, the orbit will cover everywhere dense an  $s$ -dimensional extension limited in the characteristic way mentioned above, it follows that, except in cases of degeneration, a separation of variables will not be possible for two different sets of coordinates  $q_1, \dots, q_s$  and  $q'_1, \dots, q'_s$ , unless  $q_1 = f_1(q'_1), \dots, q_s = f_s(q'_s)$ , and since a change of coordinates of this type will not affect the values of the expressions (21), it will be seen that the values of the  $I$ 's are completely determined for a given motion of the system. By putting

$$I_k = n_k \cdot h, \quad (k = 1, \dots, s) \quad (22)$$

where  $n_1, \dots, n_s$  are positive entire numbers, we obtain therefore *a set of conditions which form a natural generalisation of condition (10)* holding for a system of one degree of freedom.

Since the  $I$ 's, as given by (21), depend on the constants  $\alpha_1, \dots, \alpha_s$  only and not on the  $\beta$ 's, the  $\alpha$ 's may, in general, inversely be determined from the values of the  $I$ 's. The character of the motion will therefore, in general, be completely determined by the conditions (22), and especially the value for the total energy, which according to (17) is equal to  $\alpha_1$ , will be fixed by them. In the cases of degeneration referred to above, however, the conditions (22) involve an ambiguity, since in general for such systems there will exist an infinite number of different sets of coordinates which allow of a separation of variables, and which will lead to different motions in the stationary states, when these conditions are applied. As we shall see below, this ambiguity will not influence the fixation of the total energy in the stationary states, which is the essential factor in the theory of spectra based on (1) and in the applications of the quantum theory to statistical problems.

A well known characteristic example of a conditionally periodic system is afforded by a particle moving under the influence of the attractions from

two fixed centres varying as the inverse squares of the distance apart, if the relativity modifications are neglected. As shown by Jacobi this problem can be solved by a separation of variables if so called elliptical coordinates are used, i.e. if for  $q_1$  and  $q_2$  we take two parameters characterising respectively an ellipsoid and a hyperboloid of revolution with the centres as foci and passing through the instantaneous position of the moving particle, and for  $q_3$  we take the angle between the plane through the particle and the centres and a fixed plane through the latter points, or, in closer conformity with the above general description, some continuous periodic function of this angle period  $2\pi$ . A limiting case of this problem is afforded by an electron rotating a positive nucleus and subject to the effect of an additional homogeneous electric field, because this field may be considered as arising from a second nucleus at infinite distance apart from the first. The motion in this case will therefore be conditionally periodic and allow a separation of variables in parabolic coordinates, if the nucleus is taken as focus for both sets of paraboloids of revolution, and their axes are taken parallel to the direction of the electric force. By applying the conditions (22) to this motion Epstein and Schwarzschild have, as mentioned in the introduction, independent of each other, obtained an explanation of the effect of an external electric field on the lines of the hydrogen spectrum, which was found to be convincing agreement with Stark's measurements. To the results of these calculations we shall return in Part II.

In the above way of representing the general theory we have followed the same procedure as used by Epstein. By introducing the so called 'angle variables' well known from the astronomical theory of perturbations, Schwarzschild has given the theory a very elegant form in which the analogy with systems of one degree of freedom presents itself in a somewhat different manner. The connection between this treatment and that given above has been discussed in detail by Epstein.<sup>20</sup>

As mentioned above the conditions (22), first established from analogy with systems of one degree of freedom, have subsequently been proved generally to be *mechanically invariant for any slow transformation for which the system remains conditionally periodic*. The proof of this invariance has been given quite recently by Burgers<sup>21</sup> by means of an interesting application of the theory of contact-transformations based on Schwarzschild's introduction of angle variables. We shall not enter here on these calculations but shall

<sup>20</sup>P. Epstein, Ann. d. Phys. **51** (1916) 168. See also Note on page 33 of the present paper.

<sup>21</sup>J.M. Burgers, loc. cit. Versl. Akad. Amsterdam **25** (1917) 1055.

only consider some points in connection with the problem of the mechanical transformability of the stationary states which are of importance for the logical consistency of the general theory and for the later applications. In § 2 we saw that in the proof of the mechanical invariance of relation (10) for a periodic system of one degree of freedom, it was essential that the comparative variation of the external conditions during the time of one period could be made small. This may be regarded as an immediate consequence of the nature of the fixation of the stationary states in the quantum theory. In fact the answer to the question whether a given state of a system is stationary, will not depend only on the motion of the particles at a given moment or on the field of force in the immediate neighbourhood of their instantaneous positions, but cannot be given before the particles have passed through a complete cycle of states, and so to speak have got to know the entire field of force of influence on the motion. If thus, in the case of a periodic system of one degree of freedom, the field of force is varied by a given amount, and if its comparative variation within the time of a single period was not small, the particle would obviously have no means to get to know the nature of the variation of the field and to adjust its stationary motion to it, before the new field was already established. For exactly the same reasons it is necessary condition for the mechanical invariance of the stationary states of a conditionally periodic system, that the alteration of the external conditions during an interval in which the system has passed approximately through all possible configurations within the above mentioned  $s$ -dimensional extension in the coordinate-space can be made as small as we like. This condition forms therefore also an essential point in Burgers' proof of the invariance of the conditions (22) for mechanical transformations. Due to this we meet with a characteristic difficulty when during the transformation of the system we pass one of the cases of degeneration mentioned above, where, for every set of values for the  $\alpha$ 's, the orbit will not cover the  $s$ -dimensional extension everywhere dense, but will be confined to an extension of less dimensions. It is clear that, when by a slow transformation of a conditionally periodic system we approach a degenerate system of this kind, the time-interval which the orbit takes to pass to any possible configuration will tend to be very long and will become infinite when the degenerate system is reached. As a consequence of this *the conditions (22) will generally not remain mechanically invariant when we pass a degenerate system*, what has intimate connection with the above mentioned ambiguity in the determination of the stationary states of such systems by means of (22).

A typical case of a degenerate system, which may serve as an illustration of this point, is formed by system of several degrees of freedom for which

every motion is simply periodic, independent of the initial conditions. In this case, which is of great importance in the physical applications, we have from (5) and (21), for any set of coordinates in which a separation of variables is possible,

$$I = \int_0^\sigma (p_1 \dot{q}_1 + \dots + p_s \dot{q}_s) dt = \kappa I_1 + \dots + \kappa_s I_s, \quad (23)$$

where the integration is extended over one period of the motion, and where  $\kappa_1, \dots, \kappa_s$  are a set of positive entire numbers without a common divisor. Now we shall expect that every motion, for which it is possible to find a set of coordinates in which it satisfied (22), will be stationary.

For any such motion we get from (23)

$$I = (\kappa_1 \cdot n_1 + \dots + \kappa_s \cdot n_s) \cdot h = nh, \quad (24)$$

where  $n$  is a whole number which may take all positive values if, as in the applications mentioned below, at least one of the  $\kappa$ 's is equal to one. Inversely, if the system under consideration allows of separation of variables in an infinite continuous multitude of sets of coordinates, we must conclude that generally every motion which satisfies (24) will be stationary, because in general it will be possible for any such motion to find a set of coordinates in which it satisfied also (22). It will thus be seen that, for a periodic system of several degrees of freedom, condition (24) forms a simple generalisation of condition (1)). From relation (8), which holds for two neighbouring motions of any periodic system, it follows further that the energy of the system will be completely determined by the value of  $I$ , just as for systems of one degree of freedom.

Consider now a periodic system in some stationary state satisfying (24), and let us assume that an external field is slowly established at a continuous rate and that the motion at any moment during this process allows of a separation of variables in a certain set of coordinates. If we would assume that the effect of the field on the motion of the system at any moment could be calculated directly by means of ordinary mechanics, we would find that the values of the  $I$ 's with respect to the latter coordinates would remain constant during the process, but this would involve that the values of the  $n$ 's in (22) would in general not be entire numbers, but would depend entirely on the accidental motion, satisfying (24), originally possessed by the system. That mechanics, however, cannot generally be applied directly to determine the motion of a periodic system under influence of an increasing external field, is just what we should expect according to the singular position of degenerate system as regards mechanical transformations. In fact,

in the presence of a small external field, the motion of a periodic system will undergo slow variations as regards the shape and position of the orbit, and if the perturbed motion is conditionally periodic these variations will be of a periodic nature. Formally, we may therefore compare a periodic system exposed to an external field with a simple mechanical system of one degree of freedom in which the particle performs a slow oscillating motion. Now the frequency of a slow variation of the orbit will be seen to be proportional to the intensity of the external field, and it is therefore obviously impossible to establish the external field at a rate so slow that the comparative change of its intensity during a period of this variation is small. The process which takes place during the increase of the field will thus be analogous to that which takes place if an oscillating particle is subject to the effect of external forces which change considerably during a period. Just as the latter generally will give rise to emission or absorption of radiation and cannot be described by means of ordinary mechanics, we must expect that the motion of a periodic system of several degrees of freedom under the establishment of the external field cannot be determined by ordinary mechanics, but that the field will give rise to effects of the same kind as those which occur during a transition between two stationary states accompanied by emission or absorption of radiation. Consequently we shall expect that, during the establishment of the field, *the system will in general adjust itself in some unmechanical way* until a stationary state is reached in which the frequency (or frequencies) of the above mentioned slow variation of the orbit has a relation to the additional energy of the system due to the presence of the external field, which is of the same kind as the relation, expressed by (8) and (10), between the energy and frequency of a periodic system of one degree of freedom. As it will be shown in Part II in connection with the physical applications, this condition is just secured if the stationary states in the presence of the field are determined by the conditions (22), and it will be seen that these considerations offer a means of fixing the stationary states of a perturbed periodic system also in cases where no separation of variables can be obtained.

In consequence of the singular position of the degenerate systems in the general theory of stationary states of conditionally periodic systems, we obtain a means of *connecting mechanically two different stationary states of a given system* through a continuous series of stationary states without passing through systems in which the forces are very small and the energies in all the stationary states tend to coincide (comp. page 9). In fact, if we consider a given conditionally periodic system which can be transformed in a continuous way into a system for which every orbit is periodic and for

which every state satisfying (24) will also satisfy (22) for a suitable choice of coordinates, it is clear in the first place that it is possible to pass in a mechanical way through a continuous series of stationary states from a state corresponding to a given set of values of the  $n$ 's in (22) to any other such state for which  $\kappa_1 n_1 + \dots + \kappa_s n_s$  possesses the same value. If, moreover, there exists a second periodic system of the same character to which the first periodic system can be transformed continuously, but for which the set of  $\kappa$ 's is different, it will be possible in general by a suitable cyclic transformation to pass in a mechanical way between any two stationary states of the given conditionally periodic system satisfying (22).

To obtain an example of such a cyclic transformation let us take the system consisting of an electron which moves round a fixed positive nucleus exerting an attraction varying as the inverse square of the distance. If we neglect the small relativity corrections, every orbit will be periodic independent of the initial conditions and the system will allow of separation of variables in polar coordinates as well as in any set of elliptical coordinates, of the kind mentioned on page 23, if the nucleus is taken as one of the foci. It is simply seen that any orbit which satisfies (24) for a value of  $n > 1$ , will satisfy (22) for a suitable choice of elliptical coordinates. By imagining another nucleus of infinite small charge placed at the other focus, the orbit may further be transformed into another which satisfies (24) for the same value of  $n$  but which may have any given value for the eccentricity. Consider now a state of the system satisfying (24), and let us assume that by the above means the orbit is originally so adjusted that in plane polar coordinates it will correspond to  $n_1 = m$  and  $n_2 = n - m$  in (16). Let then the system undergo a slow continuous transformation during which the field of force acting on the electron remains central, but by which the law of attraction is slowly varied until the force is directly proportional to the distance apart. In the final state, as well as in the original state, the orbit of the electron will be closed, but during the transformation the orbit will not be closed, and the ratio between the mean period of revolution and the period of the radial motion, which in the original motion was equal to one, will during the transformation increase continuously until in the final state it is equal to two. This means that, using polar coordinates, the values of  $\kappa_1$  and  $\kappa_2$  in (22) which for the first state are equal to  $\kappa_1 - \kappa_2 = 1$ , will be for the second state  $\kappa_1 = 2$  and  $\kappa_2 = 1$ . Since during the transformation  $n_1$  and  $n_2$  will keep their values, we get therefore in the final state  $I = h \cdot (2m + (n - m)) = h \cdot (n + m)$ . Now in the latter state, the system allows a separation of variables not only in polar coordinates but also in any system of rectangular Cartesian coordinates, and by suitable choice of the direction of the axes, we can obtain that any orbit, satisfying (24) for a value of  $n > 1$ , will also satisfy (22). By an infinite small change of the force components in the axes, in such a way that the motions of these directions remain independent of each other

but possess slightly different periods, it will further be possible to transform the elliptical orbit mechanically into one corresponding to any ratio between the axes. Let us now assume that in this way the orbit of the electron is transformed into a circular one, so that, returning to plane polar coordinates, we have  $n_1 = 0$  and  $n_2 = n + m$ , and let then by a slow transformation the law of attraction be varied until again it is that of the inverse square. It will be seen that when this state is reached the motion will again satisfy (24), but this time we will have  $I = h \cdot (n+m)$  instead of  $I = nh$  as in the original state. By repeating a cyclic process of this kind we may pass from any stationary state of the system in question which satisfies (24) for a value of  $n > 1$  to any other such state without leaving at any moment the region of stationary states.

The theory of the mechanical transformability of the stationary states gives us a means to discuss the question of the *a-priori probability* of the different states of a conditionally periodic system, characterised by different sets of values for the  $n$ 's in (22). In fact from the considerations, mentioned in § 1, it follows that, if the a-priori probability of the stationary states of a given system is known, it is possible at once to deduce the probabilities for the stationary states of any other system to which the first system can be transformed continuously without passing through a system of degeneration. Now from the analogy with systems of one degree of freedom it seems necessary to assume that, for a system of several degrees of freedom for which the motions corresponding to the different coordinates are dynamically independent of each other, the a-priori probability is the same for all states corresponding to different sets of  $n$ 's in (15). According to the above we shall therefore assume that the a-priori probability is the same for all states, given by (22), of a by which can be formed in a continuous way from a system of this kind without passing through systems of degeneration. It will be observed that on this assumption we obtain exactly the same relation to the ordinary theory of statistical mechanics in the limit of large  $n$ 's as obtained in the case of systems of one degree of freedom. Thus, for a conditionally periodic system, the volume given by (11) of the element of phase-space, including all points  $q_1, \dots, q_s, p_1, \dots, p_s$  which represent states for which the value of  $I_k$  given by (21) lies between  $I_k$  and  $I_k + \delta I_k$  it seen at once to be equal to<sup>22</sup>

$$\delta W = \delta I_1 \delta I_2 \dots \delta I_s, \quad (25)$$

if the coordinates are so chosen that the motion corresponding to every

---

<sup>22</sup>Comp. A. Sommerfeld, Ber. Akad. München, 1917, p. 83.

degree of freedom is of oscillating type. The volume of the phase-space limited by  $s$  pairs of surfaces, corresponding to successive values for the  $n$ 's in the conditions (22), will therefore be equal to  $h^s$  and consequently be the same for every combination of the  $n$ 's. In the limit where the  $n$ 's are large numbers and the stationary states corresponding to successive values for the  $n$ 's differ only very little from each other, we thus obtain the same result on the assumption of equal *a-priori* probability of all the stationary states, corresponding to different sets of values of  $n_1, n_2, \dots, n_s$  in (22), as would be obtained by application of ordinary statistical mechanics.

The fact that the last considerations hold for every non-degenerate conditionally periodic system suggests the assumption that in general *a-priori probability will be the same for all the states determined by (22)*, even if it should not be possible to transform the given system into a system of independent degrees of freedom without passing through degenerate systems. This assumption will be shown to be supported by the consideration of the intensities of the different components of the Stark-effect of the hydrogen lines, mentioned in the next Part. When we consider a degenerate system, however, we cannot assume that the different stationary states are *a-priori* equally probable. In such a case the stationary states will be characterized by a number of conditions less than the number of degrees of freedom, and the probability of a given state must be determined from the number of different stationary states of some non-degenerate system which will coincide in the given state, if the latter system is continuously transformed into the degenerate system under consideration.

In order to illustrate this, let us take the simple case of a degenerate system formed by an electrified particle moving in the plane orbit in a central field, the stationary states of which are given by the two conditions (16). In this case the plane of the orbit is undetermined, and it follows already from a comparison with ordinary statistical mechanics, that the *a-priori* probability of the states characterized by different combinations of  $n_1$  and  $n_2$  in (16) cannot be the same. Thus the volume of the phase-space, corresponding to states for which  $I_1$  lies between  $I_1$  and  $I_1 + \delta I_1$  and for which  $I_2$  lies between  $I_2$  and  $I_2 + \delta I_2$ , is found by a simple calculation<sup>23</sup> to be equal to  $\delta W = 2I_2\delta I_1\delta I_2$ , if the motion is described by ordinary polar coordinates. For large values of  $n_1$  and  $n_2$ , we must therefore expect that the *a-priory* probability of a stationary state corresponding to a given combination  $(n_1, n_2)$  is proportional to  $n_2$ . The question of the *a-priori* probability of states corresponding to small values of the  $n$ 's has been dis-

---

<sup>23</sup>See A. Sommerfeld, loc. cit.

cussed by Sommerfeld in connection with the problem of the intensities of the different components in the fine structure of the hydrogen lines (see Part II). From considerations about the volume of the extensions in the phase-space, which might be considered as associated with the states characterised by different combinations  $(n_1, n_2)$ , Sommerfeld proposes several different expressions for the a-priori probability of such states. Due to the necessary arbitrariness involved in the choice of these extensions, however, we cannot in this way obtain a rational determination of the a-priori probability of states corresponding to small values of  $n_1$  and  $n_2$ . On the other hand, this probability may be deduced by regarding the motion of the system under consideration as the degeneration of a motion characterised by three numbers  $n_1, n_2$  and  $n_3$ , as in the general applications of the conditions (22) to a system of three degrees of freedom. Such a motion may be obtained for instance by imagining the system placed in a small homogeneous magnetic field. In certain respects this case falls outside the general theory of conditionally periodic system discussed in this section, but, as we shall see in Part II, it can be simply shown that the presence of the magnetic field imposes the further condition on the motion in the stationary states that the angular momentum round the axis of the field is equal to  $n'h/2\pi$ , where,  $n'$  is a positive entire number equal to or less than  $n_2$ , and which for the system considered in the spectral problems must be assumed to be different from zero. When regard is taken to the two opposite directions in which the particle may rotate round the axis of the field, we see therefore that for this system a state corresponding to a given combination of  $n_1$  and  $n_2$  in the presence of the field can be established in  $2n_2$  different ways. The a-priori probability of the different states of the system may consequently for all combinations of  $n_1$  and  $n_2$  be assumed to be proportional to  $n_2$ .

The assumption just mentioned that the angular momentum round the axis of the field cannot be equal to zero is deduced from considerations of system for which the motion corresponding to special combinations of the  $n$ 's in (22) would become physically impossible due to some singularity in its character. In such cases we must assume that no stationary states exist corresponding to the combinations  $(n_1, n_2, \dots, n_s)$  under consideration, and on the above principle of the invariance of the a-priori probability for continuous transformations we shall accordingly expect that the a-priori probability of any other state, which can be transformed continuously into one of these states without passing through cases of degeneration, will also be equal to zero.

Let us now proceed to consider the *spectrum of a conditionally periodic system*, calculated from the values of the energy in the stationary states by

means of relation (1). If  $E \cdot (n_1, \dots, n_s)$  is the total energy of a stationary state determined by (22) and if  $\nu$  is the frequency of the line corresponding to the transition between two stationary states characterised by  $n_k = n'_k$  and  $n_k = n''_k$  respectively, we have

$$\nu = \frac{1}{\hbar} \cdot [E \cdot (n'_1, \dots, n'_s) - E \cdot (n''_1, \dots, n''_s)]. \quad (26)$$

In general, this spectrum will be entirely different from the spectrum to be expected on the ordinary theory of electrodynamics from the motion of the system. Just as for a system of one degree of freedom we shall see, however, that in the limit where the motions in neighbouring stationary states differ very little from each other, there exists a close relation between the spectrum calculated on the quantum theory and that to be expected on ordinary electrodynamics. As in § 2 we shall further see, that this connection leads to certain general considerations about the probability of transition between any two stationary states and about the nature of the accompanying radiation which are found to be supposed by observations. In order to discuss this question we shall first deduce a general expression for the energy difference between neighbouring of a conditionally periodic system, which can be simple obtained by a calculation analogous to that used in § 2 in the deduction of the relation (8).

Consider some motion of a conditionally periodic system which allows of separation of variables in a certain set of coordinates  $q_1, \dots, q_s$  and let us assume that at the time  $t = \vartheta$  the configuration of the system will to a close approximation be the same as at the time  $t = 0$ . By taking  $\vartheta$  large enough we can make this approximation as close as we like. If next we consider some conditionally periodic motion, obtained by a small variation of the first motion, and which allows of separation of variables in a set of coordinates  $q'_1, \dots, q'_s$  which may differ slightly from the set  $q_1, \dots, q_s$ , we get by means of Hamilton's equations (4), using the coordinated  $q'_1, \dots, q'_s$ ,

$$\begin{aligned} \int_0^\vartheta \delta E \, dt &= \int_0^\vartheta \sum_1^s \left( \frac{\partial E}{\partial p'_k} \cdot \delta p'_k + \frac{\partial E}{\partial q'_k} \cdot \delta q'_k \right) \, dt = \\ &= \int_0^\vartheta \sum_1^s (\dot{q}_k \cdot \delta p'_k - p'_k \cdot \delta \dot{q}'_k) \, dt. \end{aligned}$$

By partial integration of the second term in the bracket this gives:

$$\int_0^\vartheta \delta E \, dt = \int_0^\vartheta \sum_1^s \delta (p'_k \cdot \dot{q}'_k) \cdot dt - \left| \sum_1^s p'_k \cdot \delta q'_k \right|_{t=0}^{t=\vartheta}. \quad (27)$$

Now we have for the unvaried motion

$$\int_0^\vartheta \sum_1^s p'_k \cdot \dot{q}_k dt = \int_0^\vartheta \sum_1^s p_k \dot{q}_k dt = \sum_1^s N_k \cdot I_k,$$

where  $I_k$  is defined by (21) and where  $N_k$  is the number of oscillations performed by  $q_k$  in the time interval  $\vartheta$ . For the varied motion we have on the other hand:

$$\int_0^\vartheta \sum_1^s p'_k \dot{q}'_k dt = \int_{t=0}^{t=\vartheta} \sum_1^s p'_k \cdot dq'_k = \sum_1^s N_k \cdot I'_k + \left| \sum_1^s p'_k \cdot \delta q'_k \right|_{t=0}^{t=\vartheta},$$

where the  $I$ 's correspond to the conditionally periodic motion in the coordinates  $q'_1, \dots, q'_s$ , and the  $\delta q'$ 's which enter in the last term are the same as those in (27). Writing  $I'_k - I_k = \delta \cdot I_k$ , we get therefore from the latter equation

$$\int_0^\vartheta \delta E dt = \sum_1^s N_k \cdot \delta I_k. \quad (28)$$

In the special case where the varied motion is an undisturbed motion belonging to the same system as the unvaried motion we get, since  $\delta \cdot E$  will be constant,

$$\delta E = \sum_1^s \omega_k \cdot \delta I_k, \quad (29)$$

where  $\omega_k = N_k / \vartheta$  is the mean frequency of oscillation of  $q_k$  between its limits, taken over a long time interval of the same order of magnitude as  $\vartheta$ . This equation forms a simple generalisation of (8), and in the general case in which a separation of variables will be possible only for one system of coordinates leading to a complete definition of the  $I$ 's it might have been deduced directly from the analytical theory of the periodicity properties of the motion of a conditionally periodic system, based on the introduction of angle variables.<sup>24</sup> From (29) it follows moreover that, if the system allows of a separation of variables in an infinite continuous of sets of coordinates, the

---

<sup>24</sup>See Charlier, die Mechanik des Himmels, bd. I abt. 2, and especially P. Epstein, Ann. d. Phys. 51 (1916) 178. By means of the well known theorem of Jacobi about the change of variables in the canonical equations of Hamilton, the connection between the notion of angle-variables and the quantities  $I$ , discussed by Epstein in the latter paper, may be briefly exposed in the following elegant manner which has been kindly pointed out to me by Mr. H.A. Kramers. Consider the function  $S(q_1, \dots, q_s, I_1, \dots, I_s)$  obtained from (20) by introducing for the  $\alpha$ 's their expressions in terms of the  $I$ 's given by the equations

(21). This function will be a many valued function of the  $q$ 's which increases by  $I_k$  if  $q_k$  described one oscillation between its limits and comes back to its original value while the other  $q$ 's remain constant. If we therefore introduce a new set of variables  $w_1, \dots, w_s$  defined by

$$w_k = \frac{\partial S}{\partial I_k}, \quad (k = 1, \dots, s) \quad (1^*)$$

it will be seen that  $w_k$  increases by one unit while the other  $w$ 's will come back to their original values if  $q_k$  described one oscillation between its limits and the other  $q$ 's remain constant. Inversely it will therefore be seen that the  $q$ 's, and also the  $p$ 's which were given by

$$p_k = \frac{\partial S}{\partial q_k}, \quad (k = 1, \dots, s), \quad (2^*)$$

when considered as functions of the  $I$ 's and  $w$ 's will be periodic functions of every of the  $w$ 's with period. according to Fourier's theorem any of the  $q$ 's may therefore be represented by an  $s$ -double trigonometric series of the form

$$q = \sum A_{\tau_1, \dots, \tau_s} \cos 2\pi \cdot (\tau_1 \cdot w_1 + \dots + \tau_s \cdot w_s - \alpha_{\tau_1, \dots, \tau_s}), \quad (3^*)$$

where the  $A$ 's and  $\alpha$ 's are constants depending on the  $I$ 's and where the summation is to be extended over all entire values of  $\tau_1, \dots, \tau_s$ . On account of this property of the  $w$ 's, the quantities  $2\pi w_1, \dots, 2\pi w_s$  are denoted as 'angle variables'. Now from (1<sup>\*</sup>) and (2<sup>\*</sup>) it follows according to the above mentioned theorem of Jacobi (see for instance Jacobi, Vorlesungen über Dynamik § 37) that the variations with the time of the  $I$ 's and  $w$ 's will be given by

$$\frac{dI_k}{dt} = -\frac{\partial E}{\partial w_k}, \quad \frac{dw_k}{dt} = \frac{\partial E}{\partial I_k}, \quad (k = 1, \dots, s) \quad (4^*)$$

where the energy  $E$  is considered as a function of the  $I$ 's and  $w$ 's. Since  $E$ , however, is determined by the  $I$ 's only we get from (4<sup>\*</sup>), besides the evident result that the  $I$ 's are constant during the motion, that the  $w$ 's will vary linearly with the time and can be represented by

$$\omega_k = \omega_k t + \delta_k, \quad \omega_k = \frac{\partial E}{\partial I_k}, \quad (k = 1, \dots, s) \quad (5^*)$$

where  $\delta_k$  is a constant, and where  $w_k$  is easily seen to be equal to the mean frequency of oscillation of  $q_k$ . From (5<sup>\*</sup>) eq. (28) follows at once, and it will further be seen that by introducing (5<sup>\*</sup>) in (3<sup>\*</sup>) we get the result that every of the  $q$ 's, and consequently also any one-valued function of the  $q$ 's, can be represented by an expression of the type (31).

In this connection it may be mentioned that the method of Schwarzschild of fixing the stationary states of a conditionally periodic system, mentioned on page 117, consists in seeking for a given system a set of canonically conjugated variables  $Q_1, \dots, Q_s, P_1, \dots, P_s$  in such a way that the positional coordinates of the system  $q_1, \dots, q_s$  and their conjugated momenta  $p_1, \dots, p_s$ , when considered as functions of the  $Q$ 's and  $P$ 's, are periodic in every of the  $Q$ 's with period  $2\pi$ , while the energy of the system depends only on the  $P$ 's. In analogy with the condition which fixes the angular momentum in Sommerfeld's theory of central systems Schwarzschild next puts every of the  $P$ 's equal to an entire multiplum of  $h/2\pi$ . In contrast to the theory of stationary states of conditionally periodic systems based on the possibility of separation of variables and the fixation of the  $I$ 's by (22), this method does not lead to an absolute fixation of the stationary states, because, as pointed out by Schwarzschild himself, the above definition of the  $P$ 's leaves an arbitrary constant undermined in every of these quantities. In many cases, however, these constants may be

total energy will be the same for all motions corresponding to the same values of the  $I$ 's, independent of the special set of coordinates used to calculate these quantities. as mentioned above and as we have already shown in the case of purely periodic systems by means of (8), the total energy is therefore also in cases of degeneration completely determined by the conditions (22).

Consider now a transition between two stationary states determined by (22) by putting  $n_k = n'_k$  and  $n_k = n''_k$  respectively, and let us assume that  $n'_1, \dots, n'_s, n''_1, \dots, n''_s$  are large numbers, and that the differences  $n'_k - n''_k$  are small compared with these numbers. Since the motions of the system in these states will differ relatively very little from each other we may calculate the difference of the energy by means of (29), and we get therefore, by means of (1), for the frequency of the radiation corresponding to the transition between the two states

$$\nu = \frac{1}{h} \cdot (E' - E'') = \frac{1}{h} \sum_1^s \omega_k \cdot (I'_k - I''_k) = \sum_1^s \omega_k \cdot (n'_k - n''_k), \quad (30)$$

which is seen to be a direct generalisation of the expression (13) in § 2.

Now, in complete analogy to what is the case for periodic systems of one degree of freedom, it is proved in the analytical theory of the motion of conditionally periodic mentioned above that for the latter systems the coordinates  $q_1, \dots, q_s$ , and consequently also the displacements of the particles in any given direction, may be expressed as a function of the time by an  $s$ -double infinite Fourier series of the form:

$$\xi = \sum C_{\tau_1, \dots, \tau_s} \cdot \cos 2\pi \{(\tau_1 \cdot \omega_1 + \dots + \tau_s \cdot \omega_s) \cdot t + c_{\tau_1, \dots, \tau_s}\}, \quad (31)$$

where the summation is to be extended over all positive and negative entire values of the  $\tau$ 's, and where the  $\omega$ 's are the above mentioned mean frequencies of oscillation for the different  $q$ 's. The constants  $C_{\tau_1, \dots, \tau_s}$  depend only on the  $\alpha$ 's in the equations (18) or, what is the same, on the  $I$ 's, while the constants  $c_{\tau_1, \dots, \tau_s}$  depend on the  $\alpha$ 's as well as on the  $\beta$ 's. In general the quantities  $\tau_1 \omega_1 + \dots + \tau_s \omega_s$  will be different for any two different sets of values for the  $\tau$ 's, and in the course of time the orbit will cover everywhere dense a

---

simple determined from considerations of mechanical transformability of the stationary states, and as pointed out by Burgers [loc. cit. Versl. Akad. Amsterdam 25 (1917) 1055] Schwaszchild method possesses on the other hand the essential advantage of being applicable to certain classes of system in which the displacements of the particles may be represented by trigonometric series of the type (31), but for which the equations of motion cannot be solved by separation of variables in any fixed set of coordinates. An interesting application of this to the spectrum of rotating molecules, given by Burgers, will be mentioned in Part IV.

certain  $s$ -dimensional extension. In a case of degeneration, however, where the orbit will be confined to an extension of less dimensions, there will exist for all values of the  $\alpha$ 's one or more relations of the type  $m_1\omega_1 + \dots + m_s\omega_s = 0$  where the  $m$ 's are entire numbers and by the introduction of which the expression (31) can be reduced to a Fourier series which is less than  $s$ -double infinite. Thus in the special case of a system of which every orbit is periodic we have  $\omega_1/\kappa_1 = \dots = \omega_s/\kappa_s = \omega$ , where the  $\kappa$ 's are the numbers which enter in eq. (23), and the Fourier series for the displacements in the different directions will in this case consist only of terms of the simple form  $C_\tau \cdot \cos 2\pi \cdot \{\tau \cdot \omega \cdot t + c_\tau\}$ , just as for a system of one degree of freedom.

On the ordinary theory of radiation, we should expect from (31) that the spectrum emitted by the system in a given state would consist of an  $s$ -double infinite series of lines of frequencies equal to  $\tau_1\omega_1 + \dots + \tau_s\omega_s$ . In general, this spectrum would be completely different from that given by (26). This follows already from the fact that the  $\omega$ 's will depend on the values for the constants  $\alpha_1, \dots, \alpha_s$  and will vary in a continuous way for the continuous multitude of mechanically possible states corresponding to different sets of values for these constants. Thus in general the  $\omega$ 's will be quite different for two different stationary states corresponding to different sets of  $n$ 's in (22), and we cannot expect any close relation between the spectrum calculated on the quantum theory and that to be expected on the ordinary theory of mechanics and electrodynamics. In the limit, however, where the  $n$ 's in (22) are large numbers, the ratio between the  $\omega$ 's for two stationary states, corresponding to  $n_k = n'_k$  and  $n_k = n''_k$  respectively, will tend to unity if the differences  $n'_k = n''_k$  are small compared with the  $n$ 's, and as seen from (30) the spectrum calculated by (1) and (22) will in this limit just tend to coincide with that to be expected on the ordinary theory of radiation from the motion of the system.

As far as the frequencies are concerned, we thus see that for conditionally periodic systems there exists a connection between the quantum theory and the ordinary theory of radiation of exactly the same character as that shown in § 2 to exist in the simple case of periodic systems of one degree of freedom. Now on ordinary electrodynamics the coefficients  $C_{\tau_1, \dots, \tau_s}$  in the expression (31) for the displacements of the particles in the different directions would in the well known determine the intensity and polarisation of the emitted radiation of the corresponding frequency  $\tau_1\omega_1 + \dots + \tau_s\omega_s$ . As for systems of one degree of freedom we must therefore conclude that, in the limit of large values for the  $n$ 's, the probability of spontaneous transition between two stationary states of a conditionally periodic system, as well as the polarisation of the accompanying radiation, can be determined directly

from the values of the coefficient  $C_{\tau_1, \dots, \tau_s}$  in (31) corresponding to a set of  $\tau$ 's given by  $\tau_k = n'_k - n''_k$ , if  $n'_1, \dots, n'_s$  and  $n''_1, \dots, n''_s$  are the numbers which characterise the two stationary states.

Without a detailed theory of the mechanism of transition between the stationary states we cannot, of course, in general obtain an exact determination of the *probability of spontaneous transition* between two such states, unless the  $n$ 's are large numbers. Just as in the case of systems of one degree of freedom, however, we are naturally led from the above considerations to assume that, also for values of the  $n$ 's which are not large, there must exist an intimate connection between the probability of a given transition and the values of the corresponding Fourier coefficient in the expressions for the displacements of the particles in the two stationary states. This allows us at once to draw certain important conclusions. Thus, from the fact that in general negative as well as positive values for the  $\tau$ 's appear in (31), it follows that we must expect that in general not only such transitions will be possible in which all the  $n$ 's decrease, but that also transitions will be possible for which some of the  $n$ 's increase while others decrease. This conclusion, which is supported by observations on the fine structure of the hydrogen lines as well as on the Stark effect, is contrary to the suggestion, put forward by Sommerfeld with reference to the essential positive character of the  $I$ 's, that every of the  $n$ 's must remain constant or decrease under a transition. Another direct consequence of the above considerations is obtained if we consider a system for which, for all values of the constants  $\alpha_1, \dots, \alpha_s$  the coefficient  $C_{\tau_1, \dots, \tau_s}$  corresponding to a certain set  $\tau_1^0, \dots, \tau_s^0$ , of values for the  $\tau$ 's is equal to zero in the expressions for the displacements of the particles in every direction. In this case we shall naturally expect that no transition will be possible for which the relation  $n'_k - n''_k = \tau_k^0$  is satisfied for every  $k$ . In the case where  $C_{\tau_1^0, \dots, \tau_s^0}$  is equal to zero in the expressions for the displacement to a certain direction only, we shall expect that all transitions, for which  $n'_k - n''_k \neq \tau_k^0$  for every  $k$ , will be accompanied by a radiation which is polarized in a plane perpendicular to this direction.

A simple illustration of the last considerations is afforded by the system mentioned in the beginning of this section, and which consists of a particle executing motions in three perpendicular directions which are independent of each other. In this case all the fourier coefficients in the expressions for the displacements in any direction will disappear if more than one of the  $\tau$ 's are different from zero. Consequently we must assume that only such transitions are possible for which only one of the  $n$ 's varies at the same time, and that the radiation corresponding to such a transition will be linearly polarized in the direction of the displacement of the corresponding coordinate. In the

special case where the motions in the three directions are simply harmonic, we shall moreover conclude that none of the  $n$ 's can vary by more than a single unit, in analogy with the considerations in the former section about a linear harmonic vibrator.

Another example which has more direct physical importance, since it includes all the special applications of the quantum theory to spectral problems mentioned in the introduction, is formed by a conditionally periodic system possessing an axis of symmetry. In all these applications a separation of variables is obtained in a set of three coordinates  $q_1, q_2$  and  $q_3$ , of which the first two serve to fix the position of the particle in a plane through the axis of the system, while the last is equal to the angular distance between this plane and a fixed through the same axis. Due to the symmetry, the expression for the total energy in Hamilton's equations will not contain the angular distance  $q_3$  but only the angular momentum  $p_3$  round the axis. The latter quantity will consequently remain constant during the motion, and the vibrations of  $q_1$  and  $q_2$  will be exactly the same as in a conditionally periodic system of two degrees of freedom only. If the position of the particle is described in a set of cylindrical coordinates  $z, \varrho, \vartheta$ , where  $z$  is the displacement in the direction of the axis,  $\varrho$  the distance of the particle from this axis and  $\vartheta$  is equal to the angular distance  $q_3$ , we have therefore

$$z = \sum C_{\tau_1, \tau_2} \cos 2\pi \cdot \{(\tau_1 \cdot \omega_1 + \tau_2 \cdot \omega_2) \cdot t + c_{\tau_1, \tau_2}\}$$

and

$$\varrho = \sum C'_{\tau_1, \tau_2} \cos 2\pi \cdot \{(\tau_1 \cdot \omega_1 + \tau_2 \cdot \omega_2) \cdot t + c'_{\tau_1, \tau_2}\}, \quad (32)$$

where the summation is to be extended over all positive and negative entire values of  $\tau_1$  and  $\tau_2$ , and where  $\omega_1$  and  $\omega_2$  are the mean frequencies of oscillation of the coordinates  $q_1$  and  $q_2$ . For the rate of variation of  $\vartheta$  with the time we have further

$$\begin{aligned} \frac{d\vartheta}{dt} &= \dot{q}_3 = \frac{\partial E}{\partial p_3} = f \cdot (q_1, q_2, p_1, p_2, p_3) = \\ &= \pm \sum C''_{\tau_1, \tau_2} \cos 2\pi \cdot \{(\tau_1 \cdot \omega_1 + \tau_2 \cdot \omega_2) \cdot t + C''_{\tau_1, \tau_2}\}, \end{aligned}$$

where the two signs correspond to a rotation of the particle in the direction of increasing and decreasing  $q_3$  respectively, and are introduced to separate the two types of symmetrical motions corresponding to these directions. This gives

$$\pm\vartheta = 2\pi \cdot \omega_3 \cdot t + \sum C'''_{\tau_1, \tau_2} \cos 2\pi \cdot \{(\tau_1 \cdot \omega_1 + \tau_2 \cdot \omega_2) \cdot t + c'''_{\tau_1, \tau_2}\}, \quad (33)$$

where the positive constant  $\omega_3 = C''_{0,0}/2\pi$  is the mean frequency of rotation round the axis of symmetry of the system. Considering now the displacement of the particle in rectangular coordinates  $x, y$  and  $z$ , and taking as above the axis of symmetry as  $z$ -axis, we get from (32) and (33) after a simple contraction of terms

$$x = \varrho \cos \vartheta = \sum D_{\tau_1, \tau_2} \cdot \cos 2\pi \cdot \{(\tau_1 \cdot \omega_1 + \tau_2 \cdot \omega_2 + \omega_3) \cdot t + d_{\tau_1, \tau_2}\}$$

and

$$y = \varrho \cos \vartheta = \pm \sum D_{\tau_1, \tau_2} \cdot \sin 2\pi \cdot \{(\tau_1 \cdot \omega_1 + \tau_2 \cdot \omega_2 + \omega_3) \cdot t + d_{\tau_1, \tau_2}\}, \quad (34)$$

where the  $D$ 's and  $d$ 's are new constants, and the summation is again to be extended over all positive and negative values of  $\tau_1$  and  $\tau_2$ .

From (32) and (34) we see that the motion in the present case may be considered as composed of a number of linear harmonic vibrations parallel to the axis of symmetry and of frequencies equal to the absolute values of  $(\tau_1 \cdot \omega_1 + \tau_2 \cdot \omega_2)$ , together with a number of circular harmonic motions round this axis equal to the absolute values of  $(\tau_1 \cdot \omega_1 + \tau_2 \cdot \omega_2 + \omega_3)$  and possessing the same direction of rotation as that of the moving particle or the opposite if the latter expression is positive or negative respectively. According to ordinary electrodynamics the radiation from the system would therefore consist of a number of components of frequency  $|\tau_1 \cdot \omega_1 + \tau_2 \cdot \omega_2|$  polarised parallel to the axis of symmetry, and a number of components of frequencies  $|\tau_1 \cdot \omega_1 + \tau_2 \cdot \omega_2 + \omega_3|$  and of circular polarisation round this axis (when viewed in the direction of the axis). On the present theory we shall consequently expect that in this case only kinds of transition between the stationary states given by (22) will be possible. In both of these  $n_1$  and  $n_2$  may vary by an arbitrary number of units, but in the first kind of transition, which will give rise to a radiation polarised parallel to the axis of the system,  $n_3$  will remain unchanged, while in the second kind of transition  $n_3$  will decrease or increase by one unit and the emitted radiation will be circularly polarised round the axis in the same direction as or the opposite of that of the rotation of the particle respectively.

In the next Part we shall see that these conclusions are supported in an instructive manner by the experiments on the effects of electric and magnetic field on the hydrogen spectrum. In connection with the discussion of the general theory, however, it may be of interest to show that the formal analogy between the ordinary theory of radiation and the theory based on (1) and (22), in case of systems possessing an axis of symmetry, can be traced not only with respect to frequency relations but also by considerations of

*conservation of angular momentum.* For a conditionally periodic system possessing an axis of symmetry the angular momentum round this axis is, with above choice of coordinates, according to (22) equal to  $I_3/2\pi = n_3 h/2\pi$ . If therefore, as assumed above for a transition corresponding to an emission of linearly polarised light,  $n_3$  is unaltered, it means that the angular momentum of the system remains unchanged, while if  $n_3$  alters by one unit, as assumed for a transition corresponding to an emission of circularly polarised light, the angular momentum will be altered by  $h/2\pi$ . Now it is easily seen that the ratio between this amount of angular momentum and the amount of energy  $h\nu$  emitted during the transition is just equal to the ratio between the amount of angular momentum and energy possessed by the radiation which according to ordinary electrodynamics would be emitted by an electron rotating in a circular orbit in a central field of force. In fact, if  $\alpha$  is the radius of the orbit,  $\nu$  the frequency of revolution and  $F$  the force of reaction due to the electromagnetic field of the radiation, the amount of energy and of angular momentum round an axis through the centre of the field perpendicular to the plane of the orbit, lost by the electron in unit of time as a consequence of the radiation, would be equal to  $2\pi\nu\alpha F$  and  $\alpha F$  respectively. Due to the principles of conservation of energy and of angular momentum holding in ordinary electrodynamics, we should therefore expect that the ratio between the energy and the angular momentum of the emitted radiation would be  $2\pi\nu$ ,<sup>25</sup> but this seen to be equal to the ratio between the energy  $h\nu$  and the angular momentum  $h/2\pi$  lost by the system considered above during a transition for which we have assumed that the radiation is circularly polarised. This agreement would seem not only to support the validity of the above considerations but also to offer a direct support, independent of the equations (22), of the assumption that, *for a atomic system possessing an axis of symmetry, the total angular momentum round this axis is equal to an entire multiple of  $h/2\pi$ .*

A further illustration of the above considerations of the relation between the quantum theory and the ordinary theory of radiation is obtained if we consider a conditionally periodic system subject to the *influence of a small perturbing field of force*. Let us assume that the original system allows of separation of variables in a certain set of coordinates  $q_1, \dots, q_s$ , so that the stationary states are determined by (22). From the necessary stability of the stationary states we must conclude that the perturbed system will possess a set of stationary states which only differ slightly from those of the original system. In general, however, it will not be possible for the per-

---

<sup>25</sup>Comp. K. Schaposchnikow, Phys. Zeitschr. **15** (1914) 454.

turbed system to obtain a separation of variables in any set of coordinates, but if the perturbing force is sufficiently small the perturbed motion will again be of conditionally periodic type and may be regarded as a superposition of a number of harmonic vibrations just as the original motion. The displacements of the particles in the stationary states of the perturbed system will therefore be given by an expression of the same type as (31) where the fundamental frequencies  $\omega_k$  and the amplitudes  $C_{\tau_1, \dots, \tau_s}$  may differ from those corresponding to the stationary states of the original system by small quantities proportional to the intensity of the perturbing forces. If now for the original motion the coefficients  $C_{\tau_1, \dots, \tau_s}$  corresponding to certain combinations of the  $\tau$ 's are equal to zero for all values of the constants  $\alpha_1, \dots, \alpha_s$ , these coefficients will therefore for the perturbed motion, in general, possess small values proportional to the perturbing forces. From the above considerations we shall therefore expect that, in addition to the main probabilities of such transitions between stationary states which are possible for the original system, there will for the perturbed system exist small probabilities of new transitions corresponding to the above mentioned combinations of the  $\tau$ 's. Consequently we shall expect that the effect of the perturbing field on the spectrum of the system will consist partly in a small displacement of the original lines, partly in the appearance of new lines of small intensity.

A simple example of this afforded by a system consisting of a particle moving in a plane and executing harmonic vibrations in two perpendicular directions with frequencies  $\omega_1$  and  $\omega_2$ . If the system is undisturbed all coefficients  $C_{\tau_1, \dots, \tau_2}$  will be zero, except  $C_{1,0}$  and  $C_{0,1}$ . When, however, the system is perturbed, for instance by an arbitrary small central force, there will in the Fourier expressions for the displacements of the particle, in addition to the main terms corresponding to the fundamental frequencies  $\omega_1$  and  $\omega_2$ , appear a number of small terms corresponding to frequencies given by  $\tau_1\omega_1 + \tau_2\omega_2$  where  $\tau_1$  and  $\tau_2$  are entire numbers which may be positive as well as negative. On the present theory we shall therefore expect that in the presence of the perturbing force there will appear small probabilities for new transitions which will give rise to radiations analogous to the so-called harmonics and combination tones in acoustics, just as it should be expected on the ordinary theory of radiation where a direct connection between the emitted radiation and the motion of the system is assumed. Another example of more direct physical application is afforded by the effect of an external homogeneous electric field in producing new spectral lines. In this case the potential of the perturbing force is a linear function of the coordinates of the particles and, whatever is the nature of the original system, it follows directly from the general theory of perturbations that the frequency of any

additional term in the expression for the perturbed motion, which is of the same order of magnitude as the external force, must correspond to the sum or difference of two frequencies of the harmonic vibrations into which the original motion can be resolved. With applications of these considerations we will meet in Part II in connection with the discussion of Sommerfeld's theory of the fine structure of the hydrogen lines and in Part III in connection with the problem of the appearance of new series in the spectra of other elements under the influence of intense external electric field.

As mentioned we cannot without more detailed theory of the mechanism of transition between stationary states obtain quantitative information as regards the general question of the intensities of the different lines of the spectrum of a conditionally periodic system given by (26), except in the limit where the  $n$ 's are large numbers, or in such special cases where for all values of the constants  $\alpha_1, \dots, \alpha_s$  certain coefficient  $C_{\tau_1, \dots, \tau_s}$  in (31) are equal to zero. From considerations of analogy, however, we must expect that it will be possible also in the general case to obtain an estimate of the intensities of the different lines in the spectrum by comparing the intensity of a given line, corresponding to a transition between two stationary states characterised by the numbers  $n'_1, \dots, n'_s$  and  $n''_1, \dots, n''_s$  respectively, with the intensities of the radiations of frequencies  $\omega_1 \cdot (n'_1 - n''_1) + \dots + \omega_s \cdot (n'_s - n''_s)$  to be expected on ordinary electrodynamics from the motions in these states; although of course this estimate becomes more uncertain the smaller the values for the  $n$ 's are. as it will be seen from the applications mentioned in the following Parts this is supposed in a general way by comparison with the observations.

*Collision of a Particle with Light Atoms. IV. An  
Anomalous Effect in Nitrogen.*

by Professor Sir E. Rutherford, F.R.S.,  
University of Cambridge.  
(Received 1919)

It has been shown in paper I that a metal source, coated with a deposit of radium C, always gives rise to a number of scintillations on a zinc sulphide screen far beyond the range of the  $\alpha$  particles. The swift atoms causing these scintillations carry a positive charge and are deflected by a magnetic field, and have about the same range and energy as the swift H atoms produced by the passage of  $\alpha$  particles through hydrogen. These "natural" scintillations are believed to be due mainly to swift H atoms from the radioactive source, but it is difficult to decide whether they are expelled from the radioactive source itself or are due to the action of  $\alpha$  particles on occluded hydrogen.

The apparatus employed to study "natural" scintillations is the same as that described in paper I. The intense source of radium C was placed inside a metal box about 3 cm from the end, and an opening in the end of the box was covered with a silver plate of stopping power equal to about 6 cm of air. The zinc sulphide screen was mounted outside, about 1 mm distant from the silver plate, to admit of the introduction of absorbing foils between them. The whole apparatus was placed in a strong magnetic field to deflect the  $\beta$  rays. The variation in the number of these "natural" scintillations with absorption in terms of cms. of air is shown in fig. 1, curve A. In this case, the air in the box was exhausted and absorbing foils of aluminium were used. When dried oxygen or carbon dioxide was admitted into the vessel, the number of scintillations diminished to about the amount to be expected from the stopping power of the column of gas.

A surprising effect was noticed, however, when dried air was introduced. Instead of diminishing, the number of scintillations was increased, and for an absorption corresponding to about 19 cm of air the number was about twice that observed when the air was exhausted. It was clear from this experiment that the  $\alpha$  particles in their passage through air gave rise to long-range scintillations which appeared to the eye to be about equal in brightness to H scintillations. A systematic series of observations was undertaken to account for the origin of these scintillations. In the first place we have seen that the passage of  $\alpha$  particles through nitrogen and oxygen gives rise to numerous scintillations which have a range of about 9 cm in air. These scintillations have about the range to be expected if they are due to swift N or O atoms, carrying unit charge, produced by collision with  $\alpha$  particles.

This job contains a bitmapped image that can only be printed on PostScript(R)-Level-2 printers. Please send this file to a PostScript Level-2 printer, or print your document again with the correct printer selected in the printer driver's dialog box.

Figure 1:

All experiments have consequently been made with an absorption greater than 9 cm of air, so that these atoms are completely stopped before reaching the zinc sulphide screen. It was found that these long-range scintillations could not be due to the presence of water vapour in the air; for the number was only slightly reduced by thoroughly drying the air. This is to be expected, since on the average the number of the additional scintillations due to air was equivalent to the number of H atoms produced by the mixture of hydrogen at 6 cm pressure with oxygen. Since on the average the vapour pressure of water in air was not more than 1 cm, the effects of complete drying would

not reduce the number by more than one sixth. Even when oxygen and carbon dioxide saturated with water vapour at 20° C were introduced in place of dry air, the number of scintillations was much less than with dry air.

It is well known that the amount of hydrogen or gases containing hydrogen is normally very small in atmospheric air. No difference was observed whether the air was taken directly from the room or from outside the laboratory or was stored for some days over water.

There was the possibility that the effect in air might be due to liberation of H atoms from the dust nuclei in the air. No appreciable difference, however, was observed when the dried air was filtered through long plugs of cotton-wool, or by storage over water for some days to remove dust nuclei.

Since the anomalous effect was observed in air, but not in oxygen, or carbon dioxide, it must be due either to nitrogen or to one of the other gases present in atmospheric air. The latter possibility was excluded by comparing the effects produced in air and in chemically prepared nitrogen. The nitrogen was obtained by the well-known method of adding ammonium chloride to sodium nitrite, and stored over water. It was carefully dried before admission to the apparatus. With pure nitrogen, the number of long-range scintillations under similar conditions was greater than in air. As a result of careful experiments, the ratio was found to be 1.25, the value to be expected if the scintillations are due to nitrogen.

The results so far obtained show that the long-range scintillations obtained from air must be ascribed to nitrogen, but it is important, in addition, to show that they are due to collision of  $\alpha$  particles with atoms of nitrogen through the volume of the gas. In the first place, it was found that the number of the scintillations varied with the pressure of the air in the way to be expected if they resulted from collision of  $\alpha$  particles along the column of gas. In addition, when an absorbing screen of gold or aluminium was placed close to the source, the range of the scintillations was found to be reduced by the amount to be expected if the range of the expelled atom was proportional to the range of the colliding  $\alpha$  particles. These results show that the scintillations arise from the volume of the gas and are not due to some surface effect in the radioactive source.

In fig. 1 curve A the results of a typical experiment are given showing the variation in the number of natural scintillations with the amount of absorbing matter in their path measured in terms of centimetres of air for  $\alpha$  particles. In these experiments carbon dioxide was introduced at a pressure calculated to give the same absorption of the  $\alpha$  rays as ordinary air. In curve B the corresponding curve is given when air at N.T.P. is introduced

in place of carbon dioxide. The difference curve C shows the corresponding variation of the number of scintillations arising from the nitrogen in the air. It was generally observed that the ratio of the nitrogen effect to the natural effect was somewhat greater for 19 cm than for 12 cm absorption.

In order to estimate the magnitude of the effect, the space between the source and screen was filled with carbon dioxide at diminished pressure and a known pressure of hydrogen was added. The pressure of the carbon dioxide and of hydrogen were adjusted so that the total absorption of  $\alpha$  particles in the mixed gas should be equal to that of the air. In this way it was found that the curve of absorption of H atoms produced under these conditions was somewhat steeper than curve C of fig. 1. As a consequence, the amount of hydrogen mixed with carbon dioxide required to produce a number of scintillations equal to that of air, increased with the increase of absorption. For example, the effect in air was equal to about 4 cm of hydrogen at 12 cm absorption, and about 8 cm at 19 cm absorption. For a mean value of the absorption, the effect was equal to about 6 cm of hydrogen. This increased absorption of H atoms under similar conditions indicated either that (1) the swift atoms from air had a somewhat greater range than the H atoms, or (2) that the atoms from air were projected more in the line of flight of the  $\alpha$  particles. While the maximum range of the scintillations from air using radium C as a source of  $\alpha$  rays appeared to be about the same, viz. 28 cm, as for H atoms produced from hydrogen, it was difficult to fix the end of the range with certainty on account of the smallness of the number and the weakness of the scintillations. Some special experiments were made to test whether, under favourable conditions, any scintillations due to nitrogen could be observed beyond 28 cm of air absorption. For this purpose a strong source (about 60 mg Ra activity) was brought within 2.5 cm of the zinc sulphide screen, the space between containing dry air. On still further reducing the distance, the screen became too bright to detect very feeble scintillations. no certain evidence of scintillations was found beyond a range of 28 cm. It would therefore appear that 92) above is the more probable explanation.

In a previous paper (III.) we have seen that the number of swift atoms of nitrogen or oxygen produced per unit path by collision with  $\alpha$  particles is about the same as the corresponding number of H atoms in hydrogen. Since the number of long-range scintillations in air is equivalent to that produced under similar conditions in a column of hydrogen at 6 cm pressure, we may consequently conclude that only one long-range atom is produced for every 12 close collisions giving rise to a swift nitrogen atom of maximum range 9 cm.

It is of interest to give data showing the number of long-range scintillations produced in nitrogen at atmospheric pressure under definite conditions. For a column of nitrogen 3.3 cm long, and for a total absorption of 19 cm of air from the source, the number due to nitrogen per milligram of activity is 0.6 per minute on a screen of 3.14 sq. mm area.

Both as regards range and brightness of scintillations, the long-range atoms from nitrogen closely resemble H atoms, and in all probability are hydrogen atoms. In order, however, to settle this important point definitely, it is necessary to determine the deflexion of these atoms in a magnetic field. Some preliminary experiments have been made by a method similar to that employed in measuring the velocity of the H atom (see paper II.) The main difficulty is to obtain a sufficiently large deflexion of the stream of atoms and yet have a sufficient number of scintillations per minute for counting. The  $\alpha$  rays from a strong source passed through dry air between two parallel horizontal plates 3 cm long and 1.6 cm apart, and the number of scintillations on the screen placed near the end of the plates was observed for different strengths of the magnetic field. Under these conditions, when the scintillations arise from the whole length of the column of air between the plates, the strongest magnetic field available reduced the number of scintillations by only 30 per cent. When the air was replaced by a mixture of carbon dioxide and hydrogen of the same stopping power for  $\alpha$  rays, about an equal reduction was noted. As far as the experiment goes, this is an indication that the scintillations are due to H atoms; but the actual number of scintillations and the amount of reduction was too small to place much reliance on the result. In order to settle this question definitely, it will probably prove necessary to employ a solid nitrogen compound, free from hydrogen, as a source, and to use much stronger sources of  $\alpha$  rays. In such experiments, it will be of importance to discriminate between the deflexions due to H atoms and possible atoms of atomic weight 2. From the calculations given in paper III, it is seen that a collision of an  $\alpha$  particle with a free atom of mass 2 should give rise to an atom of range about 32 cm in air, and of initial energy about 0.89 of that of the H atom produced under similar conditions. The deflexion of the pencil of these rays in a magnetic field should be about 0.6 of that shown by a corresponding pencil of H atoms.

### *Discussion of results*

From the results so far obtained it is difficult to avoid the conclusion that the long-range atoms arising from collision of  $\alpha$  particles with nitrogen are not nitrogen atoms but probably atoms of hydrogen, or atoms of mass 2. If this be the case, we must conclude that the nitrogen atom is disintegrated under the intense forces developed in a close collision with a swift  $\alpha$  particle, and that the hydrogen atom which is liberated formed a constituent part of the nitrogen nucleus. We have drawn attention in paper III to the rather surprising observation that the range of the nitrogen atoms in air is about the same as the oxygen atoms, although we should expect a difference of about 19 per cent. If in collisions which give rise to swift nitrogen atoms, the hydrogen is at the same time disrupted, such a difference might be accounted for, for the energy is then shared between two systems.

It is of interest to note, that while the majority of the light atoms, as is well known, have atomic weights represented by  $4n$  or  $4n + 3$  where  $n$  is a whole number, nitrogen is the only atom which is expressed by  $4n + 2$ . We should anticipate from radioactive data that the nitrogen nucleus consists of three helium nuclei each of atomic mass 4 and either two hydrogen nuclei or one of mass 2. If the H nuclei were outriders of the main system of mass 12, the number of close collisions with the bound H nuclei would be less than if the latter were free, for the  $\alpha$  particle in a collision comes under the combined field of the H nucleus and of the central mass. Under such conditions, it is to be expected that the  $\alpha$  particle would only occasionally approach close enough to the H nucleus to give it the maximum velocity, although in many cases it may give it sufficient energy to break its bond with the central mass. Such a point of view would explain why the number of swift H atoms from nitrogen is less than the corresponding number in free hydrogen and less also than the number of swift nitrogen atoms. The general results indicate that the H nuclei, which are released, are distant about twice the diameter of the electron ( $7 \times 10^{-13}$  cm) the centre of the main atom. without a knowledge from of the laws of force at such small distances, it is difficult to estimate the energy required to free the H nucleus or to calculate the maximum velocity that can be given to the escaping H atom. It is not to be expected, *a priori*, that the velocity or range of the H atom released from the nitrogen atom should be identical with that due to a collision in free hydrogen.

Taking into account the great energy of motion of the  $\alpha$  particle expelled from radium C, the close collision of such an  $\alpha$  particle with a light atom

seems to bee the most likely agency to promote the disruption of the latter; for the forces on the nuclei arising from such collisions appear to be greater than can be produced by any other agency at present available. Considering the enormous intensity of the forces brought into play, it is not so much a matter of surprise that the nitrogen atom should suffer disintegration as that the  $\alpha$  particle itself escapes disruption into its constituents. The results as a whole suggest that, if  $\alpha$  particles—or similar projectiles—of still greater energy were available for experiment, we might expect to break down the nucleus structure of many of the lighter atoms.

I desire to express my thanks to Mr. William Kay for his invaluable assistance in counting scintillations.

*University of Manchester,*  
April 1919.



Bakerian Lecture: *Nuclear Constitution of Atoms*

by Sir E. Rutherford, F.R.S., Cavendish Professor of Experimental Physics,  
University of Cambridge.

(Received June 3, – Lecture delivered June 3, 1920)

*Introduction* – The conception of the nuclear constitution of atoms arose initially from attempts to account for the scattering of  $\alpha$ -particles through large angle in traversing thin sheets of matter<sup>1</sup>. Taking into account the large mass and velocity of the  $\alpha$ -particles, these large deflexions were very remarkable, and indicated that very intense electric or magnetic field exist within the atom. To account for these results, it was found necessary to assume<sup>2</sup> that the atom consists of a charged massive nucleus of dimensions very small compared with the ordinarily accepted magnitude of the diameter of the atom. This positively charged nucleus contains most of the mass of the atom, and is surrounded at a distance by a distribution of negative electrons equal in number to the resultant positive charge on the nucleus. Under these conditions, a very intense electric field exists close to the nucleus, and the large deflection of the  $\alpha$ -particle in an encounter with a single atom happens when the particle passed close to the nucleus. Assuming that the electric forces between the  $\alpha$ -particle and the nucleus varied according to an inverse square law in the region close to the nucleus, the writer worked out the relations connecting the number of  $\alpha$ -particles scattering through any angle with the charge on the nucleus and the energy of the  $\alpha$ -particle. Under the central field of force, the  $\alpha$ -particle described a hyperbolic orbit round the nucleus, and the magnitude of the deflection depends on the closeness of approach to the nucleus. From the data of scattering of  $\alpha$ - particles then available, it was deduced that the resultant charge on the nucleus was about

<sup>1</sup>Geiger and Marsden, Roy. Soc. Proc., A, vol. 82, p. 495 (1909).

<sup>2</sup>Rutherford, Phil. Mag., vol. 21, p. 669 (1911); vol. 27, p. 488 (1914)

$\frac{1}{2}Ae$ , where  $A$  is the atomic weight and  $e$  the fundamental unit of charge. Geiger and Marsden<sup>3</sup> made an elaborate series of experiments to test the correctness of the theory, and confirmed the main conclusions. They found the nucleus charge was about  $\frac{1}{2}Ae$ , but, from the nature of the experiments, it was difficult to fix the actual value within about 20 per cent. C.G. Darwin<sup>4</sup> worked out completely the deflexion of the  $\alpha$ -particle and of the nucleus, taking into account the mass of the latter, and showed that the scattering experiments of Geiger and Marsden could not be reconciled with any law of central force, except the inverse square. The nuclear construction of the atom was thus very strongly supported by the experiments on scattering of  $\alpha$ -rays.

Since the atom is electrically neutral, the number of external electrons surrounding the nucleus must be equal to the number of units of resultant charge on the nucleus. It should be noted that, from the consideration of the scattering of X-rays by light elements, Barkla<sup>5</sup> had shown, in 1911, that the number of electrons was equal to about half the atomic weight. This was deduced from the theory of scattering of Sir J.J. Thomson, in which it was assumed that each of the external electrons in an atom acted as an independent scattering unit.

Two entirely different methods had thus given similar results with regard to the number of external electrons in the atom, but the scattering of  $\alpha$ -rays had shown in addition that the positive charge must be concentrated on a massive nucleus of small dimensions. It was suggested by Van den Broek<sup>6</sup> that the scattering of  $\alpha$ -particles by the atoms was not inconsistent with the possibility that the charge on the nucleus was equal to the atomic number of the atom, i.e., to the number of the atom when arranged in order of increasing atomic weight. The importance of the atomic number in fixing the properties of an atom was shown by the remarkable work of Moseley<sup>7</sup> on the X-ray spectra of the elements. He showed that the frequency of vibration of corresponding lines in the X-ray spectra of the elements depended on the square of a number which varied by unity in successive elements. This relation received an interpretation by supposing that the nuclear charge varied by unity in passing from atom to atom, and was given numerically by the atomic number. I can only emphasise in passing the great importance of Moseley's work, not only in fixing the number of possible elements, and

<sup>3</sup>Geiger and Marsden, Phil. Mag., vol. 25, p. 604 (1913).

<sup>4</sup>Darwin, Phil. Mag., vol. 27, p. 499 (1914).

<sup>5</sup>Barkla, Phil. Mag., vol. 21, p. 648 (1911).

<sup>6</sup>Van den Broek, Phys. Zeit., vol. 14, p. 32 (1913).

<sup>7</sup>Moseley, Phil. Mag., vol. 26, p. 1024 (1913); vol. 27, p. 703 (1914).

the position of undermined elements, but in showing that the properties of an atom were defined by a number which varied by unity in successive atoms. This gives a new method of regarding the periodic classification of the elements, for the atomic number, or its equivalent the nuclear charge, is of more fundamental importance than its atomic weight. In Moseley's work, the frequency of vibration of the atom was not exactly proportional to  $N$ , where  $N$  is the atomic number, but to  $(N-a)^2$ , where  $a$  was a constant which had different values, depending of whether the  $K$  or  $L$  series of characteristic radiations were measured. It was supposed that this constant depended on the number and position of the electrons close to the nucleus.

*Charge on the Nucleus.* — The question whether the atomic number of an element is the actual measure of its nuclear charge is a matter of such fundamental importance that all methods of attack should be followed up. Several researches are in progress in the Cavendish Laboratory to test the accuracy of this relation. The two most direct methods depend on the scattering of swift  $\alpha$ - and  $\beta$ -rays. The former is under investigation, using new methods, by Mr. Chadwick, and the latter by Dr. Crowther. The results so far obtained by Mr. Chadwick strongly support the identity of the atomic number with the nuclear charge within the possible accuracy of experiment, viz., about 1 per cent.

It thus seems clear that we are on firm ground in supposing that the nuclear charge is numerically given by the atomic number of the element. Incidentally, these results, combined with the work of Moseley, indicate that the law of the inverse square holds with considerable accuracy in the region surrounding the nucleus. It will be of great interest to determine the extent of this region, for it will give us definite information as to the distance of the inner electrons from the nucleus. A comparison of the scattering of slow and swift  $\beta$ -rays should yield important information on this point. The agreement of experiment with theory for the scattering of  $\alpha$ -rays between  $5^\circ$  and  $150^\circ$  shows that the law of inverse square holds accurately in the case of a heavy element like gold for distances between about  $36 \times 10^{-12}$  cm and  $3 \times 10^{-12}$  cm from the centre of the nucleus. We may consequently conclude that few, if any, electrons are present in this region.

An  $\alpha$ -particle in a direct collision with a gold atom of nuclear charge 79 will be turned back in its path at a distance of  $3 \times 10^{-12}$  cm, indicating that the nucleus may be regarded as a point charge even for such a short distance. Until swifter  $\alpha$ -particles are available for experiment, we are unable in the ease of heavy elements to push further the question of dimensions of heavy

atoms. We shall see later, however, that the outlook is more promising in the case of lighter atoms, where the  $\alpha$ -particle can approach closer to the nucleus.

It is hardly necessary to emphasise the great importance of the nuclear charge in fixing the physical and chemical properties of an element, for obviously the number and the arrangements of the external electrons on which the great majority of the physical and chemical properties depend, is conditioned by the resultant charge on the nucleus. It is to be anticipated theoretically, and is confirmed by experiment, that the actual mass of the nucleus exercises only a second order effect on the arrangement of the external electrons and their rates of vibration.

It is thus possible to imagine the existence of elements of almost identical physical and chemical properties, but which differ from one another in mass, for, provided the resultant nuclear charge is the same, a number of possible stable modes of combination of the different units which make up a complex nucleus may be possible. The dependence of the properties of an atom on its nuclear charge and not on its mass thus offers a rational explanation of the existence of isotopes in which the chemical and physical properties may be almost indistinguishable, but the mass of the isotopes may vary within certain limits. This important question will be considered in more detail later in the paper in the light of evidence as to the nature of the units which make up the nucleus.

The general problem of the structure of the atom thus naturally divides itself into two parts: —

1. Constitution of the nucleus itself.
2. The arrangement and modes of vibration of the external electrons.

I do not propose to-day to enter into (2), for it is a very large subject in which there is room for much difference of option. This side of the problem was first attacked by Bohr and Nicholson, and substantial advances have been made. Recently, Sommerfeld and others have applied Bohr's general method with great success in explaining the fine structure of the spectral lines and the complex modes of vibration of simple atoms involved in the Stark effect. Recently, Langmuir and others have attacked the problem of the arrangement of the external electrons from the chemical standpoint, and have emphasised the importance of assuming a more or less cubical arrangement of the electrons in the atom. No doubt each of these theories

has a definite sphere of usefulness, but our knowledge is as yet too scanty to bridge over the apparent difference between them.

I propose to-day to discuss in some detail experiments that have been made with a view of throwing light on the constitution and stability of the nuclei of some of the simpler atoms. From a study of ratio-activity we know that the nuclei of the radio-active elements consist in part of helium nuclei of charge  $2e$ . We also have strong reason for believing that the nuclei of atoms contain electrons as well as positively charged bodies, and that the positive charge on the nucleus represents the excess positive charge. It is of interest to note the very different *role* played by the electrons in the outer and inner atom. In the former case, the electrons arrange themselves at a distance from the nucleus, controlled no doubt mainly by the charge on the nucleus and the interaction of their fields. In the case of the nucleus, the electron forms a very close and powerful combination with the positively charged units and, as far as we know, there is a region just outside the nucleus where no electron is in stable equilibrium. While no doubt each of the external electrons acts as a point charge in considering the forces between it and the nucleus, this cannot be the case for the electron in the nucleus itself. It is to be anticipated that under the intense forces in the latter, the electrons are much deformed and the forces may be of a very different character from those to be expected from an undeformed electron, as in the outer atom. It may be for this reason that the electron can play such a different part in the two cases and yet from stable system.

It has been generally assumed, on the nucleus theory, that electric forces and charges play a predominant part in determining the structure of the inner and outer atom. The considerable success of this theory in explaining fundamental phenomena is an indication of the general correctness of this point of view. At the same time if the electrons and parts composing the nucleus are in motion, magnetic fields must arise which will have to be taken into account in any complete theory of the atom. In this sense the magnetic field are to be regarded as a secondary rather than a primary factor, even though such fields may be shown to have an important bearing on the conditions of equilibrium of the atom.

### *Dimensions of Nuclei*

We have seen that in the case of atoms of large nuclear charge the swiftest  $\alpha$ -particle is unable to penetrate to the actual structure of the nucleus so that it is possible to give only a maximum estimate of its dimensions. In the case of light atoms, however, when the nucleus charge is small, there is so close an approach during a direct collision with an  $\alpha$ -particle that we are able to estimate its dimensions and from some idea of the forces in operation. This is best shown in the case of a direct collision between an  $\alpha$ -particle and an atom of hydrogen. In such a case, the H atom is set in such swift motion that it travels four times as far as the colliding  $\alpha$ -particle and can be detected by the scintillation produced on a zinc sulphide scree.<sup>8</sup> The writer<sup>9</sup> has shown that these scintillations are due hydrogen atoms carrying unit positive charge recoiling with the velocity to be expected from the simple collision theory, viz., 1.6 times the velocity of the  $\alpha$ -particle. The relation between the number and velocity of these H atoms is entirely different from that to be expected if the  $\alpha$ -particle and H atom are regarded as point charges for the distances under consideration. The result of the collision with swift  $\alpha$ -particles is to produce H atoms which have a narrow range velocity, and which travel nearly in the direction of the impinging particles. It was deduced that the law of inverse square no longer holds when the nuclei approach to within a distance of  $3 \times 10^{-12}$  cm of each other. This is an indication that the nuclei have dimensions of this order of magnitude and that the forces between the nuclei vary very rapidly in magnitude and in direction for a distance of approach comparable with the diameter of the electron as ordinarily calculated. It was pointed out that in such close encounters there were enormous forces between the nuclei, and probably the structure of the nuclei was much deformed during the collision. The fact that the helium nucleus, which may be supposed to consist of four H atoms and two electrons, appeared to survive the collision is an indication that it must be a highly stable structure. Similar results<sup>10</sup> were observed in the collision between  $\alpha$ -particles and atoms of nitrogen and oxygen for the recoil atoms appeared to be shot forward mainly in the direction of the  $\alpha$ -particles and the region where special forces come into play is of the same order of magnitude as in the case of the collision of an  $\alpha$ -particle with hydrogen.

No doubt the space occupied by a nucleus and the distance at which

<sup>8</sup>Marsden, Phil. Mag., vol. 27, p. 824 (1914).

<sup>9</sup>Rutherford, Phil. Mag., vol. 37, I and II, pp. 538-571 (1919).

<sup>10</sup>Rutherford, Phil. Mag., vol. 37, III, p. 571 (1919)

the forces become abnormal increase with the complexity of the nucleus structure. WE should expect the H nucleus to be the simplest of all and, if it be the positive electron, it may have exceedingly small dimensions compared with the negative electron. In the collisions between  $\alpha$ -particles of H atoms, the  $\alpha$ -particle is to be regarded as the more complex structure.

The diameter of the nuclei of the light atoms except hydrogen are probably of the order of magnitude  $5 \times 10^{-13}$  cm and in a close collision the nuclei come nearly in contact and may possibly penetrate each other's structure. Under such conditions, only very stable nuclei would be expected to survive the collision and it is thus of great interest to examine whether evidence can be obtained of their disintegration.

#### *Long Range Particles from Nitrogen*

In previous papers, *loc. cit.*, I have an account of the effect produced by close collisions of swift  $\alpha$ -particles with light atoms of matter with the view of determining whether the nuclear structure of some of the lighter atoms could be disintegrated by the intense forces brought into play in such close collisions. Evidence was given that the passage of  $\alpha$ - particles through dry nitrogen gives rise to swift particles which closely resembled in brilliancy of the scintillations and distance of penetration hydrogen atoms set in motion by close collision with  $\alpha$ -particles. It was shown that these swift atoms which appeared only in dry nitrogen and not in oxygen or carbon dioxide could not be ascribed to the presence of water vapour or other hydrogen material but must from the collision of  $\alpha$ -particles with nitrogen atoms. The number of such scintillations due to nitrogen was small, viz., about 1 in 12 of the corresponding number in hydrogen, but was two to three times the number of natural scintillations from the source. The number observed in in nitrogen was on an average equal to the number of scintillations when hydrogen at about 6 cm pressure was added to oxygen or carbon dioxide at normal pressure. while the general evidence indicated that these long range atoms from nitrogen were charged atoms of hydrogen, the preliminary experiments to test the mass of the particles by bending them in a strong magnetic field yielded no definite results.

From the data given in my previous paper (*loc. cit.*) several theories could be advanced to account for these particles. The calculated range of a singly charged atom set in motion by a close collision with an  $\alpha$ -particle of

range R cm in air was shown to be for

Mass	1 .....	Range	3.91R
"	2 .....	"	4.6 R
"	3 .....	"	5.06R
"	4 .....	"	4.0 R

On account of the small number and weakness of the scintillations under the experimental conditions, the range of the swift atoms from nitrogen could not be determined with sufficient certainty to decide definitely between any of these possibilities. The likelihood that the particles were the original  $\alpha$ -particles which had lost one of their two charged, i.e., atoms of charge 1 and mass 4, was suggested by me to several correspondents, but there appeared to be no obvious reason why nitrogen, of all the elements examined, should be the only one in which the passage of a swift  $\alpha$ -particle led to the capture of a single electron.

If, however, a sufficient number of scintillations could be obtained under the experimental conditions, there should be no inherent difficulty in deciding between the various possibilities by examining the deflexion of the swift atoms by a magnetic field. The amount of deflexion of charged atoms in a magnetic field perpendicular to the direction of flight is proportional to  $e/mu$ . Assuming that the particles were liberated by a direct collision with an  $\alpha$ -particle, the relative values of this quantity for different recoiling masses are easily calculated. Taking values  $MV/E$  for the  $\alpha$ -particle as unity, the corresponding values of  $mu/e$  for atoms of charge 1 and mass 1, 2, 3, and 4 are 1.25, 0.75, 0.58, and 0.50 respectively. Consequently the H atoms should be more bent than the  $\alpha$ -particles which produced them while the atoms of mass 2 or 3, or 4 would be more difficult to deflect than the parent  $\alpha$ -particle.

On my arrival in Cambridge, this problem was attacked in several ways.

By the choice of objectives of wide aperture, the scintillations were increased in brilliancy and counting thus made easier. A number of experiments were also made to obtain more powerful of radiation with the radium at my command, but finally it was found best, for reason which need not be discussed here, to obtain the active source of radiation of radium C in the manner described in my previous paper. After a number of observations with solid nitrogen compounds, described later, a simple method was finally devised to estimate the mass of the particle by the use of nitrogen in the gaseous state. The use of the gas itself for this purpose had several advantages over the use of solid nitrogen compounds, for not only was

the number of scintillations greater, but the absence of hydrogen or other hydrogen compounds could be ensured.

The :  
ial point

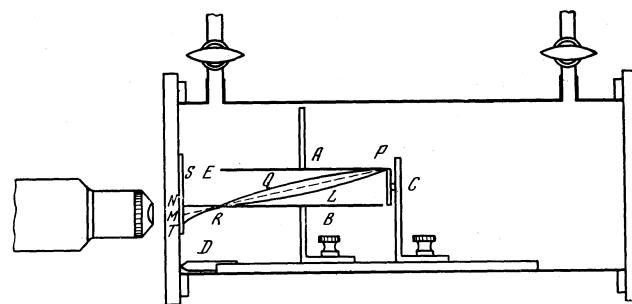


Figure 1:

lay in the use of wide slits, between which the  $\alpha$ -particles passed. Experiment showed that the ratio of the number of scintillations on the screen arising from the gas to the number of natural scintillations from the source, increased rapidly with increased depth of the slits. For plates 1 mm apart this ratio was less than unity, but for slits 8 mm apart the ratio had a value 2 to 3. such a variation is to be anticipated on theoretical grounds if the majority of the particles are liberated at an angle with the direction of the incident  $\alpha$ -particles.

The horizontal slits *A*, *B* were 6.0 cm long 1.5 cm wide, and 8 mm deep, with the source, *C* of the active deposit of radium placed at one end the zinc sulphide screen near the other. The carrier of the source and slits were placed in a rectangular brass box, through which a current of dry air or other gas was continuously passed to avoid the danger of radio-active contamination. The box was placed between the poles of a large electromagnet, so that the uniform field was parallel to the plane of the plates and perpendicular to their length. A distance piece, *D*, of length 1.2 cm, was added between the source and end of the slits, in order to increase the amount of deflexion of the radiation issuing from the slits. The zinc sulphide screen *S*, was placed on a glass plate covering the end of the box. The distance between the source and the screen was 7.4 cm. The recoil atoms from oxygen or nitrogen of range 9 cm could be stopped by inserting an aluminium screen of stopping power about 2 cm of air placed at the end of the slits.

With such deep slits it was impossible to bend the wide beam of radiation

to the sides, but the amount of deflexion of the radiation issuing near the bottom of the slit was measured. For this purpose it was essential to observe the scintillations at a fixed point of the screen near  $M$ . The method of fixing the position of the counting microscope was as follows: The source,  $C$ , was placed in position, and the air exhausted to a pressure of a few centimetres. Without the field, the bottom edge of the beam was fixed by the straight  $PM$  cutting the screen at  $M$ . The microscope was adjusted so that the boundary line of scintillations appeared above the horizontal cross wire in the microscope, making the centre of the field.

On exciting the magnet to bend the rays upward (called the + field), the path of the limiting  $\alpha$ -particles is marked by the curve  $PLRN$  cutting the screen at  $N$ , so that the boundary of the scintillations appears to be displaced downwards in the field of view. On reversing the field (called the - field), the path of the limiting  $\alpha$ -particle  $PQRT$  cuts the screen at  $T$ , and the band of scintillations appears to be bent upwards. The strength of the magnetic field was adjusted so that, with a negative field, the scintillations were observed all over the screen, while, with a positive field, they were mainly confined below the cross wide. The appearance in the field of view of the microscope for the two fields is illustrated in fig. 2, where the dots represent the scintillations.

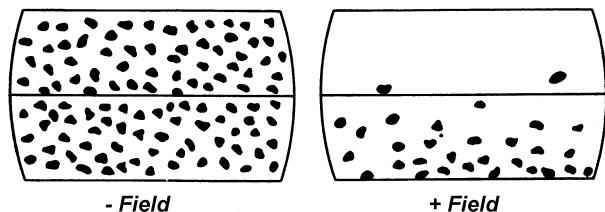


Figure 2:

The horizontal boundaries of the field of view were given by a rectangular opening in a plate fixed in the position of the cross wires. A horizontal wire, which bisected the field of view, was visible under the conditions of counting, and allowed the relative number of scintillations in the two halves of the field to be counted if required. Since the number of scintillations in the actual experiments with nitrogen was much too small to mark directly the boundary of the scintillations, in order to estimate the bending of the rays, it was necessary to determine the ratio of the number of scintillations with the + and - field.

The position of the microscope and the strength of the magnetic field were in most experiments so adjusted that this ratio was about one-third. Preliminary observations showed that this ratio was sensitive to changes of the field and it thus afforded a suitable method for estimating the relative bending of any radiations under examination.

After the position of the microscope was fixed, air was let in, and a continuous flow of dry maintained through the apparatus. The absorbing screen was introduced at  $E$  to stop the atoms from N and O of range 9 cm. The number of scintillations was then systematically counted for the two directions of the field, and a correction, if required, made for any slight radioactive contamination of the screen. The deflexion due to the unknown radiation was directly compared with that produced by a known radiation of  $\alpha$ - rays. For this purpose, after removal of the source and absorbing screen, a similar plate, coated with a weak distribution of the active deposit of thorium, was substituted for the radium source. The  $\alpha$ -particles from thorium  $C$  of range 8.6 cm produced bright scintillations in the screen after traversing the 7.4 cm of air in their path. The ratio of the number of scintillations with + and - fields was determined as before.

An example of such comparison is given below. For a current of 4.0 amp. through the electromagnet, the ratio for particles from nitrogen was found to be 0.33. The corresponding ratio for  $\alpha$ -particles from thorium  $C$  was 0.44 for a current of 4 amp. and 0.31 for current of 5 amp. It is thus seen that on the average, the particles from nitrogen are more bent in a given field than the  $\alpha$ -particles from thorium  $C$ . In order, however, to make a quantitative comparison, it is necessary to take into account the reduction in velocity of the radiations in passing through the air. The value  $mu/e$  for the  $\alpha$ -ray of range 8.6 cm from thorium  $C$  is known to be  $4.28 \times 10^5$ . Since the rays pass through 7.4 cm of air in a uniform field before striking the screen, it can be calculated that the actual deflection corresponds to  $\alpha$ -rays in a vacuum for which  $mu/e = 3.7 \times 10^5$ , about. Taking the deflection of the  $\alpha$ -particles for a current of 4.8 amp. to be the same as for the nitrogen particles for a field of 4 atm. - ratio of fields 1.17 - it is seen that the average deflexion of the nitrogen particles under the experimental conditions corresponds to a radiation in a vacuum for which the value of  $mu/e = 3.1 \times 10^5$ .

Bearing in mind that the particles under examination are produced throughout the volume of the gas between the slits, and that their distribution is unknown, and also that the particles are shot forward on an average at an angle with the incident  $\alpha$ -particles, the experimental data are quite insufficient to calculate the average value of  $mu/e$  to be expected under the experimental conditions for any assumed mass of project particles. It seems

probable that the majority of the particles which produce scintillations are generated in the first few centimetres of the air next the source. The actual deflection of a given particle by the magnetic field will depend on the distance of its point of origin from the source. These factors will obviously tend to make the average deflection of the particles to appear less than if they were all expelled with constant velocity from the source itself. Assuming that the correction for reduction of velocity of the long range particles in traversing the gas is 10 per cent., the average value of  $mu/e$  is about  $3.4 \times 10^5$ . Since the value of  $MV/E$  for the  $\alpha$ -particle from radium  $C$  is  $3.98 \times 10^5$ , it is seen that under the experimental conditions the average value of  $mu/e$  for the nitrogen particles is less than that of the  $\alpha$ -particles which produce them.

From the data given earlier in the paper, this should only be true if the particles are comparable in mass with an atom of hydrogen, for singly charged particles of mass 2, 3, or 4 should suffer less deflexion than the  $\alpha$ -particles. For example, if we assume that the particles were helium atoms carrying one charge, we should expect them to be deflected to about one-half of the extent of the  $\alpha$ -particle. The experimental results thus afford strong presumptive evidence that the particles liberated from nitrogen are atoms of hydrogen.

A far more decisive test, however, can be made by comparing the deflexion of the nitrogen particles with that of H atoms under similar conditions. For the purpose, a mixture of about one volume of hydrogen to two of carbon dioxide was stored in a gas-holder and circulated in place of air through the testing apparatus. The proportions of two gases were so adjusted that the stopping power of the mixture for  $\alpha$ -rays was equal to that of air. Under these conditions, the H atoms, like the nitrogen particles, are produced throughout the volume of the gas, and probably the relative distribution of H atoms along the path of the  $\alpha$ -rays is not very different from that of the nitrogen particles under examination. If the nitrogen particles are H atoms, we should expect the average deflexion to be nearly the same as for the H atoms liberated from the hydrogen mixture. A number of careful experiments showed that the ratio of the number of scintillations in + and - fields of equal value was so nearly identical in the two cases that the experiments were unable to distinguish between them. Since the two experiments were carried out under as nearly as possible identical conditions, the equality of the ratio shows that the long range particles liberated from nitrogen are atoms of hydrogen. The possibility that the particles may be of mass 2, 3, or 4 is definitely excluded.

In a previous paper I have given evidence that the long range particles

observed in dry air and pure nitrogen must arise from the nitrogen atoms themselves. It is thus clear that some of the nitrogen atoms are disintegrated by their collision with swift  $\alpha$ -particles and that swift atoms of positively charged hydrogen are expelled. It is to be inferred that the charged atom of hydrogen is one of the components of which the nucleus of nitrogen is built up.

While it has long been known that helium is a product of the spontaneous transformation of some of the radio-active elements, the possibility of disintegrating the structure of stable atoms by artificial methods has been a matter of uncertainty. This is the first time that evidence has been obtained that hydrogen is one of the components of the nitrogen nucleus.

It should be borne in mind that the amount of disintegration effected in nitrogen by the particles is excessively small, for probably on an average only one  $\alpha$ -particle in about 300,000 is able to get near enough to the nitrogen nucleus to liberate the atom of hydrogen with sufficient energy to be detected by the scintillation method. Even if the whole  $\alpha$ -radiation from 1 gramme of radium were absorbed in nitrogen gas, the volume of hydrogen set free would be only about  $1/300000$  of the volume of helium due to the collected  $\alpha$ -particles, viz., about  $5 \times 10^{-4}$  cub. mm per year. It may be possible that the collision of an  $\alpha$ -particle is effective in liberating the hydrogen from the nucleus without necessarily giving it sufficient velocity to be detected by scintillations. If this should prove to be the case, the amount of disintegration may be much greater than the value given above.

#### *Experiments with Solid Nitrogen Compounds*

A brief account will now be given of experiments with solid nitrogen compounds. Since the liberation of the particle from nitrogen is a purely atomic phenomenon, it was to be expected that similar particles would be liberated from nitrogen compounds in number proportional to the amount of nitrogen. To test this point, and also the nature of the particles, a number of compounds rich in nitrogen were examined. For this purpose I have employed the following substances, which were prepared as carefully as possible to exclude the presence of hydrogen in any form: —

1. Boron nitrite, kindly prepared for me by W.J. Shutt, in Manchester University.
2. Sodium nitride, titanium nitride and para-cyanogen, kindly prepared

for me by Sir William Pope and his assistants.

The apparatus used was similar in form to that given in fig. 1, except that the plates were 4 cm long. By means of a fine gauze, the powdered material was sifted uniformly as possible on a thin aluminium plate about 2 sq. cm in area. The weight of the aluminium plate was about 6 mgrm. per square centimetre, and usually about 4 to 5 mgrm. of the material per square centimetre was used. The stopping power of the aluminium plate for  $\alpha$ -particles corresponded to about 3.4 cm of air, and it was usually arranged that the average stopping power of the material was about the same as for the aluminium. In order to make the material adhere tightly to the plate, a layer of alcohol was first brushed on the material rapidly sifted into position, and the plate then dried.

Experiment showed that no detectable hydrogen contamination was introduced by the use of alcohol in this way. The zinc sulphide screen was placed outside the box close to an aluminium plate of stopping power equal to 5.2 cm of air which covered an opening in the end of the brass box. The aluminium carried was then placed in position to cover the end of the slits near the source, care being taken not to shake off any material. The air was exhausted and the number of scintillations on the screen counted.

1. With Material Facing the Source.
2. With plate reversed.

In the former case, the  $\alpha$ -particles were fired directly into the material under examination; in the latter case the  $\alpha$ -particles only fell on the material when their range was reduced to about half, when power of liberating swift atoms is much reduced. This method of reversal had the great advantage that no correction was necessary for unequal absorption of the H-particles from the source in different experiments.

In this way it was found that the nitrogen compounds examined gave a larger number of scintillations in position (1). The nature of these particles was examined by a method similar to that employed in the case of nitrogen and a direct comparison was made of the deflexion of the particles with that of H atoms liberated from a film of paraffin put in place of the nitrogen compound. In all experiments, the particles were found to be deflected to the same degree as H atoms from the paraffin and no trace of particles of mass 2, 3 or 4 was detected.

For films of equal average stopping power for  $\alpha$ -rays, it can readily be calculated from Bragg's rule that the relative stopping power of the nitrogen in the compounds is 0.67 for B.N., 0.74 for  $C_2N_2$ , 0.40 for titanium nitride,

taking the stopping power of sodium nitride as unity. Since the expulsion of long range nitrogen particles must be an atomic phenomenon, it was to be expected that the number of scintillations under the experimental conditions, after correction for the natural effect from the source, should be proportional to the relative values of stopping power given above. The observations with sodium nitride and titanium nitride were very consistent and the number of long range nitrogen particles was in the right proportion and about the same as that to be expected from the experiments with nitrogen gas. On the other hand, boron nitride and para-cyanogen gave between 1.5 and 2 times the number to be expected theoretically. In these experiments every precaution was taken to get rid of hydrogen and water vapour. Before use, the aluminium plates were heated in an exhausted quartz tube in an electric furnace nearly to its melting point to get rid of hydrogen and other gases. The films under examination were kept in a dessicator and heated in the electric furnace just before use and transferred at once to the testing vessel. Several control experiments were made, using preparations not containing nitrogen, viz., pure graphite and silica which had been kindly prepared for me by Sir William Pope. In both of these cases, the number of scintillations observed with the material facing the  $\alpha$ -rays was actually less than when the plate was reversed. This showed that some H atoms were liberated by the  $\alpha$ -rays from the heated aluminium. The control experiments were thus very satisfactory in showing that H atoms were not present in materials not containing nitrogen. Incidentally, they show that H atoms do not arise in appreciable numbers from carbon, silicon, or oxygen.

The increased effect in boron nitride and para-cyanogen naturally led to the suspicion that these preparations contained some hydrogen although every precaution was taken to avoid such a possibility. In the case of boron nitride there is also the uncertainty whether boron itself emits H atoms. This point has not yet been properly examined. On account of these uncertainties, experiments on solid nitrogen compounds were abandoned for the time, and experiments already described made directly on gaseous nitrogen.

It is of interest to note that a considerable contamination with hydrogen is required to produce the number of H atoms observed in these compounds. In the case of sodium nitride at least 50 c.c. of hydrogen must be present per gram of material. I am inclined to think that the H atoms liberated by the  $\alpha$ -rays from sodium nitride is due mainly, if not entirely to the nitrogen, and in the case of para-cyanogen, part of the effect is probably due to the presence of hydrogen or other hydrogen compound. It is hoped to examine this question in more detail later.

### *Short Range Atoms from Oxygen and Nitrogen*

In addition to the range H atoms liberated from nitrogen, the passage of  $\alpha$ -particles through oxygen as well as through nitrogen gives rise to much more numerous swift atoms, which have a range in air of about 9.0 cm compared with that of 7.0 cm for the colliding  $\alpha$ -particles. The method of determining the range and number of these atoms has been explained in a previous paper.<sup>11</sup>. It is there shown that these projected atoms arise from the passage of the  $\alpha$ -particles from radium C, the scintillations are much brighter than those due to H atoms, and more resemble  $\alpha$ -particles.

In the absence of definite information as to the nature of these atoms, it was provisionally assumed that they were atoms of oxygen or nitrogen carrying a single charge set in swift motion by close collision with  $\alpha$ - particles, for the observed range of these particles was in approximate accord with that calculated on these assumptions. At the same time it was pointed out that the agreement of the ranges of the atoms set in N and O was rather surprising, for it was to be anticipated that the range of the swifter N atoms should be about 19 per cent. greater than for the slower O atoms. The possibility that these swift atoms might prove to be fragments of disintegrated atoms was always present, but up till quite recently, I did not see any method of setting the question.<sup>12</sup>

As soon as the use of wide slits had proved successful in deciding the nature of the long range particles from nitrogen, experiments were made with the same apparatus and method to test the nature of the short range particles in O and N.

First consider the relative deflexion to be expected for an O atom which is set in motion by a direct impact with an  $\alpha$ -particle. The velocity of the O atom after the collision is  $2/5V$ , where  $V$  is the velocity of the incident  $\alpha$ -particle. The value of  $mu/e$  for the O atom carrying a single charge is easily seen to be 3.1 times that of the  $\alpha$ -particle before impact. Consequently the O atom with a single charge should be much more difficult to deflect than the  $\alpha$ -particle, and this is the case even if the former carries two charges.

To test these points, the apparatus was the same as that shown in fig. 1.

The source was 7.4 cm distant from the zinc sulphide screen, the and pieces, 1.2 cm long, being used as before to increase the deflexion of the rays.

---

<sup>11</sup>Rutherford, Phil. Mag., vol. 37, III, p. 571 (1919)

<sup>12</sup>Mr. G.S. Fulcher, of the National Research Council, U.S.A., sent me, in November, 1919, a suggestion that these atoms might to be  $\alpha$ -particles.

During an experiment, dried air or oxygen was circulated slowly through the apparatus to avoid radio-active contamination of the screen. In the case of oxygen, the scintillations observed on the screen were due to the O atoms with a small proportion of H atoms from the source. In the case of air, the scintillations on the screen were due partly to N atoms, some O atoms, and H atoms from the source and nitrogen. The actual number of short range N atoms appeared to be less than the number of O atoms under similar conditions.

The position of the microscope was fixed as before to give a convenient ratio for the number of scintillations on reversing the magnetic field. This ratio varies with the position of the microscope, and in the actual experiments had values between 0.2 and 0.4.

It was at once obvious that the atoms from O instead of being *less* deflected than the  $\alpha$ -particles, as they should be if they were O atoms, were *more* deflected. This at once excluded the possibility that the atoms from oxygen were actual atoms of oxygen carrying either one or two charges. Since helium is expelled in so many radio-active charges, it might be expected to be one of the components of light atoms, liberated by the intense collision. The deflexion of the atoms from O was, however, much too large to be accounted for in this way. To test this point, at the conclusion of the experiments with oxygen, a plate which had been exposed to thorium emanation was substituted for the radium source, and the bending of the rays of range 8.6 cm from thorium C was examined in a similar way. If an  $\alpha$ -particle were ejected from an O atom near the source, if would be bent like an  $\alpha$ -particle of range 9.0 cm; if produced near the end of the range of  $\alpha$ -rays, the amount of bending could not be more than for an  $\alpha$ -particle of range 7.0 cm, i.e., about 9 per cent. more than in the first case. Even supposing the particles were liberated uniformly along the path of the  $\alpha$ -rays and moved in the same line as the colliding particle, the average bending would not differ by 5 per cent. from that the  $\alpha$ -particle from thorium C. If, as seems probable, some of the atoms are liberated at an angle with the incident particles, the average amount of bending of the beam would be less than the above, and in all probability less than for the  $\alpha$ -particles from thorium C. Actually the bending observed was about 20 per cent. greater, showing that the hypothesis that the atoms O are charged atoms of helium is quite untenable.

If the atoms from O were H atoms, they would be more bent than the  $\alpha$ -particles, but would have a maximum range of 28 cm instead of the 9.0 cm observed. It thus seemed clear from this evidence that the atom must be of mass intermediate between 1 and 4, while from consideration of the

range of the particles and their amount of deflexion it was clear that the atom carried two units of charge.

In order to make a more decisive test, the deflexion of O atoms in a positive and negative field of given value was directly compared with the deflexion of H atoms from a mixture of hydrogen and carbonic acid, in the ratio of about 1 to 2 in volume. In order to absorb completely the O atoms from  $\text{CO}_2$  aluminium foil was placed over the zinc sulphide screen, so that the total absorption between the source and screen corresponded to slightly more than 9 cm of air. In both experiments, the atoms, under examination are produced in the gas between the slits, and probably the relative distribution along the path of the  $\alpha$ -rays is not markedly different in the two cases. The ratio for reversing the field in the two experiments were found to be nearly equal; but, as an average of a number of experiments, the H atoms were slightly more bent than the atoms from O. From a number of experiments it was concluded that the difference in deflexion did not on the average amount to more than 5 per cent., although from the nature of the observations it was difficult to fix the difference with any certainty.

From these data and the range of the atoms from O in air, we can deduce the mass of the particle liberated from oxygen.

Let  $m$  = mass of the atom O;  
 $u$  = its maximum velocity near the source,  
 $E$  = charge

Let  $M, V, E$  be the corresponding values for the incident  $\alpha$ -particles and  $m'u'e$  the values for the H atoms liberated close to the source.

Taking into account that the particle from O of range 9 cm is steadily reduced in velocity in passing through the 7.4 cm of oxygen between the source and screen, it can easily be calculated that its average deflexion by the magnetic field is proportional to  $1.14E/mu$  in place of  $E/mu$  in a vacuum.

In a similar way, the deflexion of the H atom is proportional to  $1.05e/m'u'$ , the correction in this case for change of velocity being smaller, and estimated to be about 5 per cent. Now we have seen that the experimental results showed that the atoms from O were bent about 5 per cent. less than the H atoms. Consequently

$$1.14E/mu = \frac{1.05}{1.05} \cdot e/m'u' = 1.25 \cdot E/mV,$$
$$\mathbf{1.14MV = 1.25mu,} \quad (1)$$

since it has been calculated and verified by experiment that the deflexion of the H atom in a magnetic field is 1.25 times that of the  $\alpha$ - particle which

sets it in motion (see Paper II, *loc. cit.*). Also in a previous paper, III, I have given reasons for believing that the range  $x$  of mass  $m$  and initial velocity  $u$ , carrying a double charge, is given by

$$\frac{x}{R} = \frac{m}{M} \cdot \left( \frac{u}{V} \right)^3,$$

where  $R$  is the range of the  $\alpha$ -particle of mass  $M$  and velocity  $V$ . Since  $x = 9.0$  cm for the atoms from O set in motion by collision with  $\alpha$ -particles from radium C of range 7 cm,

$$\frac{x}{R} = 1.29,$$

and taking  $M = 4$

$$mu^3 = 5.16V^3. \quad (2)$$

A formula of this type has been shown to account for the range of the H atom, and there is every reason to believe it is fairly accurate over such a short difference of range.

From (1) and (2)

$$u = 1.19V,$$

$$m = 3.1.$$

Considering the difficulty of obtaining accurate data, the value  $m = 3.1$  indicated that the atom has a mass about 3 and this value will be taken as the probable value in later discussions.

When air was substituted for oxygen it was not possible to distinguish any difference between the bending of the short range atoms in the two cases. Since the short range atoms from air arise mainly from the nitrogen, we may consequently conclude that the short range atoms liberated by the passage of particles through oxygen or nitrogen consist of atoms of mass 3, carrying a double charge, and initially projected with a velocity  $1.19V$ , where  $V$  is the velocity of the colliding  $\alpha$ -particle.

There seems to be escape from the conclusion that these atoms of mass 3 are liberated from the atoms of oxygen or nitrogen as a result of an intense collision with an  $\alpha$ -particle. It is thus reasonable to suppose that atoms of mass 3 are constituents of the structure of the nuclei of the atoms of both oxygen and nitrogen. We have shown earlier in the paper that hydrogen is also one of the constituents of the nitrogen nucleus. It is thus clear that the nitrogen nucleus can be disintegrated in two ways, one by the expulsion of an H atom and the other by the expulsion of an atom of mass 3 carrying

two charges. Since now these atoms of mass 3 are five to ten times as numerous as the H atoms, it appears that these two forms of disintegration are independent and not simultaneous. From the rareness of the collisions it is highly improbable that a single atom undergoes both types of disintegration.

Since the particles ejected from O and N are not produced at the source, but along the path of the  $\alpha$ -particles, it is difficult to determine their mass and velocity with the precision desired. To overcome this drawback, attempts were made to determine the deflection of O atoms released from a mica plate placed over the source. In consequence of hydrogen in combination in the mica, the H atoms falling on the screen were so numerous compared with the O particles, and their deflexion under the experimental conditions so nearly alike, that it was difficult to distinguish between them.

### *Energy Considerations*

In close collision between an  $\alpha$ -particle and an atom, the laws of conservation of energy and of momentum appear to hold,<sup>13</sup> but, in cases where the atoms are disintegrated, we should not necessarily expect these laws to be valid, unless we are able to take into account the charge of energy and momentum of the atom in consequence of its disintegration.

In the case of the ejection of a hydrogen atom the nitrogen nucleus, the data available are insufficient, for we do not know with certainty either the velocity of the H atom or the velocity of the  $\alpha$ -particle after the collision.

If we are correct in supposing that atoms of mass 3 are liberated from O and N atoms, it can be easily calculated that there is a slight gain of energy as a result of the disintegration. If the mass is 3 exactly, the velocity of escape of the atom is  $1.2V$ , where  $V$  is the velocity of the impinging  $\alpha$ -particle.

Thus,

$$\frac{\text{Energy of liberated atom}}{\text{energy of } \alpha\text{-particle}} = \frac{3 \cdot 1.44}{4} = 1.08,$$

or there is a gain of 8 per cent. in energy of motion, even though we disregard the subsequent motion of the disintegrated nucleus or of the colliding  $\alpha$ -particle. This extra energy must be derived from the nitrogen or oxygen nucleus in the same way that the  $\alpha$ -particle gains energy of motion in escaping from the radio-active atom.

<sup>13</sup>Rutherford, Phil. Mag., vol. 37, p. 562 (1919)

For the purpose of calculation, consider a direct collision between an  $\alpha$ -particle and an atom of mass 3. The velocity of the latter is  $8/7V$ , where  $V$  is the velocity of the  $\alpha$ -particle, and its energy is 0.96 of the initial energy of the  $\alpha$ -particle. No doubt, in the actual case of a collision with the O or N atom, in which the atom of mass is liberated, the  $\alpha$ -particle comes under the influence of the main field of the nucleus, as well as of that of the part of mass 3 immediately in its path. Under such conditions, it is not to be expected that the  $\alpha$ -particle can give 0.96 of its energy to the escaping atom, but the latter acquires additional energy due to the repulsive field of the nucleus.

In our ignorance of the constitution of the nuclei and the nature of the force in their immediate neighbourhood, it is not desirable to enter into speculations as to the mechanism of the collision at this stage, but it may be possible to obtain further information by a study of the trails of  $\alpha$ - particles through oxygen or nitrogen by the well-known expansion method of C.T.R. Wilson. In a previous paper,<sup>14</sup> I discussed the photograph obtained by Mr. Wilson, in which there is a sudden change of  $43^\circ$  in the direction of the trail, with the appearance of a short spur at the fork. Evidence was given that the relative length of the tracks of the  $\alpha$ -particle and of the spur were in rough accord with the view that the spur was due to the recoiling oxygen atom. This is quite probably the case, for the general evidence shows that the atoms of mass 3, after liberation, travel nearly in the direction of the  $\alpha$ -particle, and an oblique collision may not result in the disintegration of the atom.

Recently, Dr. Shimizu, of the Cavendish Laboratory, has devised a modification of the Wilson expansion apparatus, in which expansions can be periodically produced several times a second, so that the trails of many particles can be inspected in a reasonable time. Under these conditions, both Shimizu and myself saw on several occasions what appeared to be branching trails of an  $\alpha$ -particle in which the lengths of the two tracks were comparable. Eye observations of this kind are too uncertain to regard them with much confidence, so arrangements are being made by Mr. Shimizu to obtain photographs, so that the tracks can be examined in detail at leisure. In this way we may hope to obtain valuable information as to the conditions which determine the disintegration of the atoms, and on the relative energy communicated to the three systems involved, viz., the  $\alpha$ -particle, the escaping atom, and the residual nucleus.

So far no definite information is available as to the energy of the  $\alpha$ -particle

---

<sup>14</sup>Rutherford, Phil. Mag., vol. 37, p. 577 (1919).

required to produce disintegration, but the general evidence indicates that fast  $\alpha$ -particles, of range about 7 cm in air, are more effective than  $\alpha$ -particles of range about 4 cm. This may not be connected directly with the actual energy required to effect the disintegration of the atom itself, but rather to the inability of the slower  $\alpha$ -particle under the repulsive field to approach close enough to the nucleus to be effective in disrupting it. Possibly the actual energy required to disintegrate the atom is small compared with the energy of the  $\alpha$ -particle.

If this be the case, it may be possible for other agents of less energy than the  $\alpha$ -particle to effect the disintegration. For example, a swift electron may reach the nucleus with sufficient energy to cause its disintegration, for it moves in an attractive and not a repulsive field as in the case of the  $\alpha$ -particle. Similarly, a penetrating  $\gamma$ -ray may have sufficient energy to cause disintegration. It is thus of great importance to test whether oxygen or nitrogen or other elements can be disintegrated under the action of swift cathode rays generated in a vacuum tube. In the case of oxygen and nitrogen, this could be tested simply by observing whether a spectrum closely resembling helium is given by the gas in the tube, after an intense bombardment of a suitable substance, by electrons. Experiments of this type are being undertaken by Dr. Ishida in the Cavendish Laboratory, every precaution being taken by the heating of the vacuum tube of special glass and electrodes to a high temperature to ensure the removal of any occluded helium which may be initially in the material. Helium has previously been observed by several investigators in vacuum tubes and is known to be released from substances by bombardment with cathode rays. The proof of the actual production of helium in such cases is exceedingly difficult, but the recent improvements in vacuum tube technique may make it easier to give a decisive answer to this important question.

### *Properties of the New Atom*

We have shown that atoms of mass about 3 carrying two positive charges are liberated by  $\alpha$ -particles both from nitrogen and oxygen, and it is natural to suppose that these atoms are independent units in the structure of both gases. Since probably the charged atom during its flight is the nucleus of a new atom without any external electrons, we should anticipate that the new atom when it has gained two negative electrons should have physical

and chemical properties very nearly identical with those of helium, but with a mass 3 instead of 4. We should anticipate that the spectrum of helium and this isotope should be nearly the same, but on account of the marked difference in the relative masses of the nuclei, the displacement of the lines should be much greater than in the case of isotopes of heavy elements like lead.

It will be remembered that Bourget, Fabry, and Buisson,<sup>15</sup> from an examination of the width of the lines in the spectrum of nebular, conclude that the spectrum arises from an element of atomic mass about 2.7 or 3 in round numbers. It is difficult, however, on modern views to suppose that the spectrum of the so-called "nebulium" can be due to an element of nuclear charge 2 unless the spectrum under the conditions existing in nebular are very different from those observed in the laboratory. The possible origin of the spectrum of nebulium has been discussed at length by Nicholson<sup>16</sup> on quite other lines, it is not easy at the moment to see how the new atoms from oxygen or nitrogen can be connected with the nebular material.

Since probably most of the helium in use is derived, either directly or indirectly, from the transformation of radio-active materials, and these, as far as we know, always give rise to helium of mass 4, the presence of an isotope of helium of mass 3 is not likely to be detected in such sources. It would, however, be of great interest to examine whether the isotope may be present in cases where the apparent presence of helium is difficult to connect with radio-active; for example, in beryl, drawn attention to by Strutt.<sup>17</sup> This is based on the assumption that the atom of mass 3 is stable. The fact that it survives the intense disturbance of its structure due to a close collision with an  $\alpha$ -particle is an indication that it is a structure difficult to disintegrate by external forces.

### *Constitution of Nuclei and Isotopes*

In considering the possible constitution of the elements, it is natural to suppose that they are built up ultimately of hydrogen nuclei and electrons. On this view the helium nucleus is composed of four hydrogen nuclei and two negative electrons with a resultant charge of two. The fact that the mass of

<sup>15</sup>Bourget, Fabry and Buisson, 'C.R.', April 6, May 18 (1914).

<sup>16</sup>Nicholson, Roy. Ast. Soc., vol. 72, N 1, p. 49 (1911); vol. 74, N 7, p. 623 (1914).

<sup>17</sup>Strutt, Roy. Soc. Proc., A, vol. 80, p. 572 (1908).

the helium atom 3.997 in terms of oxygen 16 is less than the mass of four hydrogen atoms, viz., 4.032, has been generally supposed to be due to the close interaction of the fields in the nucleus resulting in a smaller electromagnetic mass than sum of the masses of the individual components. Sommerfeld<sup>18</sup> has concluded from this fact that the helium nucleus must be a very stable structure which would require intense forces to disrupt it. Such a conclusion is in agreement with experiment, for no evidence has been obtained to show that helium can disintegrated by the swift  $\alpha$ -particles which are able to disrupt the nuclei of nitrogen and oxygen. In his recent experiments on the isotopes of ordinary elements Aston<sup>19</sup> has shown that within the limit of experimental accuracy the masses of all the isotopes examined are given by whole numbers when oxygen is taken as 16. The only exception is hydrogen, which has a mass 1.008 in agreement with chemical observations. This does not exclude the probability that hydrogen is the ultimate constituent of which nuclei are composed, but indicated that either the grouping of the hydrogen nuclei and electrons is such that the average electromagnetic mass is nearly 1, or what is more probable, that the secondary units, of which the atom is mainly built up., it e.g., helium or its isotope, have a mass given nearly by a whole number when O is 16.

The experimental observations made so far are unable to settle whether the new atom has a mass exactly 3, but from the analogy with helium we may expect the nucleus of the new atom to consist of three H nuclei and one electron, and to have a mass more nearly 3 than the sum of the individual masses in the free state.

If we are correct in this assumption it seems very likely that one electron can also bind two H nuclei and possibly also one H nucleus. In the one case, this entails the possible existence of an atom of mass nearly 2 carrying one charge, which is to be regarded as an isotope of hydrogen. In the other case, it involves the idea of the possible existence of an atom of mass 1 which has zero nucleus charge. Such an atomic structure seems by no means impossible. On present views, the neutral hydrogen atom is regarded as a nucleus of unit charge with an electron attached at a distance, and the spectrum of hydrogen is ascribed to the movements of this distant electron. Under some conditions, however, it may be possible for an electron to combine much more closely with the H nucleus. forming a kind of neutral doublet. Such an atom would have very novel properties. Its external field would be practically zero, except very close to the nucleus, and in consequence it should be able

---

<sup>18</sup> Sommerfeld, Atombau und Spektrallinien, p. 538. Vieweg and Son., 1919.

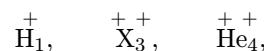
<sup>19</sup> Aston, Phil. Mag., December, 1919: April and May, 1920

to move freely through matter. Its presence would probably be difficult to detect by the spectroscope, and it may be impossible to contain it in a sealed vessel. On the other hand, it should enter readily the structure of atoms, and may either unite with the nucleus or be disintegrated by its intense field, resulting possibly in the escape of a charged H atom or an electron or both.

If the existence of such atoms be possible, it is to be expected that they may be produced, but probably only in very small numbers, in the electric discharge through hydrogen, where both electrons and H nuclei are present in considerable numbers. It is the intention of the writer to make experiments to test whether any indication of the production of such atoms can be obtained under these conditions.

The existence of such nuclei may not be confined to mass 1 but may be possible for masses 2, 3, or 4, or more, depending on the possibility of combination between the doublets. The existence of such atoms seems almost necessary to explain the building up the nuclei of heavy elements; for unless we suppose the production of charged particles of very high velocities it is difficult to see how any positively charged particle can reach the nucleus of a heavy atom against its intense repulsive field.

We have seen that so far the nuclei of three light atoms have been recognised experimentally as probable units of atomic structure, viz.,



where the subscript represents the mass of the element.

In considering the possible ways in which nuclei can be build up, difficulties at once arise, for many combinations of these units with negative electrons are possible to give an element of the required nuclear charge and mass. In our complete ignorance of the laws of force close to the nuclei, no criterion is available as to the stability or relative probability of the theoretical system. With the exception of a few elements which can exist in the gaseous state, the possible isotopes of the elements have not yet been settled. When further information is available as to the products of the disintegration of other elements than the two so far examined, and more complete data have been obtained as to the number and mass of the isotopes, it may be possible to deduce approximate rules which may serve as a guide to the mode in which the nuclei are built up from the simpler units. For these reasons it seems premature at this stage to attempt to discuss with any detail the possible structure of even the lighter and presumably less complex atoms. It may, however, be of some interest to give an example to illustrate a possible method of the formation of isotopes in the case of

the lighter elements. This is based on the view that probably in many cases a helium of mass 4 may be substituted in the complex structure for the corresponding nucleus of mass 3 without seriously interfering with the stability of the system. In such a case, the nuclear charge remains uncharged but the masses differ by unity.

For example, take the case of lithium of nuclear charge 3 and atomic mass about 7. It is natural to suppose that the nucleus is composed of helium or its isotope of mass 3 with one binding electron. The three possible combinations shown in fig. 3 (p. 398).

On this view, at least three isotopes of lithium of mass 6, 7, and 8 are theoretically probable, but even if the combinations were equally stable, the question of their relative abundance in the element lithium on the earth will be dependent on many factors of which we know nothing; for example, the mode of actual formation of such nuclei, the relative amount of the combining units present, and the probability of their combinations.

The experimental results given in the paper support, as far as they go, the view that the atoms of hydrogen and of mass 3 are important units in the nuclear structure of nitrogen and oxygen. In the latter case, one could

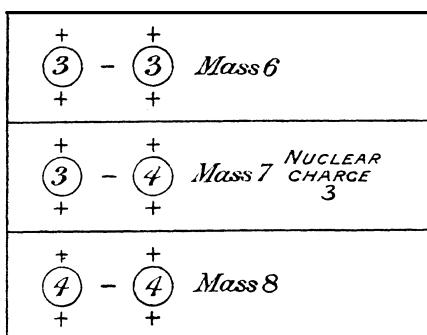


Figure 3:

*a priori* have supposed that oxygen was in some way a combination of four helium nuclei of mass 4. It seems probable that the mass 3 is an important unit of the nuclei of light atoms in general, but it is not unlikely, with increasing complexity of the nuclei and corresponding increase of the electric field, the structure of mass suffer a rearrangement and tend to revert to the presumably more stable nucleus of mass 4. This may be the reason why helium of mass 4 always appears to be expelled from the radio-active atoms,

while the isotope of mass 3 arises in the artificial disintegration of lighter atoms like oxygen and nitrogen. It has long been known that for many of the elements the atomic weights can be expressed by the formula  $4n$  or  $4n + 3$ , where  $n$  is an integer, suggesting that atoms of mass 3 and 4 are important units of the structure of nuclei.<sup>20</sup>

### *Structure of Carbon, Oxygen, and Nitrogen Nuclei*

In the light of the present experiments, it may be of interest to give some idea, however crude, of the possible formation of the above atoms to account for the experimental facts. It will be remembered that nitrogen alone gives rise to H atoms while carbon and oxygen do not. Both nitrogen and oxygen give rise to atoms of mass 3, while carbon has not yet been investigated from this point of view. A possible structure is shown in fig. 4 when the masses and charges of the combining units are indicated. Negative electrons are represented by the symbol —,

The carbon nucleus is taken to consist of four atoms of mass 3 and charge 2, and two binding electrons. The change to nitrogen is represented by the addition of two H atoms with a binding electron and an oxygen nucleus by the substitution of a helium nucleus in place of the two H atoms.

We can see from this type of structure that the chance of a direct collision with one of the four atoms of mass 3 in nitrogen is much greater than the chance of removing an H atom, for it is to be anticipated that the main nucleus would screen the H atom from a direct collision except in restricted regions facing the H atoms. This serves to illustrate why the number of H atoms of mass 3 liberated from nitrogen should be much greater than the number of H atoms released under corresponding conditions. It should be borne in mind that the structures outlined are purely illustrative and no importance is attached to the particular arrangement employed.

It is natural to inquire as to the nature of the residual atoms after the disintegration of oxygen and nitrogen, supposing that they survive the collision and sink into a new stage of temporary or permanent equilibrium.

<sup>20</sup>From these and other considerations, Harkins (Phys. Rev., vol. 15, p. 73 (1920)) has proposed a constitutional formula for all the elements. The combining units employed by him are electrons and atoms of mass 1, 3, and 4 of nuclear charges 1, 1 and 2, respectively. The unit of mass 3 is taken by him to have a nucleus charge of 1 and not 2, and is thus to be regarded as an isotope of hydrogen and not an isotope of helium.

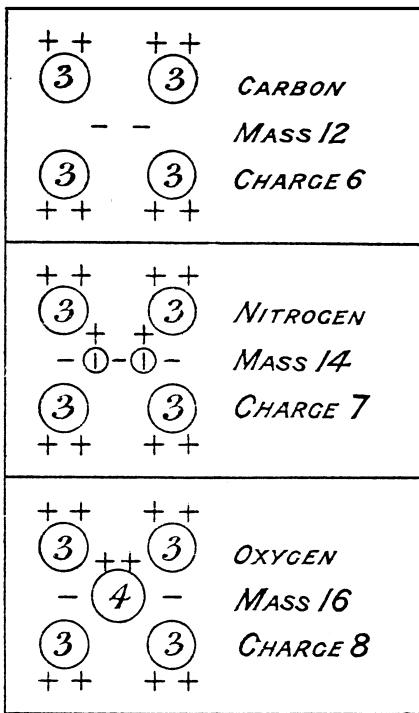


Figure 4:

The expulsion of an H atom carrying one charge from nitrogen should lower the mass by 1 and the nuclear charge by 1. The residual nucleus should thus have a nuclear charge 6 and mass 13, and should be an isotope of carbon. If a negative electron is released at the same time, the residual atom becomes an isotope of nitrogen.

The expulsion of a mass 3 carrying two charges from nitrogen, probably quite independent of the release of the H atom, lowers the nuclear charge by 2 and the mass by 3. The residual atom should thus be an isotope of boron of nuclear charge 5 and mass 11. If an electron escapes as well, there remains an isotope of carbon of mass 11. The expulsion of a mass 3 from oxygen gives rise to a mass 13 of nuclear charge 6, which should be an isotope of carbon. In case of the loss of an electron as well, there remains an isotope of nitrogen of mass 13. The data at present available are quite insufficient to distinguish between these alternatives.

It is intended to continue experiments, to test whether any evidence can

be obtained of the disintegration of other light atoms besides nitrogen and oxygen. The problem is more difficult in the case of elements which cannot be conveniently obtained in the gaseous state, since it is not an easy matter to ensure the absence of hydrogen or to prepare uniform thin films of such substances. For these reasons, and the strain involved in counting scintillations under difficult conditions, further progress is not likely to be rapid.

I am indebted to my assistant, G.A.R. Crowe, for the preparation of the radio-active sources and his help in counting; also to Mr. J. Chadwick and Dr. Ishida for assistance in counting scintillations in some of the later experiments.



# 125. On the Curvature of Space

Aleksandr Friedmann

TRANSLATED BY BRIAN DOYLE

(*Zeitschrift für Physik* 10, 377–386 [1922])

In deriving a cosmological model from his general theory of relativity, Einstein somewhat arbitrarily opted for a static universe. The mathematical consequence of this decision was a nonzero value for one of the constants of integration, the so-called cosmological constant,  $\Lambda$ . From a Newtonian analogue,  $\Lambda$  can be viewed as representing a repulsive force that increases with distance and that keeps the universe from collapsing under gravitational attraction.

In the following selection, the Russian mathematician Aleksandr Friedmann considers nonstatic models for the first time. By treating the spatial curvature of the universe as a function of time, he shows the possibility of nonstationary worlds with positive and negative curvature. These dynamic world models became especially important several years later when the universe of galaxies was found to be expanding. Although Friedmann's nonstatic cosmology was for some years overlooked by astronomers, Einstein noticed this paper and within a few months issued a one-paragraph critique in the same journal, only to retract his objection early in 1923.

Friedmann begins with the general idea that at any given instant of time the cosmological model represents a space of positive spatial curvature  $R(t)$ . If  $R$  is independent of time, then the stationary world models of Einstein and Wilhelm de Sitter follow. If  $R(t)$  depends only on the time variable, then a variety of monotonically expanding or periodically oscillating models result, depending on the value chosen for  $\Lambda$ . Friedmann notes that with  $\Lambda = 0$ , there follows an oscillating model whose period depends on the total mass of the universe.

In a second paper<sup>1</sup> Friedmann considers models with negative curvature. He finds a nonstationary world with negative spatial curvature and positive matter density, but no static model. Friedmann also notes in this later contribution that Einstein's field equations do not suffice to extract a conclusion about the finiteness of space without some supplementary assumptions.

1. A. Friedmann, *Zeitschrift für Physik* 21, 326 (1924).

## I

A. In their well-known works on general cosmological questions, Einstein<sup>1</sup> and de Sitter<sup>2</sup> arrive at two possible types of universe: Einstein obtains the so-called cylindrical world, in which space<sup>3</sup> possesses a constant curvature independent of time and in which the radius of curvature is connected with the total mass of matter existing in space. De Sitter obtains a spherical world in which not only space but also the world can be spoken of, in a certain sense, as a world of constant curvature.<sup>4</sup> In doing so, certain assumptions about the matter tensor are made by both Einstein and de Sitter; these correspond to the incoherence of matter and its being relatively at rest, e.g. the velocity of matter is assumed to be sufficiently small in comparison with the fundamental velocity,<sup>5</sup> the velocity of light.

The goal of this notice is, first, the derivation of the cylindrical and spherical worlds (as special cases) from some general assumptions and, second, the proof of the possibility of a world whose spatial curvature is constant with respect to three coordinates that are permissible spatial coordinates and that depend on time, e.g. on the fourth (time) coordinate. This new type is, as far as its remaining properties are concerned, an analogue of the Einsteinian cylindrical universe.

B. The assumptions on which we shall base our considerations break down into two classes. To the first class belong assumptions that coincide well with the assumptions of Einstein and de Sitter. They refer to the equations that the gravitational potentials satisfy and to the state and motion of matter. To the second class belong assumptions about the general, so-to-speak geometric, character of the world. From our hypothesis the cylindrical world of Einstein and the spherical world of de Sitter follow as special cases.

*Class 1* The assumptions of the first class are the following:

1. The gravitational potentials satisfy the Einstein system of equations with the cosmological term, which we may also set equal to zero:

$$R_{ik} - \frac{1}{2} g_{ik} \bar{R} + \lambda g_{ik} = -\kappa T_{ik} \quad (i, k = 1, 2, 3, 4). \quad (1)$$

Here the  $g_{ik}$  are the gravitational potentials,  $T_{ik}$  is the matter tensor,  $\kappa$  is a constant,  $\bar{R} = g^{ik} R_{ik}$ , [the cosmological constant is denoted by  $\lambda$ ] and  $R_{ik}$  is determined by the equations

$$R_{ik} = \frac{\partial^2 \log(\sqrt{g})}{\partial x_i \partial x_k} - \frac{\partial(\log \sqrt{g})}{\partial x_\sigma} \begin{Bmatrix} ik \\ \sigma \end{Bmatrix} - \frac{\partial}{\partial x_\sigma} \begin{Bmatrix} ik \\ \sigma \end{Bmatrix} + \begin{Bmatrix} i\alpha \\ \sigma \end{Bmatrix} \begin{Bmatrix} k\sigma \\ \alpha \end{Bmatrix}, \quad (2)$$

where the  $x_i$  ( $i = 1, 2, 3, 4$ ) are world coordinates and  $\begin{Bmatrix} ik \\ l \end{Bmatrix}$  are the Christoffel symbols of the second kind.<sup>6</sup>

2. The matter is incoherent and relatively at rest. Stated less strongly, the relative velocities of matter are vanishingly small in comparison with the velocity of light. In consequence of these assumptions, the matter tensor is given by the equations

$$\begin{aligned} T_{ik} &= 0 \text{ for } i \text{ and } k \neq 4 \\ T_{44} &= c^2 \rho g_{44}. \end{aligned} \quad (3)$$

Here  $\rho$  is the density of matter and  $c$  is the fundamental velocity. Moreover, the world coordinates are divided into three spatial coordinates  $x_1, x_2, x_3$  and the time coordinate  $x_4$ .

*Class 2* The assumptions of the second class are the following:

1. After distribution of the three spatial coordinates  $x_1, x_2, x_3$ , we have a space of constant curvature, which, however, may depend on  $x_4$ , the time coordinate. The interval<sup>7</sup>  $ds$ , determined by  $ds^2 = g_{ik} dx_i dx_k$ , can be brought into the following form by the introduction of suitable spatial coordinates:

$$\begin{aligned} ds^2 &= R^2(dx_1^2 + \sin^2 x_1 dx_2^2 + \sin^2 x_1 \sin^2 x_2 dx_3^2) \\ &\quad + 2g_{14} dx_1 dx_4 + 2g_{24} dx_2 dx_4 + 2g_{34} dx_3 dx_4 \\ &\quad + g_{44} dx_4^2. \end{aligned} \quad (4)$$

Here  $R$  depends only on  $x_4$  and it is proportional to the radius of curvature of space, which may therefore change with time.

2. In the expression for  $ds^2$ , the  $g_{14}, g_{24}, g_{34}$  can be made to vanish by a suitable choice of the time coordinate. In brief, time is orthogonal to space. It seems to me that no physical or philosophical grounds can be given for the second assumption. It serves exclusively to simplify the calculation. One must still notice that the worlds of Einstein and de Sitter are contained in our assumptions as special cases.

In consequence of assumptions 1 and 2,  $ds^2$  can be brought into the form

$$ds^2 = R^2(dx_1^2 + \sin^2 x_1 dx_2^2 + \sin^2 x_1 \sin^2 x_2 dx_3^2) + M^2 dx_4^2, \quad (5)$$

where  $R$  is a function of  $x_4$  and  $M$  depends, in the general case, on all four world coordinates. The Einstein universe is obtained if one replaces  $R^2$  by  $-R^2/c^2$  in equation (5) and if one also sets  $M = 1$ , whereby  $R$  signifies the constant (independent of  $x_4$ ) radius of curvature of space.

$$d\tau^2 = -\frac{R^2}{c^2}(dx_1^2 + \sin^2 x_1 dx_2^2 + \sin^2 x_1 \sin^2 x_2 dx_3^2) + dx_4^2 \quad (6)$$

## VIII. RELATIVITY AND COSMOLOGY

The universe of de Sitter is obtained if one replaces  $R^2$  by  $-R^2/c^2$  and  $M$  by  $\cos x_1$  in equation<sup>8</sup> (5)

$$d\tau^2 = -\frac{R^2}{c^2} (dx_1^2 + \sin^2 x_1 dx_2^2 + \sin^2 x_1 \sin^2 x_2 dx_3^2) + \cos^2 x_1 dx_4^2 \quad (7)$$

c. Now we must still strike an agreement about the boundaries within which the world coordinates are confined, e.g. what points of the 4-dimensional manifold we will treat as different. Without engaging in a more detailed motivation, we shall assume that the spatial coordinates are confined to the following intervals:  $x_1$  in the interval  $(0, \pi)$ ,  $x_2$  in the interval  $(0, \pi)$ , and  $x_3$  in the interval  $(0, 2\pi)$ . With respect to the time coordinate we make, for the present, no restricting assumptions, but we shall consider this question further below.

### II

a. From equations (3) and (5) it follows, if one sets  $i = 1, 2, 3$  and  $k = 4$  in equation (1), that

$$R'(x_4) \frac{\partial M}{\partial x_1} = R'(x_4) \frac{\partial M}{\partial x_2} = R'(x_4) \frac{\partial M}{\partial x_3} = 0.$$

Two cases arise. (1)  $R'(x_4) = 0$ ,  $R$  is independent of  $x_4$ . We shall designate this world as a *stationary* world. (2)  $R'(x_4) \neq 0$ ,  $M$  depends only on  $x_4$ . This shall be called a *nonstationary* world.

We consider, first, the stationary world and write the equations (1) for  $i, k = 1, 2, 3$  and moreover  $i \neq k$ . Then we obtain the following system of formulae:

$$\begin{aligned} \frac{\partial^2 M}{\partial x_1 \partial x_2} - \cotg x_1 \frac{\partial M}{\partial x_2} &= 0 \\ \frac{\partial^2 M}{\partial x_1 \partial x_3} - \cotg x_1 \frac{\partial M}{\partial x_3} &= 0 \\ \frac{\partial^2 M}{\partial x_2 \partial x_3} - \cotg x_2 \frac{\partial M}{\partial x_3} &= 0. \end{aligned}$$

The integration of these equations yields the following expression for  $M$ :

$$M = A(x_3, x_4) \sin x_1 \sin x_2 + B(x_2, x_4) \sin x_1 + C(x_1, x_4), \quad (8)$$

where  $A, B, C$  are arbitrary functions of their arguments. If we solve the equations (1) for  $R_{ik}$  and eliminate the unknown density<sup>9</sup>  $\rho$  from the still-unused equations, we obtain, if we insert for  $M$  equation (8), the following two possibilities for

$M$  after some long, but elementary calculations:

$$M = M_0 = \text{const.} \quad (9)$$

$$M = (A_0 x_4 + B_0) \cos x_1, \quad (10)$$

where  $M_0, A_0$  and  $B_0$  are constants.

If  $M$  is equal to a constant, then the stationary world is the cylindrical world. Here it is advantageous to work with the gravitational potentials of equation (6). If we determine the density and the quantity  $\lambda$ , then the well-known result of Einstein is obtained:

$$\lambda = \frac{c^2}{R^2}, \quad \rho = \frac{2}{\kappa R^2}, \quad \bar{M} = \frac{4\pi^2}{\kappa} R,$$

where  $\bar{M}$  denotes the total mass of space.<sup>8</sup>

In the second possible case, when  $\bar{M}$  is given by equation (10), we get, by means of a judicious transformation<sup>10</sup> of  $x_4$ , the spherical world of de Sitter in which  $M = \cos x_1$ . With the help of equation (7) we obtain the relations of de Sitter:

$$\lambda = 3c^2/R^2, \quad \rho = 0, \quad \bar{M} = 0.$$

We thus have the following result: *the stationary world is either the Einstein cylindrical world or the de Sitter spherical world.*

b. We now want to consider the nonstationary world.  $M$  is now a function of  $x_4$ . By an appropriate choice of  $x_4$  one can obtain  $M = 1$ , without loss of generality. In order to couple to our customary presentation, we give  $ds^2$  a form that is analogous to equations (6) and (7):

$$d\tau^2 = \frac{-R^2(x_4)}{c^2} (dx_1^2 + \sin^2 x_1 dx_2^2 + \sin^2 x_1 \sin^2 x_2 dx_3^2) + dx_4^2. \quad (11)$$

Our task is now the determination of  $R$  and  $\rho$  from the equations (1). It is clear that the equations (1) with different indices yield nothing. The equations (1) for  $i = k = 1, 2, 3$  give the relation

$$\frac{R'^2}{R^2} + \frac{2RR''}{R^2} + \frac{c^2}{R^2} - \lambda = 0. \quad (12)$$

The equations (1) with  $i = k = 4$  yield the relation

$$\frac{3R'^2}{R^2} + \frac{3c^2}{R^2} - \lambda = \kappa c^2 \rho \quad (13)$$

with

$$R' = \frac{dR}{dx_4} \quad \text{and} \quad R'' = \frac{d^2R}{dx_4^2}.$$

Because  $R' \neq 0$ , the integration of equation (12), if we write  $t$  for  $x_4$ , gives the following equation:

$$\frac{1}{c^2} \left( \frac{dR}{dt} \right)^2 = \frac{A - R + \frac{\lambda}{3c^2} R^3}{R}, \quad (14)$$

where  $A$  is an arbitrary constant. From this equation, we obtain  $R$  through the inversion of an elliptic integral, e.g. through the solution for  $R$  of the equation

$$t = \frac{1}{c} \int_a^R \sqrt{\frac{x}{A - x + \frac{\lambda}{3c^2} x^3}} dx + B \quad (15)$$

in which  $B$  and  $a$  are constants. Attention must still be paid to the usual conditions about sign variation in the square root. The mass density,  $\rho$ , may be determined from equation (13):

$$\rho = \frac{3A}{\kappa R^3}. \quad (16)$$

The constant  $A$  is expressed in terms of the total mass of space  $\bar{M}$  in the following way:

$$A = \frac{\kappa \bar{M}}{6\pi^2}. \quad (17)$$

If  $\bar{M}$  is positive, then  $A$  is also positive.

c. We must base the consideration of the nonstationary world on equations (14) and (15). The quantity  $\lambda$  is not determined by these equations. We shall postulate that it can have an arbitrary value. We now determine that value of the variable  $x$ , for which the square root of equation (15) changes its sign. If we restrict our consideration to positive radii of curvature, it will suffice to consider the interval  $(0, \infty)$  for  $x$  and in this interval the values of  $x$  that make the radicand equal to zero or  $\infty$ . One value of  $x$  for which the square root in equation (15) equals 0 is  $x = 0$ . The remaining values of  $x$ , for which the square root in equation (15) changes sign, are given by the positive roots of the equation  $A - x + (\lambda/3c^2)x^3 = 0$ . We denote  $\lambda/3c^2$  by  $y$  and consider the system of third degree curves in the  $x-y$  plane:

$$yx^3 - x + A = 0. \quad (18)$$

Here  $A$  is the parameter of the curve, which varies over the interval  $(0, \infty)$ . A curve of the system cuts the  $x$ -axis at the point  $x = A$ ,  $y = 0$  and has a maximum at the point

$$x = \frac{3A}{2}, \quad y = \frac{4}{27A^2}.$$

From figure 125.1 it is obvious that the equation  $A - x + (\lambda/3c^2)x^3 = 0$  has a positive root  $x_0$  in the interval  $(0, A)$  for negative  $\lambda$ . If one considers  $x_0$  as a function of  $\lambda$  and  $A$ , then

$$x_0 = \Theta(\lambda, A),$$

one finds that  $\Theta$  is an increasing function of  $\lambda$  and of  $A$ . If  $\lambda$  is in the interval  $[0, 4c^2/(9A^2)]$ , the equation has two positive roots  $x_0 = \Theta(\lambda, A)$  and  $x_0' = \Phi(\lambda, A)$ , where  $x_0$  is the root in the interval  $(A, 3A/2)$  and  $x_0'$  is in the interval  $(3A/2, \infty)$ .  $\Theta(\lambda, A)$  is an increasing function of  $A$  and  $\lambda$ , whereas  $\Phi(\lambda, A)$  is a decreasing function of  $A$  and  $\lambda$ . Finally, if  $\lambda$  is bigger than  $4c^2/(9A^2)$ , then the equation has no positive roots.

Let us now pass on to a discussion of equation (15) taking into consideration the following remark: Let the radius of curvature equal  $R_0$  for  $t = t_0$ . The sign of the square root in equation (15) is, for  $t = t_0$ , positive or negative depending on whether the radius of curvature is increasing or decreasing for  $t = t_0$ . Since we can replace  $t$  by  $-t$ , if need be, we can always make the square root positive, e.g. by choice of the

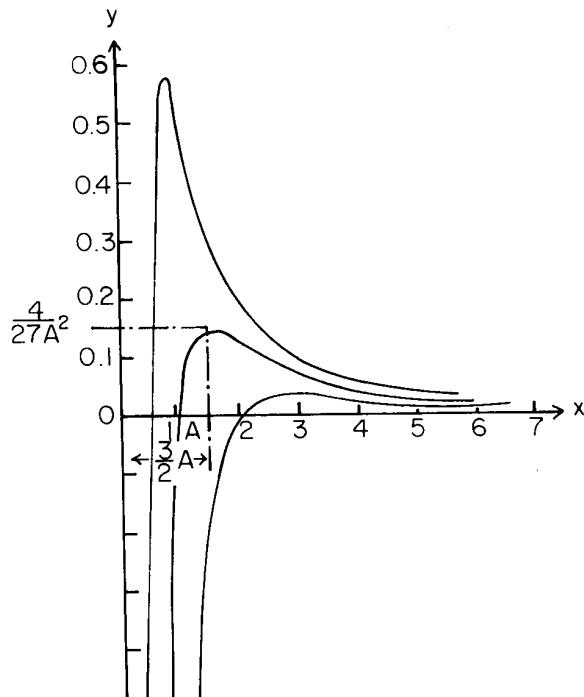


Fig. 125.1 A plot in the  $x - y$  plane of the curves satisfying the non-stationary world equation  $yx^3 - x + A = 0$  where  $y = \lambda/3c^2$  is the reciprocal of the radius  $R$  and  $A$  is a constant which is proportional to the total mass of the universe.

### VIII. RELATIVITY AND COSMOLOGY

time it can always be arranged such that the radius of curvature increases with increasing time at  $t = t_0$ .

D. We consider first the case  $\lambda > 4c^2/(9A^2)$ , e.g. the case in which the equation  $A - x + (\lambda/3c^2)x^3 = 0$  has no positive roots. Equation (15) can then be written

$$t - t_0 = \frac{1}{c} \int_{R_0}^R \sqrt{\frac{x}{A - x + \frac{\lambda}{3c^2}x^3}} dx, \quad (19)$$

where, in consequence of our remark, the square root is always positive. From that, it follows that  $R$  is an increasing function of  $t$ . The positive initial value  $R_0$  is free of any restriction.

Since the radius of curvature may not be smaller than zero, it must decrease with decreasing time,  $t$ , from  $R_0$  to the value zero at time  $t'$ . We shall call the growth time of  $R$  from 0 to  $R_0$  the time since the creation of the world.<sup>11</sup> This time,  $t'$ , is given by

$$t' = \frac{1}{c} \int_0^{R_0} \sqrt{\frac{x}{A - x + \frac{\lambda}{3c^2}x^3}} dx \quad (20)$$

We denote the world under consideration as a *monotonic world of the first kind*.

The time since the creation of the monotonic world of the first kind, considered as a function of  $R_0$ ,  $A$ ,  $\lambda$ , has the following properties:

- 1) It increases with increasing  $R_0$ .
- 2) It decreases if  $A$  increases, e.g. if the mass in space is increased.
- 3) It decreases if  $\lambda$  increases.

If  $A > 2R_0/3$ , then for an arbitrary  $\lambda$  the time elapsed since the creation of the world is finite. If  $A \leq 2R_0/3$ , then a value of  $\lambda = \lambda_1 = 4c^2/(9A^2)$  can always be found that as  $\lambda$  approaches this value, the time since the creation of the world increases without limit.

E. Now let  $\lambda$  lie in the interval  $[0, 4c^2/(9A^2)]$ ; then the initial value of the radius of curvature can lie in one of the intervals

$$(0, x_0), \quad (x_0, x_0'), \quad \text{or} \quad (x_0', \infty).$$

If  $R_0$  falls in the interval  $(x_0, x_0')$ , then the square root in equation (15) is imaginary. A space with this initial curvature is impossible.

We devote the next section to the case where  $R_0$  lies in the interval  $(0, x_0)$ . Here we consider the third case:  $R_0 > x_0'$  or  $R_0 > \Phi(\lambda, A)$ . Through considerations that are analogous

to the preceding ones, it can be shown that  $R$  is an increasing function of time, whereby  $R$  can begin with the value  $x_0' = \Phi(\lambda, A)$ . The time that has elapsed from the moment when  $R = x_0'$  to the moment that corresponds to  $R = R_0$ , we again call the time since the creation of the world. Let it be  $t'$ ; then

$$t' = \frac{1}{c} \int_{x_0'}^{R_0} \sqrt{\frac{x}{A - x + \frac{\lambda}{3c^2}x^3}} dx. \quad (21)$$

We call this world a *monotonic world of the second kind*.

F. We now consider the case that  $\lambda$  falls between the limits  $(-\infty, 0)$ . If  $R_0 > x_0 = \Theta(\lambda, A)$ , the square root in equation (15) becomes imaginary, and the space with this  $R_0$  is impossible. If  $R_0 < x_0$ , the considered case is identical with that which we have left aside in the preceding sections. We therefore assume that  $\lambda$  lies in the interval  $[-\infty, 4c^2/(9A^2)]$  and  $R_0 < x_0$ . By means of a well-known argument<sup>12</sup> one can now show that  $R$  becomes a periodic function of  $t$  with the period  $t_\pi$ , which we name the *world period*;  $t_\pi$  is given by the formula

$$t_\pi = \frac{2}{c} \int_0^{x_0} \sqrt{\frac{x}{A - x + \frac{\lambda}{3c^2}x^3}} dx. \quad (22)$$

The radius of curvature varies between 0 and  $x_0$ . We shall call this universe the *periodic world*. The period of the periodic world increases if we increase  $\lambda$  and tends to infinity if  $\lambda$  tends to the value  $\lambda_1 = 4c^2/(9A^2)$ .

For small  $\lambda$ , the period is represented by the approximate formula

$$t_\pi = \pi A/c. \quad (23)$$

With reference to the periodic world two points of view are possible: We count two events as coincident if their spatial coordinates coincide and the difference of time coordinate is an integral multiple of the period, so that the radius of curvature grows from 0 to  $x_0$  and thereafter decreases to the value 0. The time of world existence is finite. On the other hand, if the time varies between  $-\infty$  and  $+\infty$  (e.g. if we consider two events as coincident only when not only their spatial but also their world coordinates coincide), we come to a real periodicity of the space curvature.

G. Our information is completely insufficient to carry out numerical calculations and to distinguish which world our universe is. It is possible that the causality problem and the problem of centrifugal force will illuminate these questions. It remains to note that the "cosmological" magnitude  $\lambda$  remains undetermined in our formula, because it is a superfluous constant in the problem. Possibly electrodynamic considerations can lead to its evaluation. If we set  $\lambda = 0$  and

$M = 5 \times 10^{21}$  solar masses, the world period becomes of the order 10 billion years. However, these numbers are valid only as an illustration of our calculation.

1. Einstein, "Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie," *Sitzungsberichte Berl. Akad.* 1, 142 (1917).

2. De Sitter, "On Einstein's theory of gravitation and its astronomical consequences," *Monthly Notices of the R. Astronom. Soc.* 78, 3 (1917).

3. By "space" we understand here a space that is described by a manifold of 3 dimensions; the "world" corresponds to a manifold of 4 dimensions.

4. Klein, "Über die Integralform der Erhaltungssätze und die Theorie der räumlich-geschlossenen Welt," *Götting. Nachr.* (1918).

5. See this name in Eddington's book *Espace, Temps et Gravitation*, 2 Partie (Paris: Hermann, 1921), p. 10.

6. The sign of  $R_{ik}$  and  $\bar{R}$  differs here from the usual convention.

7. See, for example, Eddington, *Espace, Temps et Gravitation*, 2 Partie (Paris: Hermann, 1921).

8. The  $ds$ , which is taken to have the dimension of time, we designate  $d\tau$ ; then the constant  $\kappa$  has the dimension Length/Mass and in c.g.s. units equals  $1.87 \times 10^{-27}$ . See Laue, *Die Relativitätstheorie*, vol. 2 (Braunschweig, 1921), p. 185.

9. The density  $\rho$  is for us an unknown function of the world coordinates  $x_1, x_2, x_3$ , and  $x_4$ .

10. This transformation is given by the formula

$$dx_4 = \sqrt{A_0 x_4 + B_0} dx_4.$$

11. The time since the creation of the Universe is the time that has elapsed from the moment when space was a point ( $R = 0$ ) to the present state ( $R = R_0$ ): this term may also be infinite.

12. See, for example, Weierstrass, "Über eine Gattung reell periodischer Funktionen," *Monatsber. d. Königl. Akad. d. Wissensch.* (1866), and Horn, "Zur Theorie der kleinen endlichen Schwingungen," *ZS. f. Math. und Physik* 47, 400 (1902). In our case the considerations of these authors have to be altered appropriately. However, the periodicity can be established in our case by elementary considerations.

## RADIATION — *Waves and Quanta*<sup>1</sup>

Note of **Louis de Broglie**, presented by Jean Perrin.

(Translated from *Comptes rendus*, Vol. 177, 1923, pp. 507-510)

---

Let us consider a material moving object of rest mass  $m_0$  moving with respect to a fixed observer with a speed  $v = \beta c$  ( $\beta < 1$ ). According to the principle of the inertia of energy, it should possess an internal energy equal to  $m_0 c^2$ . On the other hand, the quantum principle suggests associating this internal energy with a simple periodic phenomenon of frequency  $v_0$  such that

$$h v_0 = m_0 c^2,$$

$c$  being, as usual, the limiting velocity of the theory of relativity and  $h$  Planck's constant.

For the fixed observer, the frequency  $v = \frac{m_0 c^2}{h\sqrt{1-\beta^2}}$  corresponds to the total energy

of the moving object. But, if this fixed observer observes the internal periodic phenomenon of the moving object, he will see it slowed down and will attribute to it a frequency  $v_1 = v_0 \sqrt{1-\beta^2}$ ; for him this phenomenon varies therefore like

$$\sin 2\pi v_1 t .$$

Now let us suppose that at the time  $t = 0$  the moving object coincides in space with a wave of frequency  $v$  defined above and propagating in the same direction as it does with the speed  $\frac{c}{\beta}$ . This wave, which has a speed greater than  $c$ , cannot correspond to

---

<sup>1</sup> Concerning the present note, see Brillouin, *Comptes rendus*, Vol. 168, 1919, p. 1318.

transport of energy; we will only consider it as a fictitious wave associated with the motion of the object.

I maintain that, if at the time  $t = 0$ , there is phase agreement between the vectors of the wave and the internal phenomenon of the object, this phase agreement will be maintained. In effect, at time  $t$  the object is at a distance from the origin equal to

$$vt = x; \text{ its internal motion is then represented by } \sin 2\pi v_1 \frac{x}{v} .$$

The wave, at this point, is represented by

$$\sin 2\pi v \left( t - \frac{x\beta}{c} \right) = \sin 2\pi vx \left( \frac{1}{v} - \frac{\beta}{c} \right) .$$

The two sines are equal and the phase agreement is realized if one has

$$v_1 = v(1 - \beta^2),$$

a condition that is clearly satisfied by the definitions of  $v$  and  $v_1$ .

The demonstration of this important result rests uniquely on the principle of special relativity and on the correctness of the quantum relationship as much for the fixed observer as for the moving observer.

Let us apply this to an atom of light. I showed elsewhere<sup>2</sup> that the atom of light should be considered as a moving object of a very small mass ( $< 10^{-50}$  g) that moves with a speed very nearly equal to  $c$  (although slightly less). We come therefore to the following conclusion: *The atom of light, which is equivalent by reason of its total energy to a radiation of frequency  $v$ , is the seat of an internal periodic phenomenon that, seen by the fixed observer, has at each point of space the same phase as a wave of frequency  $v$  propagating in the same direction with a speed very nearly equal (although very slightly greater) to the constant called the speed of light.*

<sup>2</sup> See *Journal de Physique*, 6-th series, Vol. 3, 1922, p. 422.

Let us consider now the case of an electron describing a closed trajectory with uniform speed slightly less than  $c$ . At time  $t = 0$ , the object is at point  $O$ . The associated fictitious wave, launched from the point  $O$  and describing the entire trajectory with the speed  $\frac{c}{\beta}$ , catches up with the electron at time  $\tau$  at a point  $O'$  such that  $\overline{OO'} = \beta c \tau$ .

One has then that

$$\tau = \frac{\beta}{c} [\beta c (\tau + T_r)] \quad \text{or} \quad \tau = \frac{\beta^2}{1 - \beta^2} T_r ,$$

where  $T_r$  is the period of revolution of the electron in its orbit. The internal phase of the electron, when the electron goes from  $O$  to  $O'$ , has a variation of

$$2\pi v_1 \tau = 2\pi \frac{m_0 c^2}{h} T_r \frac{\beta^2}{\sqrt{1 - \beta^2}} .$$

It is *almost necessary* to suppose that the trajectory of the electron will be stable *only if* the fictitious wave passing  $O'$  catches up with the electron in phase with it: the wave of frequency  $v$  and speed  $\frac{c}{\beta}$  has to be in resonance over the length of the trajectory. This

leads to the condition

$$\frac{m_0 \beta^2 c^2}{\sqrt{1 - \beta^2}} T_r = nh , \quad n \text{ being integer.}$$

Let us show that this stability condition happens to be that of the Bohr and Sommerfeld theories for a trajectory described by a constant speed. Let us call  $p_x, p_y, p_z$  the momenta of the electron along three rectangular axes. The general condition for stability formulated by Einstein is in effect

$$\int_0^{T_r} (p_x dx + p_y dy + p_z dz) = nh \quad (n \text{ integer})^3$$

which, in the present case, can be written

$$\int_0^{T_r} \frac{m_0}{\sqrt{1-\beta^2}} (v_x^2 + v_y^2 + v_z^2) dt = \frac{m_0 \beta^2 c^2}{\sqrt{1-\beta^2}} T_r = nh ,$$

as above.

In the case of an electron turning in a circular orbit of radius  $R$  with an angular velocity  $\omega$ , one finds again for sufficiently small speeds the original formula of Bohr:

$$m_0 \omega R^2 = n \frac{h}{2\pi} .$$

If the speed varies along the length of the trajectory, one finds again the Bohr-Einstein formula if  $\beta$  is small. If  $\beta$  assumes large values, the question becomes more complicated and necessitates a special examination.

Pursuing research along these lines we have reached important results, which will be communicated soon. We are as of today able to explain the phenomena of diffraction and of interference taking into account the quantization of light.

<sup>3</sup> The case of quasi-periodic motion does not present any new difficulty. The necessity of satisfying the condition stated in the text for an infinity of pseudo-periods leads to the conditions of Sommerfeld.

## The Spectrum of Scattered *X*-Rays

A.H. Compton  
Washington University  
(Received May, 9, 1923)

### Abstract

**The spectrum of molybdenum  $K_{\alpha}$  rays scattered by graphite at  $45^\circ$ ,  $90^\circ$  and  $135^\circ$  has been compared with the spectrum of the primary beam.** A primary spectrum line when scattered is broken up into two lines an “unmodified” line whose wave-length remains unchanged, and a “modified” line whose wave-length is greater than that of the primary spectrum line. Within a probable error of about  $0.001 \text{ \AA}$ , the difference in the wave-length ( $\lambda - \lambda_0$ ) increases with the angle  $\theta$  between the primary and the scattered rays according to the quantum relation  $(\lambda - \lambda_0) = \lambda(1 - \cos \theta)$ , where  $\lambda = h/mc = 0.0242 \text{ \AA}$ . This wave-length change is confirmed also by absorption measurements. The modified ray does not seem to be as homogeneous as the unmodified ray; it is less intense at small angles and more intense at large angles than is the unmodified ray.

**An X-ray tube of small diameter and with a water-cooled target** is described, which is suitable for giving intense *X*-rays.

The writer has recently proposed a theory of the scattering of *X*-rays, based upon the postulate that each quantum of *X*-rays is scattered by an individual electron.<sup>1</sup> <sup>2</sup> The recoil of this scattering electron, due to the change

<sup>1</sup>A.H. Compton, Bull. Natl. Coun., No 20, p. 18 (October 1922); Phys. Rev. 21, 207 (abstract) (Feb. 1923); Phys. Rev. 21, 483 (May, 1923).

<sup>2</sup>Cf. also P. Debye, Phys. Zeitschr. 24, 161 (April 15, 1923)

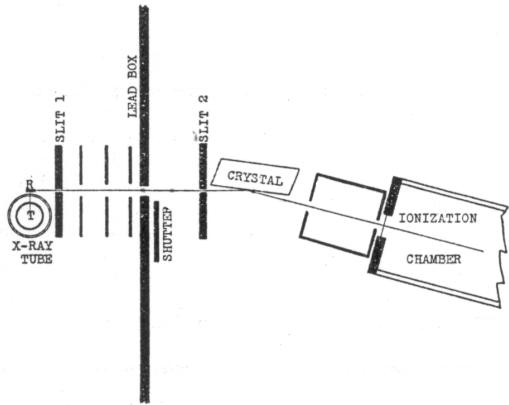


Figure 1: Measuring the wave-length of scattered  $X$ -rays.

in momentum of the  $X$ -ray quantum when its direction is altered, reduced the energy and hence also the frequency of the quantum radiation. The corresponding increase in the wave-length of the  $X$ -rays due to scattering was shown to be

$$\lambda - \lambda_0 = \gamma (1 - \cos \theta) \quad (1)$$

where  $\lambda$  is the wave-length of the ray scattered at an angle  $\theta$  with the primary ray whose wave-length is  $\lambda_0$ , and

$$\gamma = h/mc = 0.0242 A$$

where  $h$  is Planck's constant,  $m$  is the electron and  $c$  the velocity of light. It is the purpose of this paper to present more precise experimental data than has previously been given regarding this change in wave-length when  $X$ -rays scattered.

*Apparatus and method.* For the quantitative measurement of the change in wave-length it was clearly described to employ a spectroscopic method. In view of the comparatively low intensity of scattered  $X$ -rays, the apparatus had to be designed in such a manner as to secure the maximum intensity in the beam whose wave-length was measured. The arrangement of the apparatus is shown diagrammatically in Fig. 1. Rays proceeded from the molybdenum target  $T$  of an  $X$ -ray tube to the graphite scattering block  $R$ , which was placed in line with the slits 1 and 2. Lead diaphragms, suitably disposed, prevented stray radiation from leaving the lead box that surrounded the  $X$ -ray tube. Since the slit 1 and the diaphragms

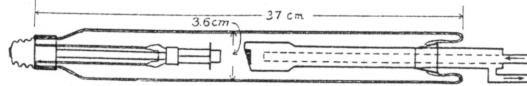


Figure 2: *X*-ray tube.

were mounted upon an insulating support, it was possible to place the *X*-ray tube close to the slit without danger of puncture. The *X*-rays, after passing through the slits, were measured by a Bragg spectrometer in the usual manner.

The *X*-ray tube was of special design. A water-cooled target was mounted in a narrow glass tube, as shown in Fig. 2, so as to shorten as much as possible the distance between the target *T* and the the radiator *R*. This distance in the experiments was about 2 cm. When 1.5 kw was dissipated in the *X*-ray tube, the intensity of the rays reaching the radiator was thus 125 times as great as it would have been if a standard Coolidge tube with a molybdenum target had been employed. The electrodes for this tube were very kindly supplied by the General Electric Company.

In the final experiments the distance between the slits was about 18 cm, their length about 2 cm, and their width about 0.01 cm. Using a crystal of calcite, this made possible a rather high resolving power even in the first order spectrum.

*Spectra of scattered molybdenum rays.* Results of the measurements, using slits of two different widths, are shown in Fig. 3 and 4. Curves *A* represent the spectrum of the  $K_{\alpha}$  line, and curves *B*, *C* and *D* are the spectra of this line after being scattered at angles of  $45^{\circ}$ ,  $90^{\circ}$  and  $135^{\circ}$  respectively with the primary beam. while in Fig. 4 the experimental points are a little erratic, it may be noted that in this case the intensity of the *X*-rays is only about  $1/25,000$  as great as if the spectrum of the primary beam were under examination, so that small variations produce a relatively large effect.

It is clear from these curves that when a homogeneous *X*-ray is scattered by graphite it is separated into two distinct parts, one of the same wave-length as the primary beam, and the other of increased wave-length. Let us call these the *modified* and the *unmodified* rays respectively. In each curve the line *P* is drawn through the peak of the curve representing the primary line, and the line *T* is drawn at the angle at which the scattered line should appear according to Eq. (1). In Fig. 4, in which the settings were made with the greater care, within an experimental error of less than 1 minute of arc, or about  $0.001 A$ , the peak of the unmodified ray falls upon the line *P*

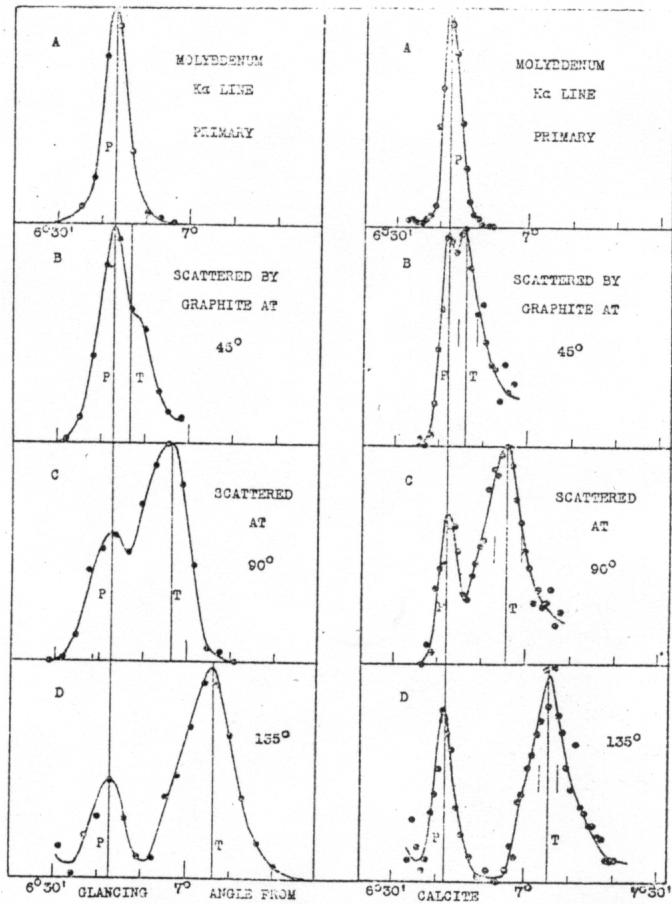


Figure 3:

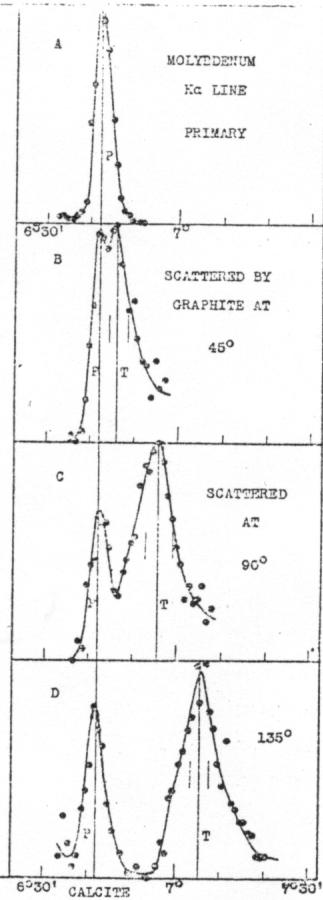


Figure 4:

and the peak of the modified ray falls upon the line  $T$ . The wave-length of the modified ray thus increases with the scattering angle as predicted by the quantum theory, while the wave-length of the unmodified ray is in accord with the classical theory.

There is a distinct difference between the widths of the unmodified and the modified lines. A part of the width of the modified line is due to the fact that the graphite radiator  $R$  subtends a rather large angle as view from the target  $T$ , so that the angles at which the rays are scattered to the spectrometer crystal vary over an appreciable range. As nearly as I can estimate, the width at the middle of the modified line due to this cause is that indicated in Fig. 4 by the two short lines above the letter  $T$ . It does not appear, however, that this geometrical consideration is a sufficient explanation for the whole increased width of the modified line, at least for the rays scattered at  $135^\circ$ . It seems more probable that the modified line is heterogeneous, even in a ray scattered at a definite angle.

The unmodified ray is usually more prominent in a beam scattered at a small angle with the primary beam, and the modified ray more prominent when scattered at a large angle. A part of the unmodified ray is doubtless due to regular reflection from the minute crystals of which the graphite is composed. If this were only source of the unmodified ray, however, we should expect its intensity to diminish more rapidly at large angles than is actually observed. The conditions which determine the distribution of energy between these two rays are those which determine whether an  $X$ -ray shall be scattered according to the simple quantum law or in some other manner. I have studied this distribution experimentally by another method, and shall discuss it in another paper,<sup>3</sup> but the reasons underlying this distribution are puzzling.

*Experiments with shorter wave-lengths.* These experiments have been performed using a single wave-length,  $\lambda = 0.711 \text{ \AA}$ . In this case we find for the modified ray a change in wave-length which increases with the angle of scattering exactly in the manner described by Eq. (1). While these experiments seem conclusive, the evidence would of course be more complete if similar experiments had been performed for other wave-lengths. Preliminary experiments similar to those here described have been performed using the  $K$  radiation from tungsten, of wave-length about  $0.2 \text{ \AA}$ . This work has shown a change in wave-length of the same order of magnitude as that observed using the molybdenum  $K_\alpha$  line. Furthermore, as described in ear-

---

<sup>3</sup>A.H. Compton, Phil.Mag. (in printer's hands)

lier papers,<sup>4</sup> absorption measurements have confirmed these results as to order of magnitude over a very wide range of wave-lengths. This satisfactory agreement between the experiments and the theory gives confidence in the quantum formula (1) for the change in wave-length due to scattering. These is, indeed, no indication of any discrepancy, for the range of wave-length investigated, when this formula is applied to the wave-length of the modified ray.

Washington University,  
Saint Louis,  
May 9, 1923.

---

<sup>4</sup>Cf. e.g., A.H.Compton, Phys.Rev. 21, pp. 494–6 ( 1923).

Investigations on  $X$  Rays and  $\beta$  Rays by the Cloud  
Method. Part 1. –  $X$  Rays.

C.T.R. Wilson

Cambridge

(Received June, 23, 1923)

*1. Introduction.*

The method used in these investigations is that which was described in papers communicated to the Royal Society in 1911 and 1912.<sup>1</sup> The ionising rays are made to pass through moist air, or other gas, in which the water-vapour has been brought into the super-saturated state by sudden expansion of the gas. Each ion liberated becomes at once the nucleus for the condensation of a visible droplet of water; the clouds of drops thus formed are immediately photographed.

Very sharply defined pictures of the tracks of ionising particles – $\alpha$  – or  $\beta$ -rays – may be obtained in this way. When the conditions are suitably arranged, the effects of diffusion of the ions before their mobility has been destroyed by condensation of water upon them, as well as that of subsequent disturbance of the cloud tracks by convection currents in the gas, are negligible: photographs of the path of the ionising particles, practically free from distortion, are obtained. The almost perfect straightness of the track of a very fast  $\beta$ -particle, when it occurs among a crowd of tracks of slower

---

<sup>1</sup>Roy.Soc.Proc., A, vol. 85, p. 285, and v. 87, p. 277.

$\beta$ -particles, gives very convincing evidence that the complicated forms of the latter are not due to instrumental distortion.

The information contained in the pictures is greatly increased when two cameras are used to take simultaneous photographs. For the purposes of exact measurement relating to some definite problem, such as the branching of  $\alpha$ -ray tracks, the arrangement used by Shimizu<sup>2</sup> and others, in which the axes of the cameras are at right angles, has undoubtedly advantages. But for the purpose of disentangling the complicated phenomena which attend the passage of  $\beta$ -rays and  $X$ -rays through air the stereoscopic method is more effective; it has been used throughout the present investigations.

For the quantitative study of  $X$ -rays the cloud method has many advantages over that in which an ordinary ionisation chamber and electrometer are used. It gives directly the number and nature of the  $\beta$ -particles ejected from atoms by the  $X$ -rays, and not merely the total ionisation; if each  $\beta$ -ray which is produced by the action of the  $X$ -rays represents the absorption of one quantum of radiation the method enables us to deal directly with individual quanta. When the cloud chamber is momentarily traversed by a beam of  $X$ -radiation of suitable intensity a picture is obtained (in three dimensions if the stereoscopic method is used) of the tracks of all the electrons ejected from the atoms in a given volume of the gas by the action of the  $X$ -rays, primary and secondary. An inspection of the picture shows at once (1) the point of origin of each  $\beta$ -ray, (2) its initial direction (i.e., the direction in which an electron has been ejected from its parent atom by the action of the radiation), (3) the total length of its path or range, (4) the form of the track, its sudden or gradual bends, and the number and direction of emission of any secondary  $\beta$ -rays (branches), and (5) the variation of the ionisation along the track; under favourable conditions the number and distribution of the ions along the tracks may be obtained by direct counting.

A number of stereoscopic pictures of the tracks of  $\alpha$ - and  $\beta$ -particles with others illustrating the effects of  $X$ -rays, were taken early in 1914, but the work was then interrupted by the War. Some of these pictures were exhibited on different occasions at the Royal Society and elsewhere. Considerable improvements in the details of the method have been introduced since that time. Most of the photographs now reproduced belong to a series of nearly 500 stereoscopic pairs which were taken between December, 1921, and the end of July, 1922.

---

<sup>2</sup>Roy.Soc.Proc., A, vol. 99, p. 432 (1921).

## *2. Improvements in the apparatus and method*

The expansion apparatus which has been used throughout these investigations is that described and figured in the 1912 paper. The following are some of the improvements which have been made in the details of the method.

*Cloud-chamber.* – In most of the experiments the cloud-chamber consisted of a thin-walled glass cylinder 1.2 mm. thick, 16.5 cm. in diameter, with a plate-glass roof. The height from the top of the brass plunger, which formed the floor, to the roof of the cloud-chamber was generally 3 cm. As in the previous experiments, a marginal ring of tinfoil cemented in between the roof and walls of the cloud-chamber made it possible to maintain a potential difference between the roof and floor. The inner surfaces of the roof and sides of the cloud-chamber were at first coated as in the earlier experiments with a thin layer of gelatine ; a small quantity of copper sulphate was added to the gelatine to prevent the growth of mould.

The gelatine lining does not remain permanently effective in preventing the formation of droplets on the interior of the roof and walls of the cloud-chamber. If the upper part of the apparatus remains for some time at a lower temperature than the base, as frequently happens at night owing to changes in room temperature, water distils rapidly from the floor to the roof and sides of the cloud-chamber and collects in drops which are not readily removed. To avoid this difficulty completely it is only necessary to keep the base of the cloud-chamber at a slightly lower temperature than the roof and sides. This is most effectively done by keeping the brass expansion cylinder at a temperature slightly below that of the room by allowing a small flow of tap water to pass continuously through the shallow receptacle in which the expansion cylinder rests.

When the base of the expansion apparatus is cooled in this way the walls and roof remain perfectly clear for an indefinite time, even when the gelatine lining is omitted altogether. Most of the recent photographs have been taken without any gelatine lining on the roof and sides of the cloud-chamber.

The floor of the cloud-chamber, formed by the upper surface of the plunger, remains as in the earlier work covered with a thick layer of gelatine; this is blackened by ink and contains a small quantity of copper sulphate in solution. As in the earlier experiments, a vertical electric field was maintained in the cloud-chamber; the field was directed upwards, and in most cases amounted to about 3 volts per cm.

*Cameras.* – Two simple box-cameras of fixed focus have been used to obtain the stereoscopic pictures; they are joined rigidly together with the centres of their lenses 5.5 cm. apart. The lenses were Beck "Isostigmars" of

maximum aperture F 5.8 and focal length 12 cm.; in many cases the aperture was reduced to F 8 or F 11. The axes of the two cameras converged to a point 40 cm. in front of the lenses for which distance they are also focussed. They have generally been used with their axes in the horizontal plane which passes through the centre of the cloud-chamber.

*Illumination.* — A Leyden jar discharge through mercury vapour at atmospheric pressure was used as before to illuminate the clouds for the purpose of obtaining the photographs. The mercury discharge tube was as shown in fig. 2 of the 1912 paper, but pointed steel rods were used to close the ends of the silica tubes; the ends of the tubes were inserted in mercury cups. The discharge tube was placed as before at the principal focus of a cylindrical lens.

*Photographic plates.* — “Imperial process” plates have been used in all the recent work. They are not appreciably less sensitive than rapid plates for the light from the mercury spark, and they have the great advantages of fineness of grain and convenience in use.

*Timing Arrangements.* — In order that sharply defined pictures may be obtained of the tracks of electrons ejected from atoms by *X*-rays, it is necessary that the rays should traverse the cloud chamber immediately after the sudden expansion of the gas, and that the drops which condense on the ions set free along the tracks should be momentarily illuminated after a very short interval. As in the early experiments, the momentary flash of *X*-radiation which acts free the ions in the cloud chamber and the mercury vapour spark which illuminates the drops condensed upon them are produced by the discharge of Leyden jars. The old arrangement in which a falling sphere brought about the two discharges in succession, has been replaced by one in which three pendulums of adjustable period are all released simultaneously. The first (the “expansion” pendulum), as it reaches the lowest point of its swing, opens communication between the vacuum chamber and the space below the plunger, the others (the “*X*-ray” pendulum and “spark” pendulum) as they reach their lowest points discharge Leyden jars through the *X*-ray bulb and mercury discharge tube respectively. By adjustment of sliding weights on these pendulums, the *X*-rays may be made to traverse the expansion apparatus immediately after the expansion is completed, while the illuminating spark follows at an interval long enough to enable the cloud particles to condense on the ions, but not long enough to allow of convection currents causing distortion of the tracks.

When the rays from radio-active substances are being studied, a somewhat larger potential difference (from 20 to 100 volts) has generally been, maintained between the roof and floor of the cloud-chamber than in exper-

iments with  $X$ -raya. No attempt has been made to confine the exposure of the cloud-chamber to these rays to the period between the production of the supersaturated condition and the passage of the illuminating spark. Such photographs, therefore, show not only the sharply defined tracks of  $\alpha$ - or  $\beta$ -particles which have passed through the super-saturated air after the expansion, but also diffuse double tracks in which the positive and negative ions have been separated by the electric field) due to ionising particles which have traversed the air before the expansion.

*X-rays.* — The source of radiation has throughout the work been an  $X$ -ray bulb of old type with a platinum anticathode. The rays were produced (in nearly all cases immediately after the sudden expansion of the air in the cloud-chamber) by the discharge of a large Leyden jar. To make the action of the  $X$ -ray bulb regular when used in this way a sufficient resistance had to be inserted in the circuit to make the discharge non-oscillatory. The maximum potential difference across the terminals of the bulb as measured by a spark gap was about 45,000 volts.

The bulb was generally placed with its anticathode at a distance of 46 cm. from the centre of the cloud-chamber and on the same horizontal level, and was surrounded by a thick lead case. A horizontal beam of  $X$ -rays, which passed through the centre of the cloud-chamber at right angles to the axis of the stereoscopic camera, was obtained by means of a horizontal lead tube attached to the lead case ; the tube was 20 cms. long and was provided at each end with a thick lead diaphragm with suitable aperture. The form and area of cross-section of the beam could be varied by changing the diaphragms, and absorbing screens could be insetted at either end of the tube.

Photographs were taken of the cloud tracks when no screens were inserted (other than a thin mica window in the wall of the cloud-chamber) and with screens of thickness ranging up to more than 2 cm. of aluminium. For a satisfactory study of the ( $\beta$ -rays produced by  $X$ -radiation it was necessary that their numbers should be kept comparatively small. When no screen was used a convenient number of  $\beta$ -rays was obtained by reducing the  $X$ -ray beam to a cylinder of about 0.5 mm. in diameter; when the thickest screens were inserted a square or rectangular beam of about 0.5 square cm. in cross-section was used. The effects of radiations of very different wave-lengths were in this way studied.

A number of experiments were made with the air at pressures considerably less than atmospheric; the lowest final pressure used was about 10 cms. of mercury.

The experiments have thus far been confined almost entirely to air.

### *3. Effects to be expected from the absorption of X-rays in air*

According to the quantum theory, when  $X$ -radiation of definite frequency  $v$  traverses the air of the cloud-chamber and is partially absorbed by it, it loses a certain number of quanta of energy each equal to  $hv$ . Each quantum absorbed by the air causes the ejection of an electron and is thus represented by a  $\beta$ -ray track in the cloud picture. Of the quantum of energy  $hv$  absorbed by an atom, a part depending on the nature of the atom and the energy level from which the electron is ejected is used in removing the electron from the atom, the rest is represented by the kinetic energy of the emergent  $\beta$ -particle. The velocities and therefore the ranges of the  $\beta$ -particles ejected by primary radiation of given frequency will thus not all be alike in a mixed gas, like air.

The air of the cloud-chamber contains, besides nitrogen and oxygen, other constituents in quantities sufficient to cause appreciable absorption, of  $X$ -rays. From what is known regarding the absorption of  $X$ -rays we may deduce that the relative number of  $\beta$ -particles ejected by  $X$ -radiation of wave length  $\lambda$  from the  $K$ -levels of atoms of each of the different constituent elements of the air is approximately proportional to  $Z^4\lambda^3N$ , where  $Z$  is the atomic number and  $N$  is the number of atoms per cubic centimetre of the air; this only applies if the frequency of the incident radiation exceeds the  $K$ -absorption limit, otherwise there is no  $K$ -absorption. Applying this we find that, while the absorption of  $X$ -rays by the hydrogen, carbon dioxide, neon, krypton and xenon present should be relatively negligible, about 60 per cent. of the  $\beta$ -rays which result from ejection of electrons from the  $K$ -level should be due to nitrogen, 25 to oxygen, 15 to argon. We should expect also a small number of  $\beta$ -particles to be ejected from the outer levels of atoms of these various elements.

The  $\beta$ -rays originating from these various sources under the action of  $X$ -rays of given frequency will have different ranges. When the frequency of the incident radiation is high the differences in the ranges of the  $\beta$ -rays will be unimportant. When, however, the frequency of the incident radiation is not many times that of the absorption limit of argon, as in the experiments with a copper target described in § 4, the range of the  $\beta$ -ray from the  $K$ -level of argon will fall far short of that from the  $K$ -level of an oxygen or nitrogen atom; the difference between the ranges of the  $\beta$ -rays from the  $K$ -levels in oxygen or nitrogen atoms, or from the  $K$ - and  $L$ -levels in either

of these atoms, also ceases to be negligible. We have in such a case a whole series of possible ranges for the  $\beta$ -ray tracks corresponding to the various lines in the magnetic spectrum in experiments such as those of de Broglie.<sup>3</sup>

The ejection of an electron from an atom by the action of the primary  $X$ -rays will in general be followed by the emission of secondary radiation. An atom from which a  $K$ -electron has been ejected will emit a quantum of one of its characteristic  $K$ -radiations. The  $K$ -radiation emitted may give evidence of its existence by ejecting an electron from the  $K$ -level of an atom of smaller atomic number or from the outer level of any atom. It would in this way be possible for  $K$ -radiation from argon to be absorbed by oxygen with ejection of an electron from the  $K$ -level and subsequent emission of oxygen  $K$ -radiation; this in turn might be absorbed in ejecting an electron from the  $K$ -level of nitrogen and thus causing the emission of nitrogen.  $K$ -radiation; the final  $\beta$ -ray track will be that of an electron ejected from the outer level of an atom.

A ray produced in air by the action of primary  $X$ -rays may have no secondary  $\beta$ -ray associated with it, if it has itself been ejected from among the outer electrons of an atom; it may be expected to have a short range  $\beta$ -ray associated with it, if it has been ejected from the  $K$ -level of nitrogen, or it may have more than one if it has arisen in the  $K$ -level of oxygen or argon.

As will appear later, the matter is made more complicated by the combined effects of the primary and secondary radiations.

#### *4. Experiments with "targets" inserted in the path of the $X$ -rays*

In these experiments a small rectangular piece of metal foil, a few square mm. in area, was fixed at the centre of the cloud-chamber. It was attached to the flattened end of a needle, which projected vertically downwards from the centre of a brass plug closing a circular hole in the glass roof. A cylindrical beam of  $X$ -rays, about 0.5 mm. in diameter, was made to strike the centre of the target approximately at right angles.

Plate 1, fig. 1 shows the effects of inserting a copper target, about  $8 \times 10^{-3}$  cm. in thickness, in the path of the narrow pencil of  $X$ -rays. The various effects due to the absorption of the primary  $X$ -radiations by the copper are well shown. A large number of  $\beta$ -rays radiate from the target;

---

<sup>3</sup>Journal de Physique, vol. 2, p. 265 (1921).

the absence of their initial portions is merely due to the heating and other disturbing effects of the copper.

Scattered about in the surrounding air outside the primary  $X$ -ray beam are to be seen numerous short  $\beta$ -ray tracks, due to electrons ejected by the secondary radiations from the copper. The greater number of these short tracks, ranging from 1.0 to 2.4 mm. in length in air at  $2/3$  of atmospheric pressure, are due to the ejection, by the characteristic  $K_\alpha$  and  $K_\beta$ -radiations of copper, of electrons from the different constituents of the air. Small, nearly spherical, clouds about 1/10 mm. in diameter, also appear—the tracks of the slow  $\beta$ -particles ejected by the  $L$ -radiations of copper.

Fig. 2, Plate 1, was obtained in an experiment in which the conditions were identical with those just described except that the  $X$ -rays had been cut down before entering the cloud-chamber by inserting in their path an aluminium screen 9.2 mm. in thickness. Here no  $\beta$ -rays have been produced by the direct action of the primary  $X$ -rays on the air. One  $\beta$ -particle has been ejected from the copper plate, and one  $\beta$ -ray track has been formed in the air, outside the  $X$ -ray beam, by the absorption of a quantum of  $K$ -radiation from the copper. To produce the quantum of  $K$ -radiation which has ejected the  $\beta$ -particle from an atom of oxygen or nitrogen an electron must have been ejected from the  $K$ -level of one of the copper atoms; the track of this one electron appears in the photograph. We almost certainly have here the tracks of the two electrons which are associated respectively with the emission and absorption of the same individual quantum of copper  $K$ -radiation.

A similar picture was obtained with a platinum target; the length of the  $\beta$ -ray ejected by the secondary radiation indicates that it is due to an  $L$ -radiation from the atom of platinum from which the primary  $\beta$ -ray was ejected.

Experiments were also made in which the primary  $X$ -ray beam and the target were outside the cloud-chamber. The target was an inclined metal plate placed immediately above an aluminium window in the centre of the roof of the cloud-chamber; a horizontal beam of  $X$ -rays was incident upon it. The photographs obtained with copper and silver targets showed well the different effects of the  $KK$ -radiations from these metals.

The ranges of the  $\beta$ -rays produced in air by the  $K$ -radiations from copper varied between about 0.6 mm. and 1.7 mm., those produced by the silver  $K$ -radiations between 8 mm. and 16 mm. In both cases the tracks could be grouped into three classes according to their ranges; these classes correspond to the different lines in a  $\beta$ -ray spectrum. The range which recurred with maximum frequency is (in accordance with tile greater

intensity of the  $K$  lines in the  $X$ -ray spectrum) taken to be due to the ejection of an electron from the  $K$ -level of an oxygen or nitrogen atom by the  $K_{\alpha}$ -radiation from the metal; the group of maximum range is taken to be due to the ejection of an electron from the same elements by the  $K_{\beta}$ -radiation. The  $\beta$ -rays of shortest range produced by the  $K$ -radiations have probably been ejected from the  $K$ -level of argon atoms; their ranges are in accordance with this view.

The ranges of the  $\beta$ -rays ejected from oxygen or nitrogen by the  $K_{\alpha}$ - and  $K_{\beta}$ -rays from copper are about 1.3 mm. and 1.7 mm. respectively; those ejected by the  $K_{\alpha}$ - and  $K_{\beta}$ -radiations of silver have ranges of about 11 mm. and 15 mm.

Putting the kinetic energy of the ejected  $\beta$ -particle equal to the difference between the energy ( $hv$ ) of the incident radiation and that required to eject an electron from the  $K$ -level of nitrogen we have in the case of the copper  $K$ -radiations,  $K_{\alpha}(\text{Cu}) - K(\text{N}) = 7.700$  volts,  $K_{\beta}(\text{Cu}) - K(\text{N}) = 8.600$  volts. Similarly the kinetic energies of the  $\beta$ -particles ejected the silver  $K$  and  $K_{\beta}$ -radiations are 21.700 and 24.600 volts approximately.

The results of these measurements of the mean ranges of  $\beta$ -particles of different kinetic energies are represented approximately by the formula  $V = 21.000R^{1/2}$ , where  $V$  is the kinetic energy of the particle expressed in volts and  $R$  its range in centimetres, measured along the track, in air at atmospheric pressure. The experimental results are thus in approximate accordance with Whiddington's fourth-power law<sup>4</sup> connecting the range and velocity of a  $\beta$ -particle. The velocity of a  $\beta$ -particle of range 1.5 cm. is nearly 1/3 of that of light and the kinetic energy given by the relativity formula.

$$m_0c^2 \left( \frac{1}{\sqrt{1 - \beta^2}} - 1 \right)$$

already exceeds  $1/2m_0v^2$  by nearly 10 per cent.; for higher velocities the relation between range and kinetic energy is likely to be less simple.

The tracks of  $\beta$ -particles, such as those due to aluminium  $K$ -radiations or copper  $L$ -radiations, of which the kinetic energy is less than 2,000 volts are of too short range (less than 1/10 mm.) to be measurable; they appear in the pictures as small, nearly spherical, clouds.

---

<sup>4</sup>Roy.Soc.Proc., A, vol. 86, p. 360 (1912).

### *5. Different classes of $\beta$ -rays produced in air by X-rays*

When the air in the cloud-chamber is exposed to hard radiation - -the wave-length of which is for example about  $0.4\text{\AA}^\circ$  - three classes of  $\beta$ -ray tracks may be distinguished: - (a) "Long" tracks having a range of several centimetres; (b) "sphere" tracks, small, almost spherical, cloudlets, 1 or 2-tenths of a millimetre in diameter; and, (c) tracks of 1 or 3 mm. in range, of which the initial direction approximately coincides with that of the primary X-ray beam. From their characteristic appearance when a number of them are present I have been in the habit of recording the tracks of this last class as "fish" tracks; the this of the "fish" is directed towards the source. Figs. 3 and 4 contain examples of the three classes of tracks.

The "long" tracks are undoubtedly the paths of electrons ejected from atoms in most cases by the direct action of the primary beam, each having absorbed one quantum of energy corresponding to the frequency of the primary radiation. The range of these tracks depends on the frequency of the primary radiation and has been used for the purpose of estimating that frequency.

The "fish" tracks are also - as the direction of ejection indicates-almost certainly due to the direct action of the primary radiation; but the energy of the ejected electron is only a small fraction of that of a quantum of the primary radiation, and these tracks do not appear when the wave-length of the incident radiation exceeds  $0.5\text{\AA}^\circ$ . They are almost certainly connected with the phenomena which have led Barkla and others to postulate the existence of a "J" - radiation; again to explain other phenomena relating to secondary radiations A.H. Compton has suggested, the possibility of the forward ejection of short range  $\beta$ -particles. . - "Sphere" tracks are probably in all cases merely the tracks of very short range  $\beta$ -rays; any  $\beta$ -particle with energy less than that corresponding to about 2,000 volts will produce such a track. Some of the sphere tracks which are formed in air exposed to X-rays of high frequency are probably produced in the same way as the fish tracks; intermediate comma-like forms are of frequent occurrence. Other sphere tracks are undoubtedly due to the ejection of electrons by secondary X-rays. A sphere track is frequently situated, close to the origin of a long track and is sometimes outside the primary beam of X-rays. The ejection of the short range  $\beta$ -particle may in this case be due to the characteristic radiation from an atom, from the K-level of which the primary X-radiation has expelled the  $\beta$ -particle which produced the long track. There are, however, other cases of association of  $\beta$ -rays in pairs which cannot be explained so simply. (See section 8.)

### *Direction of ejection of the $\beta$ -particles*

(a) *Polarisation.* – In almost every stereoscopic picture of the tracks of  $\beta$ -particles produced by a horizontal beam of  $X$ -rays, in which a convenient number (10 to 30) of tracks are present, several are found to start, as nearly as can be distinguished, in directions parallel to a vertical plane containing the axis of the  $X$ -ray beam. The cathode ray stream in the  $X$ -ray bulb was always directed vertically upwards, so that the vertical plane is that in which we should expect the electric vector in the polarised portion of the radiation to lie.

About 20 per cent. of the  $\beta$ -particles were found to be ejected in the vertical plane. Barkla found about 20 per cent. of the radiation from the  $X$ -ray bulb used by him to be polarised.

That a partially polarised beam of  $X$ -rays shows a deficiency of  $\beta$ -rays starting in directions lying in the plane of polarisation and an excess starting in direction lying in the perpendicular plane has been confirmed by some more recent experiments. In these a horizontal beam of primary  $X$ -rays was incident on a cylinder of paraffin wax placed above an aluminium window in the roof of the cloud chamber; a scattered beam partially polarised in the vertical plane passes vertically downwards into the cloud-chamber;

(b) *Forward component in velocity of ejection.* – In the photographs obtained in 1912 no systematic preponderance was evident in the number of  $\beta$ -rays which were ejected with a forward rather than a backward velocity relative to the direction of the  $X$ -rays which caused their ejection; this was perhaps mainly due to the tracks being too closely crowded together. Such a preponderance of  $\beta$ -particles which have a forward component in their initial velocity is a striking feature in most of the stereoscopic pictures which have recently been obtained, *e.g.*, fig. 6, Plate 5.

A study of individual tracks shows at once that the forward component in the velocity of a  $\beta$ -particle does not depend simply upon its velocity of ejection, *i.e.*, upon the wave-length of the radiation absorbed. The  $\beta$ -ray tracks may be divided into three easily distinguishable classes, according as the forward component is positive, zero or negative.

It is remarkable that quite a considerable proportion of the tracks belong to the 2nd class, *i.e.*, they are due to  $\beta$ -particles which have been ejected almost exactly at right angles to the  $X$ -ray beam. The tracks belonging to the first class, which is the most numerous, have in nearly all cases a comparatively large forward component, very frequently approximately equal to the lateral component; *i.e.*, the most frequent direction of ejection makes an angle of about  $45^\circ$  with that of the  $X$ -ray beam. Fig. 3, Plate 2, contains

examples of the two first classes in the centre of the picture. Again the third class consists mainly of  $\beta$ -rays with a backward component comparable with the lateral components. Cases in which there can be any doubt as to the class to which a  $\beta$ -ray is to be assigned are very rare.

The results of the examination of 1,148 tracks with regard to their initial direction are given in Table I.

Table I.

Average range (millimetres)	Energy kilovolts	Total number of tracks	Tracks with positive, zero and negative forward components		
			Number in each class		Number in each class per 100 tracks
20-30	30-36	223	+ 155	0 37	- 31
15-20	25-30	662	385	136	141
7-15	17-25	202	106	56	40
2-7	9-17	61	28	21	12
			+ 69	0 17	- 14
			58	21	21
			52	28	20
			45	35	20

The last columns of Table I show very clearly that the percentage of  $\beta$ -ray tracks which have a forward component in their initial direction increases rapidly with increasing velocity and range of the ejected electron, i.e., with increasing frequency of the incident radiation. This increase is mainly at the expense of the  $\beta$ -rays with zero forward component.

The difference between the effects of radiations of higher and lower frequencies on the average direction of ejection lies much more in the relative number of electrons which start with a forward component than in the direction of ejection of those which have the forward component. Tracks which have a forward component, even if their range is less than 1 cm., start most frequently in directions inclined at angles of about  $45^\circ$  to that of the  $X$ -ray beam. The forward inclination is in this case, it may be remarked, much larger than that deduced according to the view that the quantum of radiation absorbed passes on the whole of its momentum to the ejected electron.<sup>5</sup> On the other hand tracks of long range, 3 cm. or more, frequently start at right angles to the  $X$ -ray beam; or again their initial direction may have a larger backward component.

<sup>5</sup>Richardson, "Electron Theory of Matter."

Tracks with a very large forward inclination are, however, mainly of long range. A few such tracks start almost along the direction of the  $X$ -rays; on the other hand a few long-range tracks start almost in the opposite direction, i.e., towards the source of the  $X$ -rays.

In the above account of the phenomena relating to the direction of emission of the ejected  $\beta$ -particles it is to be noted that the special type of tracks in which the forward component in the velocity is most marked – the “fish” tracks – has not been included.

A thorough investigation of the direction of ejection of  $\beta$ -particles of all ranges with a fairly accurate measurement of the angles is likely to lead to interesting results.

### 7. Short-range $\beta$ -rays: “sphere,” “comma” and “fish” tracks

The intensity of the ionisation in the final tenth of a millimetre of the range of an ordinary  $\beta$ -ray in air is so great, and the deviations so frequent and large, that its cloud track generally ends in a more or less spherical bunch or knot, consisting of drops too closely packed for resolution. Any  $\beta$ -ray of shorter range than about  $1/10$  mm. (i.e., of energy rather less than that corresponding to 2,000 volts) is represented by a cloud track which consists of the sphere alone. If the range is slightly greater the initial portion of the track may show as a small tail projecting from the sphere; we thus get a comma-like track. If the range is a little longer the form of the track is such that, when a number appear together with their “tails” all pointing in one direction, they resemble a shoal of small fishes.

In air exposed to  $X$ -rays such “sphere” tracks and “fish” tracks together generally considerably exceed in number the long tracks, if the latter have an average range exceeding 1.5 cm. (figs. 3 and 4, Plates 2 and 3).

The fish tracks and comma tracks are absent, and the sphere tracks are relatively few, if the long tracks are all of range as short as 7 mm. (Fig. 5, Plate 4.)

When the frequency of the incident radiation is increased beyond a point where the ordinary  $\beta$ -ray tracks have a range of about 1 cm. the number of sphere tracks begins to increase rapidly and soon becomes comparable with that of the long tracks. With a further increase in the frequency of the radiation some of the spheres develop tails on the side next the source and become comma-shaped. When the  $X$ -rays are hard enough to eject  $\beta$ -particles of 1.5 cm. range, fish tracks of ranges up to about 0.4 mm. appear; their range increases as the frequency of the incident radiation is

increased, but rarely exceeds 1.5 mm., even when the long tracks have a range exceeding 3 cm.

An estimate of the energy and frequency of the radiation absorbed in ejecting the electrons which produce the comma and fish tracks may be obtained from the ranges of ordinary  $\beta$ -particles ejected at the same time. It has, however, to be remembered that the incident radiations were far from homogeneous, so that the radiation absorbed in producing the fish tracks may have been of somewhat different frequency from that which corresponds to the mean range of the long tracks. But the data are sufficient to show that the difference between the energy of a quantum of the incident radiation and the kinetic energy of the ejected electron to which a fish track is due is between 20,000 and 30,000 volts; and that the maximum wave-length of the radiation consistent with the production of the fish tracks is between 0.4 and 0.6  $\text{\AA}$ .

It is just to this region of the spectrum that Barkla<sup>6</sup> and Crowther<sup>7</sup> have assigned the wave-length of the  $\beta$ -radiations, of which they found evidence in certain anomalies encountered in the study of absorption and scattering in elements of low atomic number.

Before it is concluded that there are electrons in the atoms of oxygen or nitrogen, of which the work of ejection is between 20,000 and 30,000 volts, or on the other hand, that this difference between the energy of a quantum of the incident radiation and that of the short range  $\beta$ -ray is represented by scattered radiation—in accordance with a suggestion made by Compton—some other possibilities must be considered.

It might in the first place be objected that the short-range  $\beta$ -tracks may be due to a constituent of correspondingly long wave-length remaining in the incident radiation. The two classes of tracks (of a range of several centimetres and of a fraction of a millimetre) obtained, for example, when the radiation has previously been filtered through 2 cm. of aluminium, or its equivalent in copper, tin or lead, might be due to the residual primary radiation and to the characteristic radiations from the screening material respectively. But the short range tracks are confined almost exclusively to the primary beam as defined by the two diaphragms at the ends of the collimating tube, whether the screen is placed between the  $X$ -ray bulbs and the collimating tube or between the collimating tube and the thin glass of the cloud-chamber. In the latter case the fluorescent  $X$ -rays should radiate in all directions from the screen, not only along the direction of the primary

---

<sup>6</sup>Barkla, Phil.Trans., A, vol. 217, p. 315; Phil.Mag., vol. 34, p. 270 (1917).

<sup>7</sup>Crowther, Phil.Mag., vol. 42, p. 719 (1921).

radiation. Again the nature and relative number of the short tracks was not found to depend on the material of the screen.

The fish and comma tracks might also be attributed to the ejection of electrons from the  $K$ -levels of atoms of the rare gases of the atmosphere of higher atomic number, the relatively small number of which would be to some extent counterbalanced by the very rapid increase of absorption with increasing atomic number. The  $K$ -absorption limit of xenon falls not far below the lower limit found above for the effective wave-length. The number of the short-range tracks – which are approximately as numerous as the long tracks when the frequency of the incident radiation is high – is, however, far too great for this source to be a possible one.

A “fish” track or comma track is thus almost certainly due to the ejection of an electron from an atom of one of the common constituents of the air by radiation of the same frequency as that to which the long tracks are due. If a whole quantum of the incident radiation is absorbed in the ejection of an electron from an atom of oxygen or nitrogen, the electron must have come from an energy level below that of the  $K$ -electrons, the difference between this “ $J$ ”-level and the  $K$ -level being represented by 20 or 30 kilovolts; i.e., the electron must have been very closely attached to the nucleus.

There are great difficulties in the way of accepting the view that there is a “ $J$ ” energy level in the atom, and that the difference between the kinetic energies of the electrons producing the long and short tracks respectively represents a difference in the energies required to remove electrons from the  $J$  and  $K$ , energy levels. The fact that these short-range electrons are ejected very nearly in the forward direction is of itself sufficient to indicate that the process of ejection differs essentially from that of the ejection of the ordinary long-range electron.

To account for various phenomena relating to the wave-length and distribution of secondary  $X$ -rays A.H. Compton has suggested the possibility of just such a forward ejection of electrons as actually occurs in these “fish” tracks.

Compton<sup>8</sup> points out that if there is a type of scattering in which a whole quantum of radiation is dealt with by one electron of the atom, this electron may be expected to receive the whole momentum,  $h\nu/c$ , carried by the radiation. If we suppose that the scattered radiation is emitted by the electron in all directions—not localised in a bundle—then the electron will gain by the scattering process a momentum,  $mu = h\nu/c$ , in the forward

---

<sup>8</sup>“Secondary Radiations produced by  $X$ -rays,” Bull. Nat. Research Council, Washington (1922).

direction and a kinetic energy  $\frac{1}{2}mu^2 = \frac{1}{2}h\nu u/c$ .

Let us suppose that it is only the  $K$ -electrons which are effective in this type of scattering in elements of low atomic number like oxygen and nitrogen. There will then be no ejection of the electron from the atom unless  $\frac{1}{2}mu^2$  exceeds the energy corresponding to the  $K$ -absorption limit—the equivalent of about 380 volts in the case of nitrogen. If we calculate the energy of the incident radiation  $h\nu = cmu$  corresponding to this value of  $\frac{1}{2}mu^2$  we obtain  $h\nu = 19,800$  volts. This may be regarded as giving the calculated minimum energy of the radiation which is required to produce rudimentary “fish” tracks; at this stage such a track would be represented by a single pair of ions. Similarly 28,000 volts would give to the scattering electron twice the energy required to eject it from the atom, and a typical “sphere” track would result; about 50,000 volts would be required to give a properly developed “fish” track.

The agreement between the observed phenomena and these applications of Compton’s theory lends strong support to that theory. It is a question of great interest whether the quantum of radiation scattered by an electron is emitted in all directions (with a continuous wave front) as assumed above, or in one direction only as Compton suggests. In the latter case the direction and magnitude of the resultant momentum of the electron will depend on the direction in which it emits the radiation.

Information on this very fundamental question may possibly be obtained from a more thorough study of the initial directions and ranges of the fish tracks produced by homogeneous radiation of known wave-length.

Many of the sphere tracks which appear in the path of a beam of hard  $X$ -rays are almost certainly of the same nature as the fish tracks; they are merely of too small range to show any “tail.”

Others, however, are plainly the tracks of electrons ejected by the  $K$ -radiations emitted by atoms from which an electron has been ejected by the primary beam. They may occur outside the primary beam; they are frequently situated close to the origins of other tracks, forming one component in several of the classes of paired  $\beta$ -ray tracks considered in the next section.

### *8. Association of $\beta$ -ray tracks in pairs or groups*

The  $\beta$ -ray tracks produced in air exposed to  $X$ -rays very frequently occur in pairs or groups. The association in pairs or groups may be of any one of the following types:—

- (1) A long-range track and a short-range track start from the same point.

- (2) One, or sometimes two or more, short-range tracks, generally sphere tracks, appear close to the origin of a long-range track.
- (3) Two short-range tracks form a pair without any long-range track being associated with them.
- (4) Two long-range  $\beta$ -ray tracks, generally similar in range, have their origins near together.
- (5) Two long range tracks start from the same point.

It is, of course, only when the tracks are not too closely packed that it becomes possible to distinguish some of the above types of pairing. Unless the average distance apart of the origins of the tracks exceeds considerably the distance apart of the components of a pair, the relationship of the members of a pair will not generally be obvious.

That the association in pairs or groups is real and not accidental is, I think, made sufficiently clear in the photographs which have been reproduced. Fig. 7, Plate 6, contains examples of types (1), (2) and (5).

*Class (1). A long-range track and a short-range track from the same point. (Figs. 8, 9 and 17, Plates 7 and 11.)*

About 20 per cent. of the "long"  $\beta$ -ray tracks which have been produced by hard X-rays in air at pressures not differing much from atmospheric show distinct indications that a very short-range  $\beta$ -ray starts from the same origin. In some cases there is merely an enlarged head to the long  $\beta$ -track, no larger than some of the beads which occur along the course of the track; but very frequently the short-range track has a range of 2 or 3-tenths of a millimetre and shows quite distinctly as a lateral projection of the origin of the long track. At lower pressures the short tracks are of easily measurable length. In air at a pressure of 10 cm. of mercury many of these initial lateral tracks are from 1.5 to 2 mm, in length. A large proportion start approximately at right angles to the long track, but other very different angles, larger or smaller, occur. In several cases the short track is along the direction of the primary X-ray beam, like a "fish" track.

Even at the lower pressures a long-range track may show at its origin merely a bead (sometimes wholly or partially resolved into drops) to indicate the emission of the short-range  $\beta$ -ray. The effect of very short-range  $\beta$ -particles is more easily detected at low pressures, and it is probably mainly in consequence of this that a larger proportion of long tracks (about 30 per

cent.) have at such pressure been recorded as showing the initial short-range track.

It is possible that in many cases the two particles have been ejected from the same atom, the  $\beta$ -particle which produces the short-range track having been ejected by the faster electron in the course of its escape. On the other hand the origin of this type of pairing of  $\beta$ -ray tracks may be essentially the same as that of the other types (classes 2, 3 and 4) in which the two  $\beta$ -partides do not come from the same atom.

*Class (2). A long-range track with short-range track close to its origin. (Figs. 10, 11, 12 and 18, Plates 8, 9, and 11.)*

When air is exposed to  $X$ -radiation of sufficiently short wave-length to produce  $\beta$ -rays exceeding about 15 mm. in range, about 30 per cent. of these have short tracks associated with them, but starting from separate origins; the origins of the long and short tracks are generally at distances varying from a small fraction of a millimetre up to several mm. in air at 50 mm. pressure. The long track of such a pair may or may not have a short track starting from its origin and forming with it a pair of the first class.

There can be little doubt that in some cases the short track of a pair of the second class is due to a  $\beta$ -particle, ejected from a second atom by the action of the  $K$ -radiation of the atom from which the faster  $\beta$ -particle was ejected. The  $K$ -radiations from oxygen or nitrogen would give rise to  $\beta$ -rays of ranges indistinguishable from these sphere tracks. There are frequently two, sometimes even three, or more, short tracks associated with a long one. Some cases of this kind may represent the handing-on of energy from atom to atom of successively lower atomic number; e.g., the  $\beta$ -ray which gives rise to the long track may come from the  $K$ -level of an oxygen atom, which then emits its  $K$ -radiation; this may eject an electron from the  $K$ -level of a neighbouring nitrogen atom. The  $\beta$ -particle thus ejected from nitrogen would only have sufficient energy (about 100 volts =  $O_K - N_K$ ) to set free a small group of ions; the nitrogen atom will in turn emit its  $K$ -radiation which may be absorbed in ejecting an outer electron from a neighbouring atom. The sphere tracks are as might be expected, on this view of different dimensions, and some of the smaller which have been resolved only contain about 10 pairs of ions.

The above explanation gives, however, by no means a complete account of the phenomena, and can in fact apply only to a relatively small number of

cases of this type of pairing. In a narrow beam of soft  $X$ -rays only a small proportion, not exceeding two or three per cent. of the ordinary  $\beta$ -ray tracks are accompanied by sphere tracks (other than those which appear as beads on the  $\beta$ -ray tracks themselves); and such sphere tracks as do occur are as often as not outside the primary  $X$ -ray beam. The above explanation may apply in the case of these soft  $X$ -rays; the photographs show, however, that the greater number of the quanta of  $K$ -radiation emitted by the gas exposed to the primary beam must under these conditions either escape beyond, the volume of air under observation, before ejecting an electron and producing a sphere track or they are absorbed in the long  $\beta$ -ray tracks themselves.

With  $X$ -radiation of frequency sufficient to produce  $\beta$ -rays of 15 mm. or more in range, the proportion of those which have sphere tracks associated with them is greatly increased; the increase in the number of sphere tracks is confined almost entirely to the air lying within the primary  $X$ -ray beam. A sphere track outside the primary beam, like that in the lower part of Fig. 17, is quite rare. If the pressure is reduced they still remain within the primary  $X$ -ray beam while the average distance of each from its associated long-range track increases.

The effect of reducing the pressure shows that for the production of both the long and short components of a pair, the direct action of the primary beam is in general essential. This is also shown in a striking way when the  $X$ -ray beam is given the form of a narrow vertical sheet; the two components have then their origins very nearly in the same vertical plane.

If, as is probable, one of the components of a pair is primary and the other secondary, the secondary radiation from the point of origin of the primary  $\beta$ -ray is thus shown to be much more likely to eject a  $\beta$ -particle from an atom exposed to the primary  $X$ -rays than from one lying outside the primary beam.

The short track associated with the long one is not always a simple sphere track; its range may be sufficient to give it the comma-shaped or fish-like form. When of this form it is generally directed approximately along the direction of the  $X$ -ray beam in the same way as an independent "fish" track (figs. 12 and 18, Plates 9 and 11); it may, however, be directed with its "tail" pointing towards the origin of the longer track.

When the short track is of the simple sphere form, its most frequent situation relative to the origin of the long track is along a perpendicular to the initial direction of the long track (fig. 10, Plate 8).

When two short tracks are associated with a long track they may both be alike, and may form an obvious pair, or they may be quite unlike in range and orientation. The "long" track of fig. 12, Plate 9, has apparently

two pairs of short tracks associated with it; one pair consists of two similar spheres the other of a sphere track and a "fish" track.

*Two separate short-range tracks. (Figs. 13–16, Plates 9 and 10.)*

Short-range tracks, other than those associated with the same long track also frequently occur in pairs. About 40 per cent. of such tracks belong to pairs.

Most frequently the pair takes the form of two sphere tracks. (Fig. 14, Plate 10.) A fish track is, however, frequently accompanied by a sphere track or by another fish track. (Figs. 15 and 13, Plates 10 and 9.)

There is a tendency for the points of origin of the two components of a pair to lie nearly on a line perpendicular to the axis of the  $X$ -ray beam. (Figs. 12 and 13, Plate 9.) One of the components – a sphere track – may lie outside the primary beam, but this is exceptional.

Experiments on the effects of varying the form and area of cross-section of the beam of  $X$ -radiation and the pressure of the air led to results similar to those which were found to hold for the previous class of pairs. If one of the components of the pair of  $\beta$ -rays is to be regarded as primary the other as due to radiation from the atom from which the primary  $\beta$ -particle has been ejected; then to explain the experimental results we must conclude that this secondary radiation is much more likely to eject a  $\beta$ -particle from an atom exposed to the primary radiation than from one lying outside the primary beam.

When the primary beam was about 0.5 mm. in air at 50 cm. pressure. In a wide beam distance up to 5 or 6 m. occurred, the mean distance exceeding 2 mm., and there was a much greater tendency for the two components of a pair to have their origins in a line nearly perpendicular to the axis of the  $X$ -ray beam; the frequency of occurrence of pairs as compared with single short-range tracks was also increased.

*Class (4). Two long-range tracks from neighbouring points. (Figs. 21 and 22, Plate 12.)*

A considerable proportion of the long tracks (i.e., of all tracks other than sphere, comma and fish tracks) occur in pairs. In a wide beam of hard  $X$ -rays (of wave-length less than  $0.5 \text{ \AA}^\circ$ ) about 40 per cent. of such tracks

were found to be paired. The proportion of paired tracks was smaller, not much exceeding 20 per cent. in a narrow beam of soft radiation in which all the tracks were less than about 7 mm. in range. In such a beam as has already been pointed out fish tracks are absent and sphere tracks few.

Both members of a pair are, with very rare exceptions, within the primary  $X$ -ray beam. In a wide beam their average distance apart amounts to 2 or 3 nun. in air at 50 cm. pressure. There is, as with pairs of the preceding class, a great tendency for the line joining the points of origins of the two members of a pair to be nearly perpendicular to the primary  $X$ -ray beam.

In a narrow beam, the average distance apart is diminished, as is to be expected if both components originate within the primary beam; it is frequently about 0.5 mm, in a cylindrical beam of 0.5 mm. diameter. The line joining the origins of the two members of a pair is also in general much less inclined to the axis of the  $X$ -ray beam than in a wide beam.

At low pressures the average distance apart of the origins of the members of a pair is increased, and thus the two components of the only pairs which appear—since both rays in general start within the primary beam—have their origins in a line making only a small angle with the axis of the  $X$ -ray beam.

As with the two preceding classes of pairs, when the  $X$ -ray beam has the form of a narrow vertical sheet, the two members of a pair nearly always lie in almost the same vertical plane. This is perhaps the most striking proof that the direct action of the primary  $X$ -rays is involved in the production of both members of a pair.

Groups of three or even more long  $\beta$ -rays sometimes occur, but much less frequently than pairs. In a wide beam there is, as with a pair, tendency for all the members of such a group to have their points of origin in the same plane, nearly perpendicular to the primary beam.

*Class (5). Two long-range tracks from the same point.  
(Figs. 19 and 20, Plate 11.)*

The emission of two long-range  $\beta$ -rays from the same point (or from points too near for resolution in the stereoscopic pictures) is not uncommon. Cases of the emission of three and even of four  $\beta$ -rays from the same point have been noticed. These may all possibly form merely a particular case of the preceding class in which the two electrons have been ejected from molecules which are too near together for resolution. But in practice they form quite a distinct class. It seems, moreover, quite natural to suppose that, when the

conditions for absorption are suitable, the radiation from an atom should have a specially great chance of being absorbed by the same atom or by another atom of the same molecule.

The number of cases of two long tracks originating from the same point as compared with the whole number of cases of paired long tracks is greatest in a narrow beam of  $X$ -rays traversing air at low pressure. Of the whole number of pairs of long tracks about one-third consist of two from the same point in a 0.5 mm. beam at a pressure of about 15 cm. This is in accordance with the view that the two components of the pair originate in the same molecule. For the chance of absorption by an atom of the same molecule as that in which the quantum of secondary radiation is emitted remains unaltered by lowering the pressure or narrowing the beam, while the total number of other molecules available for absorption is proportional to both the air pressure and the cross section of the beam.

There is a tendency for the two members of a pair of long  $\beta$ -rays, whether they originate at the same or neighbouring points, to be similar in range, and in the angles which their initial directions make with the axis of the primary beam of  $X$ -rays. A striking example is that of fig. 22, Plate 12.

Two long tracks from the same or neighbouring points may be accompanied by associated short tracks. For example, in fig. 20, Plate 11, a typical "comma" track is associated with two long tracks and another short track, the last three apparently starting from the same point.

### *9. Time interval between the ejections of the two components of a pair*

Owing to the fact that a resistance was inserted in the circuit, each discharge of the Leyden jars through the  $X$ -ray bulb lasted an appreciable time—about 0.01 second. During this time the nature of the  $X$ -rays emitted may have varied much in wave-length, polarisation and otherwise. Thus the tendency of the two  $\beta$ -ray tracks of a pair to be similar might be interpreted as meaning that the  $\beta$ -particles were ejected so nearly simultaneously that the radiations effective in the ejection of both were similar in character. Some of the photographs were taken under conditions such that the expansion occurred during the  $X$ -ray discharge, so that, while some of the tracks were sharp, others (due to  $\beta$ -particles ejected before the expansion was completed) were wholly or partially separated by the electric field into positive and negative diffuse tracks. In such a picture, in accordance with the above explanation, a pair may occur in which both components are sharp, while

nearly all the other tracks are diffuse, or diffuse when most of the other tracks are sharp.

What is somewhat unexpected is the appearance in such pictures of what seem undoubtedly to be pairs, in which one component is sharp and the other diffuse; i.e; pairs of which the two components have been ejected with an appreciable time interval. A rough estimate can be made from the vertical separation which the positive and negative ions of the diffuse tracks suffered before being fixed; it is of the order of  $1/1,000$  of a second.

### *10. On the origin of the paired tracks*

In Section 5 evidence was brought forward to show that  $X$ -rays of sufficiently short wave-length in traversing air cause the ejection of two classes of  $\beta$ -particles which differ greatly in range; the difference in their kinetic energies generally corresponds to more than 20,000 volts. Let us suppose that as a result of ejection of the electron and consequent re-arrangement of the remaining electrons there follows in both cases the emission of a quantum of  $K$ -radiation from the atom.

The  $K$ -radiation from an atom of nitrogen in the absence of other influences is not able to eject an electron from the  $K$ -level of a second atom of nitrogen. The phenomena relating to the paired tracks suggest that such an atom, when exposed to the  $K$ -radiation from a similar atom, together with radiation of higher frequency, is able to absorb both radiations (with ejection of a  $K$  electron) much more readily than either separately. Let us suppose that the ejection of the electron by the primary  $X$ -rays, whether it occurs with or without the help of the  $K$ -radiation, may be either of the type which gives the long-range  $\beta$ -particle or of that which gives the short-range forward-directed  $\beta$ -particle. Then all the various types of pairs—all possible combinations of long and short  $\beta$ -ray tracks—may be accounted for.

In the case of an element of low atomic number like nitrogen, it is perhaps not impossible that the  $K$ -radiation, while not able of itself to eject an electron from the  $K$ -level of a similar atom, may be absorbed in transferring the electron to an outer level, and that the  $K$ -radiation may be then re-emitted and handed on from atom to atom like ordinary resonance radiation. The occasional occurrence of a time interval of the order of  $1/1,000$  of a second between the ejection of the primary and secondary  $\beta$ -particles is perhaps more easily understood on this view.

### *Summary of Results*

Many hundred stereoscopic pictures showing the number, distribution, direction of ejection and range of the  $\beta$ -particles emitted from atoms in air exposed to  $X$ -rays have been obtained and examined. The following are some of the conclusions to which a study of the photographs has led.

(1) The cloud-method is able to deal with individual quanta of radiation, in the sense that the track of the electron ejected from the atom which emits the quantum of radiation and that of the electron ejected from the atom which absorbs the radiation may under suitable conditions be identified.

(2) Two classes of  $\beta$ -ray tracks are produced in air by the primary action of  $X$ -radiation of wave-length less than about  $0.5\text{ }A^\circ$ —(a) those of ejected electrons with initial kinetic energy comparable to a quantum of the incident radiation, and (b) tracks of very short range. The short range electrons are ejected nearly along the direction of the primary  $X$ -rays. Their direction and range and the value of the minimum frequency of the radiation which is required to produce them are in agreement with the suggestion made by A. H. Compton, that a single electron may be effective in scattering a quantum of radiation and that in so doing it receives the whole momentum of the quantum. The short-range tracks are probably related to the phenomena which have led to the postulation of a "J"-radiation.

(3) The ordinary long-range tracks may be divided into three classes according to the direction of ejection of the electron. The majority have a large forward component comparable with the lateral component; a considerable proportion, of the order of 20 per cent., are ejected almost exactly at right angles to the primary  $X$ -ray beam; others have a large backward component.

(4) Partial polarisation of the primary beams is indicated by the direction of ejection of a number of the  $\beta$ -particles being in one plane—that containing the direction of the cathode rays in the  $X$ -ray tube.

(5)  $\beta$ -rays in air exposed to  $X$ -rays frequently occur in pairs or groups, of which five classes have been distinguished. The pairs probably consist of one  $K$ -electron ejected by the direct action of the primary  $X$ -rays, and of a second electron ejected by the combined action of primary radiation and of the  $K$ -radiation from the atom from which the first electron was ejected.

## Planck's Law and Light Quantum Hypothesis.

S.N. Bose

(Received 1924)

Planck's formula for the distribution of energy in the radiation from a black body was the starting point of the quantum theory, which has been developed during the last 20 years and has borne a wealth of fruit in energy domain of physics. Since its publication in 1901 many methods for deriving this law have been proposed. It is recognized that basic assumptions of the quantum theory are irreconcilable with the laws of classical electrodynamics. All derivations up to now use the relation

$$\rho_\nu d\nu = \frac{8\pi\nu^2 d\nu}{c^3} E,$$

that is, the relation between the radiation density and the mean energy of an oscillator, and they make assumptions about the number of degrees of freedom of the ether, which appear in the above formula (the first factor on the right-hand side). This factor, however, can be derived only from classical theory. This is the unsatisfactory feature in all derivations and it is therefore no wonder that attempts are being made to obtain a derivation that is free of this logical flaw.

Einstein has given a remarkably elegant derivation. He recognized the logical defect of all previous derivations and tried to deduce the formula independently of classical theory. From very simple assumptions about the

energy exchange between molecules and a radiation field he found the relation

$$\rho_\nu = \frac{\alpha_{mn}}{e^{\frac{\epsilon_m - \epsilon_n}{kT}} - 1}.$$

To make this formula agree with Planck's he had to use Wien's displacement law and Bohr's correspondence principle. Wien's law is based on classical theory and the correspondence principle assumes that the quantum theory and the classical theory coincide in centrum limits.

In all cases it appears to me that the derivations have not been sufficiently justified from a logical point of view. As opposed to these the light quantum hypothesis combined with statistical mechanics (as it was formulated to meet the needs of the quantum theory) appears sufficient for the derivation of the law independent of classical theory. In the following I shall sketch the method briefly.

Let the radiation be enclosed in the volume  $V$  and let its total energy be  $E$ . Let various types of quanta be present of abundances  $N_s$  and energy  $h\nu_s$  ( $s = 0$  to  $s = \infty$ ). The total energy is then

$$E = \sum_s N_s h\nu_s = V \int \rho_\nu d\nu \quad (1)$$

The solution of the problem therefore requires the determination of the  $N_s$ , which, in turn, determine  $\rho_\nu$ . If we can give the probability for each distribution characterized by arbitrary values of  $N_s$  then the solution is given by the condition that this probability is to be a maximum, keeping in mind the condition (1) which is a constraint on the problem. We now seek this probability.

The quantum has the momentum  $\frac{h\nu_s}{c}$  in the direction of its motion. The momentary state of the quantum is characterized by its coordinates  $x, y, z$  and the corresponding components of the momentum  $p_x, p_y, p_z$ . These six quantities can be considered as point coordinates in a six-dimensional space, where we have the relation

$$p_x^2 + p_y^2 + p_z^2 = \frac{h^2 \nu^2}{c^2},$$

in virtue of which point representing the quantum in our six-dimensional space is forced to lie on a cylindrical surface determined by the frequency. To the frequency range  $d_\nu$  there belongs in this sense the phase space

$$\int dx dy dz dp_x dp_y dp_z = V \cdot 4\pi (h\nu/c)^2 h d\nu / c = 4\pi \cdot h^3 \nu^3 / c^3 \cdot V \cdot d\nu$$

If we divide the total phase volume into cells of size  $h^3$ , there are then  $4\pi \cdot \nu^2/c^3 \cdot d\nu$  cells in the frequency range  $d\nu$ . Nothing definite can be said about the method of dividing the phase space in this manner. However, the total number of cells must be considered as equal to the number of possible ways of placing a quantum in this volume. To take into account polarization it appears necessary to multiply this number by 2 so that we obtain  $8\pi V\nu^2 d\nu/c^3$  as the number of cells belonging to  $d\nu$ .

It is now easy to calculate the thermodynamic probability of a (macroscopically defined) state. Let  $N^s$  be number of quanta belonging to the frequency range  $d\nu^s$ . In how many ways can these be distributed among the cells that belong to  $d\nu^s$ ? Let  $p_0^s$  be number of empty cells.,  $p_1^s$  the number containing 1 quantum,  $p_2^s$  the number containing 2 quanta, and so on. The number of possible distributions is then

$$\frac{A^s!}{p_0^s! p_1^s! \dots} \quad \text{where} \quad A^s = \frac{8\pi\nu^2}{c^3} \cdot V d\nu^s$$

and where

$$N^s = 0 \cdot p_0^s + 1 \cdot p_1^s + 2p_2^s + \dots$$

is the number of quanta belonging to  $d\nu^s$ .

The probability  $W$  of the state defined by all  $p_r^s$  is clearly

$$\prod_s \frac{A^s!}{s p_0^s! p_1^s! \dots}$$

Taking into account that the  $p_r^s$  are large numbers we have

$$\log W = \sum_s A^s \log A^s - \sum_s \sum_r p_r^s \log p_r^s$$

where

$$A^s = \sum_r p_r^s.$$

This expression must be a maximum under the constraints

$$E = \sum_s N^s h \nu^s; \quad N^s = \sum_r r p_r^s.$$

Carrying through the variations we obtain the conditions

$$\sum_s \sum_r \delta p_r^s (1 + \log p_r^s) = 0, \quad \sum_s \delta N^s h \nu^s = 0$$

$$\sum_r \delta p_r^s = 0 \quad \delta N^s = \sum_r r \delta p_r^s.$$

From this we obtain

$$\sum_r \sum_s \delta p_r^s (1 + \log p_r^s + \lambda^s) + \frac{1}{\beta} \sum_s r \delta p_r^s = 0$$

From this we first see that

$$p_r^s = B^s e^{-\frac{rh\nu^s}{\beta}}.$$

Since, however,

$$A^s = \sum_r B^s e^{-\frac{rh\nu^s}{\beta}} = B^s (1 - e^{-\frac{h\nu^s}{\beta}})^{-1}$$

then

$$B^s = A^s (1 - e^{-\frac{h\nu^s}{\beta}}).$$

We further have

$$N^s = \sum_r r p_r^s = \sum_r r A^s (1 - e^{-\frac{h\nu^s}{\beta}}) e^{-\frac{rh\nu^s}{\beta}} = \frac{A^s e^{-\frac{h\nu^s}{\beta}}}{1 - e^{-\frac{h\nu^s}{\beta}}}$$

Taking into account the value of  $A^s$  found above, we have

$$E = \sum_S \frac{8\pi h\nu^{s3} d\nu^s}{c^3} V \frac{e^{-\frac{h\nu^s}{\beta}}}{1 - e^{h\nu^s/\beta}}$$

Using the result obtained previously

$$S = k \left[ \frac{E}{\beta} - \sum_s A^s \log(1 - e^{h\nu^s/\beta}) \right]$$

and noting that

$$\frac{\partial S}{\partial E} = \frac{1}{T}$$

we obtain

$$\beta = kT$$

Hence

$$E = \sum_S \frac{8\pi h\nu^{s3}}{c^3} V \frac{1}{e^{h\nu^s/kT} - 1} d\nu^s$$

which is Planck's formula.

*Comment of translator.* Bose's derivation of Planck's formula appears to me to be an important step forward. The method used here gives also the quantum theory of an ideal gas, as I shall show elsewhere. [A. Einstein]

# On the Connexion between the Completion of Electron Groups in an Atom with the Complex Structure of Spectra

W. PAULI

Z. Physik 31, 765ff (1925).

---

Especially in connexion with Millikan and Land<sup>©</sup>'s observation that the alkali doublet can be represented by relativistic formulae and with results obtained in an earlier paper, it is suggested that this doublet and its anomalous Zeeman effect expresses a classically non-describable two-valuedness of the quantum theoretical properties of the optically active electron [Germ: *Leuchtelektron*], without any participation of the closed rare gas configuration of the atom core in the form of a core angular momentum or as the seat of the magneto-mechanical anomaly of the atom. We then attempt to pursue this point of view, taken as a temporary working hypothesis, as far as possible in its consequences also for atoms other than the alkali atoms, notwithstanding its difficulties from the point of view of principle. First of all it turns out that it is possible, in contrast to the usual ideas, to assign for the case of a strong external magnetic field, which is so strong that we can neglect the coupling between the atomic core and the optically active electrons, to those two systems, as far as the number of their stationary states, the values of their quantum numbers, and their magnetic energy is concerned, no other properties than those of the free atomic core of the optically active electron of the alkalis. On the basis of these results one is also led to a general classification of every electron in the atom by the principal quantum number  $n$  and two auxiliary quantum numbers  $k_1$  and  $k_2$  to which is added a further quantum number  $m_1$  in the presence of an external field. In conjunction with a recent paper by E. C. Stoner this classification leads to a general quantum theoretical formulation of the completion of electron groups in atoms.

---

## 1. The Permanence of Quantum Numbers (Principle of Gradual Construction [Germ: *Aufbauprinzip*]) in Complex Structures and the Zeeman Effect

In a previous paper<sup>1</sup> it was emphasised that the usual ideas, according to which the inner, completed electron shells of an atom play an essential part in the complex structure of optical spectra and their anomalous Zeeman effect in the shape of core angular momenta and as the real seat of the magneto-mechanical anomaly, are subject to several serious difficulties. It seems therefore plausible to set against these ideas that especially the doublet structure of the alkali spectra and their anomalous Zeeman effect are caused by a classically undescribable two-valuedness of the quantum theoretical properties of the optically active electron. This idea is particularly based upon the results of Millikan and Land<sup>©</sup> that the optical doublets of the alkalis are similar to the relativity doublets in X-ray spectra and that their magnitude is determined by a relativistic formula.

If we now pursue this point of view, we shall assign -- as was done by Bohr and Coster

for the X-ray spectra -- to the stationary states of the optically active electron involved in the emission of the alkali spectra two auxiliary quantum numbers  $k_1$  and  $k_2$  as well as the principal quantum number  $n$ . The first quantum number  $k_1$  (usually simply denoted by  $k$ ) has the values 1, 2, 3, ... for the s, p, d, . . . terms and changes by unity in the allowed transition processes; it determines the magnitude of the central force interaction forces of the valence electron with the atom core. The second quantum number  $k_2$  is for the two terms of a doublet (e.g.,  $p_1$  and  $p_2$ ) equal to  $k_1 - 1$  and  $k_1$ , in the transition processes it changes by  $\pm 1$  or 0 and determines the magnitude of the relativity correction (which is modified according to Landé to take into account the penetration of the optically active electron in the atom core). If we follow Sommerfeld to define the total angular momentum quantum number  $j$  of an atom in general as the maximum value of the quantum number  $m_1$  (usually simply denoted by  $m$ ) which determines the component of the angular momentum along an external field, we must put  $j = k_2 - 1$  for the alkalis. The number of stationary states in a magnetic field for given  $k_1$  and  $k_2$  is  $2j + 1 = 2k_2$ , and the number of these states for both doublet terms with given  $k_1$  is altogether  $2(2k_1 - 1)$ .

If we now consider the case of strong field (Paschen-Back effect), we can introduce apart from  $k_1$  and the just mentioned quantum number  $m_1$ , instead of  $k_2$  also a magnetic quantum number  $m_2$  which determines directly the energy of the atom in the magnetic field, that is, the component of the magnetic moment of the valence electron parallel to the field. For the two terms of the doublet it has, respectively, the values  $m_1 + 1/2$  [N.B one-half] and  $m_1 - 1/2$ . Just as in the doublet structure of the alkali spectra the "anomaly of the relativity correction" is expressed (the magnitude of which is mainly determined by another quantum number, as is the magnitude of the central force interaction energy of the optically active electron and the atom core), so appears in the deviations of Zeeman structure from the normal Lorentz triplet the "magnetomechanical anomaly" which is similar to the other anomaly (the magnitude of the magnetic moment of the optically active electron is mainly determined by another quantum number, as is the angular momentum). Clearly, the appearance of half-odd-integral (effective) quantum numbers and the thereby formally caused value  $g = 2$  of the splitting factor of the s-term of the alkalis is closely connected with the twofoldness of the energy level. We shall here, however, not attempt a more detailed theoretical analysis of this state of affairs and use the following considerations of the Zeeman effect of the alkalis as empirical data.

Without worrying about the difficulties encountered by our point of view, which we shall mention presently, we now try to extend this formal classification of the optically active electron by four quantum numbers  $n, k_1, k_2, m_1$  to atoms, more complex than the alkalis. *It now turns out that we can retain completely on the basis of this classification the principle of permanence of quantum numbers (Aufbauprinzip) also for the complex structure of the spectra and the anomalous Zeeman effect in contrast to the usual ideas.* This principle, due to Bohr, states that when a further electron is added to a -- possibly charged -- atom, the quantum numbers of the electrons which are already bound to the atom retain the same values as correspond to the appropriate state of the free atom core.

Let us first of all consider the alkaline earths. The spectrum consists in this case of a singlet and a triplet system. The quantum states with a well-defined value of the quantum number  $k_1$  of the optically active electron correspond then for the first system to altogether  $1(2k_1 - 1)$  and in the last system to  $3(2k_1 - 1)$  stationary states in an external magnetic field. Up to now this was interpreted as meaning that in strong fields the optically active electron in each case could take up  $2k_1 - 1$  positions, while the atom core was able to take up in the first case one, and in the last case three positions. The number of these positions is clearly different from the number 2 of the positions of the free atom core (alkali-like s-term) in a field. Bohr<sup>2</sup> called this state of affairs a "constraint" [Germ: *Zwang*] which is not analogous to the action of external fields of force. Now, however, we can simply interpret the total  $4(2k_1 - 1)$  states of the atom as meaning that the atom core always has two positions in a field, and the optically active electron as for the alkalis  $2(2k_1 - 1)$  states.

More generally, a branching rule formulated by Heisenberg and Landé<sup>3</sup> states that a stationary state of the atom core with  $N$  states in a field leads through the addition of one more electron to two systems of terms, corresponding to altogether  $(N + 1)(2k_1 - 1)$  and  $(N - 1)(2k_1 - 1)$  states in a field, respectively, for a given value of the quantum number  $k_1$  of the last electron. According to our interpretation, the  $2N(2k_1 - 1)$  states of the complete atom in a strong field come about through  $N$  states of the atom core and  $2(2k_1 - 1)$  states of the optically active electron. In the present quantum theoretical classification of the electrons the term multiplicity required by the branching rule is simply a consequence of the "Aufbauprinzip". According to the ideas presented here Bohr's constraint expresses itself not in a violation of the permanence of quantum numbers when the series electron is coupled to the atom core, but only in the peculiar two-valuedness of the quantum theoretical properties of each electron in the stationary states of an atom.

*We can, however, from this point of view use the "Aufbauprinzip" to calculate not only the number of stationary states, but also the energies in the case of strong fields (at least that part which is proportional to the field) additively from those of the free atom core and of the optically active electron, where the latter can be taken from the alkali spectra. Because, in this case, both the total component  $\overline{m}_1$  of the angular momentum of the atom along the field (in units [h-bar]) as well as the component  $\overline{m}_2$  of the magnetic moment of the atom in the same direction (in Bohr magnetons) are equal to the sum of the quantum numbers  $m_1$  and  $m_2$  of the single electrons:*

$$\overline{m}_1 = Sm_1, \quad \overline{m}_2 = Sm_2. \quad (1)$$

The latter can independently run through all values corresponding to the values of the angular momentum quantum numbers  $k_1$  and  $k_2$  of the electrons in the stationary state of the atom considered. ( $\overline{m}_2\omega h$  is here thus the part of the energy of the atom proportional to the field strength;  $\omega$  = Larmor frequency.)

Let us consider as an example the two s-terms (singlet- and triplet S-term) of the alkaline earths. To begin with it is sufficient to consider only the two valence electrons, as the contribution of the other electrons to the sums in (1) vanish when taken

together. According to our general assumption we must for each of the two valence electrons take (independently of the other electron) the values  $m_1 = -1/2$ ,  $m_2 = -1$  and  $m_1 = 1/2$ ,  $m_2 = 1$  of the  $s$ -terms of the alkalis. According to (1) we then get the following values for the quantum numbers  $\overline{m}_1$  and  $\overline{m}_2$  of the total atom:

$$\begin{aligned}\overline{m}_1 &= -1/2 \quad -1/2, \quad -1/2 \quad +1/2, \quad +1/2 \quad -1/2, \quad +1/2 \quad +1/2 \\ \overline{m}_2 &= \quad -1 \quad -1, \quad \quad -1 \quad +1, \quad \quad 1 \quad -1, \quad \quad 1 \quad +1,\end{aligned}$$

or

$\overline{m}_1 = -1$	0	1
$\overline{m}_2 = -2$	0, 0	2

[Corresponding to one term with  $j = 0$  and one with  $j = 1$  in weak fields.]<sup>a</sup> To obtain the  $p$ -,  $d$ -, . . . terms of the alkaline earths, one must combine in (1) the unchanged contribution of the first valence electron ( $S$ -term) in an appropriate manner with the  $m_1$ - and  $m_1$ -values of the  $p$ -,  $d$ -, . . . terms of the alkalis for the second electron.

The rule (1) leads in general exactly to the procedure for calculating the energy values in strong field proposed recently by Landé<sup>4</sup> which has been shown by this author to give correct results also in complicated cases. According to Landé this procedure leads, for instance, to the correct Zeeman terms of neon (at least in the case of strong fields) if one assumes<sup>b</sup> that in the atom core there is one active electron in a  $p$ -term (instead of in an  $s$ -term as above) and if one lets the optically active electron go through  $s$ -,  $p$ -,  $d$ -,  $f$ -, . . . terms.

This result now suggests that we *characterise in general each electron in an atom not only by a principal quantum number  $n$ , but also by the two auxiliary quantum numbers  $k_1$  and  $k_2$* , even when several equivalent electrons or completed electron groups are present. Moreover, we shall allow (also in the just-mentioned cases) in our thoughts *such a strong magnetic field that we can assign to each electron, independently of the other electrons not only the quantum numbers  $n$  and  $k_1$ , but also the two quantum numbers  $m_1$  and  $m_2$*  (where the last one determines the contribution of the electron to the magnetic energy of the atom). The connexion between  $k_2$  and  $m_2$  for given  $k_1$  and  $m_1$  must be taken from the alkali spectra.

Before we apply in the next section this quantum theoretical classification of the electrons in an atom to the problem of the completion of the electron groups, we must discuss in more detail the difficulties encountered by the here-proposed ideas of the complex structure and the anomalous Zeeman effect and the limitations of the meaning of our ideas.

First of all, these ideas do not pay proper regard to the, in many respects independent, separate appearance of the different term systems (e.g., the singlet and the triplet systems of the alkaline earths), which also play a role in the position of the terms of these systems and in the Landé interval rule. Certainly, one cannot assume two

different causes for the energy differences of the triplet levels of the alkaline earths, both the anomaly of the relativity correction of the optically active electron and the dependence of the interaction between the electron and the atom core on the relative orientation of these two systems.

A more serious difficulty, raising a matter of principle, is however the connexion of these ideas with the correspondence principle which is well known to be a necessary means to explain the selection rules for the quantum numbers  $k_1$ ,  $j$ , and  $m$  and the polarisation of the Zeeman components. It is, to be sure, not necessary according to this principle to assign in a definite stationary state to each electron an orbit uniquely determined in the sense of usual kinematics; however, it is necessary that the totality of the stationary states of an atom corresponds to a collection (class) of orbits with a definite type of periodicity properties. In our case, for instance, the above-mentioned selection and polarisation rules require according to the correspondence principle a kind of motion corresponding to a central force orbit on which is superposed a precession of the orbital plane around a definite axis of the atom to which is added in weak external magnetic fields also a precession around an axis through the nucleus in the direction of the field. The dynamic explanation of this kind of motion of the optically active electron, which was based upon the assumption of deviations of the forces between the atom core and the electron from central symmetry, seems to be incompatible with the possibility to represent the alkali doublet (and thus also the magnitude of the corresponding precession frequency) by relativistic formulae. The situation with respect to the kind of motion in the case of strong fields is similar.

The difficult problem thus arises *how to interpret the appearance of the kind of motion of the optically active electron which is required by the correspondence principle independently of its special dynamic interpretation which has been accepted up to now but which can hardly be retained*. There also seems to be a close connexion between this problem and the question of the magnitude of the term values of the Zeeman effect (especially of the alkali spectra).

As long as this problem remains unsolved, the ideas about the complex structure and the anomalous Zeeman effect suggested here can certainly not be considered to be a sufficient physical basis for the explanation of these phenomena, especially as they were in many respects better reproduced in the usually accepted point of view. It is not impossible that in the future one will succeed in merging these two points of view. In the present state of the problem it seemed of interest to us to pursue as far as possible also the first point of view to see what its consequences are. This is the sense in which one must consider our discussions in the next section of the application of the tentative point of view, presented here to the problem of the completion of electron groups in an atom, notwithstanding the objections which can be made against it. We shall here draw conclusions only about the number of possible stationary states of an atom when several equivalent electrons are present, but not about the position and relative order of the term values.

## **2. On a General Quantum Theoretical Rule for the Possibility of the Occurrence of Equivalent Electrons in an Atom**

It is well known that the appearance of several equivalent electrons, that is, electrons

which are fully equivalent both with respect to their quantum numbers and with respect to their binding energies, in an atom is possible only under special circumstances which are closely connected with the regularities of the complex structure of spectra. For instance, the ground state of the alkaline earths in which the two valence electrons are equivalent corresponds to a singlet S-term, while in those stationary states of the atom which belong to the triplet system the valence electrons are never bound equivalently, as the lowest triplet s-term has a principal quantum number exceeding that of the ground state by unity. Let us now as second example consider the neon spectrum. This consists of two groups of terms with different series limits, corresponding to different states of the atom core. The first group, belonging to the removal of an electron with the quantum numbers  $k_1 = 2, k_2 = 1$  from the atom core can be considered to be composed of a singlet and a triplet system, while the second group, belonging to the removal of an electron with  $k_1 = k_2 = 2$  from the atom core, can be said to be a triplet and quintet system. The ultraviolet resonance lines of neon have not yet been observed, but there can hardly be any doubt that the ground state of a Ne-atom must be considered to be a p-term as far as its combination with the known excited states of the atom is concerned; in accordance with the unique definiteness and the diamagnetic behaviour of the inert gas configuration there can be only *one* such term, namely with the value  $j = 0$ .<sup>c</sup> As the only p-terms with  $j = 0$  are the (lowest) triplet terms of the two groups, we can thus conclude that for Ne for the value 2 of the principal quantum number only those two triplet terms exist and moreover are identical for both groups of terms.

In general we can thus expect *that for those values of the quantum numbers n and  $k_1$  for which already some electrons are present in the atom, certain multiplet terms of spectra are absent or coincide*. The question arises what quantum theoretical rules decide this behaviour of the terms.

As is already clear from the example of the neon spectrum, this question is closely connected with the problem of the completion of electron groups in an atom, which determines the lengths 2, 8, 18, 32, . . . of the periods in the periodic table of the elements. This completion consists in that an  $n$ -quantum electron group neither through emission or absorption of radiation nor through other external influences is able to accept more than  $2n^2$  electrons.

It is well known that Bohr in his theory of the periodic table, which contains a unified summary of spectroscopic and chemical data and especially a quantum theoretical basis for the occurrence of chemically similar elements such as the platinum and iron metals and the rare earths in the later period of the table, has introduced a subdivision of these electron groups into subgroups. By characterising each electron in the stationary states of the atom by analogy with the stationary states of a central force motion by a symbol  $n_k$  with  $k$  [less than or equal to]  $n$ , he obtained in general for an electron group with a value  $n$  of the principal quantum number  $n$  subgroups. In this way Bohr was led to the scheme of the structure of the inert gases given in Table 1. He has, however, emphasised himself<sup>5</sup> that the equality, assumed here, of the number of electrons in the different subgroups of a maingroup is highly hypothetical and that for the time being no complete and satisfying theoretical explanation of the completion of the electron groups in the atom, and especially of the period lengths 2, 8, 18, 32, . . . in the periodic table

could be given.

**Table 1. Original Bohr Scheme of Inert Gas Configurations**

Element	Atomic number	Number of $n_k$ electrons														
		1 <sub>1</sub>	2 <sub>1</sub>	2 <sub>2</sub>	3 <sub>1</sub>	3 <sub>2</sub>	3 <sub>3</sub>	4 <sub>1</sub>	4 <sub>2</sub>	4 <sub>3</sub>	4 <sub>4</sub>	5 <sub>1</sub>	5 <sub>2</sub>	5 <sub>3</sub>	6 <sub>1</sub>	6 <sub>2</sub>
Helium	2	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Neon	10	2	4	4	-	-	-	-	-	-	-	-	-	-	-	-
Argon	18	2	4	4	4	4	-	-	-	-	-	-	-	-	-	-
Krypton	36	2	4	4	6	6	6	4	4	-	-	-	-	-	-	-
Xenon	54	2	4	4	6	6	6	6	6	6	-	4	4	-	-	-
Emanation	86	2	4	4	6	6	6	8	8	8	8	6	6	6	4	4

Recently essential progress was made in the problem of the completion of the electron groups in an atom by the considerations of E. C. Stoner.<sup>6</sup> This author suggests first of all a scheme for the atomic structure of the inert gases in which in contrast to Bohr no opening of a completed subgroup is allowed by letting other electrons of the same main group be added to it, so that the number of electrons in a closed subgroup depends only on the value of  $k$ , but not on the value of  $n$ , that is, on the existence of other subgroups in the same main group. This means already a large simplification which could, moreover, be supported by several experimental data. One must here assume for  $k = 1$  two, for  $k = 2$  six, for  $k = 3$  ten, and in general for a given value of  $k$   $2(2k - 1)$  electrons in the closed state of the corresponding subgroup to remain in agreement with the empirically known numbers of electrons in the inert gases.

Stoner remarked further that *these numbers of electrons agree with the number of the stationary states of the alkali atoms in an external field for the given value of  $k$ .* He therefore pushes further the analogy with the stationary states of the alkali spectra by assuming a further subdivision of the subgroups, corresponding to the complex structure of these spectra (and of the X-ray spectra), which are characterised by the two numbers  $k_1$  and  $k_2$  where  $k_1$  is the same as Bohr's  $k$ , while we must put  $k_2 = k_1 - 1$  or  $k_2 = k_1$  (except for  $k_1 = 1$ , where in accordance with the simple nature of the s-term,  $k_2 = 1$  only). Corresponding to the number  $2k_2$  of the stationary states into which a stationary state of an alkali atom with given values of the quantum numbers  $k_1$  and  $k_2$  decomposes in an external field, Stoner assumes  $2k_2$  electrons in a completed part-subgroup corresponding to the quantum numbers  $n, k_1, k_2$ . Table 2 gives the scheme of the atomic structure of the inert gases, to which Stoner was led in this way.

**Table 2. Scheme of Inert Gas Configurations According to Stoner**

Element	Atomic number	Number of $n_k$ electrons														
		1 <sub>1</sub>	2 <sub>1</sub>	2 <sub>2, (1+2)</sub>	3 <sub>1</sub>	3 <sub>2, (1+2)</sub>	3 <sub>3, (2+3)</sub>	4 <sub>1</sub>	4 <sub>2, (1+2)</sub>	4 <sub>3, (2+3)</sub>	4 <sub>4, (3+4)</sub>	5 <sub>1</sub>	5 <sub>2, (1+2)</sub>	5 <sub>3, (2+3)</sub>	6 <sub>1, (1+2)</sub>	6 <sub>2, (1+2)</sub>
Helium	2	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Neon	10	2	2	2+4	-	-	-	-	-	-	-	-	-	-	-	-	-
Argon	18	2	2	2+4	2	2+4	-	-	-	-	-	-	-	-	-	-	-
Krypton	36	2	2	2+4	2	2+4	4+6	2	2+4	-	-	-	-	-	-	-	-
Xenon	54	2	2	2+4	2	2+4	4+6	2	2+4	4+6	-	2	2+4	-	-	-	-
Emanation	86	2	2	2+4	2	2+4	4+6	2	2+4	4+6	6+8	2	2+4	4+6	2	2+4	-

We can now make this idea of Stoner's more precise and more general, if we apply the ideas about the complex structure of the spectra and the anomalous Zeeman effect, discussed in the previous section, to the case where equivalent electrons are present in an atom. In that case we arrived, on the basis of an attempt to retain the permanence of quantum numbers, at a characterisation of each electron in an atom by both the principal quantum number  $n$  and the two auxiliary quantum numbers  $k_1$  and  $k_2$ . In strong magnetic fields also an angular momentum quantum number  $m_1$  was added to this for each electron and, furthermore, one can use apart from  $k_1$  and  $m_1$  also a magnetic moment quantum number  $m_2$ , instead of  $k_2$ . First of all, we see that the use of the two quantum numbers  $k_1$  and  $k_2$  for each electron is in excellent agreement of Stoner's subdivision of Bohr's subgroup.<sup>d</sup> Secondly, by considering the case of strong magnetic fields we can reduce Stoner's result, that the number of electrons in a completed subgroup is the same as the number of the corresponding terms of the Zeeman effect of the alkali spectra, to the following more general rule about the occurrence of equivalent electrons in an atom:

*There can never be two or more equivalent electrons in an atom for which in strong fields the values of all quantum numbers  $n$ ,  $k_1$ ,  $k_2$ ,  $m_1$  (or, equivalently,  $n$ ,  $k_1$ ,  $m_1$ ,  $m_1$ ) are the same. If an electron is present in the atom for which these quantum numbers (in an external field) have definite values, this state is "occupied".*

We must bear in mind that the principal quantum number occurs in an essential way in this rule; of course, several (not equivalent) electrons may occur in an atom which have the same values of the quantum numbers  $k_1$ ,  $k_2$ ,  $m_1$ , but have different values of the principal quantum number  $n$ .

We cannot give a further justification for this rule, but it seems to be a very plausible one. It refers, as mentioned, first of all to the case of strong fields. However, from thermodynamic arguments (invariance of statistic weights under adiabatic transformations of the system)<sup>e</sup> it follows that the number of stationary states of an atom must be the same in strong and weak fields for given values of the numbers  $k_1$  and  $k_2$  of the separate electrons and a value of  $\overline{m}_1 = Sm_1$  (see (1)) for the whole atom. We can therefore also in the latter case make definite statements about the number of stationary states and the corresponding values of  $j$  (for a given number of equivalent electrons belonging to different values of  $k_1$  and  $k_2$ ). We can thus find the number of possibilities of realising various incomplete electron shells and give an unambiguous answer to the question posed at the beginning of this section about the absence or coincidence of certain multiplet terms in spectra for values of the principal quantum number for which several equivalent electrons are present in an atom. We can, however, only say something about the number of terms and the values of their

quantum numbers, but not about their magnitude and about interval relations.

We must now show that the consequences of our rule agree with experiment in the simplest cases. We must wait and see whether it will also prove itself in comparison with experiment in more complicated cases or whether it will need modifications in that case; this will become clear when complicated spectra are sorted out.

First of all, we see that Stoner's result and with it the period lengths 2, 8, 18, 32, . . . are immediately included in a natural way in our rule. Clearly, for given  $k_1$  and  $k_2$  there cannot be more equivalent electrons in an atom than the appropriate value of  $m_1$ , (that is,  $2k_2$ ) and in the completed group there corresponds exactly *one* electron to each of these values of  $m_1$ .

Secondly, it turns out that our rule has an immediate consequence that the triplet s-term with the same principal quantum number as the ground state is absent for the alkaline earths. If we investigate the possibilities for the equivalent binding of two electrons in s-terms (in that case we have thus  $k_1 = 1$  and  $k_2$  can also only have the value 1), according to our rule the cases are excluded in strong fields where both electrons have  $m_1 = 1/2$  or both have  $m_1 = -1/2$ ; rather, we can only have  $m_1 = 1/2$  for the first electron and  $m_1 = -1/2$  for the second electron, or the other way round<sup>f</sup> so that the quantum number  $\overline{m}_1 = Sm_1$  for the total atom can only have the value 0. Therefore also in weak fields (or when there is no field) only the value  $j = 0$  is possible (singlet S-term).

We now investigate the case that *one* electron is removed from a closed shell, as will occur in X-ray spectra. Clearly when an electron is missing from one of Stoner's part-subgroups, the case is always possible that *no* electron is present with the value  $m_1$ ; we call this the "hole-value" of  $m_1$ . The other electrons are then uniquely divided over the other values of  $m_1$  so that for each of those values we have one electron. The sum of these other values of  $m_1$  and thus the quantum number  $\overline{m}_1$  of the total atom is clearly in each case equal to the opposite of the hole-value of  $m_1$ . If we let it go through all possible values and take into account that an electron can be removed from every part-subgroup, we see that in strong fields the multiplicity of the hole-values of  $m_1$  and thus also that of the values of  $\overline{m}_1$  is the same as that of the  $m_1$  value of a single electron. Due to the invariance of statistical weights it follows thus also for weak fields that the numbers of stationary states and of  $j$ -values of single ionised closed electron shells (X-ray spectra) are the same as in the alkali spectra, in accordance with experiment.

This is a special case of a general *reciprocity law*: *For each arrangement of electrons there exists a conjugate arrangement in which the hole-values of  $m_1$  and the occupied values of  $m_1$  are interchanged*. This interchange may refer to a single part-subgroup while the other part-subgroups are unchanged, or to a Bohr subgroup, or to the whole of a main group, since the different part-subgroups are completely independent of one another as far as possible arrangements are concerned. *The electron numbers of the two conjugate arrangements add up to the number of electrons in the completed state of the group (or subgroup) considered, while the  $j$ -values of the two arrangements are the same*. The latter follows from the fact that the sum of the hole-values of  $m_1$  of an

arrangement always is the opposite of the sum of the occupied  $m_1$ -values. Therefore, the quantum numbers  $\bar{m}_1$  of the whole atom are the opposite of one another for conjugate arrangements. As the  $j$ -values are defined as the upper limit of the set of  $\bar{m}_1$ -values, and as this set is symmetric around zero, it follows that the  $j$ -values are the same (compare the examples discussed below). Because of this periodicity law to some extent the relations at the end of a period of the periodic table reflect those at the beginning of a period. We must emphasise, however, that this for the time being refers only to the number of stationary states of the shell in question and the values of their quantum numbers, whereas we can say nothing about the magnitude of their energies or about interval relations.<sup>9</sup>

As an application of our rule we shall discuss now the special case of the gradual formation of the eight-shell (where of the principal quantum number considered no electrons with  $k = 2$  are present in the ground state); this gives us at the same time another example of the just-derived reciprocity rule. The binding of the first two electrons in this shell has already been discussed and in what follows we shall assume for the sake of simplicity that no electron is missing from the  $k_1 = 1$  subgroup so that it is closed (compare Table 2 with Stoner's scheme). According to Stoner, for the following elements until the completion of the eight-shell (e.g., from B to Ne) the ground state will always be a  $p$ -term, in agreement with all experimental data up to now. Especially follows the alkali-like spectrum, corresponding to the binding of the third electron of the eight-shell, with the well-known absence of the  $s$ -term with the same principal quantum number as the ground state.

We can thus immediately go over to the binding of the fourth electron of the eight-shell, which appears in the not-yet analysed arc spectrum of carbon and the partially already unraffled arc spectrum of lead. According to the Land $\ddot{\text{O}}$ -Heisenberg branching rule (see previous section) the corresponding spectrum should have in general the same structure as the neon spectrum, that is, consist of a singlet-triplet group and a triplet-quintet group with different series limits, corresponding to the  $2p_1$ - and the  $2p_2$ -doublet term of the ion considered. We shall show, however, that according to our rules these spectra must differ essentially, as far as the number and  $f$ -values of the  $p$ -terms of the maximum principal quantum number ( $n = 2$  for C,  $n = 6$  for Pb) is concerned, from the Ne-spectrum (where, as we mentioned at the beginning of this section, apart from the ground state with  $j = 0$  no further  $p$ -term exists with principal quantum number 2); this is in contrast to the structure of the excited states which we expect to be similar.

We must distinguish three cases, according to the number of electrons in the two part-subgroups with  $k_1 = 2$ ,  $k_2 = 1$  and with  $k_1 = 2$ ,  $k_2 = 2$  over which we must distribute two electrons (we have already assumed that the first two electrons are bound in  $s$ -terms,  $k_1 = k_2 = 1$ ).

- (a) Two equivalent  $n_{21}$ -electrons: Corresponding to the  $p_1$ -term of the alkalis  $m_1$  can for this part-subgroup only take on the two values  $m_1 = \pm 1/2$ . It is thus closed in this case with  $\bar{m}_1 = 0$  and  $j = 0$ .
- (b) One  $n_{21}$ - and one  $n_{22}$ -electron: For the second partsubgroup  $m_1$  can, corresponding

to the  $p_2$ -term of the alkalis take on the four values  $\hat{A} \pm 1/2$ ,  $\hat{A} \pm 3/2$  and these can be combined in all possible ways with the above-mentioned values  $m_1 = +1/2$  of the first electron, since the two electrons are in different part-subgroups and are thus not equivalent.<sup>h</sup> We have thus

$$\begin{aligned}\overline{m}_1 &= (-3/2, -1/2, 1/2, 3/2) + (-1/2, 1/2) \\ &= \hat{A} \pm (3/2 + 1/2), \hat{A} \pm (3/2 - 1/2), \hat{A} \pm (1/2 + 1/2), \hat{A} \pm (1/2 - 1/2) \\ &= \hat{A} \pm 2, \hat{A} \pm 1, \hat{A} \pm 1, 0, 0\end{aligned}$$

From this we see immediately that the terms split in two series with [absolute value]  $\overline{m}_1$  [less than or equal to] 2 and with  $\overline{m}_1$  [less than or equal to] 1. In the field free case these correspond clearly to two terms: one with  $j = 2$ , and one with  $j = 1$ .

(c) Two equivalent  $n_{22}$ -electrons: According to our rule the  $m_1$ -values of the two electrons must be different and we find for the possible values of  $\overline{m}_1$ :

$$\overline{m}_1 = \hat{A} \pm (3/2 + 1/2), \hat{A} \pm (3/2 - 1/2), (3/2 - 3/2), (1/2 - 1/2) = \hat{A} \pm 2, \hat{A} \pm 1, 0, 0.$$

If there is no magnetic field we find thus one term with  $j = 2$  and one with  $j = 0$ .

*Altogether we find thus for the four-shell five different p-terms with maximum principal quantum number, of which two have  $j = 2$ , one  $j = 1$ , and two  $j = 0$ .*

We can say nothing about the energies or the interval relations of this group of terms. However, we can make definite statements about the Zeeman splittings of these terms to be expected.

By substituting the  $m_2$ -values (taken from the Zeeman terms of the alkalis in strong fields) for the separate electrons corresponding to the given  $m_1$ -values, we find from rule (1) the Zeeman splittings for the five  $p$ -terms of the four-shell in strong fields:

$\overline{m}_1$	-2	-1	0	1	2
$\overline{m}_2$	-3, -2	-2, -1, 0	0, 0, 0, 0, 0	1, 1, 2	2, 3

Using the same rule applied by Landé<sup>4</sup> to higher-order multiplets, one obtains from this for the determination of the sum of the  $g$ -values for the two  $j = 2$  terms (denoted by  $Sg_2$ ) and for the  $g$ -value for the  $j = 1$  term (denoted by  $g_1$ ) the equations

$$2Sg_2 = 2 + 3 = 5, \quad Sg_2 + g_1 = 1 + 1 + 2 = 4,$$

or

$$Sg_2 = 5/2, \quad g_1 = 3/2.$$

The earliest test of this theoretical result for the four-shell is possible for lead. Observations certainly show four  $p$ -terms, while the existence of a fifth  $p$ -term is

doubtful.<sup>7</sup> So far unpublished measurements by E. Back of a few lead lines make it, moreover, very likely that the first four  $p$ -terms have  $j$ -values 2, 2, 1, 0, and that the  $g$ -values of these terms also agree with the theoretically expected ones.

Let us now return to the discussion of the gradual construction of the eight-shell. By means of the reciprocity rule, applied to the whole of the Bohr subgroup with  $k = 2$ , which contains in its closed state six electrons, we can immediately apply the results obtained for the four-shell to the number of possibilities to realise the six-shell (from electrons with  $k_1 = 2$ ), which occurs, for instance, for O. The following cases of the six-shell are clearly conjugate to the cases (a), (b), and (c):

- (a) Four equivalent  $n_{22}$ -electrons (two empty spaces in the  $n_{21}$ -group). This part-subgroup is closed; hence as before sub (a) one term with  $j = 0$ .
- (b) One  $n_{21}$ , three equivalent  $n_{22}$ -electrons (one empty space in the  $n_{21}$ -, and one empty space in the  $n_{22}$ - group). As before: one term with  $j = 2$  and one term with  $j = 1$ .
- (c) Two equivalent  $n_{21}$ - and two equivalent  $n_{22}$ -electrons (two empty spaces in the  $n_{22}$ -group). The first partsubgroup is closed. As before: one term with  $j = 2$ , one term with  $j = 0$ .

We must thus also here, for instance for oxygen, expect five  $p$ -terms with the smallest principal quantum number. So far only three such terms have been observed for O and S, with  $j$ -values of 2, 1, 0.<sup>8</sup> We must wait and see whether two more  $p$ -terms of the same principal quantum number can be found from the observations, or whether our rule must be modified in this case.

As yet there are no observations about the five-shell (3 electrons with  $k_1 = 2$ ) and we shall therefore give only the result of the discussion; according to our rule this shell gives rise to five  $p$ -terms, one term with  $j = 5/2$ , three terms with  $j = 3/2$ , and one term with  $j = 1/2$ . For the seven-shell, realised in x-ray spectra we get -- as we mentioned before -- terms similar to the alkalis.

We shall not discuss here further special cases, before experimental data are available, but it should be clear from the examples given that in each case our rule is able to give a unique answer to the question about the possibilities of realising the different shells for a given number of equivalent electrons. To be sure, only in the simplest cases was it possible to verify that the results obtained in this way are in agreement with experiment.

In general we may note that the discussions given here are in principle based, as far as the transition from strong to weak or vanishing fields is concerned, upon the invariance of the statistical weights of quantum states. However, on the basis of the results obtained there seem to be no reasons for a connexion between the problem of the completion of electron groups in an atom and the correspondence principle, as Bohr suspected to be the case. It is probably necessary to improve the basic principles of quantum theory before we can successfully discuss the problem of a better foundation of the general rules, suggested here, for the occurrence of equivalent electrons in an atom.

## Footnotes

[N.B. A few footnotes have been omitted in this translation.]

<sup>a</sup> One notes that one must assign to the two cases  $m_1 = -1/2$  for the first and  $m_2 = 1/2$  for the second electron, or  $m_1 = +1/2$  for the first and  $m_2 = -1/2$  for the second electron two different terms (as far as the part of the energy independent of the field is concerned). This is perhaps a blemish of the classification given here. It will later on, however, turn out that if the inner and the outer valence electron are equivalent, these two terms are in fact identical.

<sup>b</sup> The replacement here of a seven-shell (atom core of neon) by one electron will be given a theoretical basis in the next section.

<sup>c</sup> As already indicated, the value of  $j$  is defined here and henceforth as the maximum value of the quantum number  $m_1$ .

<sup>d</sup> It follows directly from Millikan and Landé's results about the relativistic doublets of the X-ray spectra that this subdivision and the question about the number of electrons in the part-subgroups also makes sense for completed electron groups. These numbers appear clearly in the expression for the energy of the whole groups as function of the order number in the shape of factors of the Moseley-Sommerfeld expressions involving definite values of the screening numbers (determined by  $k_1$ ) and the relativity correction (determined by  $k_2$ ).

<sup>e</sup> This invariance is independent of the validity of classical mechanics under the transformation.

<sup>f</sup> The second case corresponds to an interchange of the two equivalent electrons and gives us therefore here no new stationary state (compare the footnote lettered a). However, in this two-fold realisability of the quantum state considered is contained the fact that its statistical weight with respect to the exchangeability of the two electrons must be multiplied by two (compare also the discussion of statistical weights by Stoner)<sup>6</sup>.

<sup>g</sup> However, because of the equality of the number of  $m_2$ -values for conjugate arrangements it follows that also in weak fields the "g-sums" (taken over terms with the same  $j$ ) of the appropriate terms are the same.

<sup>h</sup> Because of this we must count the case  $m_1 = +1/2$  for the first and  $m_1 = -1/2$  for the second electron different from the case  $m_1 = -1/2$  for the first and  $m_1 = +1/2$  for the second electron. Compare the footnote lettered a.

## References

<sup>1</sup> W. Pauli, *Z. Physik* **31**, 373 (1925).

<sup>2</sup> N. Bohr, *Ann. Physik* **71**, 228 (1923); especially p. 276.

<sup>3</sup> W. Heisenberg and A. Landé, *Z. Physik* **25**, 279 (1924).

<sup>4</sup> A. Landé, *Ann. Physik* **76**, 273 (1925); see especially section 2.

<sup>5</sup> See N. Bohr, *Drei Aufsätze über Spektren und Atombau*, Brunswick, 1924.

<sup>6</sup> E. C. Stoner, *Phil. Mag.* **48**, 719 (1924).

<sup>7</sup> V. Thorsen, *Naturwiss.* **11**, 78 (1923); W. Grotrian, *Z. Physik* **18**, 169 (1923).

<sup>8</sup> J. J. Hopfield, *Ap. J.* **58**, 114 (1924); O. Laporte, *Naturwiss.* **12**, 598 (1924).



# QUANTUM-THEORETICAL RE-INTERPRETATION OF KINEMATIC AND MECHANICAL RELATIONS

W. HEISENBERG

The present paper seeks to establish a basis for theoretical quantum mechanics founded exclusively upon relationships between quantities which in principle are observable.

It is well known that the formal rules which are used in quantum theory for calculating observable quantities such as the energy of the hydrogen atom may be seriously criticized on the grounds that they contain, as basic element, relationships between quantities that are apparently unobservable in principle, e.g., position and period of revolution of the electron. Thus these rules lack an evident physical foundation, unless one still wants to retain the hope that the hitherto unobservable quantities may later come within the realm of experimental determination. This hope might be regarded as justified if the above-mentioned rules were internally consistent and applicable to a clearly defined range of quantum mechanical problems. Experience however shows that only the hydrogen atom and its Stark effect are amenable to treatment by these formal rules of quantum theory. Fundamental difficulties already arise in the problem of 'crossed fields' (hydrogen atom in electric and magnetic fields of differing directions). Also, the reaction of atoms to periodically varying fields cannot be described by these rules. Finally, the extension of the quantum rules to the treatment of atoms having several electrons has proved unfeasible.

It has become the practice to characterize this failure of the quantum-theoretical rules as a deviation from classical mechanics, since the rules themselves were essentially derived from classical mechanics. This characterization has, however, little meaning when one realizes

*Editor's note.* This paper was published as Zs. Phys. **33** (1925) 879–893. It was signed 'Göttingen, Institut für theoretische Physik'.

that the *Einstein-Bohr* frequency condition (which is valid in all cases) already represents such a complete departure from classical mechanics, or rather (using the viewpoint of wave theory) from the kinematics underlying this mechanics, that even for the simplest quantum-theoretical problems the validity of classical mechanics simply cannot be maintained. In this situation it seems sensible to discard all hope of observing hitherto unobservable quantities, such as the position and period of the electron, and to concede that the partial agreement of the quantum rules with experience is more or less fortuitous. Instead it seems more reasonable to try to establish a theoretical quantum mechanics, analogous to classical mechanics, but in which only relations between observable quantities occur. One can regard the frequency condition and the dispersion theory of *Kramers*<sup>1</sup> together with its extensions in recent papers<sup>2</sup> as the most important first steps toward such a quantum-theoretical mechanics. In this paper, we shall seek to establish some new quantum-mechanical relations and apply these to the detailed treatment of a few special problems. We shall restrict ourselves to problems involving one degree of freedom.

1. In classical theory, the radiation emitted by a moving electron (in the wave zone, i.e., in the region where  $\mathfrak{E}$  and  $\mathfrak{H}$  are of the same order of magnitude as  $1/r$ ) is not entirely determined by the expressions

$$\mathfrak{E} = \frac{e}{r^3 c^2} [\mathbf{r}[\mathbf{r}\dot{\mathbf{v}}]], \quad \mathfrak{H} = \frac{e}{r^2 c^2} [\dot{\mathbf{r}}\mathbf{v}],$$

but additional terms occur in the next order of approximation, e.g. terms of the form  $e\ddot{\mathbf{v}}\mathbf{v}/rc^3$  which can be called 'quadrupole radiation'. In still higher order, terms such as  $e\ddot{\mathbf{v}}\mathbf{v}^2/rc^4$  appear. In this manner the approximation can be carried to arbitrarily high order. (The following symbols, have been employed:  $\mathfrak{E}$ ,  $\mathfrak{H}$  are field strengths at a given point,  $\mathbf{r}$  the vector between this point and the position of the electron,  $\mathbf{v}$  the velocity and  $e$  the charge of the electron).

One may inquire about the form these higher order terms would assume in quantum theory. The higher order approximations can easily be calculated in classical theory if the motion of the electron is

<sup>1</sup> H. A. Kramers, Nature **113** (1924) 673.

<sup>2</sup> M. Born, Zs. f. Phys. **26** (1924) 379. H. A. Kramers and W. Heisenberg, Zs. f. Phys. **31** (1925) 681. M. Born and P. Jordan, Zs. f. Phys. (in course of publication) [33 (1925) 479; paper 7a].

given in Fourier expansion, and one would expect a similar result in quantum theory. This point has nothing to do with electrodynamics but rather – and this seems to be particularly important – is of a purely kinematic nature. We may pose the question in its simplest form thus: If instead of a classical quantity  $x(t)$  we have a quantum-theoretical quantity, what quantum-theoretical quantity will appear in place of  $x(t)^2$ ?

Before we can answer this question, it is necessary to bear in mind that in quantum theory it has not been possible to associate the electron with a point in space, considered as a function of time, by means of observable quantities. However, even in quantum theory it is possible to ascribe to an electron the emission of radiation. In order to characterize this radiation we first need the frequencies which appear as functions of two variables. In quantum theory these functions are of the form

$$\nu(n, n - \alpha) = \frac{1}{h} \{W(n) - W(n - \alpha)\},$$

and in classical theory of the form

$$\nu(n, \alpha) = \alpha \nu(n) = \alpha \frac{1}{h} \frac{dW}{dn}.$$

(Here one has  $nh=J$ , where  $J$  is one of the canonical constants).

As characteristic for the comparison between classical and quantum theory with respect to frequency, one can write down the combination relations:

Classical:

$$\nu(n, \alpha) + \nu(n, \beta) = \nu(n, \alpha + \beta).$$

Quantum-theoretical:

$$\nu(n, n - \alpha) + \nu(n - \alpha, n - \alpha - \beta) = \nu(n, n - \alpha - \beta)$$

or

$$\nu(n - \beta, n - \alpha - \beta) + \nu(n, n - \beta) = \nu(n, n - \alpha - \beta).$$

In order to complete the description of radiation it is necessary to have not only the frequencies but also the amplitudes. The amplitudes may be treated as complex vectors, each determined by six independent components, and they determine both the polarization and the phase. As the amplitudes are also functions of the two variables

$n$  and  $\alpha$ , the corresponding part of the radiation is given by the following expressions:

Quantum-theoretical:

$$\operatorname{Re}\{\mathfrak{A}(n, n - \alpha) e^{i\omega(n, n-\alpha)t}\}. \quad (1)$$

Classical:

$$\operatorname{Re}\{\mathfrak{A}_\alpha(n) e^{i\omega(n)\alpha t}\}. \quad (2)$$

At first sight the phase contained in  $\mathfrak{A}$  would seem to be devoid of physical significance in quantum theory, since in this theory frequencies are in general not commensurable with their harmonics. However, we shall see presently that also in quantum theory the phase has a definite significance which is analogous to its significance in classical theory. If we now consider a given quantity  $x(t)$  in classical theory, this can be regarded as represented by a set of quantities of the form

$$\mathfrak{A}_\alpha(n) e^{i\omega(n)\alpha t},$$

which, depending upon whether the motion is periodic or not, can be combined into a sum or integral which represents  $x(t)$ :

$$x(n, t) = \sum_{-\infty}^{+\infty} \mathfrak{A}_\alpha(n) e^{i\omega(n)\alpha t} \quad \text{or} \quad (2a)$$

$$x(n, t) = \int_{-\infty}^{+\infty} \mathfrak{A}_\alpha(n) e^{i\omega(n)\alpha t} d\alpha.$$

A similar combination of the corresponding quantum-theoretical quantities seems to be impossible in a unique manner and therefore not meaningful, in view of the equal weight of the variables  $n$  and  $n - \alpha$ . However, one may readily regard the ensemble of quantities  $\mathfrak{A}(n, n - \alpha) e^{i\omega(n, n-\alpha)t}$  as a representation of the quantity  $x(t)$  and then attempt to answer the above question: how is the quantity  $x(t)^2$  to be represented?

The answer in classical theory is obviously:

$$\mathfrak{B}_\beta(n) e^{i\omega(n)\beta t} = \sum_{-\infty}^{+\infty} \mathfrak{A}_\alpha \mathfrak{A}_{\beta-\alpha} e^{i\omega(n)(\alpha + \beta - \alpha)t} \quad (3)$$

or

$$= \int_{-\infty}^{+\infty} \mathfrak{A}_\alpha \mathfrak{A}_{\beta-\alpha} e^{i\omega(n)(\alpha + \beta - \alpha)t} d\alpha, \quad (4)$$

so that

$$x(t)^2 = \sum_{-\infty}^{+\infty} \mathfrak{B}_\beta(n) e^{i\omega(n)\beta t} \quad (5)$$

or, respectively,

$$= \int_{-\infty}^{+\infty} \mathfrak{B}_\beta(n) e^{i\omega(n)\beta t} d\beta. \quad (6)$$

In quantum theory, it seems that the simplest and most natural assumption would be to replace equations (3) and (4) by:

$$\mathfrak{B}(n, n - \beta) e^{i\omega(n, n-\beta)t} = \sum_{-\infty}^{+\infty} \mathfrak{A}(n, n - \alpha) \mathfrak{A}(n - \alpha, n - \beta) e^{i\omega(n, n-\beta)t} \quad (7)$$

or

$$= \int_{-\infty}^{+\infty} \mathfrak{A}(n, n - \alpha) \mathfrak{A}(n - \alpha, n - \beta) e^{i\omega(n, n-\beta)t} d\alpha, \quad (8)$$

and in fact this type of combination is an almost necessary consequence of the frequency combination rules. On making assumptions (7) and (8), one recognizes that the phases of the quantum-theoretical  $\mathfrak{A}$  have just as great a physical significance as their classical analogues. Only the origin of the time scale and hence a phase factor common to all the  $\mathfrak{A}$  is arbitrary and accordingly devoid of physical significance, but the phases of the individual  $\mathfrak{A}$  enter in an essential manner into the quantity  $\mathfrak{B}$ .<sup>1</sup> A geometrical interpretation of such quantum-theoretical phase relations in analogy with those of classical theory seems at present scarcely possible.

If we further ask for a representation for the quantity  $x(t)^3$  we find without difficulty:

Classical:

$$\mathfrak{C}(n, \gamma) = \sum_{-\infty}^{+\infty} \sum_{-\infty}^{+\infty} \mathfrak{A}_\alpha(n) \mathfrak{A}_\beta(n) \mathfrak{A}_{\gamma-\alpha-\beta}(n). \quad (9)$$

Quantum-theoretical:

$$\mathfrak{C}(n, n - \gamma) =$$

$$= \sum_{-\infty}^{+\infty} \sum_{-\infty}^{+\infty} \mathfrak{A}(n, n - \alpha) \mathfrak{A}(n - \alpha, n - \alpha - \beta) \mathfrak{A}(n - \alpha - \beta, n - \gamma) \quad (10)$$

or the corresponding integral forms.

<sup>1</sup> Cf. also H. A. Kramers and W. Heisenberg, loc.cit. The phases enter essentially into the expressions used there for the induced scattering moment.

In a similar manner, one can find a quantum-theoretical representation for all quantities of the form  $x(t)^n$ , and if any function  $f[x(t)]$  is given, one can always find the corresponding quantum-theoretical expression, provided the function can be expanded as a power series in  $x$ . A significant difficulty arises, however, if we consider two quantities  $x(t)$ ,  $y(t)$ , and ask after their product  $x(t)y(t)$ . If  $x(t)$  is characterized by  $\mathfrak{A}$ , and  $y(t)$  by  $\mathfrak{B}$ , we obtain the following representations for  $x(t)y(t)$ :

Classical:

$$\mathfrak{C}_\beta(n) = \sum_{-\infty}^{+\infty} \mathfrak{A}_\alpha(n) \mathfrak{B}_{\beta-\alpha}(n).$$

Quantum-theoretical:

$$\mathfrak{C}(n, n - \beta) = \sum_{-\infty}^{+\infty} \mathfrak{A}(n, n - \alpha) \mathfrak{B}(n - \alpha, n - \beta).$$

Whereas in classical theory  $x(t)y(t)$  is always equal to  $y(t)x(t)$ , this is not necessarily the case in quantum theory. In special instances, e.g., in the expression  $x(t)x(t)^2$ , this difficulty does not arise.

If, as in the question posed at the beginning of this section, one is interested in products of the form  $v(t)\dot{v}(t)$ , then in quantum theory this product  $v\dot{v}$  should be replaced by  $\frac{1}{2}(v\ddot{v} + \dot{v}v)$ , in order that  $v\dot{v}$  be the differential coefficient of  $\frac{1}{2}v^2$ . In a similar manner it would always seem possible to find natural expressions for the quantum-theoretical mean values, though they may be even more hypothetical than the formulae (7) and (8).

Apart from the difficulty just mentioned, formulae of the type (7), (8) should quite generally also suffice to express the interaction of the electrons in an atom in terms of the characteristic amplitudes of the electrons.

**2.** After these considerations which were concerned with the kinematics of quantum theory, we turn our attention to the dynamical problem which aims at the determination of the  $\mathfrak{A}$ ,  $\nu$ ,  $W$  from the given forces of the system. In earlier theory this problem was solved in two stages:

1. Integration of the equation of motion

## 2. Determination of the constants for periodic motion through

$$\oint p dq = \oint m \dot{x} dx = J (= nh). \quad (12)$$

If one seeks to construct a quantum-mechanical formalism corresponding as closely as possible to that of classical mechanics, it is very natural to take over the equation of motion (11) directly into quantum theory. At this point, however, it is necessary – in order not to depart from the firm foundation provided by those quantities that are in principle observable – to replace the quantities  $\dot{x}$  and  $f(x)$  by their quantum-theoretical representatives, as given in § 1. In classical theory it is possible to obtain the solution of (11) by first expressing  $x$  as a Fourier series or Fourier integral with undetermined coefficients (and frequencies). In general, we then obtain an infinite set of equations containing infinitely many unknowns, or integral equations, which can be reduced to simple recursive relations for the  $\alpha$  in special cases only. In quantum theory we are at present forced to adopt this method of solving equation (11) since, as has been said before, it was not possible to define a quantum-theoretical function directly analogous to the function  $x(n, t)$ .

Consequently the quantum-theoretical solution of (11) is only possible in the simplest cases. Before we consider such simple examples, let us give a quantum-theoretical re-interpretation of the determination, from (12), of the constant of periodic motion. We assume that (classically) the motion is periodic:

$$x = \sum_{-\infty}^{+\infty} a_\alpha(n) e^{i\alpha\omega_n t}; \quad (13)$$

hence

$$m \dot{x} = m \sum_{-\infty}^{+\infty} a_\alpha(n) i\alpha \omega_n e^{i\alpha\omega_n t}$$

and

$$\oint m \dot{x} dx = \oint m \dot{x}^2 dt = 2\pi m \sum_{-\infty}^{+\infty} a_\alpha(n) a_{-\alpha}(n) \alpha^2 \omega_n.$$

Furthermore, since  $a_{-\alpha}(n) = \overline{a_\alpha(n)}$ , as  $x$  is to be real, it follows that

$$\oint m \dot{x}^2 dt = 2\pi m \sum_{-\infty}^{+\infty} |a_\alpha(n)|^2 \alpha^2 \omega_n. \quad (14)$$

In the earlier theory this phase integral was usually set equal to an integer multiple of  $h$ , i.e., equal to  $nh$ , but such a condition does

not fit naturally into the dynamical calculation. It appears, even when regarded from the point of view adopted hitherto, arbitrary in the sense of the correspondence principle, because from this point of view the  $J$  are determined only up to an additive constant as multiples of  $\hbar$ . Instead of (14) it would be more natural to write

$$\frac{d}{dn} (n\hbar) = \frac{d}{dn} \oint m\dot{x}^2 dt,$$

that is,

$$\hbar = 2\pi m \sum_{-\infty}^{+\infty} \alpha \frac{d}{dn} (\alpha \omega_n \cdot |a_\alpha|^2). \quad (15)$$

Such a condition obviously determines the  $a_\alpha$  only to within a constant, and in practice this indeterminacy has given rise to difficulties due to the occurrence of half-integral quantum numbers.

If we look for a quantum-theoretical relation corresponding to (14) and (15) and containing observable quantities only, the uniqueness which had been lost is automatically restored.

We have to admit that only equation (15) has a simple quantum-theoretical reformulation which is related to *Kramers'* dispersion theory:<sup>1</sup>

$$\hbar = 4\pi m \sum_0^{\infty} \{ |a(n, n + \alpha)|^2 \omega(n, n + \alpha) - |a(n, n - \alpha)|^2 \omega(n, n - \alpha) \}. \quad (16)$$

Yet this relation suffices to determine the  $a$  uniquely since the undetermined constant contained in the quantities  $a$  is automatically fixed by the condition that a ground state should exist, from which no radiation is emitted. Let this ground state be denoted by  $n_0$ ; then we should have  $a(n_0, n_0 - \alpha) = 0$  (for  $\alpha > 0$ ). Hence we may expect that the question of half-integer or integer quantization does not arise in a theoretical quantum mechanics based only upon relations between observable quantities.

Equations (11) and (16), if soluble, contain a complete determination not only of frequencies and energy values, but also of quantum-theoretical transition probabilities. However, at present the actual mathematical solution can be obtained only in the simplest cases. In many systems, e.g. the hydrogen atom, a particular complication

<sup>1</sup> This relation has already been derived from dispersion considerations by W. Kuhn, Zs. Phys. **33** (1925) 408, and W. Thomas, Naturwiss. **13** (1925) 627.

arises because the solutions correspond to motion which is partly periodic and partly aperiodic. As a consequence of this property, the quantum-theoretical series (7), (8) and equation (16) decompose into a sum and an integral. Quantum-mechanically such a decomposition into 'periodic and aperiodic motion' cannot be carried out in general.

Nevertheless, one could regard equations (11) and (16) as a satisfactory solution, at least in principle, of the dynamical problem if it were possible to show that this solution agrees with (or at any rate does not contradict) the quantum-mechanical relationships which we know at present. It should, for instance, be established that the introduction of a small perturbation into a dynamical problem leads to additional terms in the energy, or frequency, of the type found by *Kramers* and *Born* – but not of the type given by classical theory. Furthermore, one should also investigate whether equation (11) in the present quantum-theoretical form would in general give rise to an energy integral  $\frac{1}{2}m\dot{x}^2 + U(x) = \text{const.}$ , and whether the energy so derived satisfies the condition  $\Delta W = h\nu$ , in analogy with the classical condition  $\nu = \partial W / \partial J$ . A general answer to these questions would elucidate the intrinsic connections between previous quantum-mechanical investigations and pave the way toward a consistent quantum-mechanics based solely upon observable quantities. Apart from a general connection between Kramer's dispersion formula and equations (11) and (16), we can answer the above questions only in very special cases which may be solved by simple recursion relations.

The general connection between *Kramers'* dispersion theory and our equations (11) and (16) is as follows. From equation (11) (more precisely, from the quantum-theoretical analogue) one finds, just as in classical theory, that the oscillating electron behaves like a free electron when acted upon by light of much higher frequency than any eigenfrequency of the system. This result also follows from Kramers' dispersion theory if in addition one takes account of equation (16). In fact, *Kramers* finds for the moment induced by a wave of the form  $E \cos 2\pi\nu t$ :

$$M = e^2 E \cos 2\pi\nu t \frac{2}{h} \sum_0^\infty \alpha \left\{ \frac{|a(n, n + \alpha)|^2 \nu(n, n + \alpha)}{\nu^2(n, n + \alpha) - \nu^2} - \frac{|a(n, n - \alpha)|^2 \nu(n, n - \alpha)}{\nu^2(n, n - \alpha) - \nu^2} \right\}, \quad 683$$

so that for  $\nu \gg n(n+\alpha)$ ,

$$M = -\frac{2Ee^2 \cos 2\pi\nu t}{\nu^2 h} \sum_0^\infty \left\{ |a(n, n+\alpha)|^2 \nu(n, n+\alpha) - |a(n, n-\alpha)|^2 \nu(n, n-\alpha) \right\},$$

which, due to equation (16), becomes

$$M = -\frac{e^2 E \cos 2\pi\nu t}{4\pi^2 m \nu^2}.$$

3. As a simple example, the anharmonic oscillator will now be treated:

$$\ddot{x} + \omega_0^2 x + \lambda x^2 = 0. \quad (17)$$

Classically, this equation is satisfied by a solution of the form

$$x = \lambda a_0 + a_1 \cos \omega t + \lambda a_2 \cos 2\omega t + \lambda^2 a_3 \cos 3\omega t + \dots \lambda^{\tau-1} a_\tau \cos \tau \omega t,$$

where the  $a$  are power series in  $\lambda$ , the first terms of which are independent of  $\lambda$ . Quantum-theoretically we attempt to find an analogous expression, representing  $x$  by terms of the form

$$\begin{aligned} & \lambda a(n, n); \quad a(n, n-1) \cos \omega(n, n-1)t; \\ & \lambda a(n, n-2) \cos \omega(n, n-2)t; \\ & \dots \lambda^{\tau-1} a(n, n-\tau) \cos \omega(n, n-\tau)t \dots \end{aligned}$$

The recursion formulae which determine the  $a$  and  $\omega$  (up to, but excluding, terms of order  $\lambda$ ) according to equations (3), (4) or (7), (8) are:

Classical:

$$\begin{aligned} & \omega_0^2 a_0(n) + \frac{1}{2} a_1^2(n) = 0; \\ & -\omega^2 + \omega_0^2 = 0; \\ & (-4\omega^2 + \omega_0^2) a_2(n) + \frac{1}{2} a_1^2 = 0; \\ & (-9\omega^2 + \omega_0^2) a_3(n) + a_1 a_2 = 0; \\ & \dots \dots \dots \dots \dots \end{aligned} \quad (18)$$

Quantum-theoretical:

$$\begin{aligned} & \omega_0^2 a_0(n) + \frac{1}{4} [a^2(n+1, n) + a^2(n, n-1)] = 0; \\ & -\omega^2(n, n-1) + \omega_0^2 = 0; \\ & [-\omega^2(n, n-2) + \omega_0^2] a(n, n-2) + \frac{1}{2} [a(n, n-1) a(n-1, n-2)] = 0; \\ & [-\omega^2(n, n-3) + \omega_0^2] a(n, n-3) \\ & + \frac{1}{2} [a(n, n-1) a(n-1, n-3)] + \frac{1}{2} [a(n, n-2) a(n-2, n-3)] = 0; \\ & \dots \dots \dots \dots \dots \dots \dots \dots \end{aligned} \quad (19)$$

The additional quantum condition is:

Classical ( $J = nh$ ):

$$1 = 2\pi m \frac{d}{dJ} \sum_{-\infty}^{+\infty} \frac{1}{4}\tau^2 |a_\tau|^2 \omega.$$

Quantum-theoretical:

$$h = \pi m \sum_0^{\infty} [|a(n + \tau, n)|^2 \omega(n + \tau, n) - |a(n, n - \tau)|^2 \omega(n, n - \tau)].$$

We obtain in first order, both classically and quantum-mechanically

$$a_1^2(n) \quad \text{or} \quad a^2(n, n - 1) = \frac{(n + \text{const})h}{\pi m \omega_0}. \quad (20)$$

In quantum theory, the constant in equation (20) can be determined from the condition that  $a(n_0, n_0 - 1)$  should vanish in the ground state. If we number the  $n$  in such a way that in the ground state  $n$  is zero, i.e.  $n_0 = 0$ , then  $a^2(n, n - 1) = nh/\pi m \omega_0$ .

It thus follows from the recursive relations (18) that in classical theory the coefficient  $a_\tau$  has (to first order in  $\lambda$ ) the form  $\kappa(\tau)n^{\frac{1}{2}\tau}$  where  $\kappa(\tau)$  represents a factor independent of  $n$ . In quantum theory, equation (19) implies

$$a(n, n - \tau) = \kappa(\tau) \sqrt{\frac{n!}{(n - \tau)!}}, \quad (21)$$

where  $\kappa(\tau)$  is the same proportionality factor, independent of  $n$ . Naturally, for large values of  $n$  the quantum-theoretical value of  $a_\tau$  tends asymptotically to the classical value.

An obvious next step would be to try inserting the classical expression for the energy  $\frac{1}{2}m\dot{x}^2 + \frac{1}{2}m\omega_0^2x^2 + \frac{1}{3}m\lambda x^3 = W$ , because in the present first-approximation calculation it actually is constant, even when treated quantum-theoretically. Its value is given by (19), (20) and (21) as:

Classical:

$$W = nh\omega_0/2\pi. \quad (22)$$

Quantum-theoretical, from (7) and (8):

$$W = (n + \frac{1}{2})h\omega_0/2\pi \quad (23)$$

(terms of order  $\lambda^2$  have been excluded).

Thus from the present viewpoint, even the energy of a harmonic oscillator is not given by 'classical mechanics', i.e., by equation (22), but has the form (23).

The more precise calculation, taking into account higher order approximations in  $W$ ,  $a$ ,  $\omega$  will now be carried out for the simpler example of an anharmonic oscillator  $\ddot{x} + \omega_0^2 x + \lambda x^3 = 0$ .

Classically, one can in this case set

$$x = a_1 \cos \omega t + \lambda a_3 \cos 3\omega t + \lambda^2 a_5 \cos 5\omega t + \dots;$$

quantum-theoretically we attempt to set by analogy

$$a(n, n-1) \cos \omega(n, n-1)t; \quad \lambda a(n, n-3) \cos \omega(n, n-3)t; \quad \dots$$

The quantities  $a$  are once more power series in  $\lambda$  whose first term has the form, as in equation (21),

$$a(n, n-\tau) = \varkappa(\tau) \sqrt{\frac{n!}{(n-\tau)!}},$$

as one finds by evaluating the equations corresponding to (18) and (19).

If the evaluation of  $\omega$  and  $a$  from equations (18) and (19) is carried out to order  $\lambda^2$  or  $\lambda$  respectively, one obtains

$$\omega(n, n-1) = \omega_0 + \lambda \frac{3nh}{8\pi\omega_0^2 m} - \lambda^2 \frac{3h^2}{256\omega_0^5 m^2 \pi^2} (17n^2 + 7) + \dots \quad (24)$$

$$a(n, n-1) = \sqrt{\frac{nh}{\pi\omega_0 m}} \left( 1 - \lambda \frac{3nh}{16\pi\omega_0^3 m} + \dots \right). \quad (25)$$

$$a(n, n-3) = \frac{1}{32} \sqrt{\frac{h^3}{\pi^3 \omega_0^7 m^3} n(n-1)(n-2)} \cdot \left( 1 - \lambda \frac{39(n-1)h}{32\pi\omega_0^3 m} \right). \quad (26)$$

The energy, defined as the constant term in the expression

$$\frac{1}{2}m\dot{x}^2 + \frac{1}{2}m\omega_0^2 x^2 + \frac{1}{4}m\lambda x^4,$$

but this was the case for all terms evaluated) turns out to be

$$W = \frac{(n + \frac{1}{2})\hbar\omega_0}{2\pi} + \lambda \frac{3(n^2 + n + \frac{1}{2})\hbar^2}{8 \cdot 4\pi^2\omega_0^2 m} - \lambda^2 \frac{\hbar^3}{512\pi^3\omega_0^5 m^2} (17n^3 + \frac{51}{2}n^2 + \frac{59}{2}n + \frac{21}{2}). \quad (27)$$

This energy can also be determined using the *Kramers-Born* approach by treating the term  $\frac{1}{4}m\lambda x^4$  as a perturbation to the harmonic oscillator. The fact that one obtains exactly the same result (27) seems to me to furnish remarkable support for the quantum-mechanical equations which have here been taken as basis. Furthermore, the energy calculated from (27) satisfies the relation (cf. eq. 24):

$$\frac{\omega(n, n-1)}{2\pi} = \frac{1}{\hbar} [W(n) - W(n-1)],$$

which can be regarded as a necessary condition for the possibility of a determination of the transition probabilities according to equations (11) and (16).

In conclusion we consider the case of a rotator and call attention to the relationship of equations (7), (8) to the intensity formulae for the Zeeman effect<sup>1</sup> and for multiplets.<sup>2</sup>

Consider the rotator as represented by an electron which circles a nucleus with constant distance  $a$ . Both classically and quantum-theoretically, the 'equations of motion' simply state that the electron describes a plane, uniform rotation at a distance  $a$  and with angular velocity  $\omega$  about the nucleus. The 'quantum condition' (16) yields, according to (12),

$$\hbar = \frac{d}{dn} (2\pi ma^2\omega),$$

and according to (16)

$$\hbar = 2\pi m \{a^2\omega(n+1, n) - a^2\omega(n, n-1)\},$$

<sup>1</sup> S. Goudsmit and R. de L. Kronig, Naturwiss. **13** (1925) 90; H. Hönl, Zs. f. Phys. **31** (1925) 340.

<sup>2</sup> R. de L. Kronig, Zs. f. Phys. **31** (1925) 885; A. Sommerfeld and H. Hönl, Sitzungsber. d. Preuss. Akad. d. Wiss. (1925) 141; H. N. Russell, Nature **115** (1925) 835.

from which, in both cases, it follows that

$$\omega(n, n - 1) = \frac{h(n + \text{const})}{2\pi ma^2}.$$

The condition that the radiation should vanish in the ground state ( $n_0=0$ ) leads to the formula

$$\omega(n, n - 1) = \frac{hn}{2\pi ma^2}. \quad (28)$$

The energy is

$$W = \frac{1}{2}mv^2,$$

or, from equations (7), (8),

$$W = \frac{m}{2} a^2 \frac{\omega^2(n, n-1) + \omega^2(n+1, n)}{2} = \frac{h^2}{8\pi^2 ma^2} (n^2 + n + \frac{1}{2}), \quad (29)$$

which again satisfies the condition  $\omega(n, n-1) = (2\pi/h)[W(n) - W(n-1)]$ .

As support for the validity of the formulae (28) and (29), which differ from those of the usual theory, one might mention that, according to Kratzer,<sup>1</sup> many band spectra (including spectra for which the existence of an electron momentum is improbable) seem to require formulae of type (28), (29), which, in order to avoid rupture with the classical theory of mechanics, one had hitherto endeavoured to explain through half-integer quantization.

In order to arrive at the *Goudsmid-Kronig-Hönl* formula for the rotator we have to leave the field of problems having one degree of freedom. We assume that the rotator has a direction in space which is subject to a very slow precession  $\omega$  about the  $z$ -axis of an external field. Let the quantum number corresponding to this precession be  $m$ . The motion is then represented by the quantities

$$\begin{aligned} z: & \quad a(n, n - 1; m, m) \cos \omega(n, n - 1)t; \\ x + iy: & \quad b(n, n - 1; m, m - 1) e^{i[\omega(n, n-1) + \omega]t}; \\ & \quad b(n, n - 1; m - 1, m) e^{i[-\omega(n, n-1) + \omega]t}. \end{aligned}$$

The equations of motion are simply

$$x^2 + y^2 + z^2 = a^2.$$

<sup>1</sup> Cf. for example, B. A. Kratzer, Sitzungsber. d. Bayr. Akad. (1922) p. 107

Because of (7) this leads to<sup>1</sup>

$$\begin{aligned} \frac{1}{2}\{ & \frac{1}{2}a^2(n, n-1; m, m) + b^2(n, n-1; m, m-1) + b^2(n, n-1; m, m+1) \\ & + \frac{1}{2}a^2(n+1, n; m, m) + b^2(n+1, n; m-1, m) \\ & + b^2(n+1, n; m+1, m) \} = a^2. \end{aligned} \quad (30)$$

$$\begin{aligned} & \frac{1}{2}a(n, n-1; m, m)a(n-1, n-2; m, m) \\ & = b(n, n-1; m, m+1)b(n-1, n-2; m+1, m) \\ & + b(n, n-1; m, m-1)b(n-1, n-2; m-1, m). \end{aligned} \quad (31)$$

One also has the quantum condition from (16):

$$2\pi m\{b^2(n, n-1; m, m-1)\omega(n, n-1) - b^2(n, n-1; m-1, m)\omega(n, n-1)\} = (m + \text{const})\hbar. \quad (32)$$

The classical relations corresponding to these equations are

$$\begin{aligned} & \frac{1}{2}a_0^2 + b_1^2 + b_{-1}^2 = a^2; \\ & \frac{1}{4}a_0^2 = b_1 b_{-1}; \\ & 2\pi m(b_{+1}^2 - b_{-1}^2)\omega = (m + \text{const})\hbar. \end{aligned} \quad (33)$$

They suffice (up to the unknown constant added to  $m$ ) to determine  $a_0, b_1, b_{-1}$  uniquely.

The simplest solution of the quantum-theoretical equations (30), (31), (32) which presents itself is:

$$\begin{aligned} b(n, n-1; m, m-1) &= a \sqrt{\frac{(n+m+1)(n+m)}{4(n+\frac{1}{2})n}}; \\ b(n, n-1; m-1, m) &= a \sqrt{\frac{(n-m)(n-m+1)}{4(n+\frac{1}{2})n}}; \\ a(n, n-1; m, m) &= a \sqrt{\frac{(n+m+1)(n-m)}{(n+\frac{1}{2})n}}. \end{aligned}$$

These expressions agree with the formulae of *Goudsmit, Kronig and Hönl*. It is, however, not easily seen that these expressions represent the *only* solution of equations (30), (31), (32), though this would seem likely to me from consideration of the boundary conditions (vanishing

<sup>1</sup> Equation (30) is essentially identical with the *Ornstein-Burger* sum rules.

of  $a$  and  $b$  at the 'boundary'; cf. the papers of *Kronig*, *Sommerfeld* and *Hönl*, *Russell* quoted above).

Considerations similar to the above, applied to the multiplet intensity formulae, lead to the result that these intensity rules are in agreement with equations (7) and (16). This finding may again be regarded as furnishing support for the validity of the kinematic equation (7).

Whether a method to determine quantum-theoretical data using relations between observable quantities, such as that proposed here, can be regarded as satisfactory in principle, or whether this method after all represents far too rough an approach to the physical problem of constructing a theoretical quantum mechanics, an obviously very involved problem at the moment, can be decided only by a more intensive mathematical investigation of the method which has been very superficially employed here.

## On Quantum Mechanics

M. Born

Received 1925



Translation into English: *Sources of Quantum Mechanics*,  
Ed. by B. L. van der Waerden, North Holland, Amsterdam (1967) 277.



The recently published theoretical approach of Heisenberg is here developed into a systematic theory of quantum mechanics (in the first place for systems having one degree of freedom) with the aid of mathematical matrix methods. After a brief survey of the latter, the mechanical equations of motion are derived from a variational principle and it is shown that using Heisenberg's quantum condition, the principle of energy conservation and Bohr's frequency condition follow from the mechanical equations. Using the anharmonic oscillator as example, the question of uniqueness of the solution and of the significance of the phases of the partial vibrations is raised. The paper concludes with an attempt to incorporate electromagnetic field laws into the new theory.

### Introduction

The theoretical approach of Heisenberg <sup>1</sup> recently published in this Journal, which aimed at setting up a new kinematical and mechanical formalism in conformity with the basic requirements of quantum theory, appears to us of considerable potential significance. It represents an attempt to render

---

<sup>1</sup>W. Heisenberg, Zs. f. Phys. **33** (1925) 879.

justice to the new facts by selling up a new and really suitable conceptual system instead of adapting the customary conceptions in a more or less artificial and forced manner. The physical reasoning which led Heisenberg to this development has been so clearly described by him that any supplementary remarks appear superfluous. But, as he himself indicates, in its formal, mathematical aspects his approach is but in its initial stages. His hypotheses have been applied only to simple examples without being fully carried through to a generalized theory. Having been in an advantageous position to familiarize ourselves with his ideas throughout their formative stages, we now strive (since his investigations have been concluded) to clarify the mathematically formal content of his approach and present some of our results here. These indicate that it is in fact possible, starting with the basic premises given by Heisenberg, to build up a closed mathematical theory of quantum mechanics which displays strikingly close analogies with classical mechanics, but at the same time preserves the characteristic features of quantum phenomena.

In this we at first confine ourselves, like Heisenberg, to systems having *one degree of freedom* and assume these to be – from a classical standpoint – *periodic*. We shall in the continuation of this publication concern ourselves with the generalization of the mathematical theory to systems having an arbitrary number of degrees of freedom, as also to aperiodic motion. A noteworthy generalization of Heisenberg's approach lies in our confining ourselves neither to treatment of nonrelativistic mechanics nor to calculations involving Cartesian systems of coordinates. The only restriction which we impose upon the choice of coordinates is to base our considerations upon *libration coordinates*, which in classical theory are *periodic* functions of time. Admittedly, in some instances it might be more reasonable to employ other coordinates: for example, in the case of a rotating body to introduce the angle of rotation  $\varphi$ , which becomes a linear function of time. Heisenberg also proceeded thus in his treatment of the rotator; however, it remains undecided whether the approach applied there can be justified from the standpoint of a consistent quantum mechanics.

The mathematical basis of Heisenberg's treatment is the *law of multiplication* of quantum-theoretical quantities, which he derived from an ingenious consideration of correspondence arguments. The development of his formalism, which we give here, is based upon the fact that this rule of multiplication is none other than the well-known mathematical rule of *matrix multiplication*. The infinite square array (with discrete or continuous indices) which appears at the start of the next section, termed a *matrix*, is a representation of a physical quantity which is given in classical theory as

a function of time. The mathematical method of treatment inherent in the new quantum mechanics is thereby characterized through the employment of *matrix analysis* in place of the usual number analysis.

Using this method, we have attempted to tackle some of the simplest problems in mechanics and electrodynamics. A *variational principle*, derived from correspondence considerations, yields *equations of motion* for the most general Hamilton function which are in closest analogy with the classical canonical equations. The quantum condition conjoined with one of the relations which proceed from the equations of motion permits a simple matrix notation. With the aid of this, one can prove the general validity of the *law of conservation of energy* and the *Bohr frequency relation* in the sense conjectured by Heisenberg: this proof could not be carried through in its entirety by him even for the simple examples which he considered. We shall later return in more detail to one of these examples in order to derive a basis for consideration of the part played by the phases of the partial vibrations in the new theory. We show finally that the basic laws of the electromagnetic field in a vacuum can readily be incorporated and we furnish substantiation for the assumption made by Heisenberg that the squares of the absolute values of the elements in a matrix representing the electrical moment of an atom provide a measure for the transition probabilities.

## Chapter 1. Matrix Calculation

### 1. Elementary operations. Functions

We consider square infinite matrices,<sup>2</sup> which we shall denote by heavy type to distinguish them from ordinary quantities which will throughout be in light type,

$$\mathbf{a} = (a(nm)) = \begin{pmatrix} a(00) & a(01) & a(02) & \cdots \\ a(10) & a(11) & a(12) & \cdots \\ a(20) & a(21) & a(22) & \cdots \\ \dots & \dots & \dots & \dots \end{pmatrix}.$$

---

<sup>2</sup>Further details of matrix algebra can be found, e.g., in M. Bocher, Einführung in die höhere Algebra (translated from the English by Hans Beck; Teubner, Leipzig, 1910) § 22–25; also in R. Courant and D. Hilbert, Methoden der mathematischen Physik 1 (Springer, Berlin, 1924) Chapter I.

Equality of two matrices is defined as equality of corresponding components:

$$\mathbf{a} = \mathbf{b} \quad \text{means} \quad a(nm) = b(nm). \quad (1)$$

Matrix addition is defined as addition of corresponding components:

$$\mathbf{a} = \mathbf{b} + \mathbf{c} \quad \text{means} \quad a(nm) = b(nm) + c(nm). \quad (2)$$

Matrix multiplication is defined by the rule “rows times columns”, familiar from the theory of determinants:

$$\mathbf{a} = \mathbf{bc} \quad \text{means} \quad a(nm) = \sum_{k=0}^{\infty} b(nk)c(km). \quad (3)$$

Powers are defined by repeated multiplication. The associative rule applies to multiplication and the distributive rule to combined addition and multiplication:

$$(\mathbf{ab})\mathbf{c} = \mathbf{a}(\mathbf{bc}); \quad (4)$$

$$\mathbf{a}(\mathbf{b} + \mathbf{c}) = \mathbf{ab} + \mathbf{ac}. \quad (5)$$

However, the commutative rule does *not* hold for multiplication: it is not in general correct to set  $\mathbf{ab} = \mathbf{ba}$ . If  $\mathbf{a}$  and  $\mathbf{b}$  do satisfy this relation, they are said to commute.

The *unit matrix* defined by

$$\mathbf{1} = (\delta_{nm}), \quad \begin{cases} \delta_{nm} = 0 & \text{for } n \neq m, \\ \delta_{nm} = 1 & \end{cases} \quad (6)$$

has the property

$$\mathbf{a}\mathbf{1} = \mathbf{1}\mathbf{a} = \mathbf{a}. \quad (6a)$$

The *reciprocal matrix* to  $\mathbf{a}$ , namely  $\mathbf{a}^{-1}$ , is defined by<sup>3</sup>

$$\mathbf{a}^{-1}\mathbf{a} = \mathbf{aa}^{-1} = \mathbf{1} \quad (7)$$

As *mean value* of a matrix  $\mathbf{a}$  we denote that matrix whose diagonal elements are the same as those of  $\mathbf{a}$  whereas all other elements vanish:

$$\tilde{\mathbf{a}} = (\delta_{nma}(nm)). \quad (8)$$

---

<sup>3</sup>As is known,  $\mathbf{a}^{-1}$  is uniquely defined by (7) for *finite* square matrices when the determinant  $A$  of the matrix  $\mathbf{a}$  is non-zero. If  $A = 0$  there is no matrix to  $\mathbf{a}$ .

The sum of these diagonal elements will be termed the *diagonal sum* of the matrix  $\mathbf{a}$  and written as  $\mathbf{D}(\mathbf{a})$ , viz.

$$\mathbf{D}(\mathbf{a}) = \sum_n a(nm). \quad (9)$$

From (3) it is easy to prove that if the diagonal sum of a product  $\mathbf{y} = \mathbf{x}_1 \mathbf{x}_2 \cdots \mathbf{x}_m$  be finite, then it is unchanged by cyclic rearrangement of the factors:

$$\mathbf{D}(\mathbf{x}_1 \mathbf{x}_2 \cdots \mathbf{x}_m) = \mathbf{D}(\mathbf{x}_r \mathbf{x}_{r+1} \cdots \mathbf{x}_m \mathbf{x}_1 \mathbf{x}_2 \cdots \mathbf{x}_{r-1}). \quad (10)$$

Clearly, it suffices to establish the validity of this rule for *two* factors.

If the elements of the matrices  $\mathbf{a}$  and  $\mathbf{b}$  are functions of a parameter  $t$ , then

$$\frac{d}{dt} \sum_n a(nk)b(km) = \sum_k \{\dot{a}(nk)b(km) + a(nk)\dot{b}(km)\},$$

or from the definition (3):

$$\frac{d}{dt} (\mathbf{ab}) = \dot{\mathbf{a}}\mathbf{b} + \mathbf{a}\dot{\mathbf{b}}. \quad (11)$$

Repeated application of (11)

$$\frac{d}{dt} (\mathbf{x}_1 \mathbf{x}_2 \cdots \mathbf{x}_n) = \dot{\mathbf{x}}_1 \mathbf{x}_2 \cdots \mathbf{x}_n + \mathbf{x}_1 \dot{\mathbf{x}}_2 \cdots \mathbf{x}_n + \cdots + \mathbf{x}_1 \mathbf{x}_2 \cdots \dot{\mathbf{x}}_n. \quad (11')$$

From the definitions (2) and (3) we can define *functions* of matrices. To begin with, we consider as the most general function of this type,  $\mathbf{f}(\mathbf{x}_1, \mathbf{x}_2 \cdots \mathbf{x}_m)$ , one which can formally be represented as a sum of a finite or infinite number of products of powers of the arguments  $\mathbf{x}_k$ ; weighted by numerical coefficients. Through the equations

$$\begin{aligned} \mathbf{f}_1(\mathbf{y}_1, \dots, \mathbf{y}_n; \mathbf{x}_1, \dots, \mathbf{x}_n) &= 0, \\ \dots &\dots \\ \mathbf{f}_n(\mathbf{y}_1, \dots, \mathbf{y}_n; \mathbf{x}_1, \dots, \mathbf{x}_n) &= 0 \end{aligned} \quad (12)$$

we can then also define functions  $\mathbf{y}_1(\mathbf{x}_1, \dots, \mathbf{x}_n)$ ; namely, in order to obtain functions  $\mathbf{y}_1$ ; having the above form and satisfying equation (12), the  $\mathbf{y}_1$  need only be set in form of a series in increasing power products of the  $\mathbf{x}_k$  and the coefficients determined through substitution in (12). It can be seen that one will always derive as many equations as there are unknowns. Naturally, the number of equations and unknowns exceeds that which would ensue

from applying the method of undetermined coefficients in the normal type of analysis incorporating *commutative* multiplication. In each of the equations (12), upon substituting the series for the  $\mathbf{y}_1$ ; and gathering together like terms one obtains not only a sum term  $C'\mathbf{x}_1\mathbf{x}_2$  but also a term  $C''\mathbf{x}_2\mathbf{x}_1$  and thereby has to bring both  $C'$  and  $C''$  to vanish (e.g., not only  $C' + C''$ ). This is, however, made possible by the fact that in the expansion of each of the  $\mathbf{y}_1$ , two terms  $\mathbf{x}_1\mathbf{x}_2$  and  $\mathbf{x}_2\mathbf{x}_1$  appear, with two available coefficients.

## 2. Symbolic differentiation

At this stage we have to examine in detail the process of *differentiation* of a matrix function, which will later be employed frequently in calculation. One should at the outset note that only in a few respects does this process display similarity to that of differentiation in ordinary analysis. For example, the rules for differentiation of a product or of a function of a function here no longer apply in general. Only if all the matrices which occur commute with one another can one apply all the rules of normal analysis to this differentiation.

Suppose

$$\mathbf{y} = \prod_{m=1}^s \mathbf{x}_{l_m} = \mathbf{x}_{l_1} \mathbf{x}_{l_2} \dots \mathbf{x}_{l_s}. \quad (13)$$

We define

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}_k} = \sum_{r=1}^s \delta_{l_r k} \prod_{m=r+1}^s \mathbf{x}_{l_m} \prod_{m=1}^{m=r-1} \mathbf{x}_{l_m}, \quad \begin{cases} \delta_{jk} = 0 & \text{for } j \neq k, \\ \delta_{kk} = 1. & \end{cases} \quad (14)$$

This rule may be expressed as follows: In the given product, one regards all factors as written out individually (e.g., not as  $\mathbf{x}_1^3\mathbf{x}_2^2$ , but as  $\mathbf{x}_1\mathbf{x}_1\mathbf{x}_1\mathbf{x}_2\mathbf{x}_2$ ); one then picks out any factor  $\mathbf{x}_k$  and builds the product of all the factors which follow this and which precede (in this sequence). The sum of all such expressions is the differential coefficient of the product with respect to this  $\mathbf{x}_k$ .

The procedure may be illustrated by some examples:

$$\begin{aligned} \mathbf{y} &= \mathbf{x}^n, & \frac{d\mathbf{y}}{d\mathbf{x}} &= n\mathbf{x}^{n-1} \\ \mathbf{y} &= \mathbf{x}_1^n \mathbf{x}_2^m, & \frac{\partial \mathbf{y}}{\partial \mathbf{x}_1} &= \mathbf{x}_1^{n-1} \mathbf{x}_2^m + \mathbf{x}_1^{n-2} \mathbf{x}_2^m \mathbf{x}_1 + \dots + \mathbf{x}_2^m \mathbf{x}_1^{n-1}, \\ \mathbf{y} &= \mathbf{x}_1^2 \mathbf{x}_2 \mathbf{x}_1 \mathbf{x}_3, & \frac{\partial \mathbf{y}}{\partial \mathbf{x}_1} &= \mathbf{x}_1 \mathbf{x}_2 \mathbf{x}_1 \mathbf{x}_3 + \mathbf{x}_2 \mathbf{x}_1 \mathbf{x}_3 \mathbf{x}_1 + \mathbf{x}_3 \mathbf{x}_1^2 \mathbf{x}_2. \end{aligned}$$

If we further stipulate that

$$\frac{\partial(\mathbf{y}_1 + \mathbf{y}_2)}{\partial \mathbf{x}_k} = \frac{\partial \mathbf{y}_1}{\partial \mathbf{x}_k} + \frac{\partial \mathbf{y}_2}{\partial \mathbf{x}_k}, \quad (15)$$

then the derivative  $\partial \mathbf{y}/\partial \mathbf{x}$  is defined for the most general analytical functions  $\mathbf{y}$ .

With the above definitions, together with that of the diagonal sum (9), there follows the relation

$$\frac{\partial \mathbf{D}(\mathbf{y})}{\partial \mathbf{x}_k(mn)} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}_k}(mn), \quad (16)$$

on the right-hand side of which stands the  $mn$ -component of the matrix  $\partial \mathbf{y}/\partial \mathbf{x}_k$ . This relation can also be used to define the derivative  $\partial \mathbf{y}/\partial \mathbf{x}_k$ . In order to prove (16), it obviously suffices to consider a function  $\mathbf{y}$  having the form (13). From (14) and (3) it follows that

$$\begin{aligned} \frac{\partial \mathbf{y}}{\partial \mathbf{x}_k(mn)} &= \sum_{r=1}^s \delta_{l_r k} \sum_{\tau} \prod_{p=r+1}^s x_{l_p}(\tau_p \tau_{p+1}) \prod_{p=1}^{r-1} x_{l_p}(\tau_p \tau_{p+1}); \\ \tau_{r+1} &= m, \quad \tau_{s+1} = \tau_1, \quad \tau_r = n. \end{aligned} \quad (17)$$

On the other hand, from (3) and (9) ensues

$$\begin{aligned} \frac{\partial \mathbf{D}(\mathbf{y})}{\partial \mathbf{x}_k(mn)} &= \sum_{r=1}^s \delta_{l_r k} \sum_{\tau} \prod_{p=1}^{r-1} \mathbf{x}_{l_p}(\tau_p \tau_{p+1}) \prod_{p=r+1}^s x_{l_p}(\tau_p \tau_{p+1}); \\ \tau_1 &= \tau_{s+1}, \quad \tau_r = n, \quad \tau_{r+1} = m. \end{aligned} \quad (17')$$

Comparison of (17) with (17') yields (16).

We here pick out a fact which will later assume importance and which can be deduced from the definition (14): *the partial derivatives of a product are invariant with respect to cyclic rearrangement of the factors.* Because of (16) this can also be inferred from (10).

To conclude this introductory section, some additional description is devoted to functions  $\mathbf{g}(\mathbf{pq})$  of the variables. For

$$\mathbf{y} = \mathbf{p}^s \mathbf{q}^r \quad (18)$$

it follows from (14) that

$$\frac{\partial \mathbf{y}}{\partial \mathbf{p}} = \sum_{l=1}^{s-1} \mathbf{p}^{s-1-l} \mathbf{q}^r \mathbf{p}^l, \quad \frac{\partial \mathbf{y}}{\partial \mathbf{q}} = \sum_{j=1}^{r-1} \mathbf{q}^{r-1-j} \mathbf{p}^s \mathbf{q}^j. \quad (18')$$

The most general function  $\mathbf{g}(\mathbf{pq})$  to be considered is to be represented in accordance with § 1 by a linear aggregate of terms

$$\mathbf{z} = \prod_{j=1}^k (\mathbf{p}_j^s \mathbf{q}_j^r). \quad (19)$$

With the abbreviation

$$\mathbf{p}_l = \prod_{j=l+1}^k (\mathbf{p}_j^s \mathbf{q}_j^r) \prod_{j=1}^{l-1} (\mathbf{p}_j^s \mathbf{q}_j^r), \quad (20)$$

one can write the derivatives as

$$\left. \begin{aligned} \frac{\partial \mathbf{z}}{\partial \mathbf{p}} &= \sum_{l=1}^k \sum_{m=0}^{s_l-1} \mathbf{p}^{s_l-1-m} \mathbf{q}_l^r \mathbf{p}_l \mathbf{p}^m, \\ \frac{\partial \mathbf{z}}{\partial \mathbf{q}} &= \sum_{l=1}^k \sum_{m=0}^{r_l-1} \mathbf{q}^{r_l-1-m} \mathbf{p}_l \mathbf{p}^{s_l} \mathbf{q}^m. \end{aligned} \right\} \quad (21)$$

From these equations we find an important consequence. We consider the matrices

$$\mathbf{d}_1 = \mathbf{q} \frac{\partial \mathbf{z}}{\partial \mathbf{q}} - \frac{\partial \mathbf{z}}{\partial \mathbf{q}} \mathbf{q}, \quad \mathbf{d}_2 = \mathbf{p} \frac{\partial \mathbf{z}}{\partial \mathbf{p}} - \frac{\partial \mathbf{z}}{\partial \mathbf{p}} \mathbf{p}. \quad (22)$$

From (21) we have

$$\begin{aligned} \mathbf{d}_1 &= \sum_{l=1}^k (\mathbf{q}^{r_l} \mathbf{P}_l \mathbf{p}^{s_l} - \mathbf{P}_l \mathbf{p}^{s_l} \mathbf{q}^{r_l}), \\ \mathbf{d}_2 &= \sum_{l=1}^k (\mathbf{p}^{s_l} \mathbf{q}^{r_l} \mathbf{P}_l - \mathbf{P}_l \mathbf{p}^{s_l} \mathbf{q}^{r_l}). \end{aligned}$$

and thus it follows that

$$\mathbf{d}_1 + \mathbf{d}_2 = \sum_{l=1}^k (\mathbf{p}^{s_l} \mathbf{q}^{r_l} \mathbf{P}_l - \mathbf{P}_l \mathbf{p}^{s_l} \mathbf{q}^{r_l}).$$

Herein the second member of each term cancels the first member of the following, and the first and last member of the overall sum also cancel, so that

$$\mathbf{d}_1 + \mathbf{d}_2 = 0. \quad (23)$$

Because of its linear character in  $\mathbf{z}$ , this relation holds not only for expressions  $\mathbf{z}$  having the form (19), but indeed for arbitrary analytical functions  $\mathbf{g}(\mathbf{pq})$ .<sup>4</sup>

In concluding this brief survey of matrix analysis, we establish the following rule: *Every matrix equation*

$$\mathbf{F}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_r) = 0$$

*remains valid if in all the matrices  $x_j$  one and the same permutation of all rows and columns is undertaken.* To this end, it suffices to show that for two matrices  $\mathbf{a}$ ,  $\mathbf{b}$  which thereby become transposed to  $\mathbf{a}'$ ,  $\mathbf{b}'$ , the following invariance conditions apply:

$$\mathbf{a}' + \mathbf{b}' = (\mathbf{a} + \mathbf{b})', \quad \mathbf{a}'\mathbf{b}' = (\mathbf{ab})',$$

wherein the right-hand sides denote those matrices which are formed from  $\mathbf{a} + \mathbf{b}$  and  $\mathbf{ab}$  respectively by such an interchange.

We set forth this proof by replacing the procedure of permutation by that of multiplication with a suitable matrix.<sup>5</sup>

We write a permutation as

$$\begin{pmatrix} 0 & 1 & 2 & 3 & \dots \\ k_0 & k_1 & k_2 & k_3 & \dots \end{pmatrix} = \begin{pmatrix} n \\ k_n \end{pmatrix}$$

and to this we assign a *permutation matrix*,

$$\mathbf{p} = (p(nm)), \quad p(nm) = \begin{cases} 1 & \text{when } m = k_n \\ 0 & \text{otherwise.} \end{cases}$$

The transposed matrix to  $\mathbf{p}$  is

$$\tilde{\mathbf{p}} = (\tilde{p}(nm)), \quad \tilde{p}(nm) = \begin{cases} 1 & \text{when } n = k_m \\ 0 & \text{otherwise.} \end{cases}$$

---

<sup>4</sup>More generally, for function of  $r$  variables, one has

$$\sum_r \left( \mathbf{x}_r \frac{\partial \mathbf{g}}{\partial \mathbf{x}_r} - \frac{\partial \mathbf{g}}{\partial \mathbf{x}_r} \mathbf{x}_r \right) = 0.$$

<sup>5</sup>The method of proof adopted here possesses the merit of revealing the close connection of permutations with an important class of more general transformations of matrices. The validity of the rule in question can however also be established directly on noting that in the definitions of *equality*, as also of *addition* and *multiplication* of matrices, no use was made of order relationships between the rows or the columns.

On multiplying the two together, one has

$$\mathbf{p}\tilde{\mathbf{p}} = \left( \sum_k p(nk)\tilde{p}(km) \right) = (\delta_{nm}) = \mathbf{1},$$

since the two factors  $p(nk)$  and  $\tilde{p}(km)$  differ from zero simultaneously only if  $k = k_n = k_m$ , i.e., when  $n = m$ . Hence  $\tilde{\mathbf{p}}$  is reciprocal to  $\mathbf{p}$ :

$$\tilde{\mathbf{p}} = \mathbf{p}^{-1}.$$

If now  $\mathbf{a}$  be any given matrix, then

$$\mathbf{p}\mathbf{a} = \left( \sum_k p(nk)a(km) \right) = (a(k_n, m))$$

is a matrix which arises from the permutation  $\begin{pmatrix} n \\ k_n \end{pmatrix}$  of the rows of  $\mathbf{a}$  and equivalently

$$\mathbf{a}\mathbf{p}^{-1} = \left( \sum_k a(mk)\tilde{p}(km) \right) = (a(n, k_m))$$

is the matrix arising from permutation of the columns of  $\mathbf{a}$ . One and the same permutation applied both to the rows and the columns of  $\mathbf{a}$  thus yields the matrix

$$\mathbf{a}' = \mathbf{p}\mathbf{a}\mathbf{p}^{-1}.$$

Thence follows directly

$$\begin{aligned} \mathbf{a}' + \mathbf{b}' &= \mathbf{p}(\mathbf{a} + \mathbf{b})\mathbf{p}^{-1} &= (\mathbf{a} + \mathbf{b})', \\ \mathbf{a}'\mathbf{b}' &= \mathbf{p}\mathbf{a}\mathbf{b}\mathbf{p}^{-1} &= (\mathbf{a}\mathbf{b})' \end{aligned}$$

which proves our original contention.

It is thus apparent that from matrix equations one can never determine any given sequence or order of rank of the matrix elements. Moreover, it is evident that a much more general rule applies, namely that every matrix equation is invariant with respect to transformations of the type

$$\mathbf{a}' = \mathbf{b}\mathbf{a}\mathbf{b}^{-1},$$

where  $\mathbf{b}$  denotes an *arbitrary* matrix. We shall see later that this does not necessarily always apply to matrix differential equations.

## Chapter 2. Dynamics

### 3. The basic laws

The dynamic system is to be described by (lie spatial coordinate  $\mathbf{q}$  and the momentum  $\mathbf{p}$ , these being represented by matrices

$$\mathbf{q} = (q(nm)e^{2\pi i\nu(nm)t}, \quad \mathbf{p}(p(nm)e^{2\pi i\nu(nm)t}). \quad (24)$$

Here the  $\nu(nm)$  denote the quantum-theoretical frequencies associated with transitions between states described by the *quantum numbers*  $n$  and  $m$ . The matrices (24) are to be Hermitian, e.g., on transposition of the matrices, each element is to go over into its complex conjugate value, a condition which should apply for all real  $t$ . We thus have

$$q(nm)q(mn) = |q(nm)|^2 \quad (25)$$

and

$$\nu(nm) = -\nu(mn). \quad (26)$$

If  $\mathbf{q}$  be a *Cartesian* coordinate, then the expression (25) is a measure of the *probabilities*<sup>6</sup> of the transitions  $n \leftrightarrow m$ .

Further, we shall require that

$$\nu(jk) + \nu(kl) + \nu(lj) = 0. \quad (27)$$

This can be expressed together with (26) in the following manner: there exist quantities  $W_n$  such that

$$h\nu(nm) = W_n - W_m. \quad (28)$$

From this, with equations (2), (3), it follows that a function  $\mathbf{g}(\mathbf{pq})$  invariably again takes on the form

$$\mathbf{g} = (g(nm)e^{2\pi i\nu(nm)t}) \quad (29)$$

and the matrix  $(g(nm))$  therein results from identically the same process applied to the matrices  $(q(nm)), (p(nm))$  as was employed to find  $\mathbf{g}$  from  $\mathbf{q}, \mathbf{p}$ . For this reason we can henceforth abandon the representation (24) in favour of the shorter notation

$$\mathbf{q} = (q(nm)), \quad \mathbf{p} = (p(nm)). \quad (30)$$

---

<sup>6</sup>In this connection see §8.

For the *time derivative* of the matrix  $\mathbf{g} = (g(nm))$ , recalling to mind (24) or (29), we obtain the matrix

$$\dot{\mathbf{g}} = 2\pi i(\nu(nm)g(nm)). \quad (31)$$

If  $\nu(nm) \neq 0$  when  $n \neq m$ , a condition which we wish to assume, then the formula  $\dot{\mathbf{g}} = 0$  denotes that  $\mathbf{g}$  is a diagonal matrix with  $g(nm) = \delta_{nm}(nn)$ .

A matrix differential equation  $\dot{\mathbf{g}} = \mathbf{a}$  is invariant with respect to that process in which the same permutation is carried out on rows and columns of all the matrices and also upon the numbers  $W_n$ . In order to realize this, consider the diagonal matrix

$$\mathbf{W} = (\delta_{nm}W_n).$$

Then

$$\begin{aligned}\mathbf{W}\mathbf{g} &= \left( \sum_k \delta_{nk} W_n g(km) \right) = (W_n g(nm)), \\ \mathbf{g}\mathbf{W} &= \left( \sum_k g(nk) \delta_{km} W_k \right) = (W_m g(nm)),\end{aligned}$$

i.e., according to (31),

$$\dot{\mathbf{g}} = \frac{2\pi i}{h}((W_n - W_m)g(nm)) = \frac{2\pi i}{h}(\mathbf{W}\mathbf{g} - \mathbf{g}\mathbf{W}).$$

If now  $\mathbf{p}$  be a permutation matrix, then the transform of  $\mathbf{W}$ ,

$$\mathbf{W}' = \mathbf{p}\mathbf{W}\mathbf{p}^{-1} = (\delta_{nkm}W_{nk})$$

is the diagonal matrix with the permuted  $W_n$  along the diagonal. Thence one has

$$\mathbf{p}\dot{\mathbf{g}}\mathbf{p}^{-1} = \frac{2\pi i}{h}(\mathbf{W}'\mathbf{g}' - \mathbf{g}'\mathbf{W}') = \dot{\mathbf{g}}',$$

where  $\mathbf{g}' = \mathbf{p}\mathbf{g}\mathbf{p}^{-1}$  and  $\dot{\mathbf{g}}'$  denotes the time derivative of  $\mathbf{g}'$  constructed in accordance with the rule (31) with permuted  $W_n$ .

The rows and columns of  $\dot{\mathbf{g}}$  thus experience the same permutation as those of  $\mathbf{g}$ , and hence our contention is vindicated.

It is to be noted that a corresponding rule does *not* apply to arbitrary transformations of the form  $\mathbf{a}' = \mathbf{bab}^{-1}$  since for these  $\mathbf{W}'$  is no longer a diagonal matrix. Despite this difficulty, a thorough study of these general transformations would seem to be called for, since it offers promise of insight

into the deeper connections intrinsic to this new theory: we shall later revert to this point.<sup>7</sup>

In the case of a Hamilton function having the form

$$\mathbf{H} = \frac{1}{2m}\mathbf{p}^2 + \mathbf{U}(\mathbf{q})$$

we shall assume, as did Heisenberg, that the *equations of motion* are just of the same form as in classical theory, so that using the notation of §2 we can write:

$$\left. \begin{aligned} \dot{\mathbf{q}} &= \frac{\partial \mathbf{H}}{\partial \mathbf{p}} = \frac{1}{m}\mathbf{p}, \\ \dot{\mathbf{p}} &= -\frac{\partial \mathbf{H}}{\partial \mathbf{q}} = -\frac{\partial \mathbf{U}}{\partial \mathbf{q}}. \end{aligned} \right\} \quad (32)$$

We now use correspondence considerations to try more generally to elucidate the equations of motion belonging to an arbitrary Hamilton function  $\mathbf{H}(\mathbf{pq})$ . This is required from the standpoint of relativistic mechanics and in particular for the treatment of electron motion under the influence of magnetic fields. For in this latter case, the function  $\mathbf{H}$  cannot in a Cartesian coordinate system any longer be represented by the sum of two functions of which one depends only on the momenta and the other on the coordinates.

Classically, equations of motion can be derived from the action principle

$$\int_{t_0}^{t_1} L dt = \int_{t_0}^{t_1} \{p\dot{q} - H(pq)\} dt = \text{extremum.} \quad (33)$$

If we now envisage the Fourier expansion  $L$  substituted in (33) and the time interval  $t_1 - t_0$  taken sufficiently large, we find that only the constant term of  $L$  supplies a contribution to the integral. The form which the action principle thence acquires suggests the following translation into quantum mechanics:

The diagonal sum  $\mathbf{D}(\mathbf{L}) = \sum_k L(kk)$  is to be made an extremum:

$$\mathbf{D}(\mathbf{L}) = \mathbf{D}(\mathbf{p}\dot{\mathbf{q}} - \mathbf{H}(\mathbf{pq})) = \text{extremum,} \quad (34)$$

namely, by suitable choice of  $\mathbf{p}$  and  $\mathbf{q}$ , with  $\nu(nm)$  kept fixed.

Thus, by setting the derivatives of  $\mathbf{D}(\mathbf{L})$  with respect to the elements of  $\mathbf{p}$  and  $\mathbf{q}$  equal to zero, one obtains the equations of motion

$$2\pi i\nu(nm)q(nm) = \frac{\partial \mathbf{D}(\mathbf{H})}{dp(nm)},$$

---

<sup>7</sup>Cf. the continuation of this work, to be published forthwith.

$$2\pi i\nu(mn)p(mn)\frac{\partial \mathbf{D}(\mathbf{H})}{\partial q(mn)}.$$

From (26), (31) and (16) one observes that these equations of motion can always be written in *canonical* form,

$$\left. \begin{aligned} \dot{\mathbf{q}} &= \frac{\partial \mathbf{H}}{\partial \mathbf{p}}, \\ \dot{\mathbf{p}} &= -\frac{\partial \mathbf{H}}{\partial \mathbf{q}}. \end{aligned} \right\} \quad (35)$$

For the quantization condition, Heisenberg employed a relation proposed by Thomas<sup>8</sup> and Kuhn.<sup>9</sup> The equation

$$\mathbf{J} = \oint p dq = \int_0^{1/\nu} p \dot{q} dt$$

of “classical” quantum theory can, on introducing the Fourier expansions of  $p$  and  $q$ ,

$$p = \sum_{\tau=-\infty}^{\infty} p_{\tau} e^{2\pi i \nu \tau t}, \quad q = \sum_{\tau=-\infty}^{\infty} q_{\tau} e^{2\pi i \nu \tau t},$$

be transformed into

$$1 = 2\pi i \sum_{\tau=-\infty}^{\infty} \tau \frac{\partial}{\partial \mathbf{J}} (q_{\tau} p_{-\tau}). \quad (36)$$

If therein one has  $p = m\dot{q}$ , one can express the  $p_{\tau}$  in terms of  $q_{\tau}$  and thence obtain that classical equation which on transformation into a difference equation according to the principle of correspondence yields the formula of Thomas and Kuhn. Since here the assumption that  $\mathbf{p} = m\dot{\mathbf{q}}$  should be avoided, we are obliged to translate equation (36) directly into a difference equation.

The following expressions should correspond:

$$\begin{aligned} &\sum_{\tau=-\infty}^{\infty} \tau \frac{\partial}{\partial \mathbf{J}} (q_{\tau} p_{-\tau}) \quad \text{with} \\ &\frac{1}{h} \sum_{\tau=-\infty}^{\infty} q(n+\tau, n)p(n, n+\tau) - q(n, n-\tau)p(n-\tau, n); \end{aligned}$$

---

<sup>8</sup>W. Thomas, Naturwiss. **13** (1925) 627.

<sup>9</sup>W. Kuhn, Zs. f. Phys. **33** (1925) 408.

where in the right-hand expression those  $q(nm)$ ,  $p(nm)$  which take on a negative index are to be set equal to zero. In this way we obtain the quantization condition corresponding to (36) as

$$\sum_k (p(nk)q(kn) - q(nk)p(kn)) = \frac{h}{2\pi i}. \quad (37)$$

This is a system of infinitely many equations, namely one for each value of  $n$ .

In particular, for  $\mathbf{p} = m\dot{\mathbf{q}}$  this yields

$$\sum_k \nu(kn)|q(nk)|^2 = \frac{h}{8\pi^2 m},$$

which, as may easily be verified, agrees with Heisenberg's form of the quantization condition, or with the Thomas-Kuhn equation. The formula (37) has to be regarded as the appropriate generalization of this equation.

Incidentally one sees from (37) that the diagonal sum  $\mathbf{D}(\mathbf{pq})$  necessarily becomes infinite. For otherwise one would have  $\mathbf{D}(\mathbf{pq}) - \mathbf{D}(\mathbf{qp}) = 0$  from whereas (37) leads to  $\mathbf{D}(\mathbf{pq}) - \mathbf{D}(\mathbf{qp}) = \infty$ . Thus the matrices under consideration are never finite.<sup>10</sup>

#### 4. Consequences. Energy-conservation and frequency laws

The content of the preceding paragraphs furnishes the basic rules of the new quantum mechanics in their entirety. All other laws of quantum mechanics, whose general validity is to be verified, must be *derivable* from these basic tenets. As instances of such laws to be proved, the law of energy conservation and the Bohr frequency condition primarily enter into consideration. The law of conservation of energy states that if  $\mathbf{H}$  be the energy, then  $\dot{\mathbf{H}} = 0$ , or that  $\mathbf{H}$  is a *diagonal matrix*. The diagonal elements  $H(nn)$  of  $\mathbf{H}$  are interpreted, according to Heisenberg, as the *energies of the various states of the system* and the Bohr frequency condition requires that

$$h\nu(nm) = H(nn) - H(mm),$$

or

$$W_n = H(nn) + \text{const.}$$

---

<sup>10</sup>Further, they do not belong to the class of "bounded" infinite matrices hitherto almost exclusively investigated by mathematicians.

We consider the quantity

$$\mathbf{d} = \mathbf{pq} - \mathbf{qp}.$$

From (11), (35) one finds

$$\dot{\mathbf{d}} = \dot{\mathbf{p}}\mathbf{q} + \mathbf{p}\dot{\mathbf{q}} - \dot{\mathbf{q}}\mathbf{p} - \mathbf{q}\dot{\mathbf{p}} = \mathbf{q}\frac{\partial \mathbf{H}}{\partial \mathbf{q}} - \frac{\partial \mathbf{H}}{\partial \mathbf{q}}\mathbf{q} + \mathbf{p}\frac{\partial \mathbf{H}}{\partial \mathbf{p}} - \frac{\partial \mathbf{H}}{\partial \mathbf{p}}\mathbf{p}.$$

Thus from (22), (23) it follows that  $\dot{\mathbf{d}} = 0$  and  $\mathbf{d}$  is a diagonal matrix. The diagonal elements of  $\mathbf{d}$  are, however, specified just by the quantum condition (27). Summarizing, we obtain the equation

$$\mathbf{pq} - \mathbf{qp} = \frac{h}{2\pi i} \mathbf{1}, \quad (38)$$

on introducing the unit matrix  $\mathbf{1}$  defined by (6). We term the equation (38) the “stronger quantum condition” and base all further conclusions upon it.

From the form of this equation, we deduce the following: If an equation (*A*) be derived from (38), then (*A*) remains valid if  $\mathbf{p}$  be replaced by  $\mathbf{q}$  and simultaneously  $h$  by  $-h$ . For this reason one need for instance derive only one of the following two equations from (38), which can readily be performed by induction

$$\mathbf{p}^n \mathbf{q} = \mathbf{q} \mathbf{p}^n + n \frac{h}{2\pi i} \mathbf{p}^{n-1}, \quad (39)$$

$$\mathbf{q}^n \mathbf{p} = \mathbf{p} \mathbf{q}^n - n \frac{h}{2\pi i} \mathbf{q}^{n-1}. \quad (39')$$

We shall now prove the energy-conservation and frequency laws, as expressed above, in the first instance for the case

$$\mathbf{H} = \mathbf{H}_1(\mathbf{p}) + \mathbf{H}_2(\mathbf{q}).$$

From the statements of §1, it follows that we may formally replace  $\mathbf{H}_1(\mathbf{p})$  and  $b f H_2(q)$  by power expansions

$$\mathbf{H}_1 = \sum_s a_s \mathbf{p}^s, \quad \mathbf{H}_2 = \sum_s b_s \mathbf{q}^s.$$

Formulae (39) and (39') indicate that

$$\left. \begin{aligned} \mathbf{H}\mathbf{q} - \mathbf{q}\mathbf{H} &= \frac{h}{2\pi i} \frac{\partial \mathbf{H}}{\partial \mathbf{p}}, \\ \mathbf{H}\mathbf{p} - \mathbf{p}\mathbf{H} &= -\frac{h}{2\pi i} \frac{\partial \mathbf{H}}{\partial \mathbf{p}}. \end{aligned} \right\} \quad (40)$$

Comparison with the equations of motion (35) yields

$$\left. \begin{aligned} \dot{\mathbf{q}} &= \frac{2\pi i}{\hbar} (\mathbf{H}\mathbf{q} - \mathbf{q}\mathbf{H}), \\ \dot{\mathbf{p}} &= \frac{2\pi i}{\hbar} (\mathbf{H}\mathbf{p} - \mathbf{p}\mathbf{H}). \end{aligned} \right\} \quad (41)$$

Denoting the matrix  $\mathbf{Hg} - \mathbf{gH}$  by  $\begin{vmatrix} \mathbf{H} \\ \mathbf{g} \end{vmatrix}$  for brevity, one has

$$\begin{vmatrix} \mathbf{H} \\ \mathbf{ab} \end{vmatrix} = \begin{vmatrix} \mathbf{H} \\ \mathbf{a} \end{vmatrix} \begin{vmatrix} \mathbf{b} + \mathbf{a} \\ \mathbf{b} \end{vmatrix}; \quad (42)$$

from which generally for  $\mathbf{g} = \mathbf{g}(\mathbf{pq})$  one may conclude that

$$\dot{\mathbf{g}} = \frac{2\pi i}{\hbar} \begin{vmatrix} \mathbf{H} \\ \mathbf{g} \end{vmatrix} = \frac{2\pi i}{\hbar} (\mathbf{Hg} - \mathbf{gH}). \quad (43)$$

To establish this result, one need only conceive  $\dot{\mathbf{g}}$  as expressed in function of  $\mathbf{p}, \mathbf{q}$  and  $\dot{\mathbf{p}}, \dot{\mathbf{q}}$  with the aid of (11), (11'), and  $\begin{vmatrix} \mathbf{H} \\ \mathbf{g} \end{vmatrix}$  as evaluated by means of (42) in function of  $\mathbf{p}, \mathbf{q}$  and  $\begin{vmatrix} \mathbf{H} \\ \mathbf{p} \end{vmatrix}, \begin{vmatrix} \mathbf{H} \\ \mathbf{q} \end{vmatrix}$  followed by application of the relations (41). In particular, if in (43) one sets  $\mathbf{g} = \mathbf{H}$ , one obtains

$$\dot{\mathbf{H}} = 0. \quad (44)$$

Now that we have verified the energy-conservation law and recognized the matrix  $\mathbf{H}$  to be diagonal, equation (41) can be put into the form

$$h\nu(nm)q(nm) = (H(nn) - H(mm))q(nm),$$

$$h\nu(nm)p(nm) = (H(nn) - H(mm))p(nm),$$

from which the frequency condition follows.

If we now go over to consideration of more general Hamilton functions  $\mathbf{H}^* = \mathbf{H}^*(\mathbf{pq})$ , it can easily be seen that in general  $\dot{\mathbf{H}}^*$  no longer vanishes (examples such as  $\mathbf{H}^* = \mathbf{p}^2\mathbf{q}$ , readily reveal this). It can however be observed that the Hamilton function  $\mathbf{H} = \frac{1}{2}(\mathbf{p}^2\mathbf{q} + \mathbf{qp}^2)$  yields the same equations of motion as  $\mathbf{H}^*$  and that  $\dot{\mathbf{H}}$  again vanishes. In consequence we may express the energy-conservation and frequency laws in the following way: To each function  $\mathbf{H}^* = \mathbf{H}^*(\mathbf{pq})$  there can be assigned a function  $\mathbf{H} = \mathbf{H}(\mathbf{pq})$  such

that as Hamiltonians  $\mathbf{H}^*$  and  $\mathbf{H}$  yield the same equations of motion and that for these equations of motion  $\mathbf{H}$  assumes the role of an energy which is constant in time and which fulfills the frequency condition.

On bearing in mind the considerations discussed above, it suffices to show that the function  $\mathbf{H}$  to be specified satisfies not only the conditions

$$\frac{\partial \mathbf{H}}{\partial \mathbf{p}} = \frac{\partial \mathbf{H}^*}{\partial \mathbf{p}}, \quad \frac{\partial \mathbf{H}}{\partial \mathbf{q}} = \frac{\partial \mathbf{H}^*}{\partial \mathbf{q}}, \quad (45)$$

but in addition satisfies equations (40). From §1, the matrix  $\mathbf{H}^*$  is formally to be represented as a sum of products of powers of  $\mathbf{p}$  and  $\mathbf{q}$ . Because of the linearity of equations (40), (45) in  $\mathbf{H}, \mathbf{H}^*$  we have simply to specify the commensurate sum term in  $\mathbf{H}$  a counterpart to each individual sum term in  $\mathbf{H}^*$ . Thus we need consider solely the case

$$\mathbf{H}^* = \prod_{j=1}^k (\mathbf{p}^{s_j}, \mathbf{q}^{r_j}). \quad (46)$$

It follows from the remarks of §2 that equations (45) can be satisfied by specifying  $\mathbf{H}$  as a linear form of those products of powers of  $\mathbf{p}, \mathbf{q}$  which arise from  $\mathbf{H}^*$  through cyclic interchange of the factors; herein the sum of the coefficients must be held to unity. The question as to how these coefficients are to be chosen so that equations (40) may also be satisfied is less easy to answer. It may at this juncture suffice to dispose of the case  $k = 1$ , namely

$$\mathbf{H}^* = \mathbf{p}^s \mathbf{q}^r. \quad (47)$$

The formula (39) can be generalized<sup>11</sup> to

$$\mathbf{p}^m \mathbf{q}^n - \mathbf{q}^n \mathbf{p}^m = m \frac{h}{2\pi i} \sum_{l=0}^{n-1} \mathbf{q}^{n-1-l} \mathbf{p}^{m-1} \mathbf{q}^l. \quad (48)$$

---

<sup>11</sup>A different generalization is furnished by the formulae

$$\begin{aligned} \mathbf{p}^m \mathbf{q}^n &= \sum_{j=0}^{m,n} j! \binom{m}{j} \binom{n}{j} \left(\frac{h}{2\pi i}\right)^j \mathbf{q}^{n-j} \mathbf{p}^{m-j}, \\ \mathbf{q}^n \mathbf{p}^m &= \sum_{j=0}^{m,n} j! \binom{m}{j} \binom{n}{j} \left(\frac{h}{2\pi i}\right)^j \mathbf{p}^{m-j} \mathbf{q}^{n-j}, \end{aligned}$$

where  $j$  runs to the lesser of the two integers  $m, n$ .

For  $n = 1$  this reverts to (39); in general (48) ensues from the fact that because of (39) one has

$$\mathbf{p}^m \mathbf{q}^{n+1} - \mathbf{q}^{n+1} \mathbf{p}^m = (\mathbf{p}^m \mathbf{q}^n - \mathbf{q}^n \mathbf{p}^m) \mathbf{q} + m \frac{h}{2\pi i} \mathbf{q}^n \mathbf{p}^{m+1}.$$

The new formula

$$\mathbf{p}^m \mathbf{q}^n - \mathbf{q}^n \mathbf{p}^m = n \frac{h}{2\pi i} \sum_{j=0}^{m-1} \mathbf{p}^{m-1-j} \mathbf{q}^{n-1} \mathbf{p}^j \quad (48')$$

is obtained on interchanging  $\mathbf{p}$  and  $\mathbf{q}$  and reversing the sign of  $h$ .

Comparison with (48) yields

$$\frac{1}{s+1} \sum_{l=0}^s \mathbf{p}^{s-l} \mathbf{q}^r \mathbf{p}^l = \frac{1}{r+1} \sum_{j=0}^r \mathbf{q}^{r-j} \mathbf{p}^s \mathbf{q}^j. \quad (49)$$

We now assert: The matrix  $\mathbf{H}$  belonging to  $\mathbf{H}^*$  as given by (47) is:

$$\mathbf{H} = \frac{1}{s+1} \sum_{l=0}^s \mathbf{p}^{s-l} \mathbf{q}^r \mathbf{p}^l. \quad (50)$$

We need only prove equations (40), to which end we recall the derivatives, (18') §2.

From (50), we now obtain the relation

$$\mathbf{H}\mathbf{p} - \mathbf{p}\mathbf{H} = \frac{1}{s+1} (\mathbf{q}^r \mathbf{p}^{s+1} - \mathbf{p}^{s+1} \mathbf{q}^r),$$

and according to (48) this is equivalent to the lower of equations (40).

Further, using (49) we find

$$\mathbf{H}\mathbf{q} - \mathbf{q}\mathbf{H} = \frac{1}{r+1} (\mathbf{p}^s \mathbf{q}^{r+1} - \mathbf{q}^{r+1} \mathbf{p}^s),$$

and by (48') this is equivalent to the upper of equations (40). This completes the requisite proof.

Whereas in classical mechanics energy conservation ( $\dot{\mathbf{H}} = 0$ ) is directly apparent from the canonical equations, the same law of energy conservation in quantum mechanics,  $\mathbf{H} = 0$  lies, as one can see, more deeply hidden beneath the surface.

That its demonstrability from assumed postulates is far from being trivial will be appreciated if, following more closely the classical method of

proof, one sets out to prove  $\mathbf{H}$  to be constant simply by evaluating  $\dot{\mathbf{H}}$ . To this end, one first has to express  $\dot{\mathbf{H}}$  as function of  $\mathbf{p}, \mathbf{q}$  and  $\dot{\mathbf{p}}, \dot{\mathbf{q}}$  with the aid of (11), (11'), whereupon for  $\dot{\mathbf{p}}$  and  $\dot{\mathbf{q}}$  the values  $-\partial\mathbf{H}/\partial\mathbf{q}$ ,  $\partial\mathbf{H}/\partial\mathbf{p}$  have to be introduced. This yields  $\dot{\mathbf{H}}$  in function of  $\mathbf{p}$  and  $\mathbf{q}$ . Equation (38) or the formulae quoted in the footnote to equation (48) which were derived from (38) permit this function to be converted into a sum of terms of the type  $a\mathbf{p}^s\mathbf{q}^r$  and one then has to prove that the coefficient  $a$  in each of such terms vanishes. This calculation for the most general case, as considered above along different lines, becomes so exceedingly involved<sup>12</sup> that it seems hardly feasible. The fact that nonetheless energy-conservation and frequency laws could be proved in so general a context would seem to us to furnish strong grounds to hope that this theory embraces truly deep-seated physical laws.

In conclusion, we append a result here which can easily be derived from the formulae of this section, namely: *Equations (35), (37) can be replaced by (38) and (44) (with  $\mathbf{H}$  representing the energy); the frequencies are thereby to be derived from the frequency condition.*

In the continuation to this paper, we shall examine the important applications to which this theorem gives rise.

### Chapter 3. Investigation of the Anharmonic Oscillator

The anharmonic oscillator, having

$$\mathbf{H} = \frac{1}{2}\mathbf{p}^2 + \frac{1}{2}\omega_0^2\mathbf{q}^2 + \frac{1}{2}\lambda\mathbf{q}^3 \quad (51)$$

has already been considered in detail by Heisenberg. Nevertheless, its investigation will here be renewed with the aim of determining the *most general* solution of the fundamental equations for this case. If the basic equations of the present theory are indeed complete and do not require to be supplemented any further, then the absolute values  $|q(nm)|$ ,  $|p(nm)|$  of the elements of the matrices  $\mathbf{q}$  and  $\mathbf{p}$  must *uniquely* be determined by these equations, and thus it becomes important to check this for the example (51). On the other hand, it is to be expected that an uncertainty will still persist with respect to the phases  $\phi_{nm}, \varphi_{nm}$  in the relations

$$q(nm) = |q(nm)|e^{i\phi_{nm}},$$

---

<sup>12</sup>For the case  $\mathbf{H} = (1/2m)\mathbf{p}^2 + \mathbf{U}(\mathbf{q})$  it can immediately be carried out with the aid of (39').

$$p(nm) = |p(nm)| e^{i\varphi_{nm}}.$$

For the statistical theory, e.g., of the interaction of quantised atoms with external radiation fields, it becomes of fundamental importance to ascertain the precise degree of such uncertainty.

## 5. Harmonic oscillator

The starting point in our considerations is the theory of the harmonic oscillator; for small  $\lambda$ , one can regard the motion as expressed by equation (51) to be a perturbation of the normal harmonic oscillation having energy

$$\mathbf{H} = \frac{1}{2}\mathbf{p}^2 + \frac{1}{2}\omega_0^2\mathbf{q}^2. \quad (52)$$

Even for this simple problem it is necessary to supplement Heisenberg's analysis. This latter employs correspondence considerations to arrive at significant deductions as to the form of the solution: namely, since classically only a *single* harmonic component is present, Heisenberg selects a matrix which represents transitions between adjacent states only, and which thus has the form

$$\mathbf{q} = \begin{pmatrix} 0 & q^{(01)} & 0 & 0 & 0 & \dots \\ q^{(10)} & 0 & q^{(12)} & 0 & 0 & \dots \\ 0 & q^{(21)} & 0 & q^{(23)} & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}. \quad (53)$$

We here strive to build up the entire theory self-dependently, without invoking assistance from classical theory on the basis of the principle of correspondence. We shall therefore investigate whether the form of the matrix (53) cannot itself be derived from the basic formulae or, if this proves impossible, which additional postulates are required.

From what has been stated in §3 regarding the invariance with respect to permutation of rows and columns, one can see right away that the exact form of the matrix (53) can never be deduced from the fundamental equations, since if rows and columns be subjected to the same permutation, the canonical equations and the quantum condition remain invariant and thereby one obtains a new and apparently different solution. But all such solutions naturally differ only in the notation, i.e., in the way the elements are numbered. We seek to prove that through a mere renumbering of its elements, the solution can always be brought into the form (53). The equation of motion

$$\ddot{\mathbf{q}} + \omega_0^2\mathbf{q} = 0 \quad (54)$$

runs as follows for the elements:

$$(\nu^2(nm) - \nu_0^2)q(nm) = 0, \quad (55)$$

where

$$\omega^0 = 2\pi\nu_0, \quad h\nu(nm) = W_n - W_m.$$

From the stronger quantum condition

$$\mathbf{pq} - \mathbf{qp} = \frac{h}{2\pi i} \mathbf{1}, \quad (56)$$

it follows that for each  $n$  there must exist a corresponding  $n'$  such that  $q(nn') \neq 0$ , since if there were a value of  $n$  for which all  $q(nn')$  were equal to zero, then the with diagonal element of  $\mathbf{pq} - \mathbf{qp}$  would be zero, which contradicts the quantum condition. Hence equation (55) implies that there is always an  $n'$  for which

$$|W_n - W_{n'}| = h\nu_0.$$

But since we have assumed in our basic principles that when  $n \neq m$ , the energies are always unequal ( $W_n \neq W_m$ ), it follows that at most two such indices  $n'$  and  $n''$  can exist, for the corresponding  $W_{n'}, W_{n''}$  are solutions of the quadratic equation

$$(W_n - x)^2 = h^2\nu_0^2;$$

and if indeed *two* such indices  $n', n''$  exist, it follows that the corresponding frequencies must be related as:

$$\nu(nn') = -\nu(nn''). \quad (57)$$

Now from (56) we get

$$\sum_k \nu(kn)|q(nk)|^2 = \nu(n'n)\{|q(nn')|^2 - |q(nn'')|^2\} = h/8\pi^2, \quad (58)$$

and the energy (52) ensues as

$$\begin{aligned} H(nm) &= \frac{1}{2} \times 4\pi^2 \sum_k \{-\nu(nk)\nu(km)q(nk)q(km) + \nu_0^2 q(nk)q(km)\} \\ &= 2\pi^2 \sum q(nk)q(km)\{\nu_0^2 - \nu(nk)\nu(km)\}. \end{aligned}$$

In particular, for  $m = n$  we have

$$H(nn) = W_n = 4\pi^2\nu_0^2(|q(nn')|^2 + |q(nn'')|^2). \quad (59)$$

Moreover, we can now distinguish between three possible cases:

- (a) no  $n''$  exists and one has  $W_{n'} > W_n$ ;
- (b) no  $n''$  exists and one has  $W_{n'} < W_n$ ,
- (c)  $n''$  exists.

In case (b) we now consider  $n'$  in place of  $n$ ; to this there belong at most two indices  $(n')'$  and  $(n' )''$  and of these, one has to equal  $n$ . We thereby revert to one of the cases (a) or (c) and can accordingly omit further consideration of (b).

In case (a),  $\nu(n'n) = +\nu_0$  and from (58) it follows that

$$\nu_0|q(nn')|^2 = h/8\pi^2, \quad (60)$$

and thus from (59) that

$$W_n = H(nn) = 4\pi^2\nu_0^2|q(nn')|^2 = \frac{1}{2}\nu_0h.$$

Because of the assumption that  $W_n \neq W_m$  for  $n \neq m$  there is thus at most *one* index  $n = n_0$  for which the case (a) applies.

If such an  $n_0$  exists, we can specify a series of numbers  $n_0, n_1, n_2, n_3, \dots$ , such that  $(n_k)' = n_{k+1}$  and  $W_{k+1} > W_k$ . Then invariably  $(n_{k+1})'' = n_k$ . Hence for  $k > 0$ , equations (58) and (59) give

$$H(n_k n_k) = 4\pi^2\nu_0^2\{|q(n_k, n_{k+1})|^2 + |q(n_k, n_{k-1})|^2\}, \quad (61)$$

$$\frac{1}{2}h = 4\pi^2\nu_0\{|q(n_k, n_{k+1})|^2 - |q(n_k, n_{k-1})|^2\}. \quad (62)$$

From (60) and (62) it follows that

$$|q(n_k, n_{k+1})|^2 = \frac{h}{8\pi^2\nu_0}(k+1), \quad (63)$$

and thence from (61) that

$$W_{n_k} = H(n_k, n_k) = \nu_0h(k + \frac{1}{2}). \quad (64)$$

Now, we still have to check whether it be possible that there is no value of  $n$  for which case (a) applies. Beginning with an arbitrary  $n_0$  we can then build  $n'_0 = n$ , and  $n''_0 = n_{-1}$  and with each of these latter write  $n'_1 = n_2$ ,  $n''_1 = n_0$  and  $n'_{-1} = n_0$ ,  $n''_{-1} = n_{-2}$  etc. In this manner we obtain a series of numbers  $\dots n_{-2}, n_{-1}, n_0, n_1, n_2 \dots$ , and equations (61), (62) hold for every

$k$  between  $-\infty$  and  $+\infty$ . But this is impossible, since by (62) the quantities  $x_k = |q(n_{k+1}, n_k)|^2$  form an equispaced series of numbers, and since they are positive, there must be a least value. The relevant index can then again be designated as  $n_0$  and we thereby revert to the previous case – thus here also, the formulae (63), (64) apply.

One can further see that every number  $n$  must be contained within the numbers  $n_k$ , since otherwise one could construct a new series (65) proceeding from  $n$ , and for this formula (60) would again hold. The starting terms of both series would then have the same value  $W_n = H(nn)$ , which is not possible.

This proves that the indices 0, 1, 2, 3... can be rearranged into a new sequence  $n_0, n_1, n_2, n_3 \dots$  such that formulae (63), (64) apply: with these new indices, the solution then takes on Heisenberg's form (53). Hence this appears as the “normal form” of the general solution. By virtue of (64), it possesses the property that

$$W_{n_{k+1}} > W_{n_k}.$$

If, inversely, one stipulate that  $W_n = H(nn)$  should always increase with  $n$ , then it necessarily follows that  $n_k = k$ ; this principle thus uniquely establishes the normal form of the solution. But thereby only the notation becomes fixed and the calculation more transparent: nothing new is conferred *physically*.

Therein lies the big difference between this and the previously adopted semiclassical methods of determining the stationary states. The classically calculated orbits merge into one another continuously; consequently the quantum orbits selected at a later stage have a particular sequence right from the outset. The new mechanics presents itself as an essentially discontinuous theory in that herein there is no question of a sequence of quantum states defined by the physical process, but rather of quantum numbers which are indeed no more than distinguishing indices which can be ordered and normalized according to any practical standpoint whatsoever (e.g., according to increasing energy  $W_n$ ).

## 6. Anharmonic oscillator

The equations of motion

$$\ddot{\mathbf{q}} + \omega_0^2 \mathbf{q} + \lambda \mathbf{q}^2 = 0, \quad (66)$$

together with the quantum condition yield the following system of equations for the elements:

$$\begin{aligned} (\omega_0^2 - \omega^2(nm))q(nm) + \lambda \sum_k q(nk)q(km) &= 0, \\ \sum_k \omega(nk)q(nk)q(kn) &= -h/4\pi. \end{aligned} \quad (67)$$

We introduce series expansions

$$\begin{aligned} \omega(nm) &= \omega^0(nm) + \lambda\omega^{(1)}(nm) + \lambda^2\omega^{(2)}(nm) + \dots \\ \dot{q}(nm) &= q_0(nm) + \lambda q^{(1)}(nm) + \lambda^2 q^{(2)}(nm) + \dots \end{aligned} \quad (68)$$

in seeking the solution.

When  $\lambda = 0$ , one has the case of the harmonic oscillator considered in the previous section; we write the solution (53) in the form

$$q_0(nm) = a_n \delta_{n,m-1} + \overline{a_m} \delta_{n-1,m}, \quad (69)$$

where the bar denotes the conjugate complex value. If one builds the square or higher powers of the matrix  $\mathbf{q}^0 = (q^0(nm))$ , one arrives at matrices of similar form, being composed of sums of terms

$$(\xi_{nm}^{(p)}) = \xi_n \delta_{n,m-p} + \overline{\xi_m} \delta_{n-p,m}. \quad (70)$$

This prompts us to try a solution of the form

$$\begin{aligned} q^0(nm) &= (a)_{nm}^{(1)}, \\ q^{(1)}(nm) &= (x)_{nm}^0 + (x')_{nm}^{(2)}, \\ q^{(2)}(nm) &= (y)_{nm}^{(1)} + (y')_{nm}^{(3)}, \\ &\dots \end{aligned} \quad (71)$$

$n$  which odd and even values of the index  $p$  always alternate. If one actually inserts this in the approximation equations

$$\lambda : \left\{ \begin{array}{l} (\omega_0^2 - \omega^0(nm)^2)q^{(1)}(nm) - 2\omega^0(nm)\omega^{(1)}(nm)q^0(nm) \\ \quad + \sum_k q^0(nk)q^0(km) = 0, \\ \sum_k \{\omega^0(nk) (q^0(nk)q^{(1)}(kn) + q^{(1)}(nk)q^0(kn)) \\ \quad + \omega^{(1)}(nk)q^0(nk)q^0(kn)\} = 0, \end{array} \right\} \quad (72)$$

$$\lambda^2 : \left\{ \begin{array}{l} (\omega_0^2 - \omega^0(nm)^2)q^2)(nm) - 2\omega^0(nm)\omega^{(1)}(nm)q^{(1)}(nm) \\ -(\omega^{(1)}(nm)^2 + 2\omega^0(nm)\omega^{(2)}(nm))q^0(nm) \\ + \sum_k (q^0(nk)q^{(1)}(km) + q^{(1)}(nk)q^0(km)) = 0, \\ \sum_k \{\omega^0 nk) (q^0(nk)q^{(2)}(km) + q^{(1)}(nk)q^{(1)}(km) \\ + q^{(2)}(nk)q^0(km)) + \omega^{(1)}(nk)(q^0(nk)q^{(1)}(km) \\ + q^{(1)}(nk)q^0(km)) + \omega^{(2)}(nk)q^0(nk)q^0(km)\} = 0 \end{array} \right\} \quad (73)$$

and notes the multiplication rule

$$\begin{aligned} \sum_k \Omega_{nkm}(\xi)_{nk}^{(p)}(\eta)_{km}^{(q)} &= \Omega_{n,n+p,n+p+q}\xi_n\eta_{n+p}\delta_{n,m-p-q} \\ &+ \Omega_{n,n+p,n+p-q}\xi_n\bar{\eta}_{n+p-q}\delta_{n,m-p+q} \\ &+ \Omega_{n,n-p,n-p+q}\bar{\xi}_{n-p}\eta_{n-p}\delta_{n,m+p-q} \\ &+ \Omega_{n,n-p,n-p-q}\bar{\xi}_{n-p}\bar{\eta}_{n-p-q}\delta_{n,m+p+q}, \end{aligned} \quad (74)$$

one sees, in setting each of the factors of  $\delta_{n,m-s}$  singly to zero, that through the substitution (71) all conditions can in fact be satisfied and that higher terms in (71) would identically vanish.

In detail, the calculation yields the following:

The first of the equations (72) gives, after substitution of the expressions (71),

$$\left. \begin{array}{l} 2\omega_0^2 x_n + |a_n|^2 + |a_{n-1}|^2 = 0, \\ -3\omega_0^2 x'_n + a_n a_{n+1} = 0, \\ \omega_{n,n-1}^{(1)} = 0, \end{array} \right\} \quad (75)$$

and the second is identically satisfied. One thus has

$$\left. \begin{array}{l} x_n = -\frac{|a_n|^2 + |a_{n-1}|^2}{2\omega_0^2}, \\ x'_n = \frac{a_n a_{n+1}}{3\omega_0^2}. \end{array} \right\} \quad (76)$$

The first of the equations (73) yields

$$\left. \begin{array}{l} 2\omega_0 a_n \omega_{n,n+1}^{(2)} + 2a_n x_{n+1} + 2a_n x_n + \tilde{a}_{n-1} x'_{n-1} + \tilde{a}_{n+1} x'_n = 0, \\ -8\omega_0^2 y'_n + a_n x'_{n+1} = a_{n+2} x'_n = 0, \\ \omega_{n,n-2}^{(1)} = 0, \end{array} \right\} \quad (77)$$

whereas the second equation is not identically satisfied, but furnishes a relation from which  $y_n$  can be determined:

$$\begin{aligned} a_n \tilde{y}_n + \tilde{a}_n y_n - a_{n-1} \tilde{y}_{n-1} - \tilde{a}_{n-1} y_{n-1} + 2|x'_n|^2 - 2|x'_{n-2}|^2 \\ - \frac{\omega_{n,n+1}^{(2)}}{\omega_0} |a_n|^2 - \frac{\omega_{n,n-1}^{(2)}}{\omega_0} |a_{n-1}|^2 = 0. \end{aligned} \quad (78)$$

The solution is:

$$\left. \begin{aligned} \omega_{n,n+1}^{(2)} &= \frac{1}{3\omega_0^3} (|a_{n+1}|^2 + |a_{n-1}|^2 + 3|a_n|^2), \\ y'_n &= \frac{1}{12\omega_0^4} a_n a_{n+1} a_{n+2}. \end{aligned} \right\} \quad (79)$$

Further, if for brevity one introduces

$$\eta_n = a_n \tilde{y}_n + \tilde{a}_n y_n, \quad (80)$$

then the  $\eta$  determined by the equation

$$\eta_n - \eta_{n-1} = \frac{1}{\omega_0^4} (|a_n|^4 - |a_{n-1}|^4 + \frac{1}{9} |a_n|^2 |a_{n+1}|^2 - \frac{1}{9} |a_{n-1}|^2 |a_{n-2}|^2). \quad (81)$$

Expressions (76) and (79) show that the quantities  $x_n, x'_n, y'_n$  can be expressed through the solution of the zero-th order approximation  $a_n$ . Thus their phases are determined by those of the harmonic oscillator. For the quantities  $y_n$ , the situation seems to be different, since although  $\eta_n$  can uniquely be determined from (81),  $y_n$  cannot be obtained absolutely from (80). It is probable that the next higher order of approximation gives rise to an auxiliary determining equation for  $y_n$ . We have to leave this question open here but we should like to indicate its significance as a point of principle in regard to the completeness of Hie entire theory. All questions of statistics invariably depend finally upon whether or not our supposition that of the phases of the  $q(nm)$  one in each row (or each column) of the matrix remains undetermined be valid.

In conclusion we present the explicit formulae which are obtained by substituting the solution of the harmonic oscillator found previously (§5). In normal form, by (63), this runs as follows:

$$a_n = \sqrt{C(n+1)} e^{i\varphi_n}, \quad C = h/4\pi\omega_0 = h/8\pi^2\nu_0. \quad (82)$$

Thence, using (76), (79), (81) one obtains

$$\left. \begin{aligned} x_n &= -\frac{C}{2\omega_0^2}(2n+1), \\ x'_n &= \frac{C}{3\omega_0^2}\sqrt{(n+1)(n+2)}e^{i(\varphi_n+\varphi_{n+1})} \end{aligned} \right\} \quad (83)$$

$$\left. \begin{aligned} y'_n &= \frac{\sqrt{C^3}}{12\omega_0^4}\sqrt{(n+1)(n+2)(n+3)}e^{i(\varphi_n+\varphi_{n+1}+\varphi_{n+2})} \\ \omega_{n,n-1}^{(1)} &= 0, \quad \omega_{n,n-2}^{(1)} = 0, \\ \omega_{n,n-1}^{(2)} &= -\frac{5C}{3\omega_0^3}n; \end{aligned} \right\} \quad (84)$$

that is,

$$\begin{aligned} \eta_n - \eta_{n-1} &= \frac{11C^2}{9\omega_0^4}(2n+1), \\ \eta_n &= a_n \tilde{y}_n + \tilde{a}_n y_n = \frac{11C^2}{9\omega_0^4}(n+1)^2. \end{aligned}$$

If one sets  $y_n = |y_n|e^{i\varphi_n}$ , then

$$|y_n| \cos(\varphi_n - \psi_n) = \frac{\eta_n}{2|a_n|} = \frac{11\sqrt{C^3}}{18\omega_0^4} \sqrt{n+1^3}. \quad (85)$$

In this approximation,  $y_n$  cannot be specified any more closely than this.

However, we should like to write out the final equations when one makes the assumption that  $\psi_n = \varphi_n$ . These are as follows (up to terms of higher than second order in  $\lambda$ ):

$$\left. \begin{aligned} \omega(n, n-1) &= \omega_0 - \lambda^2 \frac{5C}{3\omega_0^3}n + \dots, \\ \omega(n, n-2) &= 2\omega_0 + \dots; \end{aligned} \right\} \quad (86)$$

$$\left. \begin{aligned} q(n, n) &= -\lambda \frac{C}{\omega_0^2} (2n + 1) + \dots, \\ q(n, n-1) &= \sqrt{Cn} e^{i\varphi_{n-1}} \left( 1 + \lambda^2 \frac{11Cn}{18\omega_0^4} + \dots \right), \\ q(n, n-2) &= \lambda \frac{C}{3\omega_0^2} \sqrt{n(n-1)} e^{i(\varphi_{n-1} + \varphi_{n-2})} + \dots, \\ q(n, n-3) &= \lambda^2 \frac{\sqrt{C^3}}{12\omega_0^4} \sqrt{n(n-1)(n-2)} e^{i(\varphi_{n-1} + \varphi_{n-2} + \varphi_{n-3})} + \dots \end{aligned} \right\} \quad (87)$$

We have also calculated the energy directly and derived the following formula;

$$W_n = h\nu_0 \left( n + \frac{1}{2} \right) - \lambda^2 \frac{5C^2}{3\omega_0^2} \left( n(n+1) + \frac{17}{30} \right) + \dots \quad (88)$$

The frequency condition is actually satisfied, since, remembering (82), we have

$$\begin{aligned} W_n - W_{n-1} &= h\nu_0 - \lambda^2 \frac{2C^2}{\omega_0^2} n + \dots = \frac{h}{2\pi} \omega(n, n-1), \\ W_n - W_{n-2} &= 2h\nu_0 + \dots = \frac{h}{2\pi} \omega(n, n-2). \end{aligned}$$

With the formula (88) we can associate the observation that already in terms of lowest order there occurs a discrepancy from classical theory which can formally be removed by the introduction of a “half-integer” quantum number  $n' = n + 1/2$ . This has already been remarked by Heisenberg. Incidentally, our expressions  $\omega(n, n-1)$  as given by (86) agree *exactly* with the classical frequencies in all respects. For comparison, we note the classical energy to be<sup>13</sup>

$$W_n^{(e1)} = h\nu_0 n - \lambda^2 \frac{5C^2}{3\omega_0^2} n^2 + \dots,$$

and thus the classical frequency to be:

$$\begin{aligned} \omega_{e1} &= \frac{1}{h} \frac{\partial W_n^{(e1)}}{\partial n} = h\nu_0 - \lambda^2 \frac{5C^2}{3\omega_0^2} n + \dots \\ &= \omega_{qu}(n, n-1) = \frac{1}{h} (W_n^{(qu)} - W_{n-1}^{(qu)}). \end{aligned}$$

---

<sup>13</sup>See M. Born, Atommechanik (Berlin, 1925), Chapter 4, §42, p. 294; one has to set  $a = 1/3$  in the formula (6) in order to obtain agreement with the present treatment.

We have, finally checked that the expression (88) can also be derived from the Kramers-Born perturbation formula (up to an additive constant).

**On Quantum Mechanics II**

M. Born

Received November 16, 1925

— — ◇ ◇ — —

Translation into English: *Sources of Quantum Mechanics*, Ed. by B.L. van der Waerden, North Holland, Amsterdam (1967) 321.

— — ◇ ◇ — —

**Abstract:** The quantum mechanics developed in Part 1 of this paper from Heisenberg's approach is here extended to systems having arbitrarily many degrees of freedom. Perturbation theory is carried through for nondegenerate and for a large class of degenerate systems, and its connection with the eigenvalue theory of Hermitian forms is demonstrated. The results so obtained are employed in the derivation of momentum and angular momentum conservation laws, and of selection rules and intensity formulae. Finally, the theory is applied to the statistics of eigenvibrations of a black body cavity.

**Introduction**

The present paper sets out to develop further a general quantum-theoretical mechanics whose physical and mathematical basis has been treated in two previous papers by the present authors.<sup>1</sup> It was found possible to extend

---

<sup>1</sup>W. Heisenberg, Zs. f. Phys. 33 (1925) 879.  
M. Born and P. Jordan, Zs. f. Phys. 34 (1925) 858.  
Henceforth designated as (Part) 1.

the above theory to systems having several degrees of freedom <sup>2</sup> (Chapter 2), and by the introduction of “canonical transformations” to reduce the problem of integrating the equations of motion to a known mathematical formulation. From this theory of canonical transformations we were able to derive a perturbation theory (Chapter 1, § 4) which displays close similarity to classical perturbation theory. On the other hand we were able to trace a connection between quantum mechanics and the highly-developed mathematical theory of quadratic forms of infinitely many variables (Chapter 3). Before we go on to discuss the presentation of this further development in the theory, we first endeavour to define its physical content more precisely.

The starting point of our theoretical approach was the conviction that the difficulties which have been encountered at every step in quantum theory in the last few years could be surmounted only by establishing a mathematical system for the mechanics of atomic and electronic motions, which would have a unity and simplicity comparable with the system of classical mechanics, and which would entirely consist of relations between quantities that are in principle observable. Admittedly, such a system of quantum-theoretical relations between observable quantities, when compared with the quantum theory employed hitherto, would labour under the disadvantage of not being directly amenable to a geometrically visualizable interpretation, since the motion of electrons cannot be described in terms of the familiar concepts of space and time. A characteristic feature of the new theory lies in the modification it imposes upon kinematics as well as upon mechanics; a notable advantage, however, of this quantum mechanics consists in the fact that the basic postulates of quantum theory form an inherent organic constituent of this mechanics, e.g., that the existence of discrete stationary states is just as natural a feature of the new theory as, say, the existence of discrete vibration frequencies in classical theory (cf. Chapter 3). If one reviews the fundamental differences between classical and quantum theory, differences which stem from the basic quantum theoretical postulates, then the formalism proposed in the two above-mentioned publications and in this paper, if proved to be correct, would appear to represent a system of quantum mechanics as close to that of classical theory as could reasonably be hoped. In this context we merely recall the validity of energy and momentum conservation laws and the form of the equations of motion (Chapter 1, § 2). This similarity of the new theory with classical theory also precludes

---

<sup>2</sup>A paper by P.A.M. Dirac (Proc. Roy. Soc. London 109 (1925) 642), which has appeared in the meantime, independently gives some of the results contained in Part I and the present paper, together with further new conclusions to be drawn from the theory.

any question of a separate correspondence principle outside the new theory; rather, the latter can itself be regarded as an exact formulation of Bohr's correspondence considerations. In the further development of the theory, an important task will lie in the closer investigation of the nature of this correspondence and in the description of the manner in which symbolic quantum geometry goes over into visualizable classical geometry. With regard to this question, a particularly important trait in the new theory would seem to us to consist of the way in which both continuous and line spectra arise in it on an equal footing, i.e., as solutions of one and the same equation of motion and closely connected with one another mathematically (cf. Chapter 3, § 3); obviously, in this theory, any distinction between "quantized" and "unquantized" motion ceases to be at all meaningful, since the theory contains no mention of a quantization condition which selects only certain types of motion from among a large number of possible types: rather, in place of such a condition one has a basic quantum mechanical equation (Chapter 1, § 1) which is applicable to all possible types of motion and which is essential if the dynamic problem is to be given a definite meaning at all.

Now, although we should like to be able to conclude that because of its mathematical simplicity and unity, the proposed theory might reproduce essential characteristics of the actual conditions inherent in problems of atomic structure, we nevertheless have to realize, that the theory is not yet able to furnish a solution to the principal difficulties in quantum theory. The theory has not yet incorporated the forces which in classical theory would be associated with radiation resistance, and in connection with the question of how the coupling problem is to be related to the quantum mechanics postulated here, there exist but a few indistinct indications (cf. Chapter 1, § 5). Nevertheless it would seem that these basic quantum-theoretical difficulties assume an altogether different aspect in the new theory than hitherto and that one might indeed now be more justified in hoping that these problems will in due course be solved. We consider, for instance, the question of collision processes. Recently, Bohr<sup>3</sup> called attention to the basic difficulties which (in the theory as employed hitherto) confronted all attempts to reconcile the fundamental postulates of quantum theory with the law of conservation of energy in fast collisions. In the present theory, however, the fundamental principles of quantum theory and the principle of conservation of energy follow mathematically from the quantum-mechanical equations, and hence the results of the Franck-Hertz collision studies would seem to be natural mathematical consequences of the theory. One may thus hope that a

---

<sup>3</sup>N. Bohr, Zs.f.Phys. 34 (1925) 142.

future treatment of collision problems based on the new quantum mechanics may, just because of this organic relationship between the basic postulates and this mechanics, avoid difficulties of the type mentioned above.

The question of the anomalous Zeeman effect seems to be hardly different when handled by the theory proposed here than it was before. It is true that the intimate connection between the “aperiodic” and the “periodic” orbits inherent in the basic assumptions of this theory entails the fact that we cannot be certain that Larmor’s Theorem holds generally (Chapter 4, § 2); the assumptions for the validity of the theorem are satisfied by an oscillator, but not necessarily by a nuclear atom. It is not likely, however, that this standpoint can lead to an interpretation of anomalous Zeeman effects; rather the present quantum mechanics may in the case of Zeeman effects have to content with the same difficulties as the previous theory. Recently, though, the problem of anomalous Zeeman effects has entered a new phase as a result of a Note published by Uhlenbeck and Goudsmit.

<sup>4</sup> These authors make the assumption that the electron itself possesses a mechanical and a magnetic moment (whose ratio should be twice as large as for atoms), so that there should actually be no anomalous Zeeman effects. By this assumption, difficulties as to statistical weights are eliminated and a qualitative explanation of various phenomena connected with problems of multiplet structure and Zeeman effects ensues. The question as to whether it can already furnish a quantitative explanation of these phenomena can, of course, be answered only after more rigorous investigations using the methods of quantum mechanics. Some of the results contained in Chapter 4 appear, as regards the Zeeman effects, to substantiate this hope of finding a quantitative interpretation at some later date.

Finally, we have also attempted to treat a well-known statistical problem by means of the methods furnished by the present theory. It is well known that by quantizing the vibrations of a cavity within reflecting walls and using classical methods one can arrive at results which display a certain similarity with the hypotheses in a theory of light quanta and which permit a derivation of Planck’s formula. However, as Einstein <sup>5</sup> has always stressed, this semiclassical treatment of cavity radiation yields an erroneous value for the mean square deviation of the energy in a volume element. This result must be regarded as a particularly serious objection to earlier methods in quantum theory, since we are concerned here with a breakdown of the theory even for the simple problem of a harmonic oscillator. On the other hand,

---

<sup>4</sup> G. Uhlenbeck and S. Goudsmit, Naturwiss. 13 (1925) 953.

<sup>5</sup> A. Einstein, Phys. Zs. 10 (1909) 185, 817.

the above difficulty would arise in the statistical treatment of the eigenvibrations of any mechanical system whatsoever, e.g., a crystal lattice. Now, we have found that with the kinematics and mechanics inherent in the theory presented here, the corresponding calculation leads to a correct value for the mean square deviation and also to Planck's formula, a result which may well be regarded as significant evidence in favour of the quantum mechanics put forward here.

## CHAPTER 1. SYSTEMS HAVING ONE DEGREE OF FREEDOM

### 1. Fundamental principles

I. A quantum-theoretical quantity  $a$ , whether representing a coordinate or a momentum or any function of both, is depicted by a set of quantities

$$a(nm)e^{2\pi l\nu(nm)t} \quad (1)$$

or (on leaving off the factor  $e^{2\pi l\nu(nm)t}$ ) which is the same for all quantities belonging to a given system and which depends only upon the indices  $n$  and  $m$  by the set of numbers

$$a(nm). \quad (2)$$

We can thus speak of an infinite "matrix"  $a$ .

II. Elementary operations such as addition and multiplication of quantum-theoretical quantities are defined in accordance with the operational rules of matrix calculus.

III. Consider a given function  $f(x_1, x_2, \dots, x_s)$  defined through addition and multiplication of given matrices, with  $x_1, x_2, \dots, x_s$  denoting quantum-theoretical quantities. We then introduce two types of derivatives of  $f$  with respect to one of the quantities  $x$  (say,  $x_1$ ):

(a) Differential coefficient of the first type:

$$\frac{\partial f}{\partial x_1} = \lim_{\alpha \rightarrow 0} \frac{f(x_1 + \alpha \mathbf{1}, x_2, \dots, x_s) - f(x_1, x_2, \dots, x_s)}{\alpha}, \quad (3)$$

where  $\alpha$  represents a number and  $\mathbf{1}$  the unit matrix defined by

$$\mathbf{1} = (\delta_{nm}), \quad \delta_{nm} = \begin{cases} 1 & \text{for } n = m \\ 0 & , , n \neq m. \end{cases}$$

(b) Differential coefficient of the second type: Defined through <sup>6</sup>

$$\frac{\partial f}{\partial x_1} (nm) = \frac{\partial D(f)}{\partial x_1(mn)}, \quad (4)$$

where  $D(f)$  represents the diagonal sum of the matrix  $f$ .

These two forms of differentiation will be distinguished typographically by different fraction strokes [thick stroke for (a), thin for (b)].

The treatment in Part I employed differentiation of the *second* type exclusively since this leads to a simple formulation of the variational principle of quantum mechanics and hence appears to be the more natural. However, for some calculations derivatives of the first type are more convenient to employ. It might be mentioned generally that the introduction of a differential coefficient into quantum mechanics is somewhat of an artifice and that the operations on the left-hand side of the formula (6) which follow represent the natural counterpart to differential coefficients in classical theory. For the formulation of canonical equations it is important to establish the fact that both species of differentiation (3) and (4) become identical in the case of the energy function <sup>7</sup>  $H(pq)$ .

IV. Calculations involving quantum-theoretical quantities would yield non-unique results because of the inapplicability of the commutative rule in multiplication unless the value of  $pq - qp$  were prescribed. <sup>8</sup> Hence we

---

<sup>6</sup>Cf. Part I [paper 13 in this volume].

<sup>7</sup>For the energy function  $H$  of Part I, instead of arbitrary functions such as

$$H^* = \sum a_{sr} p^s q^r,$$

only those symmetrized functions giving rise to the same Hamilton equations were permitted:

$$H = \sum a_{sr} \frac{1}{s+1} \sum_{l=0}^s p^{s-l} q^r p^l.$$

Now, for these symmetrized functions  $H$  the following relations, derived in Part I, apply:

$$\begin{aligned} \frac{\partial H}{\partial p} &= \sum a_{sr} \frac{1}{s+1} \left\{ \sum_{l=0}^{s-1} (s-l)p^{s-1-l} q^r p^l + \sum_{l=1}^s l p^{s-l} q^r p^{l-1} \right\} \\ &= \sum a_{sr} \sum_{l=0}^{s-1} p^{s-1-l} q^r p^l = \frac{\partial H}{\partial p}. \\ \frac{\partial H}{\partial q} &= \sum a_{sr} \frac{r}{s+1} \sum_{l=0}^s p^{s-l} q^{r-1} p^l - \sum a_{sr} \sum_{j=0}^{r-1} q^{r-1-j} p^s q^j = \frac{\partial H}{\partial q}. \end{aligned}$$

<sup>8</sup>The equations of motion merely indicate that this difference has to be a diagonal matrix.

introduce the following basic quantum-mechanical relation:

$$pq - qp = \frac{h}{2\pi i} \mathbf{1}. \quad (5)$$

We shall later discuss the physical significance of this relation according to the correspondence principle. At this stage it would appear important to stress that eq.(5), ch. 1, is the only one of the basic formulae in the quantum mechanics here proposed which contains Planck's constant  $h$ . It is satisfying that the constant  $h$  already enters into the basic tenets of the theory at this stage in so simple a form. Furthermore, one can see from eq. (5), ch. 1, that in the limit  $h = 0$ , the new theory would converge to classical theory, as is physically required.

A relation which will later prove important can also be derived from eq. (5), ch. 1, namely:

If  $f(pq)$  be any function of  $p$  and  $q$ , then

$$\begin{aligned} fq - qf &= \frac{\partial f}{\partial p} \frac{h}{2\pi i}, \\ pf - fp &= \frac{\partial f}{\partial q} \frac{h}{2\pi i}, \end{aligned} \quad (6)$$

since, if we assume these formulae to be valid for some given pair of functions,  $\varphi$  and  $\psi$ , then they must also hold for  $\varphi + \psi$  and  $\varphi \cdot \psi$ . The former case,  $\varphi + \psi$  is trivial; for the latter,  $\varphi \cdot \psi$ , a simple calculation yields:

$$\begin{aligned} \varphi \cdot \psi q - q\varphi\psi &= \varphi(\psi q - q\psi) + (\varphi q - q\varphi)\psi \\ &= \varphi \left( \frac{\partial \psi}{\partial p} + \frac{\partial \varphi}{\partial p} \psi \right) \frac{h}{2\pi i} = \frac{\partial(\varphi\psi)}{\partial p} \frac{h}{2\pi i}; \end{aligned}$$

for  $p\varphi\psi - \varphi\psi p$ . The treatment is similar.

Now, the relations (6) hold for  $p$  and  $q$ . They must accordingly also apply to every function  $f$  which can formally be expressed as a power series in  $p$  and  $q$ .

## 2. The canonical equations, energy conservation and frequency condition

Let an energy function  $H(pq)$  be given, together with the associated canonical equations

$$\dot{p} = -\frac{\partial H}{\partial q}; \quad \dot{q} = \frac{\partial H}{\partial p}. \quad (7)$$

It follows from the frequency combination principle

$$\nu(nm) + \nu(mk) = \nu(nk) \quad (8)$$

that  $\nu$  can be expressed in the form

$$\nu(nm) = \frac{(W_n - W_m)}{h}. \quad (9)$$

We now introduce a quantum-theoretical quantity  $W$ , as “term”, defined through

$$W(nm) = \begin{cases} W_n & \text{for } n = m \\ 0 & \text{for } n \neq m. \end{cases}$$

Thus  $W$  is a diagonal matrix.

Then for any quantum-theoretical quantity whatsoever, the following relation holds:

$$\dot{a} = \frac{2\pi i}{h} (Wa - aW). \quad (10)$$

In fact  $\dot{a}$  was (cf. Part I) defined through

$$a(nm) = 2ni\nu(nm)a(nm).$$

Among the main tenets of the theory we here seek to build up, we class the law of conservation of energy ( $H = \text{constant}$ ) and the frequency condition

$$\left( \nu(nm) = \frac{H_n - H_m}{h}; \quad H_n = W_n + \text{const} \right).$$

We carry the proof through for both these conditions by inserting eqs. (6) and (10) into eq. (7), ch. I. This yields

$$Wq - qW - Hq - qH$$

$$Wp - qW = Hp - qH \quad (11)$$

or, equivalently,

$$(W - H)q - q(W - H) = 0,$$

$$(W - H) - q(W - H) = 0.$$

The entity  $W - H$  commutes with  $p$  and  $q$ , and hence also with every function of  $p, q$ , in particular with  $H$ :

$$(W - H)H - H(W - H) = 0.$$

Thence from (10), ch. 1, one has

$$\dot{H} = 0. \quad (12)$$

Thereby the law of conservation of energy is proved, and  $H$  is established as a diagonal matrix,  $H(nm) = \delta_{nm}H_n$ .

The frequency condition now follows directly from (11), ch. 1:

$$q(nm)(H_n - H_m) = q(nm)(W_n - W_m), \quad (13)$$

i.e.,

$$\frac{(H_n - H_m)}{h} = \nu(nm) \quad (14)$$

Thus far, we have proved energy-conservation and the frequency condition from the canonical equations and the basic equation (5), ch. 1. In corollary, we can, however, also invert the proof. We know energy conservation and the frequency condition to be correct. Hence if the energy function  $H$  be given as an analytical function of any variables  $P, Q$  then, provided that

$$PQ - QP = \frac{h}{2\pi i} \mathbf{1},$$

the following canonical equations always apply:

$$\dot{Q} = \frac{\partial H}{\partial P}, \quad P = -\frac{\partial H}{\partial Q}. \quad (15)$$

This follows directly from the fact that the quantities  $PH - HP$  or  $HQ - QH$  can be interpreted in a twofold manner, namely according to (6), ch. 1 and according to (10), ch. 1.

### 3. Canonical transformations

By a “canonical transformation” of the variables  $p, q$  into new variables  $P, Q$ , we understand a transformation in which

$$pq - qp = PQ - QP = \frac{h}{2\pi i}, \quad (16)$$

as is suggested by the preceding considerations, since then the same canonical equations (7), ch. 1, or (15), ch. 1, apply to  $P, Q$  as to  $p, q$ . A general transformation which satisfies this condition is

$$\begin{aligned} P &= SpS^{-1} \\ Q &= SqS^{-1}, \end{aligned} \quad (17)$$

wherein  $S$  stands for an arbitrary quantum-theoretical quantity. We would surmise that eq. (17), ch. 1, represents in fact the most general canonical transformation. The transformation (17), ch. 1, also has the simple property that for any function  $f(P, Q)$  it follows that

$$f(P, Q) = Sf(p, q)S^{-1}, \quad (18)$$

wherein  $f(p, q)$  is formed from  $f(P, Q)$  on replacing  $P$  by  $p$  and  $Q$  by  $q$ , retaining the functional form. The proof of this contention for functions in the sense of our above definition follows directly from the observation that the rule holds for sum and product with sum terms or factors  $p, q$ .

The importance of the canonical transformation is due to the following theorem: If any pair of values  $p_0, q_0$  be given which satisfy eq. (15), ch. 1, then the problem of integrating the canonical equations for an energy function  $H(pq)$  can be reduced to the following: A function  $S$  is to be determined, such that when

$$p = Sp_0S^{-1}, \quad q = Sq_0S^{-1} \quad (19)$$

the function

$$H(pq) = SH(p_0q_0)S^{-1} = W \quad (20)$$

becomes a diagonal matrix. Equation (20), ch. 1, is the analogue to the Hamilton partial differential equation, and in a sense stands for the action function.

#### 4. Perturbation theory

We consider a given mechanical problem defined by the energy function

$$H = H_0(pq) + \lambda H_1(pq) + \lambda^2 H_2(pq) + \dots \quad (21)$$

and assume the mechanical problem defined by the energy function  $H_0(pq)$  to be solved. Thus solutions  $p_0, q_0$  of this problem are known; they satisfy the condition  $p_0q_0 - q_0p_0 = (h/2\pi i)\mathbf{1}$  and cause  $H_0(p_0q_0) = W_0$  to be a diagonal matrix. We then seek a transformation function  $S$  such that

$$p = Sp_0S^{-1}, \quad q = Sq_0S^{-1}, \quad (22)$$

and that

$$H(pq) = SH(p_0q_0)S^{-1} = W,$$

e.g., that the matrix  $H$  becomes diagonalized. To arrive at a solution we try setting

$$S = \mathbf{1} + \lambda S_1 + \lambda S_2 + \dots . \quad (23)$$

Then

$$S^{-1} = \mathbf{1} - \lambda S_1 + \lambda^2 (S_1^2 - S_2) + \lambda^3 \dots . \quad (24)$$

If for  $H$  we take the expression (21), ch. 1, we can collect together powers of  $\lambda$  to obtain the following equations of approximation:

$$\begin{aligned} H_0(p_0 q_0) &= W_0 \\ S_1 H_0 - H_0 S_1 + H_1 &= W_1 \\ S_2 H_0 - H_0 S_2 + H_0 S_1^2 - S_1 H_0 S_1 + S_1 H_1 - H_1 S_1 + H_2 &= W_2 \\ \dots &\dots \\ S_r H_0 - H_0 S_r + F_r(H_0, \dots, H_r, S_0, \dots, S_{r-1}) &= W_r \end{aligned} \quad (25)$$

where  $H_0, H_1, \dots$  are throughout to be taken as having arguments  $p_0, q_0$

The first of the eqs. (25), ch. 1, is already satisfied. The others can be resolved in sequence, actually in just the same manner as in classical theory, namely by first building the mean value in order to determine the energy constant, after which the solution can straightway be written down:

$$W_r = \bar{F}_r, \quad (26)$$

$$S_r(mn) = \frac{F_r(mn)}{\hbar \nu_0(mn)} (1 - \delta_{nm}),$$

where  $\nu_0(nm)$  are the frequencies of the unperturbed motion. This solution satisfies the condition

$$S \cdot \tilde{S}^* = \mathbf{1}, \quad (27)$$

wherein the tilde represents interchange of rows and columns (transposition) and the star denotes that we take the complex conjugate quantity. Since we shall later return to this condition from a more general standpoint we confine ourselves at this stage merely to verifying it to the first order of approximation, which we shall evaluate right away. To this order, the relation runs

$$S_1 + \tilde{S}_1^* = 0. \quad (28)$$

The significance of eq. (27), ch. 1, lies in the fact that the Hermitian character of the matrices  $p, q$  follows from it, since use of (22), ch. 1, shows<sup>9</sup>

---

<sup>9</sup>On noting the rule  $(\tilde{a}\tilde{b}) = \tilde{b}\tilde{a}$ .

that

$$q^* = S^* q_0^* S^{*-1} = \tilde{S}^{-1} \tilde{q}_0 \tilde{S} = \tilde{q},$$

and analogously for  $p$ .

To first approximation it follows from (26), ch. 1, as also classically, that

$$W_1 = \overline{H}_1, \quad (29)$$

so that

$$S_1(mn) = \frac{H_1(mn)}{h\nu_0(mn)} (1 - \delta_{mn}). \quad (30)$$

This expression indeed satisfies the requirements (28), ch. 1, because  $H_1$  is assumed to be a Hermitian form. We can now evaluate the energy to the second order of approximation and find

$$W_2 = \overline{H}_2 + \frac{1}{h} \sum_l' \frac{H_1(nl)H_1(ln)}{\nu_0(nl)}, \quad (31)$$

where the prime on the summation indicates that terms having a vanishing denominator ( $l = n$ ) are to be excluded.

One can progress in this way and successively determine all terms of the  $W$  and  $S$  series. If we substitute the  $S$  series in (22), ch. 1, we obtain the expansions

$$q = q_0 + \lambda q_1 + \lambda^2 q_2 + \dots,$$

$$p = p_0 + \lambda p_1 + \lambda^2 p_2 + \dots$$

with known coefficients. Thus, for example, the first-order approximation runs

$$q_1 = S_1 q_0 - q_0 S_1,$$

$$p_1 = S_1 p_0 - p_0 S_1;$$

or, explicitly,

$$\begin{aligned} q_1(mn) &= \frac{1}{h} \sum_k' \left( \frac{H_1(mk)q_0(kn)}{\nu_0(mk)} - \frac{q_0(mk)H_1(kn)}{\nu_0(kn)} \right) \\ p_1(mn) &= \frac{1}{h} \sum_k' \left( \frac{H_1(mk)p_0(kn)}{\nu_0(mk)} - \frac{p_0(mk)H_1(kn)}{\nu_0(kn)} \right) \end{aligned} \quad (32)$$

The formulae (32), ch. 1, represent the outcome of Kramers' dispersion theory<sup>10</sup> in the limit of an infinitely low-frequency external field; this possibility

---

<sup>10</sup>H. A. Kramers, Nature 113 (1924) 673; 114 (1924) 310; cf. also R. Ladenburg. Zs. f. Plphys. 4 (1921) 451; R. Ladenburg and F. Reiche, Naturwiss. 11 (1923) 584.

of attaining a simple derivation of formulae otherwise obtained only on the basis of correspondence considerations seems to provide a strong argument in favour of the theory put forward here. Born<sup>11</sup> has derived eq. (31), ch. 1, on reinterpreting the respective classical formulae. The terms with  $m = n$  in eq. (32), ch. 1, correspond to Kramers' formula for normal dispersed light and the remaining terms ( $m \neq n$ ) correspond to the formulae of Kramers and Heisenberg<sup>12</sup> for "scattered light of combination frequencies". The latter expressions were used by Pauli<sup>13</sup> to evaluate the intensities of transitions in Hg which take place in presence of external electric fields and which would otherwise be "forbidden". In order to derive the general dispersion formulae (if the frequency of the external field does not vanish), one needs more general considerations regarding the action of external fields which change in function of time. We now pass over to such considerations.

## 5. Systems for which time-variables enter explicitly into the "energy function"

Treatment of the quantum-mechanical influence of external forces which explicitly depend upon time seems to us to be of especial interest in that therein some characteristic differences crop up between classical and quantum mechanics. The problem of the action of time-dependent external forces can be regarded as a limiting case of the interaction between two systems in which the influence of the interaction on one of the two systems (termed system  $A$ ) is so small that the action upon the other system (system  $B$ ) remains unaffected by this influence. If we now consider the coupling of two systems  $A, B$  from the standpoint of quantum mechanics, the Hamilton function decomposes into three parts,  $H_A$ ,  $\lambda H_B$ , and  $\varepsilon \lambda H_{AB}$  (with  $\lambda$  at this stage an arbitrary parameter and  $\varepsilon$  a small quantity). We take system  $A$  be known. For calculating the motion of  $B$  according to classical theory it suffices to establish the equations of motion [from the Hamilton function  $\lambda(H_B + \varepsilon H_{AB})$  for the coordinates of  $B$ , whereby for the coordinates of  $A$  one substitutes their solutions in function of time (for the definite given values of the constants in  $A$ ). By this means, apart from the constants of  $A$  only the time enters as a new variable into the perturbation problem for  $B$  then the reaction is neglected. In the quantum-mechanical calculation the situation is just the same, providing we restrict ourselves to first-order per-

---

<sup>11</sup>M. Born, Zs. f. Phys. 26 (1924) 379.

<sup>12</sup>H.A. Kramers and W. Heisenberg, Zs. f. Phys. 31 (1925) 681.

<sup>13</sup>W. Pauli, Verh. d. Dän. Acad. d. Wiss (in press).

turbations (i.e., terms proportional to  $\varepsilon$  in the coordinates and momenta of the system  $B$ ). It is altogether otherwise, however, for higher-order perturbations, since in the evaluation of higher-order perturbations we encounter products of quantities in which more than one implicitly contains the coordinates of  $A$ . But this means that according to the quantum-mechanical rule for building a product it by no means suffices to know the “external forces in function of time” merely for the *given* values of the constants in  $A$ , but these external forces must be known for *all* values of the constants. Thereby, however, the concept of external forces appears in fact to become devoid of meaning. This difficulty seems to us to be overcome on observing that the reaction itself gives rise to terms of order  $\lambda\varepsilon^2$  in the coordinates of  $B$ , and thus that simultaneous neglect of the reaction and evaluation of terms in  $B$  containing  $\varepsilon^2$  is meaningful only if  $\lambda$  can also be taken to be very small, i.e., physically, if variation of the quantities in  $A$  by amounts of the same order as the associated quantities in  $B$  does not bring about any perceptible change in the influence of  $A$  upon  $B$ . However, in this approximation the quantum-mechanical construction of products and thereby the calculation of the perturbations to higher orders in  $\varepsilon$  can again be effected. In fact, the rules for this building of products reduce simply to those of classical multiplication, as in this approximation the coordinates, amplitudes and frequencies which enter into  $H_{AB}$  do not depend on the constants in  $A$ . In this sense one could, for example, treat the action of a strong alternating electromagnetic field on an atom entirely as the influence of an “external force” with neglect of the reaction, since the field energy can be regarded as infinitely large compared with that of the atom. The action of  $\alpha$ -particles upon the electrons of an atom could also be regarded as an “external force”, as in classical theory, because of the relatively large energy of the  $\alpha$ -particles, so that in this approximation the Fourier expansion of the force thereby exerted upon the electrons would also be that of classical theory. However, the action of forces due to one atom upon another can never be treated as an operation of external forces - i.e., it can thus be regarded only in the first-order terms, for which such an approach is always possible – since the neglect of the reaction would in the higher-order terms lead to false results.

We can summarize the outcome of our considerations thus: It is meaningful under certain assumptions in quantum as in classical theory to speak of the action of time-dependent forces upon an atom. In such instances, the classical calculation rules can be applied to expressions in which the time parameter figures explicitly: e.g., if the external field of force be periodic

with a period  $\nu_0$ , then the general term of a coordinate  $q$  can be written as

$$q(mn, \tau) e^{2\pi \mathbf{1}[\nu(mn) + \tau\nu_0]l} \quad (33)$$

and the general term of  $q^2$  as

$$\sum_{k,\tau'} q(mk, \tau - \tau') q(kn, \tau') e^{2\pi \mathbf{1}[\nu(mn) + \tau\nu_0]l}. \quad (34)$$

For this reason the case of external forces which vary with time seems in our view to provide a striking illustration of the transition from theoretical quantum kinematics into classical kinematics according to the principle of correspondence.

If one is concerned with the evaluation of the operation of external forces to first order only, the results which ensue from the calculations which follow remain correct even if the assumptions listed at the outset are not obeyed - in exact analogy with the situation in classical theory.

From the preceding considerations it follows that the mathematical treatment of systems in which (provided the assumptions mentioned above are valid) time enters explicitly is simply to be handled in a manner analogous to the corresponding classical procedures. If we again assume the external force to be periodic in time, with period  $\nu_0$ , the Hamilton function becomes <sup>14</sup>

$$H = H(p_k, q_k, \cos 2\pi\nu_0 t). \quad (35)$$

We then introduce a new degree of freedom with the variables  $q', p'$  and take the following as the Hamiltonian of the new problem, in which time no longer figures explicitly:

$$H' = H(p_k, q_k; q') + 2\pi\nu_0 \sqrt{1 - q'^2} p'. \quad (36)$$

Thereby the canonical equations for  $p_k, q_k$  remain as hitherto, except that  $q'$  is throughout written for  $\cos 2\pi\nu_0 t$ . The new equations are:

$$\begin{aligned} \dot{q}' &= \frac{\partial H'}{\partial p'} = 2\pi\nu_0 \sqrt{1 - q'^2}, \\ \dot{p}' &= \frac{\partial H'}{\partial q'} = -\frac{\partial H}{\partial q'} + 2\pi\nu_0 \frac{q'}{\sqrt{1 - q'^2}} p'. \end{aligned} \quad (37)$$

The first of these equations asserts that  $q'$  indeed becomes equal to  $\cos 2\pi\nu_0 t$  (up to an arbitrary choice of origin in the time scale), so that the canonical

---

<sup>14</sup>Here we anticipate for a moment in availing ourselves of results derived in the next chapter for systems having several degrees of freedom.

equations for  $p_k, q_k$  take on the same form as in the earlier problem; the second equation (37), ch. 1, provides a determination of  $p'$ . Thus through (36), ch. 1, the problem (35), ch. 1, is really led back to cases already treated.

Of paramount interest is the question as to the manner in which the perturbation formulae (25), ch. 1, have to be modified if time enters explicitly into  $H_1, H_2, \dots$  but *not* into  $H_0$ . Simple considerations show that for this case the perturbation formulae ensue from those cited earlier on replacing every term of the form  $H_0 S_r - S_r H_0$  by

$$H_0 S_r - S_r H_0 + \frac{h}{2\pi i} \frac{\partial S_r}{\partial t}$$

(note that  $H_0$  occurs only in such combinations). Thus the lowest orders of the new perturbation formulae run:

$$\begin{aligned} H_0(p_0 q_0) &= W_0, \\ S_1 H_0 - H_0 S_1 - \frac{h}{2\pi i} \frac{\partial S_1}{\partial t} + H_1 &= W_1, \\ S_2 H_0 - H_0 S_2 - \frac{h}{2\pi i} \frac{\partial S_2}{\partial t} + \left( H_0 S_1 - S_1 H_0 + \frac{h}{2\pi i} \frac{\partial S_1}{\partial t} \right) S_1 &= W_2, \\ \dots &\dots \end{aligned} \tag{38}$$

We should like to assume that even if the assumption that the external forces are periodic in time does not apply, these formulae (38), ch. 1, nevertheless remain valid – even though this assumption was incorporated into the derivation of the formulae.

The first-order equations in the formulae (38), ch. 1, which of course remain correct even if the assumptions regarding “external forces” are no longer valid, taken together with eqs. (22), ch. 1, viz.

$$q = q_0 + \lambda(S_1 q_0 - q_0 S_1),$$

$$p = p_0 + \lambda(S_1 p_0 - p_0 S_1),$$

furnish an answer to problems of dispersion theory in a general sense. In actual fact, if we set:

$$H_1 = E e q_0 \cos 2\pi\nu_0 t,$$

then

$$H_1(mn, 1) = \frac{Ee}{2} q_0(mn), \quad H_1(mn, -1) = \frac{Ee}{2} q_0(mn),$$

$$S_1(mn, 1) = \frac{Ee}{2h} \frac{q_0(mn)}{\nu_0(mn) + \nu_0}, \quad (39)$$

$$S_1(mn, -1) = \frac{Ee}{2h} \frac{q_0(mn)}{\nu_0(mn) - \nu_0}.$$

Thence follows (cf. (22), ch. 1):

$$q_1(mn, +1) = \frac{Ee}{2h} \sum_k \left( \frac{q_0(mk)q_0(kn)}{\nu_0(mk) + \nu_0} - \frac{q_0(mk)q_0(kn)}{\nu_0(kn) + \nu_0} \right). \quad (40)$$

If we assume that we have Cartesian coordinates, i.e.,  $p = m\dot{q}$ , then

$$q_1(mn, 1) = \frac{Ee}{2h \cdot 2\pi im} \sum_k \frac{q_0(mk)p_0(kn) - p_0(mk)q_0(kn)}{(\nu_0(mk) + \nu_0)(\nu_0(kn) + \nu_0)}; \quad (41)$$

and similarly

$$q_1(mn, -1) = \frac{Ee}{2h \cdot 2\pi im} \sum_k \frac{q_0(mk)p_0(kn) - p_0(mk)q_0(kn)}{(\nu_0(mk) - \nu_0)(\nu_0(kn) - \nu_0)}. \quad (42)$$

The eqs. (40), (41), (42), ch. 1 agree with the formulae obtained from Kramers' dispersion theory.<sup>15</sup> A further particularly interesting case would seem to be that for incident light of very high frequency,  $|\nu_0| \gg |\nu_0(mk)|$  or  $|\nu_0(kn)|$ . Then to first-order approximation one finds

$$q_1 = -\frac{Ee}{h2\pi i\nu_0^2 m} (p_0 q_0 - q_0 p_0) \cos 2\pi\nu_0 t,$$

or, because of (5), ch. 1,

$$q_1 = +\frac{Ee}{4\pi^2 m \nu_0^2} \cos 2\pi\nu_0 t. \quad (43)$$

This finding indicates that in fact the quantum-mechanical commutation relation (5), ch. 1, ultimately entails the fact that for sufficiently high frequencies the electron behaves on scattering like a free electron. The scattered light of frequency  $\nu_0(mn) + \nu_0(m \neq n)$  vanishes and that of frequency  $\nu_0$  has the intensity to be expected for scattering by a free electron.<sup>16</sup>

## CHAPTER 2. FUNDAMENTALS OF THE THEORY FOR SYSTEMS HAVING AN ARBITRARY NUMBER OF DEGREES OF FREEDOM

---

<sup>15</sup>Cf. the discussion at the end of § 4 of results obtained for  $\nu_0 = 0$ .

<sup>16</sup>Cf. the articles by W. Kuhn, Zs. f. Phys. 33 (1925) 408; W. Thomas. Naturwiss. 13 (1925) 627; F. Reiche and W. Thomas, Zs. f. Phys. 34 (1925) 510.

## 1. The canonical equations of motion; perturbation theory for nondegenerate systems

For several degrees of freedom ( $f > 1$ ) it rather suggests itself that we replace the representation of quantum-theoretical quantities by two-dimensional matrices by one in terms of  $2f$ -dimensional matrices, corresponding with the  $2f$ -dimensional manifold of stationary states in the classical  $j$ -space:

$$\begin{aligned} q_k &= (q_k(n_1 \cdots n_f, m_1 \cdots m_f)), \\ p_k &= (p_k(n_1 \cdots n_f, m_1 \cdots m_f)). \end{aligned} \quad (1)$$

Nevertheless this representation, albeit under certain circumstance very convenient and clear, is by no means essential. Even for several degrees of freedom the fundamental dynamical equations assume the form of *matrix equations*, but these matrices can as heretofore also be written in two-dimensional form. It became apparent even for one degree of freedom that the *sequence* of the stationary states as given by the ordering of the matrix rows is (in contradistinction to the theory employed hitherto) purely fortuitous and is not governed by any intrinsic property of the system. This observation can now directly be referred to many-imensional matrices too; one can carry out any arbitrary rearrangements and in particular transform the  $2f$ -dimensional matrices into two-imensional ones. This is justified by the fact that the basic definitions of addition and multiplication, as also of differentiation with respect to time, are clearly independent of any *ordering relations* between the basis systems of indicates  $n_1, n_2, \dots, n_f$ , which taken singly specify the states and in pairs specify the transitions.

It is thence also clear that the general rules of matrix analysis, as presented in chapter 1 of Part I and in chapter 1 of this present paper, can be employed in the theory of systems having *several* degrees of freedom also. One can similarly take over the derivation of the equation of motion from the variational principle in I directly, so that we can in like manner write

$$\dot{q}_k = \frac{\partial H}{\partial p_k}; \quad \dot{p}_k = \frac{\partial H}{\partial q_k}. \quad (2)$$

The principal new feature distinguishable from those obtaining for systems with just one degree of freedom lies in the general commutation relations for  $p_k$  and  $q_k$  in the case of several degrees of freedom. Just as in the calculations for but one degree of freedom, so here also calculations with quantum-theoretical quantities would be to some extent indefinite if the “commutation relations” were not specified.

As a plausible generalization of eqs. (5), ch. 1, the following equations suggest themselves:

$$\begin{aligned} p_k q_l - q_l p_k &= \frac{\hbar}{2\pi i} \delta_{kl}, \\ p_k p_l - p_l p_k &= 0, \end{aligned} \tag{3}$$

$$q_k q_l - q_l q_k = 0,$$

if  $H$  denotes the (symmetrized) energy function, one can in consequence of these relations replace eqs. (2), ch. 2, by

$$\dot{q}_k = \frac{\partial H}{\partial p_k}, \quad \dot{p}_k = -\frac{\partial H}{\partial q_k}. \tag{2'}$$

Further, it follows from these relations,<sup>17</sup> as in chapter 1 of I the present paper, that

$$\begin{aligned} p_k f(q_1 \cdots q_f, p_1 \cdots p_f) - f p_k &= \frac{\hbar}{2\pi i} \frac{\partial f}{\partial q_k}, \\ f q_k - q_k f &= \frac{\hbar}{2\pi i} \frac{\partial f}{\partial p_k} \end{aligned} \tag{4}$$

The proof of energy conservation and the frequency condition then follows from (2') and (4), ch. 2, as shown in ch. 1. Similarly one can show with the aid of (3) and (4) that the canonical equations (2'), ch. 2, apply whenever the relations (3), ch. 2, are satisfied for a system  $P_k, Q_k$  and the energy function is given as an analytical function of the  $P_k$  and  $Q_k$ .

Thus a transformation of the variables  $p_k, q_k$  into new variables  $P_k, Q_k$  is termed “canonical” if it leaves the relations (3), ch. 2, unaltered.

A very general class of such transformations is again given by the formulae

$$\begin{aligned} P_k &= S p_k S^{-1}, \\ Q_k &= S q_k S^{-1}. \end{aligned} \tag{5}$$

This transformation again has the property of converting every function  $f(PQ)$  into

$$f(P_1, \dots, Q_1, \dots, Q_f) = S f(p_1, \dots, p_f, q_1, \dots, q_f) S^{-1}. \tag{6}$$

If a system  $p_1^0, \dots, p_f^0, q_1^0, \dots, q_f^0$  is known, and satisfies the relations (3), ch. 2, then the problem of integrating eqs. (2), ch. 2, again reduces itself to

---

<sup>17</sup>The physical significance of these relations for dispersion theory is discussed by H. A. Kramers, Physika, December 1925.

the simpler problem: A function  $S$  is to be sought, such that it satisfies the equations

$$\begin{aligned} p_k &= Sp_k^0 S^{-1}, \\ q_k &= Sq_k^0 S^{-1} \end{aligned} \quad (5a)$$

and transforms  $H$  into a diagonal matrix,

$$H(pq) = SH(p^0 q^0)S^{-1} = W. \quad (7)$$

Equation (7) again represents the counterpart to the Hamilton partial differential equation.

Equations (3), ch. 2 would, together with (2), ch. 2, obviously entail too extensive a set of requirements for the  $p_k, q_k$ , if all these equations were *independed* of one another. As an interesting mathematical problem must rank (the derivation of eqs. (3) using the least number of independent and mutually consistent assumptions; nevertheless, this question will not be handled here. We shall content ourselves with mentioning that

$$\frac{d}{dt} \sum_k (p_k q_k - q_k p_k) = 0$$

is a general outcome of the equations of motion (1), ch. 2. On the other hand, it will be shown generally that the eqs. (3), ch. 2, together with the equations of motion (2), ch. 2, or the equivalent requirement (7), ch. 2, *can* be satisfied (singular discrepancies apart, of course).

This proof is to be supplied in connection with the generalization of the perturbation theory presented in ch. 1 § 4, when extended to arbitrarily many degrees of freedom. We consider the energy function  $H(pq)$  such that it can be written as

$$H = H_0(pq) + \lambda H_1(pq) + \lambda^2 H_2(pq) + \dots, \quad (8)$$

so that

$$H_0(pq) = \sum_{k=1}^f H^{(k)}(p_k q_k).$$

Thus for  $\lambda = 0$  we have  $f$  uncoupled systems, each having a single degree of freedom; the/cases

$$H = H^{(k)}(p_k q_k)$$

can be solved with

$$q_k = q_k^0, \quad q_k = p_k^0,$$

wherein  $q_k^0, p_k^0$  are *two-dimensional* matrices,

$$q_k^0 : (q_k^0(nm)) ; \quad p_k^0 : (p_k^0(nm)). \quad (10)$$

if we formally regard these / uncoupled systems as a single system having  $f$  degrees of freedom, then  $q_k^0, p_k^0$  would be represented as  $2f$ -dimensional matrices,

$$\left. \begin{aligned} q_k^0 &= (q_k^0(n_1 \cdots n_f; m_1 \cdots m_f)), \\ p_k^0 &= (p_k^0(n_1 \cdots n_f; m_1 \cdots m_f)), \end{aligned} \right\} \quad (11)$$

for which

$$\begin{aligned} q_k^0(n_1 \cdots n_f; m_1 \cdots m_f) &= \delta_k q_k^0(n_k m_k), \\ p_k^0(n_1 \cdots n_f; m_1 \cdots m_f) &= \delta_k p_k^0(n_k m_k), \end{aligned}$$

where  $\delta_k = 1$  if  $n_j = m_j$  for all  $j$  except  $j = k$  and  $\delta_k = 0$  if for any  $j (j \neq k)$ ,  $n_j$  is not equal to  $m_j$ . Thence, however, one sees: firstly, that the equations

$$p_k^0 q_k^0 - q_k^0 p_k^0 = \frac{\hbar}{2\pi i} \mathbf{1} \quad (12)$$

which originally obtained for the *two-dimensional* matrices (10), ch. 2, also hold for the  $2f$ -dimensional matrices (11), ch. 2; secondly, that the following relations ensue:

$$\begin{aligned} p_k^0 q_l^0 - q_l^0 p_k^0 &= 0 \quad \text{for } l \neq k, \\ p_k^0 p_l^0 - p_l^0 p_k^0 &= q_k^0 p_l^0 - q_l^0 q_k^0 = 0. \end{aligned} \quad (13)$$

Hence for  $\lambda = 0$  the eqs. (13), ch 2, indeed apply. It is to be shown that  $p, q$  can be determined in such a manner that (3), ch. 2, is satisfied simultaneously with  $H = W$  for higher-order approximations also. One again assumes the system  $H_0$  to have been chosen as *nondegenerate*, i.e., that on substituting  $q = q^0, p = p^0$  no two diagonal elements of  $H_0$  become identical. In this case we again have to set

$$q_k = S q_k^0 S^{-1}; \quad p_k = S p_k^0 S^{-1} \quad (14)$$

as in eq. (5a), ch. 2, and to determine

$$S = \mathbf{1} + \lambda S_1 + \lambda^2 S_2 + \dots$$

in such a way as to satisfy the relation  $H = W$ . The eqs. (3), ch. 2, are then jointly also satisfied, since by virtue of (14) they go over into (12), (13). This completes the required proof.

Equations (3) are invariant with respect to a linear orthogonal transformation of the  $q_k$  and  $p_k$ , for if one sets

$$\begin{aligned} q'_k &= \sum_l a_{kl} q_l, \\ p'_k &= \sum_l a_{kl} p_l, \\ \sum_l a_{kl} a_{lj} &= \delta_{kj}, \end{aligned}$$

then

$$p'_k q'_l - q'_l p'_k = \sum_{hf} a_{kh} a_{lj} (p_h q_j - q_j p_h) = \delta_{kl} \frac{h}{2\pi i}$$

and similarly for the other respective relations. If then the conditions (3), ch. 2, hold for a given Cartesian coordinate system, they will also be valid in every other Cartesian coordinate system.

By way of supplement, now that we have established (3), ch. 2, we demonstrate that a well-known law of classical mechanics is also compatible with the new theory.

Let

$$H = E_{\text{kin}} + E_{\text{pot}} = \frac{1}{2} \sum_k \frac{p_k^2}{m_k} + E_{\text{pot}}, \quad (15)$$

and let  $E_{\text{pot}}$  be a homogeneous function of the coordinates of order  $n$ . Then from (3), ch. 2,

$$E_{\text{pot}} = \frac{1}{n} \sum_k \frac{\partial E_{\text{pot}}}{\partial q_k} q_k \quad (16)$$

and

$$\frac{d}{dt} \sum_k p_k q_k = \sum_k (\dot{p}_k q_k + p_k \dot{q}_k) = 2E_{\text{kin}} - nE_{\text{pot}},$$

so that for the *mean* values,

$$\bar{E}_{\text{kin}} = \frac{1}{2} n \bar{E}_{\text{pot}}, \quad (17)$$

Hence, e.g., for  $n = 2$  (harmonic oscillations),  $\bar{E}_{\text{kin}} = \bar{E}_{\text{pot}}$  and for  $n = -1$  (Coulomb force),  $\bar{E}_{\text{kin}} = -\frac{1}{2} \bar{E}_{\text{pot}}$ .

## 2. Degenerate systems

We now turn to examination of degenerate systems. If we permit some of the frequencies  $\nu(nm)$  to vanish (or simplicity, we imagine the matrices to be in

two-dimensional representation), then *energy conservation*,  $\dot{H} = 0$  can still be derived from the considerations employed here and in Part I concerning the equations of motion and the commutation rules (3), ch. 2. But the relation  $\dot{H} = 0$  no longer necessarily implies that  $H$  be a diagonal matrix and in consequence the proof of the frequency condition cannot be carried through. Thus for degenerate systems the equations of motion together with (3), ch. 2, do not alone suffice for the unique determination of the properties of a system: we need to strengthen these basic equations. An obvious assumption as to the form of this “increase in rigour” is:

*For basic equations, one should be able generally to choose the commutation relations and the property*

$$H = W = \text{diagonal matrix.} \quad (18)$$

This requirement manifestly ensures the validity of the frequency condition for degenerate systems as well. Very probably, the energy  $W$  is also thereby uniquely determined (apart from singular instances). On the other hand, the *coordinates*  $q_k$  are *not* uniquely determined. Given a solution  $p_k, q_k$  of  $H(pq) = W$ , we can get new solutions from

$$\begin{aligned} p' &= SpS^{-1}, \\ q' &= SqS^{-1}. \end{aligned} \quad (19)$$

Thence

$$H(p'q') = W' = SWS^{-1},$$

and the requirement  $W' = W$  yields

$$WS - SW = \dot{S} \frac{h}{2\pi i} = 0,$$

and thus

$$S = \text{constant.} \quad (20)$$

Let us at this stage examine this result as regards its implications for non-degenerate systems. From (2), ch. 2, the matrix  $S$  has to become a diagonal matrix, and the eqs. (19), ch. 2, imply that

$$\begin{aligned} p'(nm) &= p(nm)S_nS_m^{-1}, \\ q'(nm) &= q(nm)S_nS_m^{-1}, \end{aligned} \quad (19')$$

writing  $S_n$  for  $S_{(nn)}$  for the sake of conciseness.

The uncertainty in the solution indicated hereby can significantly be reduced by the requirement that the new solution  $p', q'$  should also represent “real” motion, expressed in terms of Hermitian matrices, since this yields also represent “real” motion, expressed in terms of Hermitian matrices, since this yields

$$|S_n S_n^{-1}| = |S_m S_m^{-1}|,$$

or

$$|S_n| = |S_m|. \quad (21)$$

Thus the indeterminacy which has here come to light represents an arbitrariness of the *phase constants*. We namely here find proof of the contention put forward in Part I that in each problem for every state  $n$  a phase  $\varphi_n$  always remains undetermined. From (19') one can perceive the manner in which these phases enter into the elements of the matrices  $p, q$ . It was further conjectured in Part I that apart from the above-mentioned arbitrariness of phase for non-degenerate systems, no additional non-uniqueness is to be expected. It is clear that we could still add a constant matrix to each of the “periodic” matrices  $S_n$  in the perturbation calculations of ch. 1, § 4. However, this obviously does not imply that new phases which remain undetermined enter into each approximation. It is easy to see that utilization of this possibility cannot provide any more general solution  $p, q$  provided that  $p^0, q^0$  were right from the first taken to have undetermined phases.

If we now go to degenerate systems, we cannot any longer infer from (20) that  $S$  is a diagonal matrix, and accordingly, using (19), we do indeed have the possibility of deriving solutions  $p', q'$  which are significantly different from  $p, q$ . This indeterminacy seems to lie in the very nature of things. Apparently, degenerate systems possess a *lability* by virtue of which arbitrarily small perturbations can bring about finite changes in coordinates, and this finds its mathematical expression in that in complete absence of perturbations, the solution of the dynamic equations remains partly indeterminate. Naturally, for every actual atom the coordinates which specify the physical properties of the system, in particular the transition probabilities, are always fixed uniquely either by external perturbations or by the previous history of the system.

Now we set out to examine the influence of arbitrary perturbations upon the degenerate system. We set

$$H(pq) = H_0 + \lambda H_1 + \lambda^2 H_2 + \dots, \quad (22)$$

and let  $p^0, q^0$  be an arbitrary, but definite, solution of the unperturbed

## problem

$$H_0(p^0 q^0) = W_0. \quad (23)$$

Then with

$$p = Sp^0S^{-1},$$

$$q = Sq^0S^{-1},$$

and with

$$S = S_0(1 + \lambda S_1 + \lambda^2 S_2 + \dots), \quad (24)$$

$$S^{-1} = (\mathbf{1} - \lambda(S_1 + \lambda S_2 \dots) + \lambda^2 \dots) S_0^{-1}, \quad (25)$$

We find, on leaving out the arguments  $p^0, q^0$  from  $H_0, H_1, \dots$ :

$$S_0 H_0 S_0^{-1} = W_0, \quad (26)$$

$$S_0 S_1 H_0 S_0^{-1} - S_0 H_0 S_1 S_0^{-1} + S_0 H_1 S_0^{-1} = W_1, \quad (27)$$

$$S_0 S_2 H_0 S_0^{-1} - S_0 H_0 S_2 S_0^{-1} + S_0 F_2(H_0 H_1 H_2; S_1) S_0^{-1} = W_2, \quad (28)$$

..... (29)

.....

$$S_0 S_r H_0 S_0^{-1} - S_0 H_0 S_r S_0^{-1} + S_0 F_r (H_0 H_1 \cdots H_r, S_1 \cdots S_{r-1}) S_0^{-1} = W_r. \quad (30)$$

Thus we almost repeat eqs. (26), ch. 1, but with the difference that the left-hand sides are throughout multiplied on the left by  $S_0$  and on the right by  $S_{-1}^0$ .

Equation (26), ch. 2, has already been cited above;  $S_0(nm)$  becomes zero except for vanishing  $\nu_0(nm)$ . The remaining arbitrariness in  $S_0$  now has to be used to advantage so far as possible in order to render the next equation soluble. Naturally, one cannot expect that every solution of  $H = H_0$ , and thus in particular the chosen solution  $p^0, q^0$ , will provide the limiting case  $\lambda = 0$  of the solution  $p, q$  of the problem (22), ch. 2. The function  $S_0$  should serve to obtain from  $p^0, q^0$  that solution of the degenerate problem which possesses this desired property.

We can rewrite eq. (27) as

$$S_1 H_0 - H_0 S_1 + H_1 = S_0^{-1} W_1 S_0. \quad (31)$$

To make this soluble, one has to determine  $S_0$  such that

$$\bar{H}_1 = S_0^{-1} W_1 S_0 \quad (32)$$

for a *diagonal matrix*  $W_1$ . An indication as to how one can simultaneously satisfy this eq. (31) and the requirements dictated by (26), ch. 2, can here naturally just as little be given as that for the determination of secular perturbations in classical theory. We shall, however, later use a new algebraic method to arrive at a simple treatment of an extensive class of degeneracies (ch. 3).

If (31), ch. 2, is satisfied, (30), ch. 2, can be solved as in ch. 1. Thereby those terms  $S_1(nm)$  of  $S_1$  for which  $\nu_0(nm)$  vanishes remain arbitrary, and this indeterminacy has to be utilized in order to solve the next higher order approximation formula, which can be transcribed as

$$S_2 H_0 - H_0 S_2 + F_2 = S_0^{-1} W_2 S_0 \quad (33)$$

in order to fulfil the necessary relation

$$\overline{F_2(H_0, H_1, H_2; S_1)} = S_0^{-1} W_2 S_0 \quad (31')$$

with  $W_2$  a diagonal matrix. This has to be satisfied for the problem to be soluble. The continuation of the procedure is clear.

The difficulty lies in the fact that at each order of approximation equations have to be satisfied by matrices which are already fixed to a large extent, so that it is not perceptible whether or not these equations will really prove soluble. In classical theory there is, though, an altogether analogous difficulty. These difficulties can, at least in the higher orders of approximation, be removed if in some approximation the system becomes nondegenerate.

Suppose, for example, that  $p^{(1)}$  and  $q^{(1)}$  in

$$q = q^0 + \lambda q^{(1)} + \dots,$$

$$p = p^0 + \lambda p^{(1)} + \dots$$

have really been determined, so that with

$$Q = q^0 + \lambda q^{(1)}$$

$$p = p^0 + \lambda p^{(1)}$$

one has

$$H(PQ) = W_0 + \lambda W_1 + \lambda^2 H'_2 + \lambda^3 H'_3 + \dots,$$

and suppose

$$\nu_0(nm) + \lambda \nu_1(nm) \neq 0 \quad \text{for} \quad n \neq m.$$

If for brevity we write  $H'_0$  for  $W_0 + \lambda W_1$  and set

$$p = SPS^{-1}$$

$$q = SQS^{-1},$$

then we have to build the following relation,

$$S(H'_0 + \lambda^2 H'_2 + \lambda^3 H'_3 + \dots) S^{-1} = W,$$

which, with the procedures of ch. 1, can be achieved with

$$S = \mathbf{1} + \lambda^2 S_2 + \lambda^3 S_3 + \dots$$

The generalization of these considerations for the case in which only in the  $r$ th approximation can one attain a nondegenerate system  $W = W_0 + \lambda W_1 + \dots + \lambda^r W_r$  follows of itself.<sup>18</sup>

In conclusion, we deem it important to point out that the notorious convergence difficulties encountered in the classical perturbation series, which play so decisive a role in the discussion of the three-body problem, do not arise here in quantum-mechanical perturbation theory; rather, one would here in general expect finite orbits to be periodic also.

### CHAPTER 3. CONNECTION WITH THE THEORY OF EIGENVALUES OF HERMITIAN FORMS

#### 1. General method

The treatment in the preceding sections has aimed at solving the basic quantum-theoretical equations in a manner as closely parallel to classical theory as possible. But behind the formalism of this perturbation theory there lurks a very simple, purely algebraic connection and it is well worth while to bring this into the limelight. Apart from the deeper insight into the mathematical structure of the theory, we thereby gain the advantage of being able to use the methods and results developed earlier in mathematics. We shall thus arrive at a new definition of the energy constants ("terms") which remains valid in the case of aperiodic motion also, i.e., of continuously-varying indices. Thereby we attain the prospect of finding

---

<sup>18</sup>Analogous cases in classical mechanics have been discussed by M. Born and W. Heisenberg, Ann. d. Phys. 74 (1924) 1.

methods for direct calculation of the energy without explicitly solving the problem of motion: methods which correspond to Sommerfeld's method of complex integration. We shall then be able to treat perturbations of an extensive class of degenerate systems completely, which the above-mentioned perturbation methods were not yet able to handle.

In considering a problem of  $f$  degrees of freedom specified by the energy function  $H(pq)$ , we can first select any system of matrices  $p_k^0, q_k^0$  whatsoever such that at all events the commutation relations (3), ch. 2, are satisfied: for example, we can take the  $p_k, q_k$  for a system of noncoupled harmonic oscillators.

Then, as mentioned in ch. 2 § 1, the dynamic problem, e.g., the determination of the  $p_k, q_k$  can be formulated as: A transformation  $(p_k^0 q_k^0) \rightarrow (p_k q_k)$  to be found which leaves eqs. (3), ch. 2, invariant and at the same time reduces the energy to a diagonal matrix.

The transformation of matrices can most easily be grasped if one regards them as a system of coefficients for linear transformations or bilinear forms. We therefore premise some known results of the algebra of such forms.

To every matrix  $a = (a(nm))$  there belongs a *bilinear form*

$$A(xy) = \sum_{nm} a(nm)x_n y_m \quad (1)$$

of two series of variables  $x_1, x_2, \dots$  and  $y_1, y_2, \dots$ . If the matrix be Hermitian, i.e., if the *transposed, matrix*  $\tilde{a} = (a(mn))$  be equal to the complex conjugate of the original matrix,

$$\tilde{a} = a^*, \quad a(mn) = a^*(nm), \quad (2)$$

then the form  $A$  assumes real values if in place of the variables  $y_n$  one substitutes the complex conjugate values  $x_n$ :

$$A(xx^*) = \sum_{nm} a(nm)x_n x_m^*. \quad (1a)$$

We recall the readily demonstrable transposition rule

$$(ab) = \tilde{b}\tilde{a} \quad (3)$$

and now subject the  $x_n$  to a linear transformation

$$x_n = \sum_l v(ln)y_l \quad (4)$$

with the aid of the (complex) matrix  $\nu = (v(ln))$ . Then the form  $A$  goes over into

$$A(xx^*) = B(yy^*) = \sum_{nm} b(nm)y_n y_m^*, \quad (5)$$

with

$$b(nm) = \sum_{kl} v(nk)a(kl)v^*(ml),$$

or, in matrix notation,

$$b = \nu a \nu^*. \quad (6)$$

This is termed the generation of a matrix  $b$  by the transformation  $\nu$  applied to  $a$ .

The matrix  $b$  is again of Hermitian type, for, with (3), ch. 3,

$$\tilde{b} = \nu^* \tilde{a} \nu = \nu^* a^* \nu = b^*. \quad (7)$$

The matrix  $\nu$  is called *orthogonal* if the respective transformation leaves the Hermitian unit form

$$E(xx^*) = \sum_n x_n x_n^*$$

invariant; from the result derived above, this is the case if and only if

$$\nu \tilde{\nu}^* = \mathbf{1}, \quad \text{or} \quad \tilde{\nu}^* = \nu^{-1}. \quad (8)$$

Thus, for instance, the permutation matrices mentioned in ch. 1 § 2 are real orthogonal matrices.

As is known, it is always possible for a finite number of variables to effect an orthogonal transformation of a form into a sum of squares (transformation to principal axes).<sup>19</sup>

$$A(xx^*) = \sum_n W_n y_n y_n^*. \quad (9)$$

For matrices, this means: a matrix exists for which

$$\nu \tilde{\nu}^* = \mathbf{1} \quad \text{and} \quad \nu a \tilde{\nu}^* = \nu a \nu^{-1} = W, \quad (10)$$

where  $W = (W_n \delta_{nm})$  is a diagonal matrix.

For infinite matrices, all the cases investigated so far have been found to obey an analogous rule; it can however occur that the index  $n$  on the right-hand side runs not only through a set of discrete numbers but also

---

<sup>19</sup>We write the coefficients of the transformed form  $W_n$  because in quantum mechanics they stand for the “energy”.

through a continuous range of values; this would correspond <sup>20</sup> to an integral constituent of (9) and the transformation (4).

The quantities  $W_n$  are termed “eigenvalues”, their ensemble is the “mathematical spectrum” of the form, made up of “point-” and “continuous” spectrum. As we shall see, this is identical with the “termspectrum” in physics, whereas the “frequency spectrum” is obtained from this by forming differences.

This transformation to principal axes now directly presents us with the solution of our dynamic problem which consists in seeking a transformation  $(p^0 q^0) \rightarrow (pq)$  such that the eqs. (3), ch. 2 are left invariant and at the same time the energy is brought into diagonal matrix form.

By the above rules of algebra, there exists an orthogonal matrix  $S$  for which

$$S\tilde{S}^* = \mathbf{1}, \quad \tilde{S}^*S = \mathbf{1} \quad (11)$$

and for which the transformations

$$\begin{aligned} p_k &= Sp_k^0 \tilde{S}^* = Sp_k^0 S^{-1}, \\ q_k &= Sq_k^0 \tilde{S}^* = Sq_k^0 S^{-1} \end{aligned} \quad (12)$$

leave

- (i) the Hermitian character
- (ii) the eqs. (3), ch. 2, invariant;
- (iii) the energy

$$H(pq) = SH(p^0 q^0) S^{-1} = W \quad (13)$$

converted into diagonal matrix form.

We wish to discuss the question of the uniqueness of this solution and in particular whether one could not generate other energy values through another orthogonal transformation  $T$ . Let us assume that  $W'$ , as given by

$$TH(p^0 q^0) T^{-1} = W',$$

is a diagonal matrix which differs from  $W$ . One would then have

$$TS^{-1}SHS^{-1}ST^{-1} = TS^{-1}W(TS^{-1})^{-1},$$

---

<sup>20</sup>Up till now, the theory of quadratic (or Hvrinitian) forms of infinitely many variables has been developed mainly for a special class (“bounded” forms) (D. Hilbert, Grundzü der Theorie der linearen Integralgleichungen; E. Hellinger, Crelles Journ. 136 (1910) 1). But here we are concerned just with non-bounded forms. We may nevertheless assume that in the main the rules run likewise.

and our question is equivalent to asking whether it is possible, starting from a diagonal matrix  $W$  to build another,  $W'$ , through transformation

$$W' = MWM^{-1}, \quad M\tilde{M}^* = \mathbf{1} \quad (14)$$

such that  $W'$  can not be derived from  $W$  by a permutation of the diagonal elements.

However, eq. (14), ch. 3, can be written

$$W'M - MW = 0.$$

and thus implies

$$M(nm)(W'_n - W_m) = 0. \quad (14a)$$

From the orthogonality of  $M$ , it follows in particular for  $m = n$  that

$$\sum_k |M(nk)|^2 = 1, \quad \sum_k |M(kn)|^2 = 1;$$

and consequently for a fixed  $n$  neither all the  $M(nk)$  nor all the  $M(kn)$  can vanish. But then (14a), ch. 3, shows that for every  $n$  there is certainly an  $m$  for which  $W'_n = W_m$ , i.e., all the  $W'_n$  appear among the  $W_m$ . The same holds inversely.

Thus all solutions derived from (12), ch. 3, lead (for given  $p_k^0, q_k^0$ ) to the same values for the energies of the stationary states, in accord with the conjecture stated in ch. 2 that the energies are always uniquely determined by the fundamental dynamic equations.

Degenerate systems will be characterized by the fact that multiple eigenvalues occur. The multiplicity of the eigenvalue  $W_n$ , i.e., the number of linear independent solutions  $v(ln)$  of eq. (4), ch. 3, yields the statistical weight of the respective state.

The importance of eq. (9), ch. 3, for our physical theory lies in the fact that various methods<sup>21</sup> exist in the algebra of finite or bounded infinite forms for determining the eigenvalues of a form without actually carrying the transformation through. It is to be hoped that such methods will prove of much avail in the future treatment of certain physical systems.

---

<sup>21</sup>For finite forms, the eigenvalues are the roots of an algebraic equation. Here, and also for bounded infinite matrices, they can be determined, e.g., by the method of Graeffe and Bernoulli; see, for example, R. Courant and D. Hilbert, Methoden der mathematischen Physik 1 (Springer, Berlin, 1924) § 3, pp. 14, 15.

## 2. Application to perturbation theory

In the following, we show that our present algebraic conception of the dynamic problem not only leads to exactly those formulae which were previously derived in ch. 1 § 4 in connection with perturbation theory in classical mechanics, but that when applied to degenerate systems it is considerably superior to the theory used hitherto.

We thus again assume that  $H$  has the form

$$H = H_0 + \lambda H_1 + \lambda^2 H_2 + \dots,$$

and that the dynamic problem specified by  $H_0$  has the solution  $p_k^0, q_k^0$ . We take these quantities as our starting coordinates from which the  $p_k, q_k$  are to be found, using an orthogonal transformation  $S$ . Naturally, the form assumed for  $H$  does not basically represent any limitation in generality, inasmuch as one can obviously separate off from  $H$  a component  $H_0$  of any desired form; however, the convergence of the power series in  $\lambda$  will depend essentially upon an apposite choice of  $H_0$ .

To undertake a principal-axes transformation of the Hermitian form

$$\sum_{mn} H_{mn} x_m x_n^*$$

we can, as is known, proceed as follows:

We attempt to find a solution of the linear equations

$$W x_k - \sum_l H(kl) x_l = 0; \quad (15)$$

this is possible only for certain values of the parameter  $W$ , namely  $W = W_n$  when  $W_n$  again denotes the eigenvalues (energy values). We first assume that no degeneracy is present, so that all  $W_n$  are different. Then to each  $W_n$  there corresponds a solution  $x_k = x_{kn}$  (determined except for a multiplicative factor), and hence the identities

$$W_n x_{kn} - \sum_l H(kl) x_{ln} = 0,$$

$$W_m x_{km}^* - \sum_l H^*(kl) x_{lm}^* = 0$$

obtain. On multiplying the former by  $x_{km}^*$ , the latter by  $x_{kn}$  and summing over  $k$ , it follows on subtraction (because of the Hermitian character of  $H$ ) that

$$(W_n - W_m) \sum_k x_{kn} x_{km}^* = 0.$$

By choosing the proportionality factor suitably, one can normalize to

$$\sum_k x_{kn} x_{kn}^* = 1.$$

Hence the  $x_{kn}$  form an orthogonal matrix

$$S = (x_{kn}).$$

It is precisely this which transforms the given form to a sum of squares, since if we substitute

$$x_k = \sum_n x_{kn} y_n$$

into the form, we obtain

$$\begin{aligned} \sum_{kl} H(kl) x_k x_l^* &= \sum_{kl} \sum_{mn} H(kl) x_{km} x_{ln}^* y_m y_n^* \\ &= \sum_{mn} \sum_l W_m x_{lm} x_{ln}^* y_m y_n^* \\ &= \sum_m W_m y_m y_m^*. \end{aligned}$$

From our assumption as to the form of  $H$ , the coefficients of eq. (15), ch. 3, now have the form

$$H(kl) = \delta_{kl} W_l^0 + \lambda H_1(kl) + \lambda^2 H_2(kl) + \dots$$

We thus seek to find the solution of (15), ch. 3, through expansions of the form

$$\begin{aligned} W &= W^0 + \lambda W^{(1)} + \lambda^2 W^{(2)} + \dots \\ x_k &= x_k^0 + \lambda x_k^{(1)} + \lambda^2 x_k^{(2)} + \dots \end{aligned} \tag{16}$$

If we substitute the above in (15), ch. 3, we obtain the approximation equations

$$\begin{aligned} (a) \quad x_k^0 (W^0 - W_k^0) &= 0, \\ (b) \quad x_k^{(1)} (W^0 - W_k^0) &= -x_k^0 W^{(1)} + \sum_l H^{(1)}(kl) x_l^0, \\ (c) \quad x_k^{(2)} (W^0 - W_k^0) &= -(x_k^{(1)} W^{(1)} + x_k^0 W^{(2)}) \\ &\quad + \sum_l (H^{(1)}(kl) x_l^{(1)} + H^{(2)}(kl) x_l^0). \end{aligned} \tag{17}$$

It follows from (17a), ch. 3, that  $W$  has to become equal to one of the  $W_k$ , since otherwise all  $x_k^0$  would vanish and we could then also infer the vanishing of  $x_k^{(1)}, x_k^{(2)}, \dots$  in sequence from the subsequent approximation equations.

If, then, we take our starting system as nonclcgcnract, and thus all the  $W_k^0$  as different from one another, the solution of (17a), ch. 3, is

$$W = W_n^0; \quad x_{nn}^0 = y_n^0; \quad x_{kn}^0 = 0 \quad \text{for } k \neq n. \quad (18)$$

Herein,  $y_n^0$  is an arbitrary number.

If we substitute this in (17b), ch. 3, we find, depending upon whether  $k = n$  or  $k \neq n$ ,

$$0 = y_n^0 \left( -W^{(1)} + H^{(1)}(nn) \right),$$

$$x_k^{(1)} (W_n^0 - W_k^0) = H^{(1)}(kn) y_n^0, \quad k \neq n.$$

Thus the solution runs

$$W^{(1)} = H^{(1)}(nn); \quad x_{nn}^{(1)} = y_n^{(1)}; \quad (19)$$

$$x_{kn}^{(1)} = -\frac{H^{(1)}(kn)}{\hbar\nu_0(kn)} y_n^0 \quad \text{for } k \neq n,$$

where again  $y_n^{(1)}$  is an arbitrary number.

Hence it similarly follows from (17c), ch. 3, that

$$W^{(2)} = H^{(2)}(nn) - \frac{1}{\hbar} \sum_l' \frac{H^{(1)}(nl)H^{(1)}(ln)}{\nu_0(ln)},$$

$$x_{nn}^{(2)} = y_n^{(2)} \quad (20)$$

$$x_{kn}^{(2)} = \left( \frac{1}{\hbar^2} \sum_l' \frac{H^{(1)}(kl)H^{(1)}(ln)}{\nu_0(kn)\nu_0(ln)} - \frac{H^{(1)}(nn)H^{(1)}(kn)}{\hbar^2\nu_0(kn)^2} \right. \\ \left. - \frac{H^{(2)}(kn)}{\hbar\nu_0(kn)} \right) y_n^0 - \frac{H^{(1)}(kn)}{\hbar\nu_0(kn)} y_n^{(1)}.$$

The solution of the third-order approximation can be derived just as easily; we cite only the energy value:

$$W^{(3)} = H^{(3)}(nn) - \frac{1}{\hbar} \sum_l' \frac{H^{(1)}(nl)H^{(2)}(ln) + H^{(2)}(nl)H^{(1)}(ln)}{\nu_0(ln)}$$

$$+ \frac{1}{h^2} \left( \sum'_{kl} \frac{H^{(1)}(nl)H^{(1)}(lk)H^{(1)}(kn)}{\nu_0(ln)\nu_0(kn)} - H^{(1)}(nn) \sum'_l \frac{H^{(1)}(nl)H^{(1)}(ln)}{\nu_0(ln)^2} \right).$$

The quantities  $y_n^{(0)}, y_n^{(1)}, \dots$ , which for the present are arbitrary, serve to normalize the solution (it is orthogonal of itself); the condition

$$\sum_k x_{kn}x_{kn}^* = 1$$

yields, for

$$x_{kn} = x_{kn}^0 + \lambda x_{kn}^{(1)} + \lambda^2 x_{kn}^{(2)} + \dots,$$

the equations

$$\sum_k x_{kn}^0 x_{kn}^{*0} = 1$$

$$\sum_k (x_{kn}^0 x_{kn}^{*(1)} + x_{kn}^{(1)} x_{kn}^{*0}) = 0$$

.....

On substituting the solution just obtained, it follows successively that

$$|y_n^0|^2 = 1$$

$$y_n^0 y_n^{*(1)} + y_n^{*0} y_n^{(1)} = 0$$

.....

If we now set

$$y_n^{(p)} = a_n^{(p)} e^{l\varphi_n(p)}, \quad (21)$$

we obtain

$$a_n^0 = 1$$

$$2a_n^{(1)} \cos(\varphi_n^0 - \varphi_n^{(1)}) = 0$$

$$\dots \dots \dots$$

$$2a_n^{(r)} \cos(\varphi_n^0 - \varphi_n^{(r)}) = F^{(r)}(a^{(r-1)}, \varphi^{(r-1)}, \dots).$$

Thus the phase constants  $\varphi_n^0, \varphi_n^{(1)}, \dots$  can be chosen arbitrarily; the  $a_n^0, a_n^{(1)}, \dots$  can be evaluated in sequence and determined uniquely. This stands in agreement with the result we found earlier (§ 3), namely that the phases of the diagonal terms of  $S$  remain undetermined.

On substituting the values  $a_n^0 = 1, \dots$  obtained above into (21), ch. 3, and this in turn into (18), (19), (20), ch. 3, we see that the “perturbation procedure” carried through earlier yielded just the solution for which the phases  $\varphi_n^{(p)}$  vanish, i.e., for which the diagonal terms of  $S$  are real.

We now turn to consideration of the case in which the starting system is degenerate and in which  $W_n^0$  is an  $r$ -fold eigenvalue. This means that eq. (17a), ch. 3, has the solution

$$\begin{aligned} W = W_n; \quad x_{nn}^0 &= y_{1,n}^0, \quad x_{n,n+1}^0 = y_{2,n}, \quad \dots \\ &x_{n,n+r-1} = y_{r,n}, \end{aligned} \tag{23}$$

$$x_{kn}^0 = 0 \quad \text{for} \quad k \neq n, n+1, \dots, n+r-1.$$

The left-hand side of (17b), ch. 3, then vanishes for

$$k = n, n+1, \dots, n+r-1;$$

this yields ( $r$ ) equations:

$$W^{(1)} y_{kn}^0 - \sum_{l=1}^r -l = 1 H^{(1)}(n+k, n+l) y_{ln}^0 = 0; \quad k = 1, 2, \dots, r, \tag{24}$$

whose array of coefficients is again of Hermitian type.

On setting the determinant to zero, one obtains a secular equation of the  $r$ th order for  $W^{(1)}$ :

$$\det \left( W^{(1)} \delta_{kl} - H^{(1)}(n+k, n+l) \right) = 0. \tag{25}$$

whose roots are certainly real. To each root there belong one or more independent solutions of eqs. (24), ch. 3.

If one selects one of these solutions, the perturbation proccdun1 can be pursued: we shall, however, not go into this further here.

It suffices to have recognized that our algebraic method is able to handle all degeneracies of finite multiplicity, i.e., that it can reduce the problem to the solution of algebraic equations. If, for example, each eigenvalue occurs twice, so that to each there belongs a vanishing frequency  $\nu_0(nm)$ , the perturbation problem leads to a quadratic equation:

$$\begin{vmatrix} W^{(1)} - H^{(1)}(n, n) & -H^{(1)}(n, n+1) \\ -H^{(1)}(n+1, n) & W^{(1)} - h^{(1)}(n+1, n+1) \end{vmatrix} = 0.$$

This case obtains which two originally identical nondegenerate systems (in which all frequencies in each of the respective systems are to be different) are coupled through some force.

Further, the orthogonality relation

$$\sum_k x_{kn}^0 x_{kn}^{*0} = 1$$

has an interesting meaning in the case of degenerate systems. Because of (23), this relation goes over into

$$\sum_{l=1}^r y_{ln}^0 y_{ln}^{*0} = 1.$$

From this it follows that, if  $m$  denotes any number in the series  $n, n+1, \dots, n+r-1$ , and  $k$  denotes any number outside this set, the sums

$$\begin{aligned} & \sum_{m=n}^{n+r-1} p^0(mk) p^{*0}(mk), \\ & \sum_{m=n}^{n+r-1} q^0(mk) q^{*0}(mk) \end{aligned}$$

are uniquely determined, even for degenerate systems, e.g., the summations are invariant with respect to those transformations which, by (19), ch. 2, allow new and altogether different solutions  $p', q'$  to arise from certain solutions  $p, q$  in the case of degeneracy. This result provides a mathematical representation of the so-called spectroscopic stability, which has played an important part in the more recent theories of fine-structure intensities (cf. ch. 4).

### 3. Continuous spectra

The simultaneous appearance of both continuous and line spectra as solutions of the same equations of motion and the same commutation relations seemed to us to represent a particularly significant feature of the new theory. In spite of this close connection between the two kinds of spectra, there nevertheless are characteristic distinctions, both mathematically and physically, between continuous and discrete spectra, corresponding to the

difference between Fourier series and Fourier integrals in classical theory; it therefore strikes us as desirable to indicate the rough outlines of the treatment of continuous spectra here. The mathematical theory of continuous spectra which occur for infinite quadratic forms has, starting from the fundamental investigations of Hilbert, explicitly been developed by Hellinger (*loc.cit.*) for the case of bounded quadratic forms. If we here permit ourselves to take over Hellinger's results to the unbounded forms which appear in our case, we feel ourselves to be justified by the fact that Hellinger's methods obviously conform exactly to the physical content of the problem posed.

Let us first briefly examine the classical analogue to our problem, namely aperiodic motion and its Fourier integral. Whereas in a Fourier series a certain amplitude  $a(\nu)$  always belongs to an oscillation  $\exp(2\pi i\nu t)$ , in the case of a Fourier integral one has a quantity of the form  $\varphi(\nu)d\nu$  in place of  $a(\nu)$ , where  $\varphi(\nu)$  might in a sense be conceived as an amplitude-density per frequency interval  $d\nu$ . In a similar and physically immediately obvious manner, one can always relate all quantities such as intensity, polarisation, etc. to a frequency interval  $d\nu$  between  $\nu$  and  $\nu + d\nu$ , but never to a definite frequency itself. We shall have to expect quite similar conditions to apply in quantum mechanics. Instead of quantities  $q(kl)$  we shall have quantities of the form  $q(k, W)dW$  or  $q(W, W')dWdW'$ , depending upon whether one or both of the two indices lie in the continuous region. Indeed, in place of the energy  $W$  itself, there will have to be a "total energy" per interval  $dW$ , since the probability for an atom to have an absolutely definite energy  $W$  in the continuous region is zero. To elucidate these questions we shall in the following briefly sketch Hellinger's mathematical theory.

For infinite quadratic forms, the case may arise that the form

$$\sum_{mn} H(mn)x_m x_n^*$$

cannot be converted into the expression  $\sum_n W_n y_n y_n^*$ , by an orthogonal substitution. We may then assume, in analogy with the results for bounded forms, that a representation with a continuous spectrum exists,

$$\sum_{mn} H(mn)x_m x_n^* = \sum_n W_n y_n y_n^* + \int W(\varphi) y(\varphi) y^*(\varphi) d\varphi, \quad (26)$$

in which the original variables are connected with new variables  $y_n, y(\varphi)$  through an "orthogonal transformation"; one only has to specify more clearly what is here understood by an orthogonal transformation.

If we again consider the linear equations (15), ch. 3,

$$Wx_k - \sum_l H(kl)x_l = 0, \quad (27)$$

the case under review in which (26), ch. 3, contains an integral component will occur when there are not only discrete values  $W_n$ , for which these equations can be solved, but also a continuum of such values comprising one or more “segments” on the  $W$ -axis (continuous spectrum). For any given point  $W$  of this continuum, there exists a solution  $x_l(W)$  (or several, which we for simplicity wish to exclude); for two such  $W$ -values,  $W'$  and  $W''$ , the equations

$$\begin{aligned} W'x_k(W') - \sum_l H(kl)x_l(W') &= 0, \\ W''x_k^*(W'') - \sum_l H^*(kl)x_l^*(W'') &= 0 \end{aligned} \quad (28)$$

obtain, from which, as above, we conclude that

$$(W' - W'') \sum_k x_k(W')x_k(W'') = 0. \quad (29)$$

If one tries imposing the normalization condition

$$\sum_k |x_k(W)|^2 = 1$$

on top of these orthogonality relations, one observes that the function of two variables

$$\sum_k x_k(W')x_k(W'')$$

becomes wildly irregular, if it exists at all. The above sum does not in fact converge and therefore does not represent a function.

Accordingly, a different type of normalization is required. With Hellinger, we set

$$\sum_k |\int x_k(W)dW|^2 = \varphi(W). \quad (30)$$

The series on the left-hand side is in general convergent and represents a monotonous function  $\varphi(W)$ , which apart from certain restrictions can be chosen arbitrarily, since the  $x_k(W)$  are of course determined only up to a factor which is independent of  $k$ . We shall later discuss the physical significance of this function  $\varphi(W)$ , by which the solutions  $x_k(W)$  are defined. Hellinger has termed  $\varphi(W)$  the “basis function” and has shown that the

orthogonality conditions can be derived in the following form: If  $\Delta_1$  and  $\Delta_2$  be any two intervals of the continuous spectrum and  $\Delta_{12}$  the interval common to them both (which may also be absent), then

$$\begin{aligned} \sum_k \int_{\Delta_1} x_k(W') dW' \int_{\Delta_2} x_k(W'') dW'' &= \int_{\Delta_{12}} dW \\ &= \varphi(W^{(2)}) - \varphi(W^{(1)}), \end{aligned} \quad (31)$$

where  $W^{(1)}, W^{(2)}$  are the end-points of  $\Delta_{12}$ . I fence if there is no overlap between the intervals  $\Delta_1, \Delta_2$ , a zero stands on the right-hand side.

If one conceives the intervals  $\Delta_1, \Delta_2, \Delta_{12}$  to be very small, one can symbolically write

$$\sum_k x_k(W') dW' \cdot x_k(W'') dW'' = d\varphi(W). \quad (32)$$

This relation prompts the suggestion that one operate generally with the quantities  $x_k(W)dW$  as “differential solutions” of (27), ch. 3, whereby one has to note that the respective equations are always to be interpreted in the sense of (31), ch. 3. These differential solutions are orthogonal in the usual way, but instead of being normalized to unity, are normalized to the differential of the basis function  $\varphi(W)$ .

The totality of discrete values  $x_{kn}$ , and of values  $x_k(W)$  which are discrete in one index and have a continuous distribution in the other, comprises the elements of the “orthogonal” matrix

$$S = (x_{kn}, x_k(W)dW),$$

which can schematically be represented as:

$$S = \left( \begin{array}{ccccccccc} & k & \rightarrow & & & & & & \\ n & \cdot \\ | & \cdot \\ \downarrow & \cdot \\ W & | & | & | & | & | & | & | & | \\ \downarrow & | & | & | & | & | & | & | & | \end{array} \right) \quad (33)$$

The orthogonality and normalization equations for the entire matrix split

into four different groups:

$$\begin{aligned} \sum_k x_{km} x_{kn}^* &= \delta_{mn}; \\ \sum_k x_{kn} x_k^*(W) dW &= 0; \quad \sum_k x_k(W) dW \cdot x_{kn}^* = 0; \\ \sum_k x_k(W') dW' \cdot x_k^*(W'') dW'' &= d\varphi. \end{aligned} \quad (34)$$

We can also write the orthogonality relations for the columns, which read

$$\begin{aligned} \sum_n x_{kn} x_{ln}^* + \int \frac{x_k(W) dW \cdot x_l^*(W) dW}{d\varphi} \\ = \sum_n x_{kn} x_{ln}^* + \int \frac{dW}{\varphi'} x_k(W) x_l^*(W) = \delta_{kl}, \end{aligned} \quad (35)$$

where the prime denotes differentiation,  $\varphi' = d\varphi/dW$ .

With the aid of this matrix, we have to transform the variables  $x_k$  into new ones,  $y_n, y(\varphi) d\varphi$ . We set:

$$\begin{aligned} y_n &= \sum_k x_{kn} \cdot x_k, \\ y(\varphi) d\varphi &= \sum_k x_k(W) dW \cdot x_k. \end{aligned} \quad (36)$$

A simple calculation then yields

$$\sum_n W_n y_n y_n^* + \int W(\varphi) y(\varphi) y^*(\varphi) d\varphi = \sum_{kl} H(kl) x_k x_l^*. \quad (37)$$

The principal-axes transformation has thereby been carried through. Let us now investigate which representation of coordinate and momentum matrices is obtained with the aid of this orthogonal transformation, e.g., what is meant here by the equations

$$\begin{aligned} p &= Sp_0 S^{-1}, \\ q &= Sq_0 S^{-1}, \end{aligned} \quad (38)$$

or, generally, by

$$f(pq) = S f(p_0 q_0) S^{-1}. \quad (39)$$

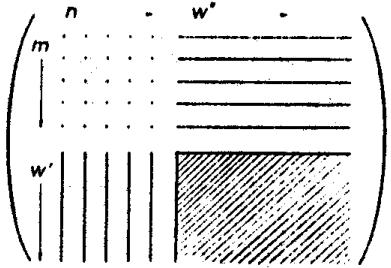
We find, for example, four types of elements for  $p$ :

$$\begin{aligned}
 p(mn) &= \sum_{kl} x_{km}^* p^0(kl) x_{ln} \\
 p(m, W) dW &= \sum_{kl} x_{km}^* p^0(kl) x_l(W) dW, \\
 p(W, n) dW &= \sum_{kl} x_k^*(W) dW \cdot p^0(kl) x_{ln}, \\
 p(W', W'') dW' dW'' &= \sum_{kl} x_k^*(W') dW' p^0(kl) x_l(W'') dW''.
 \end{aligned} \tag{40}$$

In a similar manner, instead of the amplitudes  $p(mn)$ , “amplitude densities”  $p(mW), dW$  (which refer to an interval  $dW$ ) occur generally in the case of a continuously variable index. This accords with our previously declared expectation. It is, however, not necessary to take just the energy as the continuously variable index. In place of the energy, one could, for example, introduce the quantity  $\varphi(W)$ . Then in place of  $p(mW)dW$  one would have  $p(m\varphi)(dW/d\varphi)d\varphi$ . Finally, in the continuous case the energy  $W_n$  is replaced by the quantity  $W(\varphi)d\varphi$ . In place of the energy of the individual atom, we get a sort of total energy per interval  $dW$ . Thence  $d\varphi$  essentially represents the number of atoms having an energy which lies between  $W$  and  $W + dW$ , or the a priori probability that the energy of the atom lies between  $W$  and  $W + dW$ . We here most clearly observe the difference between the cases with discrete stationary states on the one hand and those with a continuous manifold of states on the other hand, and we can see a simple connection between the problem of statistical weights and the question of the normalization of the solution of (27) ch. 3. In the case of discrete states when there are no multiple eigenvalues, we make the simple physical contention that each slate should have the statistical weight I. This was ensured by the fact that we normalized the  $x_{kn}$  on the basis of the requirement

$$\sum_k x_{kn} x_{kn}^* = 1.$$

In the case of continuous manifolds of states, it was not possible to fix the a priori probabilities so simply; more detailed investigations of the problem in question are necessary for their determination and hence also for the evaluation of the function  $q_i$ . Hence the connection between transition probabilities and the amplitudes might also assume a somewhat more complicated aspect in the case of continuous spectra than for line spectra.



The matrices of  $p, q$  or  $f(p, q)$  represented by (40), ch. 3, and corresponding forms, can for the general case be made clear by the adjoining scheme:

The physical meaning of this scheme is self-evident.

There are four types of “transitions” which to some extent furnish a simple analogue to the “transitions” postulated hitherto in the theory of the hydrogen atom, viz. (1) from ellipse to ellipse; (2) from ellipse to hyperbola; (3) from hyperbola to ellipse; (4) from hyperbola to hyperbola.

One can still raise the objection against the formulae (38) and (40), ch. 3, that manifestly in some instances the infinite sums on the right-hand sides do not converge, and hence do not represent a function, since of course in classical theory also, the representation of a function  $f(p, q)$  by Fourier integrals is sometimes impossible, as for instance if the respective functions  $f$  increase linearly with time at large times (as is in general the case with coordinates). To this objection, one may, however, rejoin that the observable effects of the atom (such as radiation, the force upon another atom, etc.) do not in general belong to this type of function, and thus that the appropriate sums of the same type as the formulae (40), ch. 3, might indeed converge.

#### CHAPTER 4. PHYSICAL APPLICATIONS OF THE THEORY

## 1. Laws of conservation of momentum and angular momentum; Intensity formulae and selection rules

By way of applying the general theory as established in the foregoing sections, we now derive the known features concerning “quantization” of angular momentum and some associated principles.

We shall thereby at the same time become acquainted with some characteristic examples involving *integration* of the quantum-mechanical equations of motion. The previously-discussed perturbation methods can, of course, be applied successfully only when a set of particularly simple examples, which can be selected as unperturbed systems  $H_0$ , has been integrated in some other way. Now, the quantum-mechanical equations of motion coming from the decomposition of matrix equations into components present the special difficulty that – apart from the instance of the harmonic oscillator – infinitely many unknowns occur in each of the separate equations. A technique frequently employed in overcoming (his difficulty in the following and, as it seems, of wide applicability, consists of the following procedure: By analogy with classical theory, one first seeks integrals of the equations of motion, i.e., functions  $\mathbf{A}(p, q)$  which on the basis of the equations of motion and the commutniiion rules are constant in function of time and consequently become diagonal matrices in the case of nondegenerate periodic systems. Now if  $\varphi(p, q)$  be any function whatsoever, the difference

$$\varphi\mathbf{A} - \mathbf{A}\varphi = \psi$$

can be evaluated with the help of the commutation rules; if  $\mathbf{A}$  is a diagonal matrix, a system of equations results, each of which contains only a finite number of unknowns, namely a single component of the matrices  $\varphi$  and  $\psi$  (and two diagonal terms of  $\mathbf{A}$ ) in each.

If in Cartesian coordinates,  $H = H'(p) + H''(q)$ , which includes the case of relativistic mechanics, then one can sec immediately that the components of the angular momentum  $\mathfrak{M}$ , viz.

$$\begin{aligned} \mathbf{M}_x &= \sum_{k=1}^{f/3} (p_{ky}q_{kz} - q_{ky}p_{kz}), \\ \mathbf{M}_y &= \sum_{k=1}^{f/3} (p_{kz}q_{kx} - q_{kz}p_{kx}), \\ \mathbf{M}_z &= \sum_{k=1}^{f/3} (p_{kx}q_{ky} - q_{kx}p_{ky}) \end{aligned} \quad (1)$$

become constant under the same general conditions as in classical theory. This is because a sum,

$$\mathbf{M}_z = \varphi(q) + \psi(p),$$

ensues for the derivative of, say,  $\mathbf{M}_z$  with respect to time, and since all the  $p$  commute with one another, as do all the  $q$ , the quantities  $\varphi, \psi$  vanish under the same conditions as in classical theory.

The same remarks are to be applied to the linear momentum

$$\mathfrak{p} = \sum_{k=1}^{f/3} \mathfrak{p}_k; \quad \text{i.e.,} \quad p_x = \sum_{k=1}^{f/3} p_{kx}, \dots, \quad (2)$$

which likewise becomes constant. Thus the centre-of-mass theorem holds just as in classical theory.

We immediately note here a formula which will be used later and which can be derived from the commutation relations (3), ch. 2. We find

$$\begin{aligned} \mathbf{M}_x \mathbf{M}_y - \mathbf{M}_y \mathbf{M}_x &= \sum_{kl} \{(p_{ky}q_{kz} - q_{ky}p_{kz})(p_{lz}q_{lx} - q_{lz}p_{lx}) \\ &\quad - (p_{kz}q_{kx} - q_{kz}p_{kx})(p_{ly}q_{lz} - q_{ly}p_{lz})\}, \\ &= \sum_{kl} \{p_{ky}q_{lx}(q_{kz}p_{lz} - p_{lz}q_{kz}) \\ &\quad + q_{ky}p_{lx}(p_{kz}q_{lz} - q_{lz}p_{kz})\}, \\ &= \frac{h}{2\pi i} \sum_k (p_{kx}q_{ky} - q_{kx}p_{ky}), \end{aligned}$$

i.e.,

$$\mathbf{M}_x \mathbf{M}_y - \mathbf{M}_y \mathbf{M}_x = \varepsilon M_z, \quad (\text{where } \varepsilon = h/2\pi i). \quad (3)$$

Incidentally, one can directly see from this formula that the theorem of conservation of angular momentum invariably holds for at most one or alternatively for all three axes, as in classical theory.

In the following we shall assume that on treating the problem with which we are confronted by the methods developed in the preceding chapter we are led to obtain *discrete* energy values (point spectrum). If then  $\mathbf{M}_z = 0$  for a *nondegenerate* system – this will for instance be the case if forces which are symmetrical about the  $z$ -axis act upon the atom –  $\mathbf{M}_z$  has to become a *diagonal matrix*: the separate diagonal terms are to be regarded as the angular moments of the atom about the  $z$ -axis for the individual *states* of

the atom. For the investigation of the motions of the electrons in this case, we first note that the relation

$$q_{lz}\mathbf{M}_z - \mathbf{M}_z q_{lz} = 0 \quad (4)$$

follows from ch. 4, and since  $\mathbf{M}_z(nm) = \delta_{nm}\mathbf{M}_{zn}$ , this means that

$$q_{lz}(nm)(\mathbf{M}_{zn} - \mathbf{M}_{zm}) = 0. \quad (5)$$

One sees that: *For a quantum jump in which there is a change in the angular momentum  $\mathbf{M}_z$ , the “plane of vibration” of the generated “spherical wave” lies perpendicular to the  $z$ -axis.*

$$\begin{aligned} q_{lx}\mathbf{M}_z - \mathbf{M}_z q_{lx} &= -\varepsilon \mathbf{q}_{ly}, \\ q_{ly}\mathbf{M}_z - \mathbf{M}_z q_{ly} &= \varepsilon \mathbf{q}_{lx}, \end{aligned} \quad (6)$$

i.e.,

$$\begin{aligned} q_{lx}(nm)(\mathbf{M}_{zn} - \mathbf{M}_{zm}) &= -\varepsilon q_{ly}(nm), \\ q_{ly}(nm)(\mathbf{M}_{zn} - \mathbf{M}_{zm}) &= \varepsilon q_{lx}(nm). \end{aligned} \quad (7)$$

*Thus for jumps in which no change in  $M_z$  occurs, the emitted light is linearly polarized parallel to the  $z$ -axis.*

Further, from (7), ch. 4, it follows that

$$\left\{ (M_{zn} - M_{zm})^2 - (h^2/4\pi^2) \right\} q_{l\eta}(nm) = 0; \quad \eta = x, y. \quad (8)$$

One finally concludes: *For every quantum jump  $M_{zn}$  changes by 0, or by  $\pm h/2\pi$ . The light emitted in the latter case is circularly polarized, as follows from (7), ch. 4.*

In accordance with the above finding concerning the possible changes in  $M_z$ , the quantity  $M_{zn}$  can be represented in the form

$$M_{zn} = \frac{h}{2\pi} (n_1 + c), \quad n_1 = \dots, -2, -1, 0, 1, 2, \dots . \quad (9)$$

If there were states whose angular momentum did not fit into this set, no transitions and no interactions whatsoever could occur between these and the states depicted by (9), ch. 4. Equation (9), ch. 4, can be taken as a motive for splitting  $n$  into two components, one of which is the number  $n_1$ , introduced in (9), ch. 4, whereas the other,  $n_2$ , counts off the various  $n$  with

the same  $n_1$ . Our matrices then become four-dimensional, and the results we found for the motions of electrons may be summarised as:

$$q_{lz}(nm) = \delta_{n_1, m_1} q_{lz}(nm); \quad (10)$$

$$q_{lx}(nm) = \delta_{1, |n_1 - m_1|} q_{lx}(nm),$$

$$q_{ly}(nm) = \delta_{1, |n_1 - m_1|} q_{ly}(nm); \quad (10')$$

$$q_{lx}(n, n_2; n_1 \pm 1, m_2) \pm i q_{ly}(n_1, n_2; n_1 \pm 1, m_2) = 0. \quad (10'')$$

Further, from (4) and (6), ch. 4, it follows that if we set

$$\mathbf{q}_l^2 = q_l^2 = q_{lx}^2 + q_{ly}^2 + q_{lz}^2,$$

then

$$q_l^2 \mathbf{M}_z - \mathbf{M}_z q_l^2 = 0. \quad (11)$$

This relation means that  $q_l^2$  is a diagonal matrix with respect to the “quantum number”  $n_1$ .

The relations (4) to (7), ch. 4 and (10), (11), ch. 4, also hold if in place of the  $q_{lx}, q_{ly}, q_{lz}$ , we insert  $p_{lx}, p_{ly}, p_{lz}$  or alternatively  $\mathbf{M}_x, \mathbf{M}_y, \mathbf{M}_z$ . Thus in particular we have:

$$\begin{aligned} M_x(nm) &= \delta_{1, |n_1 - M_1|} M_x(nm); & M_y(nm) &= \delta_{1, |n_1 - m_1|} M_y(nm), \\ M_x(n_1, n_2; n_1 \pm 1, m_2) &\pm i M_y(n_1, n_2; n_1 \pm 1, m_2) = 0. \end{aligned} \quad (12)$$

Furtlicr (cf. eq. (1), ch. 4),  $\mathfrak{M}^2 = \mathbf{M}^2 = \mathbf{M}_x^2 + \mathbf{M}_y^2 + \mathbf{M}_z^2$  is a diagonal matrix with respect to  $n_1$ , since

$$\mathbf{M}^2 \mathbf{M}_z - \mathbf{M}_z \mathbf{M}^2 = 0. \quad (13)$$

For a system in which all three angular momentum conservation theorems apply, the constant components of  $\mathfrak{M}$  certainly cannot collectively be diagonal matrices, since otherwise the above considerations for  $\mathbf{M}_z$  to be a diagonal matrix could be applied to each of these components, which would lead to discrepancies. Hence such a system is necessarily *degenerate*.

We now set out to consider a system  $H = H_0 + \lambda H_1 + \dots$  of the following type: All three angular momentum theorems *are to apply for*  $\lambda = 0$ . For  $\lambda \neq 0$  *the system is to be nondegenerate; the constancy of*  $\mathbf{M}_z$  *is to remain undisturbed. The energy*  $H_0$  *is to be independent of*  $n_1$ . The results we shall obtain from this investigation of the case  $\lambda \neq 0$  can in part also be carried over to the degenerate system  $H_0$ , namely insofar as *they are independent firstly of*  $\lambda$  *and secondly of the distinguished direction*  $z$ .

The assumed degeneracy of the system for  $\lambda = 0$  is expressed by the fact that  $\dot{\mathbf{M}}_x, \dot{\mathbf{M}}_y, (d/dt)(\mathbf{M}^2)$  contain no terms of zeroth order in  $\lambda$ .

Thus

$$\begin{aligned}\nu_0(nm)M_\eta(nm) &= 0, & \eta &= x, y; \\ \nu_0(nm)M^2(nm) &= 0.\end{aligned}\quad (14)$$

Since  $W_0$  is independent of the quantum number  $n_1$  introduced earlier, whence  $\nu_0(n_1, n_2; m_1, m_2) = 0$ , where as  $\nu_0(n_1, n_2; !m_1, m_2) \neq 0$  is invariably non-zero for  $n_2 \neq m_2$ , it follows from (14), ch. 4, that

$$\begin{aligned}M_\eta^0(nm) &= \delta_{n_2 m_2} M_\eta^0(nm), \\ m^{0^2}(nm) &= \delta_{n_2 m_2} M^{0^2}(nm),\end{aligned}\quad (15)$$

The square of the total momentum  $(\mathbf{M}^0)^2$  is a diagonal matrix in consequence of (13), (15), ch. 4. The double sum representing an element of the matrix  $\mathbf{M}_x^0, \mathbf{M}_y^0$  reduces to a simple sum

$$\begin{aligned}\sum_{k_1 k_2} M_x^0(n_1 n_2; k_2 k_2) M_y^0(k_1 k_2; m_1 m_2) \\ \delta_{n_2 m_2} \sum_{k_1} M_x^0(n_1 n_2; k_1 n_2) M_x^0(k_1 n_2; m_1 n_2),\end{aligned}\quad (16)$$

which contains only a finite number of summation terms because of the finite number of possible  $n_1$  at fixed  $n_2$  (the terms of

$$M^{0^2} = M_x^{0^2} + M_y^{0^2} + M_z^{0^2} \geq M_z^2$$

do not depend on  $n_1$ ). In (3), ch. 4, applied to  $M_x^0, M_y^0, M_z^0$ , we can at any given time sum the equations which belong to a given  $n_2$  over  $n_1$  and obtain,<sup>22</sup> for fixed  $n_2$ :

$$\sum_{n_1} M_z(n_1, n_2; m_1 n_2) = \sum_{n_1} (n_2 + C) \frac{h}{2\pi} = 0. \quad (17)$$

On noting additionally that, by (12) and (16), ch. 4, the sum (17), ch. 4 vanishes for *every* single uninterrupted sequence of the  $n_1$  it follows that at fixed  $n_2$  the possible values of  $n_1 + C$  form an unbroken series and the symmetrically with respect to zero. Hence they must necessarily constitute either *integer* or *half-integer* numbers, the latter being numbers in the series  $\dots, -3/2, -1/2, 1/2, 3/2 \dots$ . If for the moment  $\mathbf{M}_z$ , about the  $z$ -axis we

---

<sup>22</sup>In I we already noted that in the case of a finite diagonal sum  $D(ab)$  we always have  $D(ab) = D(ba)$ .

now introduce the notation usually used in the literature, namely  $m(h/2\pi)$  in place of  $(n_1 + C)9h/2\pi$ , this result accordingly shows that the selection rule  $m \rightarrow (m + 1, m - 1)$  applies to  $m$  and that  $m$  is either “integer” or “half-integer”.

Our result demonstrates further that exclusion of individual states, such as was, for example, necessary in the past theory of the hydrogen atom in order to prevent collisions between the electron and the nucleus, has no place in the theory proposed here.

We now attempt to derive the selection principle (or the “total momentum quantum number”, as also the intensities for the Zeeman effect, from our theory, proceeding from (5) and (8), ch. 4.

Let us recall the derivation of these selection rules in classical theory: There it is only necessary to introduce a coordinate system whose  $z$ -axis coincides with the direction of the total angular momentum; in the new coordinates the same results can be derived for  $\mathbf{M}$  as were previously obtained for  $M_z$ . Let us accordingly set up such a coordinate system  $x', y', z'$ . The relation

$$z' = x \frac{\mathbf{M}_x}{\mathbf{M}} + y \frac{\mathbf{M}_y}{\mathbf{M}} + z \frac{\mathbf{M}_z}{\mathbf{M}}$$

has to hold anyway in order that the  $z'$ -axis lie in the direction of the total momentum. (In the following, we shall again drop the index  $0$  for simplicity in all momenta and coordinates: the calculations throughout refer to the limiting case  $\lambda = 0$ ). Further, we can so arrange it that the  $x'$ -axis lies in the  $x, y$ -plane. Everything is thereby fixed, and we have

$$\begin{aligned} x' &= y \frac{\mathbf{M}_x}{\sqrt{(\mathbf{M}_x^2 + \mathbf{M}_y^2)}} - x \frac{\mathbf{M}_y}{\sqrt{(\mathbf{M}_x^2 + \mathbf{M}_y^2)}} \\ y' &= \frac{z(\mathbf{M}_x^2 + \mathbf{M}_y^2) - x\mathbf{M}_z\mathbf{M}_x - y\mathbf{M}_z\mathbf{M}_y}{\mathbf{M}\sqrt{(\mathbf{M}_x^2 + \mathbf{M}_y^2)}}. \end{aligned}$$

Now let us try a similar procedure in quantum mechanics. We introduce the three quantities

$$\begin{aligned} \mathbf{Z}_l &= q_{lx}\mathbf{M}_x + q_{ly}\mathbf{M}_y + q_{lz}\mathbf{M}_z, \\ \mathbf{X}_l &= q_{ly}\mathbf{M}_x - \mathbf{M}_y q_{lx}, \\ \mathbf{Y}_l &= \mathbf{M}_x q_{lz}\mathbf{M}_x + \mathbf{M}_y q_{lz}\mathbf{M}_y - q_{lx}\mathbf{M}_z\mathbf{M}_x - \mathbf{M}_y\mathbf{M}_z q_{ly}. \end{aligned} \tag{18}$$

In order to derive the desired selection rules, we still need some commutation relations, which result from (4) and (6), ch. 4 ( $\varepsilon = h/2\pi i$ ):

$$q_{lx}\mathbf{M}^2 - \mathbf{M}^2 q_{lx} = 2\varepsilon(q_{lz}\mathbf{M}_y - \mathbf{M}_z q_{ly}) \tag{19}$$

and the equations for  $q_{ly}, q_{lz}$ , which ensue from this on cyclic permutation. It then follows <sup>23</sup> from (3), (4), (6) and (19), ch. 4, that

$$\begin{aligned} X_l M^2 - M^2 X_l &= -2\varepsilon Y_l, \\ Y_l M^2 - M^2 Y_l &= \varepsilon (X_l M^2 + M^2 X_l), \\ Z_l M^2 - M^2 Z_l &= 0. \end{aligned} \quad (20)$$

These equations are fully analogous to Hie relations (4) and (6), ch. 4, which determine the selection rules for  $M_z$ ; since we shall later show that the  $q_{lx}, q_{ly}, q_{lz}$  really can be expressed as linear functions of the  $X_l, Y_l, Z_l$ , with coefficients which for  $\lambda = 0$  are constant with time, we can determine the selection rules for  $M$  directly from (20), ch. 4. As  $M^2$  is a diagonal

---

<sup>23</sup>The first and third formulae in eq. (20) ch. 4, result from a quite simple calculation. The second of eqs. (20), ch. 4, can be derived in the following way:

From (18), ch. 4,

$$\mathbf{Y}_l = \mathbf{M}_x q_{lz} \mathbf{M}_x + \mathbf{M}_y q_{ly} \mathbf{M}_y - q_{lx} \mathbf{M}_z \mathbf{M}_x - \mathbf{M}_y \mathbf{M}_z q_{ly},$$

and because of (6), ch. 4,

$$\begin{aligned} \mathbf{Y}_l &= q_{lz} (\mathbf{M}_x^2 + \mathbf{M}_y^2) - \varepsilon q_{ly} \mathbf{M}_x + \varepsilon \mathbf{M}_y q_{lx} + \varepsilon^2 q_{lz} \\ &\quad - q_{lx} \mathbf{M}_z \mathbf{M}_x - \mathbf{M}_y \mathbf{M}_z q_{ly} \\ &= q_{lz} (\mathbf{M}^2 - \mathbf{M}_z^2) - \varepsilon X_l + \varepsilon^2 q_{lz} - q_{lx} \mathbf{M}_z \mathbf{M}_x - \mathbf{M}_y \mathbf{M}_z q_{ly}. \end{aligned}$$

In the evaluation of  $\mathbf{Y}_l \mathbf{M}^2 - \mathbf{M}^2 \mathbf{Y}_l$  we now have to note that  $M^2$  commutes with  $\mathbf{M}_x, \mathbf{M}_y, \mathbf{M}_z$ . Hence for the second part of the formula for  $\mathbf{Y}_l$  written above, it follows that (cf. (19), ch. 4)

$$\begin{aligned} &(q_{lx} \mathbf{M}_z \mathbf{M}_x + \mathbf{M}_y \mathbf{M}_z q_{ly}) \mathbf{M}^2 - \mathbf{M}^2 (q_{lx} \mathbf{M}_z \mathbf{M}_x + \mathbf{M}_y \mathbf{M}_z q_{ly}) \\ &= 2\varepsilon (q_{lz} \mathbf{M}_y \mathbf{M}_z \mathbf{M}_x - \mathbf{M}_z q_{ly} \mathbf{M}_z \mathbf{M}_x + \mathbf{M}_y \mathbf{M}_z q_{lx} \mathbf{M}_z - \mathbf{M}_y \mathbf{M}_z \mathbf{M}_x q_{lz}). \end{aligned}$$

On noting that (eq. (19), ch. 4)  $q_{lz} M^2 - M^2 q_{lz} = 2\varepsilon X_l$ , it follows from the commutation relations that

$$\begin{aligned} q_{lz} \mathbf{M}_y \mathbf{M}_z \mathbf{M}_x - \mathbf{M}_y \mathbf{M}_z \mathbf{M}_x q_{lz} &= \varepsilon (M_y M_z q_{ly} - q_{lz} \mathbf{M}_z \mathbf{M}_x), \\ \mathbf{M}_y \mathbf{M}_z q_{lx} \mathbf{M}_z - \mathbf{M}_z q_{ly} \mathbf{M}_z \mathbf{M}_x &= -X_l \cdot \mathbf{M}_z^2 - \varepsilon (M_z q_{ly} \mathbf{M}_y - q_{lx} \mathbf{M}_x \mathbf{M}_z). \end{aligned}$$

and finally we obtain the desired formula (20), ch. 4:

$$\begin{aligned} \mathbf{Y}_l \mathbf{M}^2 - \mathbf{M}^2 \mathbf{Y}_l &= 2\varepsilon \mathbf{X}_l (\mathbf{M}^2 - \mathbf{M}_z^2 + \varepsilon^2) - \varepsilon (\mathbf{X}_l \mathbf{M}^2 - \mathbf{M}^2 \mathbf{X}_l) + 2\varepsilon \mathbf{X}_l \mathbf{M}_z^2 \\ &\quad - 2\varepsilon^2 (q_{lx} \mathbf{M}_x \mathbf{M}_z - q_{lx} \mathbf{M}_z \mathbf{M}_x + \mathbf{M}_y \mathbf{M}_z q_{ly} - \mathbf{M}_z \mathbf{M}_y q_{ly}) \\ &= 2\varepsilon \mathbf{X}_l (\mathbf{M}^2 - \mathbf{M}_z^2 + \varepsilon^2) - \varepsilon (X_l \mathbf{M}^2 - \mathbf{M}^2 \mathbf{X}_l) + 2\varepsilon \mathbf{X}_l \mathbf{M}_z^2 - 2\varepsilon^3 X_l \\ &= \varepsilon (\mathbf{X}_l \mathbf{M}^2 + \mathbf{M}^2 \mathbf{X}_l). \end{aligned}$$

matrix, it follows from (20), ch. 4, that

$$\begin{aligned} X_l(nm) (M_m^2 - M_n^2) &= -2\varepsilon Y_l(nm), \\ Y_l(nm) (M_m^2 - M_n^2) &= \varepsilon X_l(nm) (M_m^2 + M_n^2), \\ Z_l(nm) (M_m^2 - M_n^2) &= 0. \end{aligned} \quad (21)$$

The last of the eqs. (21), ch. 4, states that no vibrations take place in  $Z$  which could entail a change in  $M^2$ . It follows from the first two equations that

$$X_l(nm) \left\{ (M_m^2 - M_n^2)^2 - \frac{\hbar^2}{2\pi^2} (M_m^2 + M_n^2) \right\} = 0. \quad (22)$$

If we now set  $M_m^2 = (\hbar/2\pi)^2(a_m^2 - 1/4)$ , where  $a_m$  denotes any function of the quantum numbers, eq. (22), ch. 4 yields

$$X_l(nm)((a_n - a_m)^2 - 1)((a_n + a_m)^2 - 1) = 0.$$

or, if  $X_l(nm)$  does not vanish,

$$a_n = \pm a_m \pm 1. \quad (23)$$

There is no sacrifice of generality in taking  $a_m$  as positive and  $\geq 1/2$  throughout. The  $a_m$  thus constitute a series of the form  $C, 1+C, 2+C, \dots$  where  $C$  denotes a constant which is  $\geq 1/2$ . Setting  $a_m = j + 1/2$  yields

$$M^2 = j(j+1)(\hbar/2\pi)^2, \quad (24)$$

and the following selection rule holds for  $j$ :

$$j \rightarrow \begin{cases} j+1 \\ j \\ j-1 \end{cases}$$

This result is formally reminiscent of the values of  $M^2$  which enter the Lande  $g$ -formula.

If for  $\mathbf{M}_z$  we now again introduce the designation  $m(\hbar/2\pi)$ , we find from (12), ch. 4, and the relations

$$\mathbf{M}^2 = \mathbf{M}_x^2 + \mathbf{M}_y^2 + \mathbf{M}_z^2$$

and

$$(\mathbf{M}_x + i\mathbf{M}_y)(\mathbf{M}_x - i\mathbf{M}_y) = \mathbf{M}_x^2 + \mathbf{M}_y^2 - i\varepsilon\mathbf{M}_z = \mathbf{M}^2 - \mathbf{M}_z^2 - i\varepsilon\mathbf{M}_z$$

that

$$\begin{aligned}
M_x(j, m - 1; m) + iM_y(j, m - 1; j, m) \\
= \frac{h}{2\pi} \sqrt{(j(j+1) - m(m-1))}, \\
M_x(j, m; m - 1) - iM_y(j, m; j, m - 1) \\
= \frac{h}{2\pi} \sqrt{(j(j+1) - m(m-1))}.
\end{aligned} \tag{25}$$

For a given value of  $j$ , the maximum value  $M_{\max}$  of  $m$  is characterized by the absence of the jumps  $m_{\max} \rightarrow m_{\max} + 1$ , i.e., the right-hand side of (24), ch. 4, for example vanishes for such jumps. This gives

$$j = m_{\max}.$$

Hence  $j$  also can be “integer” or “half-integer” only.

The calculation of the intensity formula (or the Zeeman effect, e.g., the dependence of  $q_{lx}, q_{ly}, q_{lz}$  upon  $m$ , now appears very simple. From (18), ch. 4, we derive the relations

$$\begin{aligned}
q_{lz} &= (Z_l \mathbf{M}_z + \varepsilon X_l + Y_l) \mathbf{M}^{-2}, \\
q_{lx} + iq_{ly} &= [Z_l - q_{lz} (\mathbf{M}_z + i\varepsilon) + iX_l] (\mathbf{M}_x - i\mathbf{M}_y)^{-1}, \\
q_{lx} - iq_{ly} &= [Z_l - q_{lz} (\mathbf{M}_z - i\varepsilon) - iX_l] (\mathbf{M}_x + i\mathbf{M}_y)^{-1},
\end{aligned} \tag{26}$$

by solving for  $q_{lx}, q_{ly}, q_{lz}$ . These equations also furnish the previously postponed proof that the  $q_{lx}, q_{ly}, q_{lz}$ , can be represented as linear functions of the  $\mathbf{X}_l, \mathbf{Y}_l, \mathbf{Z}_l$ ; with coefficients which for  $\lambda = 0$  are constant with time. At the same time, eqs. (26) ch. 4, include the desired intensity formulae. This can be seen by first noting that the  $X_l, Y_l, Z_l$  are diagonal matrices with respect to  $m$ , since

$$\begin{aligned}
\mathbf{X}_l \mathbf{M}_z - \mathbf{M}_z \mathbf{X}_l &= 0, \\
\mathbf{Y}_l \mathbf{M}_z - \mathbf{M}_z \mathbf{Y}_l &= 0, \\
\mathbf{Z}_l \mathbf{M}_z - \mathbf{M}_z \mathbf{Z}_l &= 0.
\end{aligned} \tag{27}$$

Our problem now resolves itself into two parts, namely discussion of intensities for jumps  $j \rightarrow j$  and  $j \rightarrow j - 1$  (the jumps  $j \rightarrow j + 1$  then do not provide anything new). We first consider the transitions  $j \rightarrow j$ . For these, equation (20), ch. 4 shows that only terms in  $Z_l$  are present. We shall call these terms  $Z_l(j, m)$ . Then, on setting  $M_z = m(h/2\pi)$  and taking note of

(24), ch. 4, eqs. (26), ch. 4 yield:

$$\begin{aligned}
q_{lz}(j, m) &= \frac{2\pi}{\hbar} Z_l(j, m) \frac{m}{j(j+1)}, \\
(q_{lx} + iq_{ly})(j, m-1; j, m) \\
&= \frac{2\pi}{\hbar} Z_l(j, m-1) \sqrt{\frac{j(j+1) - m(m-1)}{j(j+1)}}, \quad (28) \\
(q_{lx} - iq_{ly})(j, m; j, m-1) \\
&= \frac{2\pi}{\hbar} Z_l(j, m) \sqrt{\frac{j(j+1) - m(m-1)}{j(j+1)}},
\end{aligned}$$

Finally, to establish the dependence of the quantity  $Z_l(j, m)$  upon  $m$ , we might use the relation

$$\mathbf{M}_x q_{ly} - q_{ly} \mathbf{M}_x = \varepsilon q_{lz}; \quad (29)$$

it demonstrates in our case that  $Z_l(j, m)$  does not depend on  $m$ . For the transitions  $j \rightarrow j$  we thus obtain:

$$\begin{aligned}
q_{lz}(j, m) : (q_{lx} + iq_{ly})(j, m-1; j, m) : (q_{lx} - iq_{ly})(j, m; m-1) \\
= m : \sqrt{\{j(j+1) - m(m-1)\}} : \sqrt{\{j(j+1) - m(m-1)\}}. \quad (30)
\end{aligned}$$

We treat the jumps  $j \rightarrow j-1$  analogously. For these, according to (21), ch. 4, we have  $X_l(j, m; j-1, m) = (\varepsilon/j) Y_l(j, m; j-1, m)$ . If, using (26), ch. 4, we express the intensities in terms of  $X_l(j, m; j-1)$ , we obtain:

$$\begin{aligned}
q_{lz}(j, m; j-1, m) &= i \frac{2\pi}{\hbar} X_l(j, m; j-1, m) \frac{1}{j}, \\
(q_{lx} + iq_{ly})(j, m-1; j-1, m) \\
&= i \frac{2\pi}{\hbar} X_l(j, m-1; j-1, m-1) \frac{\sqrt{(j-m)}}{j\sqrt{(j+m-1)}}, \quad (31) \\
(q_{lx} - iq_{ly})(j, m; j-1, m-1) \\
&= -i \frac{2\pi}{\hbar} X_l(j, m; j-1, m) \frac{\sqrt{(j+m-1)}}{j\sqrt{(j-m)}}.
\end{aligned}$$

In conclusion, to establish the dependence of the quantity  $X_l(j, m; j - 1, m)$  upon  $m$ , we again use the relation (29), ch. 4, which by way of a simple calculation here yields:

$$X_l(j, m; j - 1, m) = A(j, j - 1) \sqrt{(j^2 - m^2)}. \quad (32)$$

We thus find that

$$\begin{aligned} q_{lz}(j, m; j - 1, m) &: (q_{lx} + iq_{ly})(j, m - 1; j - 1, m) \\ &: (q_{lx} - iq_{ly})(j, m; j - 1, m - 1) = \sqrt{(j^2 - m^2)} : \sqrt{((j - m)(j - m + 1))} \\ &: \sqrt{((j + m)(j + m - 1))}. \end{aligned} \quad (33)$$

The jumps  $j \rightarrow j + 1$  essentially give the same intensities; we here find that

$$\begin{aligned} q_{lz}(j, m; j + 1, m) &: (q_{lx} + iq_{ly})(j, m; j + 1, m + 1) \\ &: (q_{lx} - iq_{ly})(j, m + 1; j + 1, m) = \sqrt{((j + 1)^2 - m^2)} \\ &: \sqrt{((j + m + 2)(j + m + 1))} : -\sqrt{((j - m + 1)(j - m))}. \end{aligned} \quad (34)$$

The formulae (30), (33), (34), ch. 4 agree with the intensity formulae derived from correspondence considerations.<sup>24</sup>

We wish just to draw attention to a simple deduction from (21), ch. 4: *The jumps  $\Delta j = 0$  occur only in the “ $Z_l$ -direction”.* If we consider the motion of a single electron about a nucleus, that is, examine the hydrogen atom, it follows directly from (1), ch. 4, that  $Z$  vanishes.

Hence in this case the jumps  $\Delta j = 0$  never take place.

## 2. The Zeeman effect

If one carries the Lorentz force  $(e/c)[v\mathfrak{H}]$  exerted by a magnetic field  $\mathfrak{H}$  upon an electron over into quantum mechanics, it seems obvious at first that the normal Zeeman Effect ensues for atoms, since under exactly the same assumptions as are introduced to derive Larmor's Theorem classically for the nuclear atom – namely, neglect of terms with  $\mathfrak{H}$  one can derive this theorem

---

<sup>24</sup>S. Goudsmith and R. de L. Kronig, Naturwiss, 13 (1925) 90; H. Hönl, Zs. f. Phys. 32 (1925) 340.

here. There is, nevertheless, a certain difference between classical theory and quantum mechanics insofar as the justification for dropping terms in  $\mathfrak{H}$  is concerned. The neglect of  $\mathfrak{H}$  in classical theory is certainly permissible for orbits of small dimensions and certainly *impermissible* for very large orbits or indeed, hyperbolic orbits. In quantum mechanics all these orbits, be they the innermost or the outermost, are so closely connected with one another as a result of the kinematics specific to quantum mechanics, that indication of the neglect of the quantity  $\mathfrak{H}$  is not immediately apparent. The probabilities of transitions to free electrons are indeed considerable, even from the ground state.

For an oscillator, we are thus sure of the normal Zeemam effect; on the other hand for the nuclear atom it does not seem to be entirely excluded that the intimate connection between innermost and outermost orbits leads to findings which differ somewhat from the normal Zeeman effect. However, we must emphasize that a whole set of weighty reasons speak against tlic possibility of explaining the anomalous Zeeman effects on this basis. Rather, one might perhaps hope that the hypothesis of Uhlenbeck and Goudsmid might later provide a quantitative description of the above-mentioned phenomena.

### 3. Coupled harmonic resonators. Statistics of wave fields

A system of coupled harmonic oscillators given by

$$H = \frac{1}{2} \sum_{k=1}^f \frac{p_k^2}{m_k} + Q(q), \quad (35)$$

with a quadratic form  $Q(q)$  of the coordinates (with numerical coefficients) represents the simplest conceivable system having several degrees of freedom. As was establislied in ch. 2 § 1, the commutation rules remain invariant on simultaneous orthogonal transformation of coordinates and momenta. Therefore, as in classical theory, the system (35), ch. 4, can be transformed into a system of uncoupled oscillators. In particular, the vibrations of a crystal lattice can be analyzed into *eigenvibratons*, just as in classical theory. Each individual eigenvibration is to be treated as a simple linear oscillator according to the manner discussed previously in detail, and the synthesis of the various uncoupled oscillators to a single system is to be undertaken in the way explained in ch. 2 § 1. The same will also apply if we go over to the limiting case of a system with infinitely many degrees of freedom and for

instance consider the vibrations of an elastic body idealized to a continuum or finally of an electromagnetic cavity.

In the previous quantum theory also, vibrations of an electromagnetic cavity constituted the subject of many detailed investigations, since on the one hand the problem of the harmonic oscillator represents just about the simplest problem which can be treated with the methods used hitherto, and on the other hand the familiar result that the energy of an eigenvibration should be an integer multiple of  $h\nu$  exhibits a formal similarity to the fundamental assumptions of the theory of light quanta, so that one might hence expect to gain insight into the nature of light quanta through the consideration of black-body radiation. To be sure, it is clear from the very outset that attacking the problem of light quanta from the above standpoint cannot by any means account for the most important aspect of this problem, namely the phenomenon of coupling of distant atoms, for this problem does not enter at all into the formulation of our questions regarding the vibrations of a cavity. So strong an association between the eigen vibrations of a cavity and the light quanta postulated formerly can nonetheless be drawn that every statistics of cavity eigen vibrations corresponds to a definite statistics of light quanta, and conversely.

Debye<sup>25</sup> has attempted to arrive at such a form of statistics, starting from a distribution of individual light quanta among the eigen vibrations of the cavity. In this manner he was able to derive Planck's formula. However such a mixture of theoretical wave and light-quantum considerations would seem to us hardly to accord with the real nature of the problem. Rather, we believe it to be consistent to separate the theoretical wave-aspect of the problem completely from the theory of light quanta, that is to say, to treat the wave-statistics of black-body radiation throughout by the more general statistical rules applying e.g., to the quantum theory of atomic systems. The statistics applicable to light quanta is then, as we shall show, Bose statistics.<sup>26</sup> This finding hardly seems unnatural, since this statistics has nothing to do with the hypothesis of independent light-corpuscles, but rather is to be regarded as carried over from the statistics of eigen vibrations – which just shows that the assumption of statistically independent light-corpuscles would not meet the case correctly.

However, in each such treatment of cavity radiation by quantum theory hitherto, one encountered the fundamental difficulty that although it led to

---

<sup>25</sup>P. Debye, Ann. d. Phys. 33 (1910) 1427; cf. also P. Ehrenfest, Phys. Zs. 7 (1906) 528.

<sup>26</sup>S.N. Bose, Zs. f. Phys. 26 (1924) 178.

Planck's law of radiation, it did not yield the correct mean square deviation of energy in an element of volume. One thus finds that a consistent treatment of the natural vibrations of a mechanical system or an electromagnetic cavity in accordance with past theory leads to most serious contradictions. This caused us to hope that the modified kinematics which forms an inherent feature of the theory proposed here would yield the correct value for the interference fluctuations, thus precluding the above contradictions and opening the possibility of setting up a consistent system of statistics for black-body radiation.

The states of the system of oscillators can be characterized by "quantum numbers"  $n_1, n_2, n_3 \dots$  of the individual oscillators, so that apart from an additive constant the energies of the individual states are given by

$$E_n = h \sum_m \nu_k n_k. \quad (36)$$

The additive constant, the so-called *zero-point* energy is

$$C = \frac{1}{2} h \sum_k \nu_k \quad (36')$$

(in particular, for the limiting case of infinitely many degrees of freedom, it would be infinitely large). From now on, let us simply call the quantity  $E_n$  in (36), ch. 4, the *thermal energy*. In accordance with what was stated in Part I, the same statistical weight is to be attached to each of the states of the system characterized by a certain set of values  $n_1, n_2, n_3, \dots$ . The consequences of this can immediately be perceived on the basis of the following remark: If waves are propagated with a phase velocity  $v$  in an  $s$ -dimensional isotropic part of space  $V = l^3$  the number of *eigen vibrations* for the frequency range  $d\nu$  is equal to the number of "cells" for  $d\nu$  (in the Bose-Einstein sense), and this in fact holds for arbitrary  $s$ , hence e.g. also for *vibrating membranes* or *strings*. This follows from the fact that, if we omit consideration of polarization properties, etc., the number of eigen vibrations for the range  $d\nu$  is furnished by the solution of the problem of determining the number of ways in which one can choose a set of positive integers  $m_1 \dots m_s$  such that the  $\nu$  determined by the relation

$$\frac{2l}{v} \nu = \sqrt{(m_1^2 + \dots + m_s^2)}$$

falls within the interval  $d\nu$ . If  $K_s(a)$  be the volume of an  $s$ -dimensional sphere of radius  $a$ , there are  $(V/v^s)K_s(\nu)$  eigen vibrations which have a frequency less than  $\nu$ . On the other hand, the number of cells for the range  $d\nu$

can be determined as follows: The momentum components  $p_1, \dots, p_s$  of the quantum satisfy the equation

$$h\nu/v = \sqrt{(p_1^2 + \dots + p_s^2)},$$

and the size of the cells is  $h^s$  in the  $2s$ -dimensional phase space. One can see from this that the number of cells belonging to a frequency lower than  $\nu$  is also equal to  $(V/v^s)K_s(\nu)$ . Hence, as mentioned above, one can effect a one-to-one correspondence of cells to eigen vibrations such that the individual pairs always belong to the same  $d\nu$ . This correspondence can, incidentally, be so carried out that the *directions* of an eigen vibration and those of the light quanta in the respective cell fall within the same infinitesimal angular range. From (36), ch. 4, the quantum number of an *oscillator* is then to be set equal to the *number of quanta in the appropriate cell*. Every system of light-quanta statistics yields an associated statistics of natural vibrations and conversely. It can be seen that the statement made above concerning the weighting of the states of the system of oscillators goes directly over into the basic postulate of Bose-Einstein statistics because of this association. The equally probable complexions are defined through a declaration of the number of quanta sitting in each cell. <sup>27</sup>

In Debye statistics, the number of oscillators involving  $\nu$  quanta is (except for a factor which depends on  $\nu$  only) equal to

$$\frac{1}{r} e^{-r(h\nu/kT)}, \quad (37)$$

and Planck's law arises from

$$\sum_{r=1}^{\infty} e^{-r(h\nu/kT)} = \frac{1}{e^{h\nu/kT} - 1}.$$

It is unsatisfactory that eq. (37), ch. 4, holds only for  $r > 0$  and does not also give the number of oscillators involving no quanta. From the new point of view, we have to replace (37), ch. 4, according to Bose <sup>28</sup> by

$$(1 - e^{-h\nu/kT}) e^{-r(h\nu/kT)}, \quad (38)$$

---

<sup>27</sup>A. Einstein, Sitzungsber.d.preuss. Akad.d.Wiss. (1925) p. 3. Our considerations naturally cannot yield any fresh viewpoint for the valuation of Einstein's hypothesis that this form of statistics is also applicable to an ideal gas.

<sup>28</sup>This expression naturally has to be assumed for example also in the case of elastic waves in a continuum, which necessitates a certain modification to considerations by Schrödinger (Phys. Zs. 25, (1924) 89) concerning the thermal equilibrium between light- and sound-beams. This modification can easily be carried out in analogy with the probability theorem for the Compton effect on assuming Einstein's gas theory to be valid, as has earlier been pointed out (P. Jordan, Zs. f. Phys. 33 (1925) 649.)

winch (to use the terminology of the theory of light quanta) give's the number of "r-fold occupied cells", and Planck's formula results from

$$\sum_{r=0}^{\infty} r(1 - e^{h\nu/kT}) e^{-r(h\nu/kT)} - \frac{1}{e^{h\nu/kT} - 1}.$$

The light-quanta statistics corresponding to Debye's vibration statistics is represented by the theory developed by Wolfke <sup>29</sup> and Bothe. <sup>30</sup> To be sure, these authors do not speak of r-fold occupied cells, but designate (37), ch. 4, as the number of "r-quantal light-quanta molecules".

As is known, the above-mentioned shortcomings of classical wave theory become evident in the study of energy deviations in the radiation field as follows: If there is communication between a volume  $V$  and a very large volume such that waves having frequencies which lie within a small range  $\nu$  to  $\nu + d\nu$  can pass unhindered from one to the other, whereas for all other waves the volumes remain detached, and if  $E$  be the energy of the waves with frequency  $\nu$  in  $V$ , then according to Einstein the mean square deviation  $\bar{\Delta}^2 = (\bar{E} - E)^2$  can be calculated by an inversion of the Boltzmann Principle. If  $z_\nu d\nu$  be the number of eigen vibrations (cells) in the range  $d\nu$  per unit volume, so that

$$\bar{E} = \frac{z_\nu h\nu}{e^{h\nu/kT} - 1} \cdot V, \quad (39)$$

then it follows that

$$\bar{\Delta} = h\nu\bar{E} + \frac{\bar{E}^2}{z_\nu V}. \quad (40)$$

If, however, one calculates the energy deviations from interferences in the wave field, classical theory yields only the second summation term in (40), ch. 4, as has explicitly been shown by Lorentz. <sup>31</sup> This discrepancy naturally also exists (plite generally for such waves as those in a crystal lattice or in an elastic continuum. According to Ehrenfest, <sup>32</sup> its origins are to be sought in the fact that in the Einstein treatment, *additivity of the entropies* of  $V$  and of the large volume was assumed. However, this additivity of entropies applies, according to the classical theory of natural vibrations, only in the

---

<sup>29</sup>M. Wolfke, Phys. Zs. 22 (1921) 375.

<sup>30</sup>W. Bothe, Zs. f. Phys. 20 (1923) 145; 23 (1924) 214.

<sup>31</sup>H.A. Lorentz, Les Theories Statistiques en Thermodynamique (Leipzig. 1916). p. 59.

<sup>32</sup>P. Ehrenfest, Lecture in the Gö ttingen seminar on the Structure of Matter, Summer 1925. The contents of this lecture were of great assistance to our present consideratoins. Meanwhile published in Zs. f. Phys. 34 (1925) 362.

region of validity of the Rayleigh–Jeans Law. Precisely the nonexistence of statistical independence of the volume elements in the general case is so unnatural a result of the theory of cavity radiation to date that one is obliged to conclude that this theory breaks down even in the simple problem of the harmonic oscillator.

We now calculate the mean square deviation  $\Delta^2$  from the interferences using quantum mechanics. To avoid calculational complications which have no bearing upon the nature of the case, we base ourselves on the simplest conceivable case, namely a *vibrating string* fastened at its ends. Incidentally, all essential points of the calculation can immediately be taken over in more general instances. We first cite the classical approach.

Let the length of the string be  $l$  and its lateral displacement be  $u(x, t)$ . On introducing the Fourier coefficients  $q_k(t)$  as given by

$$u(x, t) = \sum_{k=1}^{\infty} q_k(t) \sin k \frac{\pi}{l} x, \quad (41)$$

or

$$q_k(t) = \frac{2}{l} \int_0^l u(x, t) \sin k \frac{\pi}{l} x dx \quad (41')$$

as coordinates, the energy of the string goes over into a sum of squares. Namely, for suitable choice of units,

$$\begin{aligned} H = \frac{1}{2} \int_0^l & \left\{ u^2 + \left( \frac{\partial u}{\partial x} \right)^2 \right\} dx = \frac{l}{4} \sum_{k=1}^{\infty} \left\{ \dot{q}_k(t)^2 + \right. \\ & \left. \left( k \frac{\pi}{l} \right)^2 q_k(t)^2 \right\}. \end{aligned} \quad (42)$$

More generally, for the energy  $E$  in a segment  $(0, a)$  of the string, we obtain

$$\begin{aligned} E = \frac{1}{2} \int_0^a & \sum_{j,k=1}^{\infty} \left\{ \dot{q}_j \dot{q}_k \sin j \frac{\pi}{l} x \sin k \frac{\pi}{l} x \right. \\ & \left. + q_j q_k k \left( \frac{\pi}{l} \right)^2 \cos j \frac{\pi}{l} x \cos k \frac{\pi}{l} x \right\} dx. \end{aligned} \quad (43)$$

If in (43), ch. 4, we take only the terms with  $j = k$ , we find (under the explicit assumption that all wavelengths which come into consideration are small with respect to  $a$ ) just the value  $(a/l)H$ . From this one sees: The difference

$$\Delta = E - \bar{e},$$

wherein the bar represents an average over the phases  $\varphi_k$  in

$$q_k = a_k \cos(\omega_k t + \varphi_k); \quad \omega_k = k \frac{\pi}{l}, \quad (44)$$

can be derived from (43), ch. 4, by omitting terms of the sum which have  $j = k$ . This phase average is identical with the time average. On carrying out the integration, one then finds

$$\Delta = \frac{1}{4} \sum_{j,k=1}^{\infty} \sum_{j \neq k} \left\{ \dot{q}_j \dot{q}_k K_{jk} + j k q_j q_k \left( \frac{\pi}{l} \right)^2 K'_{jk} \right\}, \quad (45)$$

with

$$\begin{aligned} K_{jk} &= \frac{\sin(j-k) \frac{\pi}{l} a}{(j-k) \frac{\pi}{l}} - \frac{\sin(j+k) \frac{\pi}{l} a}{(j+k) \frac{\pi}{l}} \\ &= \frac{\sin(\omega_j - \omega_k)a}{\omega_j - \omega_k} - \frac{\sin(\omega_j + \omega_k)a}{\omega_j + \omega_k}, \\ K'_{jk} &= \frac{\sin(j-k) \frac{\pi}{l} a}{(j-k) \frac{\pi}{l}} + \frac{\sin(j+k) \frac{\pi}{l} a}{(j+k) \frac{\pi}{l}} \\ &= \frac{\sin(\omega_j - \omega_k)a}{\omega_j - \omega_k} + \frac{\sin(\omega_j + \omega_k)a}{\omega_j + \omega_k}. \end{aligned} \quad (45')$$

In consideration of later quantum-mechanical calculations, we write out the mean square deviation  $\overline{\Delta^2}$  explicitly. It is

$$\Delta^2 = (\Delta_1 + \Delta_2)^2 = \Delta_1^2 + \Delta_2^2 + \Delta_1 \Delta_2 + \Delta - 2\Delta_1, \quad (46)$$

with

$$\begin{aligned} \Delta_1^2 + \Delta_2^2 &= \frac{1}{16} \sum_{j,k=1}^{\infty} \sum_{j \neq k} \sum_{i,\kappa=1}^{\infty} \sum_{i \neq \kappa} \left\{ \dot{q}_j \dot{q}_k \dot{q}_i \dot{q}_{\kappa} K_{jk} K_{i\kappa} \right. \\ &\quad \left. + j k i \kappa \left( \frac{\pi}{l} \right)^4 q_j q_k q_i q_{\kappa} K''_{jk} K'_{i\kappa} \right\}; \end{aligned} \quad (46')$$

$$\Delta_1 + \Delta_2 + \Delta_2 \Delta_1 = \frac{1}{16} \sum_{j,k=1}^{\infty} \sum_{j \neq k} \sum_{i,\kappa=1}^{\infty} \sum_{i \neq \kappa} \left( \frac{\pi}{l} \right)^2 \left\{ j k q_j q_k \dot{q}_i \dot{q}_{\kappa} K'_{jk} K_{i\kappa} \right.$$

$$\left. + i \kappa \dot{q}_j \dot{q}_k q_i q_{\kappa} K_{jk} K'_{i\kappa} \right\} \quad (46'')$$

Equation (44), ch. 4, implies  $\overline{\Delta_1 \Delta_2 + \Delta_2 \Delta_1} + 0$  and

$$\overline{\Delta^2} = \overline{\Delta_1^2} + \overline{\Delta_2^2} = \frac{1}{8} \sum_{j,k=1}^{\infty} \left\{ \overline{\dot{q}_j^2 \dot{q}_k^2} K_{jk}^2 + j^2 k^2 \left( \frac{\pi}{l} \right)^4 \overline{q_j^2 q_k^2} K'_{jk}^2 \right\}. \quad (47)$$

If we now let the string's length  $l$  become very large, the  $\omega_k$  get ever closer together, according to (44), ch. 4, so that the sum (47), ch. 4, goes over into an integral:

$$\overline{\Delta^2} = \overline{\Delta_1^2} + \overline{\Delta_2^2} = \frac{1}{8} \int_0^\infty \int_0^\infty d\omega_k \frac{l^2}{\pi^2} \left\{ \overline{q_j^2 q_k^2} K_{jk}^2 + j^2 k^2 \left(\frac{\pi}{l}\right)^4 \overline{q_j^2 q_k^2} K'_{jk}^2 \right\}. \quad (47')$$

Finally, we also assume the “volume”  $a$  to become very large and employ the relation

$$\lim_{a \rightarrow \infty} \frac{1}{a} \int_{-\Omega}^{\Omega'} \frac{\sin^2 \omega_\alpha}{\omega^2} / (\omega) d\omega = \pi/(0) \quad \text{for} \quad \Omega, \Omega' > 0. \quad (48)$$

We then see that only the first sum terms  $(\sin(\omega_j - \omega_k)a)/(\omega_j - \omega_k)$  in (45), ch. 4, provide an appreciable contribution, and we find for (47'), ch. 4,

$$\overline{\Delta^2} = \frac{al}{8\pi} \int_0^\infty d\omega \left\{ (\overline{\dot{q}_\omega^2})^2 + (\omega^2 \overline{q_\omega^2})^2 \right\}. \quad (49)$$

On the other hand, by (42), ch. 4, the mean energy in the volume  $a$  becomes equal to

$$E = \frac{a}{l} \cdot \frac{l}{4} \int_0^\infty d\omega \frac{l}{\pi} \cdot \left\{ \overline{\dot{q}_\omega^2} + \omega^2 \overline{q_\omega^2} \right\} = \frac{al}{4\pi} \int_0^\infty \omega \left\{ \overline{\dot{q}_\omega^2} + \omega^2 \overline{q_\omega^2} \right\}. \quad (50)$$

Therein we have

$$\overline{\dot{q}_\omega^2} = \omega^2 \overline{\dot{q}_\omega^2}, \quad (51)$$

a relation which, as we would recall, remains valid in quantum theory too, according to ch. 1. In order to obtain the quantities  $\overline{\Delta^2}, E$  employed in (39), (40), ch. 4, we have merely to extract those parts referring to  $\nu = d\omega/2\pi$  from (49), (50), ch. 4, and to divide these by  $d\nu$ . With  $v = a$  we then obtain

$$\overline{\Delta^2} = \frac{\overline{E^2}}{2v} \quad (52)$$

We see from (44), ch. 4, that in our case  $z_\nu = 2$ , since

$$d\omega_k = 2\pi d\nu_k = \frac{\pi}{l} dk.$$

Thence (52), ch. 4, in fact gives precisely the second term in (40), ch. 4.

On going over to quantum mechanics, we have to regard (41), (41'), (42), (43), ch. 4, as matrix equations for  $u, H, q, E$ . The quantity  $x$ , however, remains a number, since if in place of the continuous string we consider an elastic series of points,  $x$  would denote the *number* (multiplied by the lattice constant) of any given point.

The matrix  $q_k$  has  $2f$  dimensions if  $f$  be the number of eigen vibrations, i.e., infinitely many in the case of an elastic string. Each of the components  $q_k(nm)$  of  $q_k$  vanishes except for those with

$$\left. \begin{aligned} n_j - m_j &= 0 & \text{for } j \neq k, \\ n_k - m_k &= \pm 1. \end{aligned} \right\} \quad (53)$$

The phase average of a matrix is that diagonal matrix which coincides with the diagonal of the respective matrix. From (53), ch. 4, in part similar conclusions can be drawn to those derivable from (44), ch. 4. The considerations which formerly led to (46), (46'), (46''), ch. 4, remain valid in quantum theory. The formulae (47), (47'), ch. 4 with matrices  $q_k$  also hold for the diagonal matrix  $\overline{\Delta_1^2 + \Delta_2^2}$  and finally, according to (52), ch. 4, if we denote those parts of  $\overline{\Delta^2}$  which belong to a given frequency  $\nu$  as  $\overline{\Delta^2}$ , we find

$$\overline{\Delta_1^2 + \Delta_2^2} = \frac{E^{*2}}{2v}. \quad (52')$$

The quantity  $E^*$  in (52'), ch. 4, is, by (49), (50), (51), ch. 4, no longer the mean thermal energy, but rather the sum of this and the zero-point energy: from the elementary oscillator formulae, we have

$$\overline{E^*} = h\nu \cdot V \bar{E},$$

$$\overline{\Delta_1^2 + \Delta_2^2} = \frac{1}{2}(h\nu)^2 V + h\nu \bar{E} + \frac{\overline{E^2}}{2V}, \quad (54)$$

since for  $d\nu$  the zero-point energy becomes equal to

$$\frac{v}{l} \cdot \frac{h\nu}{2} \cdot l z_\nu d\nu = h\nu \cdot V d\nu.$$

We now still have to consider  $\overline{\Delta_1 \Delta_2 + \Delta_2 \Delta_1}$ . In treating this quantity in just the same way as  $\overline{\Delta_1^2 + \Delta_2^2}$  we obtain, in accordance with (49), ch. 4, the expression:

$$\overline{\Delta_1 \Delta_2 + \Delta_2 \Delta_1} = \frac{al^2}{8\pi} \int d\omega \cdot \omega^3 \{(q_\omega \dot{q}_\omega)^2 + (\dot{q}_\omega q_\omega)^2\}.$$

However, since the quantity  $1/2l$  is, from (42), ch. 4, to be regarded as the “mass” of the resonators, the commutation rules give us

$$-q_j \dot{q}_j(nn) = \dot{q}_j q_j(nn) = \frac{1}{2} \cdot \frac{2}{l} \cdot \frac{h}{2\pi i} = \frac{h}{2l\pi i}.$$

Hence the part  $\overline{\Delta_1\Delta_2 + \Delta_2\Delta_1}$  of  $\overline{\Delta_1\Delta_2 + \Delta_2\Delta_1}$  which belongs to  $d\nu$  is, after division by  $d\nu$ , equal to

$$\overline{\Delta_1\Delta_2 + \Delta_2\Delta_1} = -\frac{1}{2}(h\nu)^2 V,$$

and, with (54), ch. 4, we have in fact

$$\overline{\Delta^2} = h\nu \bar{E} + \frac{\overline{E^2}}{z_\nu V}, \quad (55)$$

which agrees with (40), ch. 4.

If one bears in mind that the question considered here is finally somewhat remote from the problems whose investigation led to the growth of quantum mechanics, the result (55), ch. 4, can be regarded as particularly encouraging for the further development of the theory.

From Ehrenfest's finding mentioned above, one could save oneself calculation of energy deviations involving interference considerations and at the same time acquire the assurance that also for other similar problems no contradictions are possible – if the additivity of the entropies of volume elements could directly be proved in the quantum mechanics of wave fields. Our above findings lead us to expect this additivity to hold generally.

The reasons leading to the appearance in (55), ch. 4, of a term which is not provided by the classical theory are obviously closely connected with the reasons for occurrence of a zero-point energy. The basic difference between the theory proposed here and that used hitherto in both instances lies in the characteristic kinematics and not in a disparity of the mechanical laws. One could indeed perceive one of the most evident examples of the difference between quantum-theoretical kinematics and that existing hitherto on examining the formula (55), ch. 4, which actually involves no mechanical principles whatsoever.

If the quantum mechanics proposed here should prove to be correct even in its essential features, we might quite generally designate the following as constituting the most important advance of this as against the past theory: that in our theory, kinematics and mechanics have again been brought into as close a relationship as that prevailing in classical theory, and that the new

fundamental viewpoints, stemming as they do from the basic postulates of quantum theory for the mechanical concepts together with the concepts of space and time, find adequate expression in kinematics just as in mechanics and in the connection between kinematics and mechanics.





