A Survey of the Facts and the Theories of the Aurora

E. W. HEWSON*
University of Toronto, Toronto, Canada

CONTENTS

INTRODUCTION

FROM earliest times the phenomena of polar lights have been a source of wonderment and conjecture, but no completely satisfactory explanation of their form and cause has yet been suggested. It is the purpose of the present paper to summarize briefly the auroral observations and to describe the several theories which have been advanced in explanation.

Continuing the investigations started during the first polar year of 1882–83, a row of arctic and antarctic stations made observations throughout the second polar year 1932–33. Auroral investigation figured prominently on the program of nearly all the expeditions. Only a fraction of the information gained has been published yet, but it is expected that the organization and comparison of data from all stations will result in the solution of a number of problems.

I. FORM AND OCCURRENCE OF AURORA

A. Types

There are many types of auroral displays, and occasionally several types appear simultaneously. Störmer[61] has separated aurorae into two main classes: those with ray structure, including draperies, rays and the corona, and those without ray structure, comprising homogeneous arcs, homogeneous bands, and pulsating surfaces. Fig. 1 shows draperies with ray structure, photographed at Kongsberg, Norway, on January 24, 1936. For further photographs of these different types, reference may be made to the atlas of aurora published by the International

* At present, Beit Scientific Research Fellow, Imperial College of Science and Technology, London.

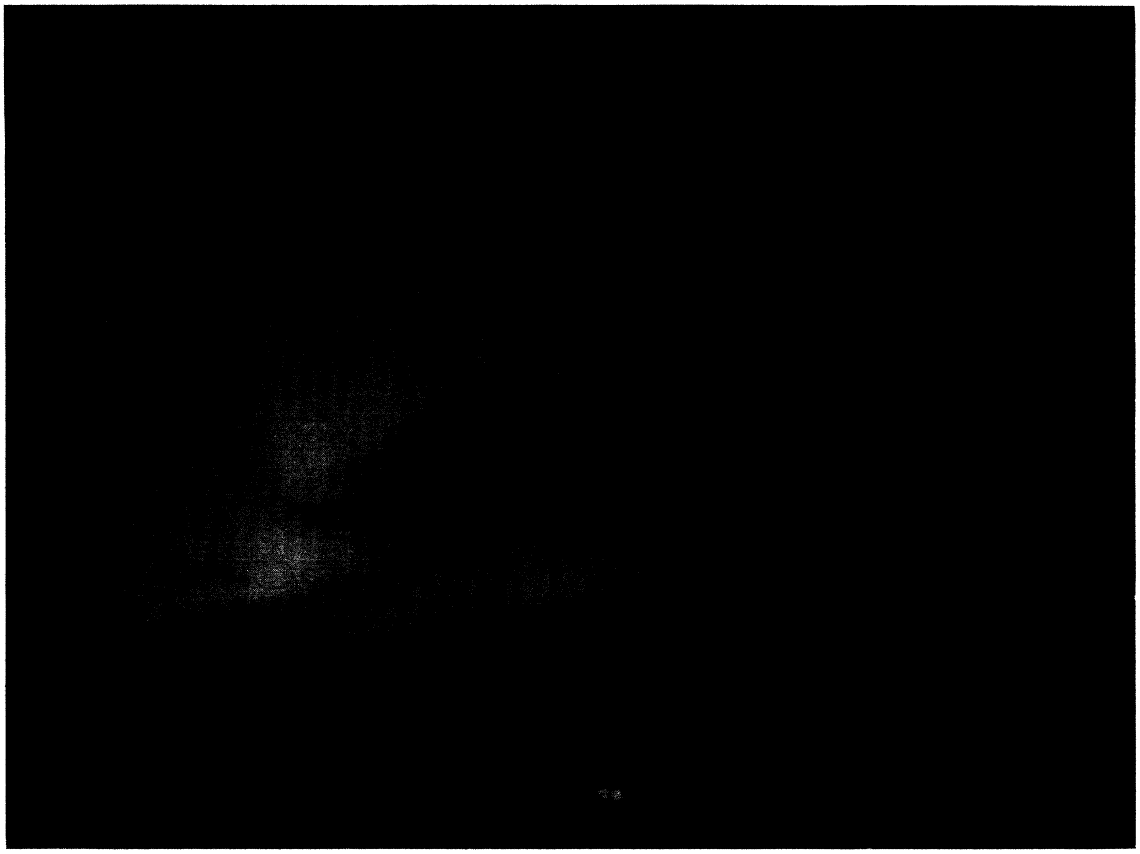

Fig. 1. Draperies with ray structure, Kongsberg, Norway, January 24, 1936. Phot. Busengdal.

Union of Geodesy and Geophysics.[58] To the average American or European observer the most common forms are arcs and rays.

The arc has its highest point on the magnetic meridian, and often seems to extend nearly to the horizon in the northeast and northwest. Since the lower limits of the arc are more sharply defined, the sky below appears darker than that above. When the arc is bright, the lower part will usually be red, the middle yellow and the upper part green. The arc may be visible alone, but frequently from it, rays appear to spread out like the spokes of a fan, giving a combination of ray and nonray structure. Sometimes the arc may be pulsating.

The rays can appear in great masses, in bundles, or singly. Sometimes they are stationary, simply appearing and disappearing without seeming to move. At other times there is a rapid cross motion, and again they seem to shoot rapidly upward and then recede.

Further north the direction of the lines of force of the earth's magnetic field become more nearly perpendicular to the earth's surface, and the corona effect is seen, with rays spreading in all directions from this central ring of light, which is situated near the observer's zenith. This is the most striking and beautiful of all types of aurora.

The draperies have very long rays which hang like a curtain and may have irregular outlines, with the lower bands intensely lighted. The draperies have the appearance of a fan when the rays follow the magnetic lines of force of the earth's field. These are rarely seen in the lower latitudes.

The homogeneous bands are similar to the draperies, but without ray structure. There may be several parallel bands in the sky at the same time.

The pulsating surfaces look like diffuse clouds illuminated by a source emitting light of rapidly changing intensity.

B. Geographical distribution

Auroral displays are seen in the vicinity of both poles. In the northern hemisphere, they are known as the aurora borealis, and in the southern hemisphere as aurora australis. Since few observations have been taken in Antarctica[19] compared with the great number taken in the northern hemisphere, these latter necessarily form the basis for any complete treatment of aurora.

Figure 2 shows the well-known chart of the lines of equal frequency drawn by Fritz. The thickest line corresponds to the zone of maximum frequency, the auroral zone strictly speaking. Northern Canada and Norway are most favorably located for the study of northern lights. The exact situation of the similar zone in antarctic regions has not yet been determined.

The aurora seems to be connected in some way with magnetic storms and solar activity, and sometimes shows similar periods of frequency and intensity. Especially during more severe magnetic storms, the auroral zones extend toward the equator, and the aurora can be seen over nearly the whole world, sometimes even in the tropics.

C. Variations in activity

Periodicities have been noted in auroral appearances. A clearly defined diurnal frequency maximum was observed by the Canadian International Polar Year party which was situated at Chesterfield in Northern Canada.[15] This maximum occurred at 7 hours, Greenwich mean time, i.e., 1 hour local mean time. This is practically at local magnetic midnight for Chesterfield. It was noted that quiet bands and arcs predominated in the early evening hours, and forms with ray structure just before daylight in the morning. The French Polar Year group which was situated at Scoresby Sound on the east coast of Greenland found a very definite periodicity, coinciding with the sun's rotation period of twenty-seven days.[18] This confirms the similar result which earlier observers have noticed. It has been well known for some time that auroral activity closely follows the eleven year cycle of sunspot activity.

The Byrd Expedition to Antarctica in 1929 from auroral observations taken during the seven months that such observations were possible there, found definite periods.[19] Analysis of the 1415 separate aurorae observed showed a diurnal maximum at about 2 to 3 hours, local mean time. At that location, local mean time corresponded very nearly to 165th west meridian time. A short period of from 4 to 6 days appeared, possibly due to the accidental spacing of sunspot groups at these intervals, and a 26 to 30 day period, corresponding to the sun's rotation, was clearly evidenced as well.

D. Connection between auroral activity and magnetic storms and abnormal earth currents

Comparisons of magnetic and auroral records have in many cases indicated a lack of relationship between the two phenomena.[66] Some observers think, however, that they have good evidence that a relationship does exist. Dauvillier at Scoresby Sound found in the first quarter of 1933 a very definite periodicity in auroral activity corresponding to the 27 day sun rotation period. Comparing the auroral and magnetic curves with the solar activity curve a remarkable fact appeared—the two curves were opposite in phase.[18] During this time a very intense focus of activity had developed during three rotations of the sun, presenting a strong maximum during the first half of February. Yet it was at this time that the only two blank periods of auroral activity were observed. On the other hand, during the second half of February, when solar activity was nil, exceptionally intense auroral displays, the most brilliant of the year, were observed (February 20, 21 and 22). Although lags of from 10 hours to 3 days have been recorded before, Dauvillier believes that this is first time that a lag of one-half the period of the sun's rotation has been so clearly demonstrated.

The Byrd Expedition found marked evidence that an auroral maximum occurred on the same day or one day after a maximum on the magnetic character curve.[19]

Rooney, comparing auroral observations and earth current records from the College Fairbanks Station in Alaska during the Polar Year, found considerable agreement between aurorae and disturbances in the earth current records.[46] Records for 80 clear nights gave coefficients of linear correlation from 0.71 to 0.76. Effects associated with brilliant isolated auroral displays at College were readily detected at Tucson, Arizona, and in exceptional cases as far south as Huancayo, Peru. The highest correlation is between oscillatory disturbances in the earth current records and moving types of aurora.

II. Method of Determination of Height

The height of an aurora may be determined by the usual method of making simultaneous observations on a certain luminous point therein from two stations whose distance apart is known. Since such observations have been made photographically with considerable success, it is well to consider the advances which have been made in photographing the aurora.

A. Auroral photography

The great difficulty in this work has been to procure a combination of lens and plate which was fast enough to record the instantaneous position of the aurora. In 1892, the German scientists Brendel and Baschin succeeded in obtaining a photograph of an intense drapery aurora with an exposure of only seven seconds. In 1909 Störmer, after trying a series of lenses and plates, succeeded in obtaining good photographs of intense displays with exposures as short as one second, by using a motion-picture camera lens of diaphragm 25 mm and focal length 50 mm and Lumiere etiquette-violette plates. Between 1909 and 1931 many new combinations were tried by Störmer, Krogness and by Harang, the director of the Northern Light Observatory at Tromsö,[59] some of which gave improved results. From December 1931 to April 1932, intensive research on this problem was undertaken at the Tromsö Observatory.[60] By comparing different lenses, it was shown that the Astro RK lens is so much superior to the next best lens obtainable and hitherto employed for photographing the aurora borealis, that it was possible to reduce the time of exposure to one-third. By selecting suitable photographic emulsions, and sensitizing the plate material according to the Schmiescheck principle, the time of exposure was again reduced to a fraction as compared with ordinary material. Through these improvements it was possible to obtain instantaneous photographs of even moderately bright aurora with time exposures of less than half a second, a speed about ten times greater than any previously attained for the same intensity.

This great reduction in exposure has made color photography, motion pictures and a more complete spectral study of aurora possible.

B. General statement of height observations

The first determinations of the height of aurorae were made in Norway, when Störmer working at Bossekop in 1910 and 1913, took photographs of aurorae simultaneously from two stations connected by telephone.[52] Much better results were obtained in 1913 by increasing the base line of four and a half kilometers, the distance between stations in 1910, to twenty-seven and a half kilometers. Since then, the photographic determination of height has been continued by Störmer from a network of auroral stations in Southern Norway, and a great deal of observational data collected. The base lines vary between 26 and 400 kilometers, and frequently simultaneous photographs have been taken from three or four stations to check the height calculations. Vegard and Krogness[67] and Harang and Tønsberg[23] also made height determinations, the former at the Haldde Observatory near Bossekop, and the latter at the Northern Light Observatory at Tromsö.

Northern Canada is very favorably situated for auroral study. In 1931, McLennan, Wynne-Edwards and Ireton published the results of height determinations made from a main station situated at about 50°–40′ north latitude, and 81°–25′ west longitude.[39] More comprehensive height measurements were made by the Canadian Polar Year groups at Chesterfield, which is situated in the maximal frequency zone,[15] and at Saskatoon, Saskatchewan.[1]

C. Computation of height from photographic plates

An outline will now be given of the method introduced by Störmer[52a] for determining the height of a point in an

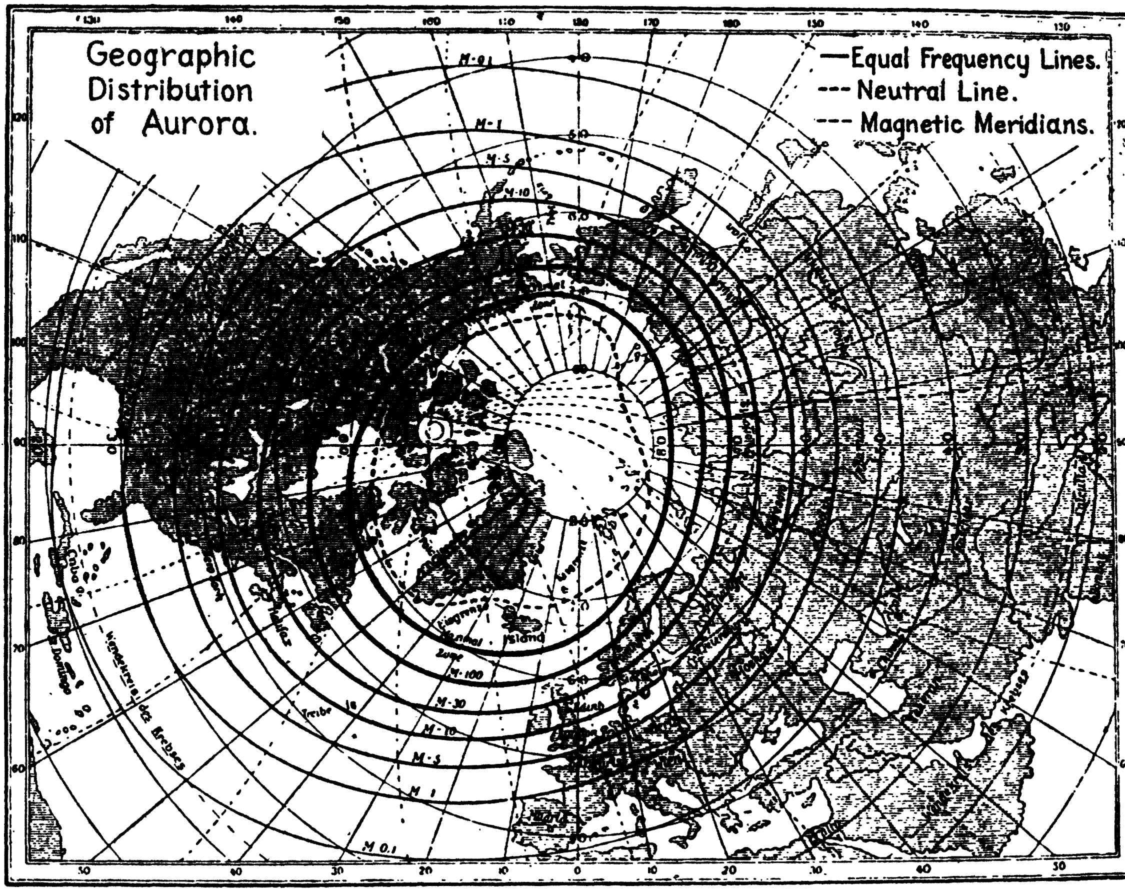

FIG. 2. Map showing the geographic distribution of aurora. (The star indicates approximately the point of intersection in 1900 of the earth's magnetic axis with the Northern Hemisphere.)

aurora, and slight modifications thereof by Vegard and Krogness[67] and Harang and Tønsberg.[23]

It is required to find the position in space of a certain point in an aurora relative to the principal station. The point is specified by three coordinates, the altitude h, the azimuth a and the distance r from the station.

These three quantities must be calculated from the following observations.

1. A pair of photographs taken from two stations, one of which is the principal station. Besides the aurora, the pictures must show at least two known stars, and preferably three.

2. The time at which the exposures were made.

The angle corresponding to 1 mm on any part of the photograph may be found from the known angular distance between the two stars which have been identified on the plate, and thus the angular distance between any two points on the plates obtained. To get absolute values of the three quantities in question, the coordinates of some point, say a star, on the plate, and the direction along which the angular differences are to be measured, must be known.

The process of determining these standard data has been called the orientation of the picture. When this has been done, the various points of the aurora may be measured, and thus their position in space found.

(a) *Theory of the method*

The orientation of the negatives.—The plates are arranged in corresponding pairs, and the identified stars, three if possible, as well as the center of each plate, should be marked for future identification, perhaps with ink (on the glass side of the plate in order not to spoil the negative). The declination and right ascension are found from astronomical tables, and recorded, and the hour angles computed in the usual way.

Figure 3 shows a geocentric celestial sphere, with its center at the principal station A, and with the celestial pole, the zenith, and the points where the direction from the principal station A to the secondary station B cuts the sphere denoted respectively by P, Z and D and D'. The straight line between A and B is known as the base-axis, and the points D and D' the base-poles. If S denotes a star, then the base-distance u means the angular distance from D to S measured along the great circle through these points. The base-height ω, i.e., the angle EDS, is the complement of the angle between the vertical plane

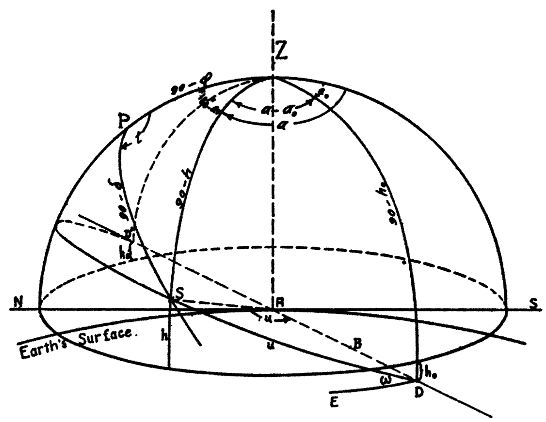

FIG. 3.

through ADZ and the plane through ADS, which may be called the plane of displacement. The base-distance u and the base-height ω represent a pair of spherical coordinates which are referred to the plane through ADE which is the base-plane, and the base-pole D itself.

By using spherical trigonometry, Störmer obtained the following equations:[52e]

$$\sin h = \sin \varphi \sin \delta + \cos \varphi \cos \delta \cos t, \quad (1)$$

$$\sin a = \cos \delta \sin t \sec h, \quad (2)$$

$$\cos a = -\cos \varphi \sin \delta \sec h + \sin \varphi \cos \delta \cos t \sec h, \quad (3)$$

$$\cos u = \sin h_0 \sin h + \cos h_0 \cos h \cos (a - a_0). \quad (4)$$

When $h_0 = 0$

$$\cos u = \cos h \cos (a - a_0), \quad (4')$$

$$\cos \omega = \cos h \sin (a - a_0) \operatorname{cosec} u. \quad (5)$$

In these equations the letters signify:

φ—geographical latitude of the principal station,
δ—declination of the star,
t—hour-angle of the star,
h—altitude of the star,
a—azimuth of the star,
u—base-distance of the star,
ω—base-height of the star,
h_0—altitude of the base-pole,
a_0—azimuth of the base-pole.

Thus these equations determine the angular distance between D and S, and also the direction of DS. Knowing the values of these two fixes the position of the star with reference to the base-axis AB at any given time. The positions of other stars may be similarly fixed at given times.

Consider in Fig. 4 that S_1 and S_2 represent two stars, with angular distance a between them. B_1 and B_2 are the angles which the declination circles form with arc S_1S_2.

Spherical trigonometry gives the following equations

$$\tan \tfrac{1}{2}(B_1 - B_2) = \sin \tfrac{1}{2}(\delta_1 + \delta_2) \sec \tfrac{1}{2}(\delta_1 - \delta_2) \tan \tfrac{1}{2}(\Delta t),$$

$$\tan \tfrac{1}{2}(B_1 + B_2) = \cos \tfrac{1}{2}(\delta_1 + \delta_2) \operatorname{cosec} \tfrac{1}{2}(\delta_1 - \delta_2) \tan \tfrac{1}{2}(\Delta t),$$

where Δt is the difference in hour-angle between the two stars. The angular distance, which is needed for the measurement of angular distances between points on the plate, will be given by the expression

$$\tan \tfrac{1}{2}a = \cos \tfrac{1}{2}(B_1 - B_2) \sec \tfrac{1}{2}(B_1 + B_2) \tan \tfrac{1}{2}(\delta_1 - \delta_2). \quad (6)$$

Determination of the position of a point in the aurora.—In Fig. 5, A, B, D and C denote respectively the principal and second auroral stations, the base pole and an auroral point. The plane through $ABCD$ has been called the plane of displacement, because an auroral point C, when viewed from B, will be displaced in this plane by an angle $p = u_2 - u_1$ relative to the direction from A. This displacement of an auroral point relative to the infinitely distant stars becomes evident by a glance at two corresponding photos. If the geographical coordinates of the two auroral stations are known, the distance between them g may be calculated.

Also, if r_1 denote the distance from A to C, from the law of sines

$$r_1 = g \sin u_2 / \sin (u_2 - u_1). \quad (7)$$

For the determination of the position of an auroral point C relative to the principal station A, consider a vertical plane through A and C, as indicated in Fig. 6. If the radius

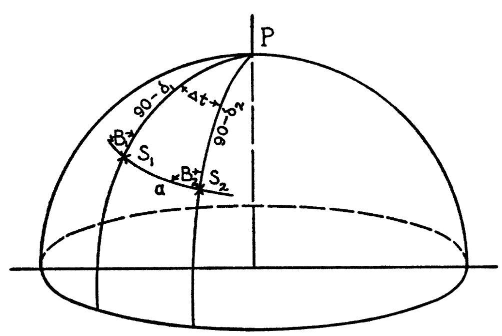

FIG. 4.

of the earth R, the height of the principal station above sea level H_0, the distance r_1 from A to C, and the altitude h of the point C are known, D and H may be calculated from the following equations:

$$H = [(R + H_0 + y)^2 + x^2]^{\frac{1}{2}} - R \quad (8)$$

and

$$D = R \arcsin [x/(H + R)], \quad (9)$$

where $x = r_1 \cos h$; $y = r_1 \sin h$; or approximately

$$H = H_0 + y + \tfrac{1}{2}x^2(R + H_0 + y)^{-1},$$
$$D = Rx/(R + H).$$

(b) *The method in practice*

The "nets."—For the determination of the angles h, a, u_1 and u_2, Störmer introduced a graphic aid, the "nets," which highly simplified the method of determination. The idea of the nets is as follows:

Suppose circles of constant declination δ and constant right ascension α are drawn on the celestial sphere at 2° intervals, and imagine that a part of these intersecting curves have been photographed by an auroral camera. An enlargement of this photograph would give a "net."

Of course this net of curves does not exist in the sky, and cannot be photographed, but the stars appear on the photo, and by means of their declination and right ascension, the net may be constructed.

The photos of the stars used for the construction of the nets have been taken in the following manner by Harang and Tønsberg.[23]

The optic axis of the camera is brought approximately into the meridian plane, and directed towards a star having a pole-distance $\gamma = 90° - \delta$. The center of the picture will have exactly the same pole-distance as that of the star if the image of the star falls on the exact center of the photographic plate. This also gives the center of the net. The γ value of the net, i.e. the pole-distance of its center, solely characterizes the net. The next picture is taken with the optic axis directed towards a point with pole-distance $\gamma + 2°$. For the construction of nets it is desirable to have as many stars on the picture as possible. With the camera in a fixed position several exposures are made, generally four, with suitable time intervals of 8 and 16 minutes of sidereal time, corresponding to a movement of the stars of respectively 2° and 4°. An exposure of exactly one minute has proved suitable.

The construction of the nets will now be explained. By means of a lantern the net photo is projected onto a sheet of thin white paper fastened to a vertical screen. The same lantern is later used when projecting and enlarging the auroral photos. As to the enlargement, the scale originally chosen by Störmer is a very suitable one. According to this scale, 1 cm corresponds to 1° of arc in the central part of the drawing where the deformation caused by the lenses of the camera and the lantern is negligible. To obtain this scale the angular distance between stars situated in the central part of the picture must be computed from formula (6).

If it be supposed that the angular distance between some centrally situated stars on the picture is known, the distance between lantern and screen must be arranged so that distances in centimeters between the stars are equal to their angular distances in degrees. This position of the lantern must be carefully determined and marked. The scale having been adjusted, the center and stars are plotted on the paper.

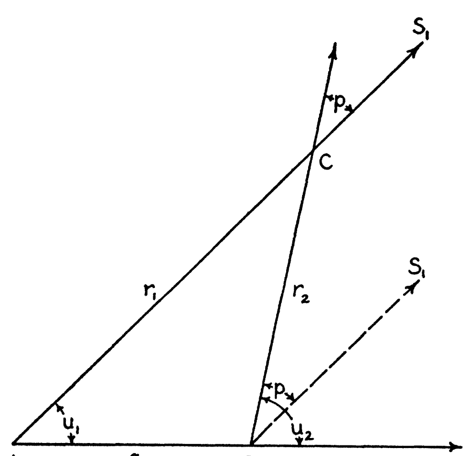

FIG. 5.

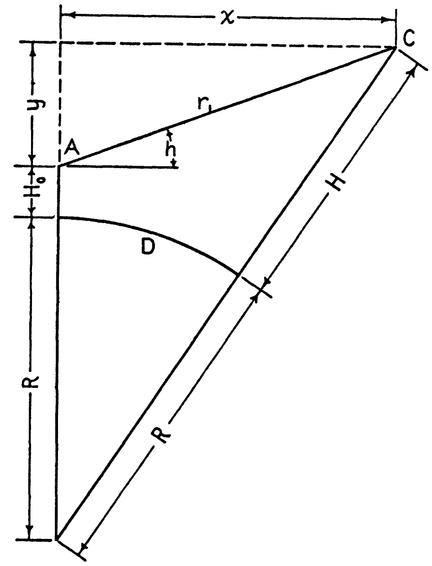

FIG. 6.

By means of the spherical coordinates (δ, α) and the movement of the stars, the curves of constant declination and right ascension can be drawn by interpolation, and are usually drawn with 2° intervals. On nets with small values of the pole-distance γ, however, the curves of constant right ascension are drawn with intervals of 4° and even 8°. Fig. 7 shows a net with $\gamma = 14°$. The nets are drawn on strong transparent tracing paper. When the nets are produced in this way every deformation caused by the lenses is also included in the construction. The lenses in the two cameras used by Harang and Tønsberg deform the pictures by the same amount, permitting the use of the one set of nets for negatives obtained with each camera.

The projection of corresponding auroral points.—Störmer advises that the two corresponding photos be projected simultaneously, by means of two identical lanterns, on sheets of thin, white paper, with the previously chosen enlargement. The use of two lanterns at the same time permits an advantageous direct comparison of corresponding auroral photos. On the sheets of paper the centers, the selected stars and the essential outlines of the aurora are carefully drawn in pencil.

The application of the nets. "Artificial stars."—The drawing from the principal station, let it be called the A drawing, is placed on a glass plate, which is illuminated from below. On the drawing are put the values of the declination and hour angle for the selected stars. Among the nets the one is selected whose value of γ coincides as nearly as possible with the pole-distance $\gamma = 90° - \delta$ for the center of the drawing. The γ value in question may be estimated from the pole-distances of the selected stars. On the "light table" the drawing is adjusted upon the net in such a way that the

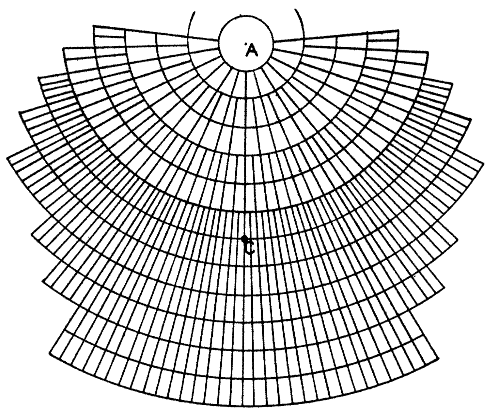

FIG. 7. Spherical coordinate net with pole distance $\gamma = 14°$.

declinations δ and differences of hour angles Δt of the selected stars can be read off *on the net.* When this has been done, the net curves indicate respectively all even values of declination and hour angle inside the border of the drawing. Also, the centers of the net and drawing should approximately coincide. Proceed in the following way. When the right position of one star has been found, stick a pin through this point, perpendicular to the surface of the drawing, so that the pin also pierces the net, and rotate the drawing about this center until the next star comes into its correct position. The coordinates of the third star will agree with those shown on the net if no errors have been made.

The points of intersection on the net, indicating even values of the coordinates δ and t can now be easily transferred to the drawing. If for these points the altitude h and the azimuth a with respect to the principal station, and the base-distance u and base-height ω for a fixed base-line have been calculated, from Eqs. (1), (2), (4) and (5), then on the drawing there are a number of points with all the coordinates required for further treatment. These points with even values of declination and hour angles, for which h, a, u and ω referred to a certain principal station and base-line, have been computed, are known as "artificial stars." Harang and Tønsberg have calculated and tabulated 500 "artificial stars" for the base-line Tromsö-Tennes. Experience has shown that it is sufficient to calculate artificial stars at 6° or 8° hour angle intervals.

On the drawing, which has been adjusted on the net with respect to δ and t, plot at least three artificial stars, selected from the table of values computed for the principal station and base-line from which observations were taken. It is advantageous to draw up such a table if many height measurements are to be made. Now the A drawing is furnished with all required coordinates.

Place the A drawing upon the B drawing, and if the selected real stars on the A drawing exactly cover the same stars on the B drawing, transfer the drawn outlines of the aurora on the B drawing to the A drawing as dotted lines, and the B drawing is not needed any more. However, if the selected real stars on corresponding drawings lie at different distances from the center, the deformation makes it impossible to adjust the A drawing on the B drawing so that the stars on the former exactly cover those on the latter. In such cases the B drawing must be treated separately.

Supposing the favorable case, return to the modified A drawing to determine corresponding points on the continuous and dotted auroral outlines. Among the nets, select the one whose γ value coincides as nearly as possible with the base-distance u_c, or its supplement $180-u_c$, of the center of the drawing. The γ value in question may be estimated from the base-distances of the artificial stars. On the "light table," adjust the drawing upon the net in such a way that the base-distances u and the differences of the base-heights $\Delta\omega$ of the artificial stars can be read off *on the net.* When this has been done accurately, the net curves indicate all even values of base-distances and base-heights inside the border of the drawing. The centers of the drawing and net must approximately coincide. The direction of the base-pole, being the direction of the displacement itself, may be drawn through one artificial star. Having now obtained by means of the net curves both the magnitude and the direction of the displacement, we can read off *on the net* the base-distances u_1 and u_2, and if necessary the base-height ω for any selected auroral point. In the selection of the auroral points the negatives must be at hand for comparison.

As mentioned before, the B drawing must be considered separately if the real stars on the A drawing will not exactly coincide with the corresponding ones on the B drawing. The corresponding auroral point on the B drawing is determined by the intersection between the outline considered and the direction of displacement fixed by the known ω. The base-distance u_2 is read off on the net.

Now it remains to determine the altitude h and the azimuth a of the selected auroral points from the principal station. The suitable net is selected and applied in the same way as those for (δ, t) and (u, ω) were. Finally, all the quantities required for the determination of the position in space of an auroral point are known.

Determination of r_1, H and D for an auroral point.—The values of r_1, H and D are now found by substituting the values obtained by the above net method in Eqs. (7), (8) and (9). From the distance D and the azimuth a of the auroral points, a map of their foot points (horizontal projections) on the earth's surface, may be plotted, giving an idea of the geographical distribution of the aurora.

D. A proposed mechanical-optical method for analyzing auroral photographs

It is interesting to note that Chapman has suggested a shorter method of determining the position of aurorae by a mechanical-optical method of analyzing the photographic plates.[13] The method seems the simplest possible in principle because it is a direct reconstruction, on a reduced scale, of the original situation of cameras and aurora at the time of the simultaneous photographic exposures. The reduced reproduction can be achieved optically by placing the photographic plates in two projectors, the optical centers of whose lens systems correspond to the positions of the auroral cameras at the principal and secondary stations. It simplifies matters to assume that the two cameras, and also the projectors are identical. The projectors should be capable of orientation in any direction, the center of rotation coinciding with the optical center.

The following method is suggested for setting the projectors in the appropriate directions. The stars which are capable of being photographed on the plates taken at any relevant auroral station are marked on the internal surface of a rigid framework, representing the celestial sphere, or a part of it. This framework should be capable of rotation about an axis corresponding with that of the actual celestial sphere, and there should be a divided circle permitting the sphere to be set at any hour angle. The axis would also have to be adjustable in a vertical N–S plane so that it could be set at any angle with the horizon by means of a divided arc. Thus this celestial sphere could be adjusted to the instantaneous position which the astronomical celestial sphere had, considered either from the principal or secondary station, at the instant when the photographs were taken.

The projector corresponding to the principal station camera is placed at the center of the celestial sphere, which has previously been set at the position appropriate to the time of exposure and location of the principal station, and then the orientation of the lantern is adjusted so that the projected star images coincide with their representations on the sphere. The instrument, clamped in this position, is removed, and the second projector, containing the plate taken at the secondary station is similarly placed at the center of the celestial sphere. When the sphere has been adjusted for the geographical situation of the secondary station, the projector is oriented so that as before the projected stars coincide with the corresponding ones on the sphere, and clamped. The two projectors, kept in the same orientation relative to each other, are placed a suitable distance apart, representing on a greatly reduced scale the actual distance between principal and secondary station, and the point where the auroral images from the two lanterns coincide gives the position in space of the aurora on the same scale.

Locating identical auroral points on each image would no doubt be the greatest source of difficulty. The images from the two projectors could not be made to coincide exactly, for they would not lie in the same plane, and considerable blurring would result.

A similar scheme is now in use for analyzing aerial photographs. No doubt many good ideas for the construction and use of the apparatus could be obtained from existing practice in aerial topography. Chapman suggests that the necessary apparatus might be located at some appropriate center, where auroral plates might be sent from any country, for reduction by a small staff specially trained and practiced in this work.

III. Intensity Measurements of Aurorae

A. Along ray-streamers

In a number of publications Vegard has pointed out that the law which governs the variation of the luminosity along a ray-streamer may give valuable information with regard to the physical properties of the rays which produce auroral phenomena.[64]

The following method has been used by Vegard and Krogness[67] to determine the variation in intensity along a ray as a function of the distance from the bottom edge.

Call l_1 the distance from the bottom edge to the maximum point, l_2 the distance from the bottom to the point where the intensity becomes low, and l_3 the whole range of luminosity. The quantities l_1 and l_2 are in most cases quite well defined, but l_3 is usually somewhat arbitrary, because the upper limit is frequently uncertain. Fig. 8 shows results of some calculations, in which l_1, l_2 and l_3 have been plotted against distance from the base. Average values found for l_1, l_2 and l_3 are used, and the upper limit is taken to indicate a point where the intensity falls below a certain value.

Certain characteristics of the distribution of luminosity can be seen from the figure. On passing from the bottom edge the light intensity

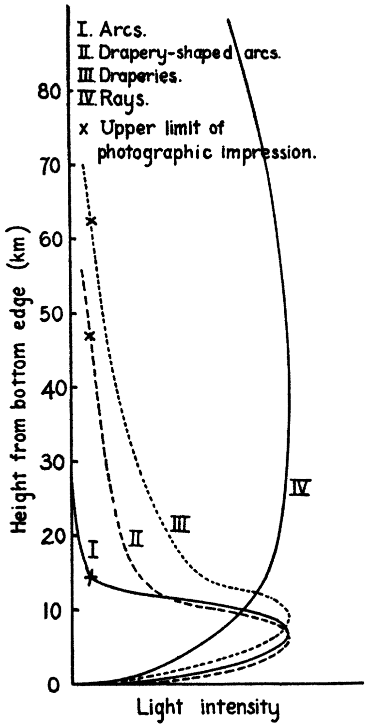

FIG. 8. Intensity distribution in the vertical (after Vegard and Krogness).

increases very rapidly to a maximum, and then usually diminishes more slowly to a faint luminosity, which thereafter gradually decreases with increasing height.

B. Total intensity measurements

In the great majority of intensity studies which have been made, the determination of the total intensity has been done visually, and expressed in terms of international coefficients. While this is useful, it is naturally empirical and individual differences in observers cause inaccuracies.

The French Polar Year party at Scoresby Sound established an electrical device for recording aurorae.[17, 18] This apparatus detected aurorae during foggy weather, as well as in the hours of dawn and twilight. It has shown that the intensity of the luminous phenomena may vary in the ratio of 1 : 1000, and so permits precise comparison between auroral and magnectic curves.

IV. Characteristics of Ordinary and Sunlit Aurorae

A. Ordinary aurorae

The majority of the determinations of the heights of aurorae have been made in Norway. As a result of polar year and other observations, several good groups of measurements have also been made in Canada.

The lower limit of an aurora is usually quite distinct, and can be determined with considerable accuracy. Fig. 9 shows the heights of 1927 auroral points which were believed to mark the lower edge of aurorae. These were calculated by Vegard and Krogness[67] from observations made at the Haldde Observatory during 1913 and 1914. This shows the lower limits of all aurorae photographed, irrespective of type. The numbers in 2 km intervals were counted and plotted. By far the greater number of aurorae, 94 percent, descend to a height between 90 and 128 km.

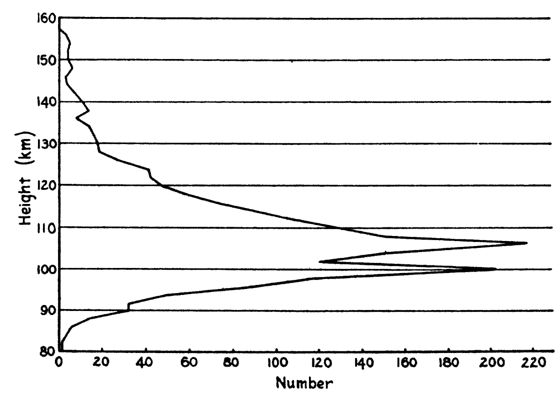

FIG. 9. Height of aurora (after Vegard and Krogness).

The mean height of the lower limit regardless of form is 107.94 km. Two very marked maxima appear, one at 100 km and one at 106 km.

It appears from measurements at this time that luminosity can be observed as high up as 335 km.

For an aurora with a sharp bottom edge, the lower limit may be determined with an accuracy of about 1 or 2 percent, provided the zenith distance does not considerably exceed 45°. In more unfavorable cases the difficulty in fixing corresponding points may cause an error as great as 20 percent.

Figure 10 gives the height distribution for ray and nonray forms from Störmer's 1913 Bossekop Expedition,[53] from Krogness and Vegard's 1913–14 Haldde Observatory measurements[67] and from Harang and Tønsberg's observations at Tromsö.[23] All three frequency curves of the *diffuse arcs and bands* show a close resemblance, with the two maxima at 100 km and 106 km appearing in each. In the Tromsö curve, the lower maximum at 100 km is the predominant one. Regarding the curve of *draperies and arcs with ray-structure*, there is close agreement between the Tromsö curve and Störmer's, while the lower limits in Tromsö on an average have a greater value than those at Haldde. Canadian auroral height findings are indicated in Fig. 11. Measurements made at Onakawana, northern Ontario, of 39 auroral points are shown.[39] The feature of this investigation is the sharp maximum of lower edge points found in the height interval 90–95 km. Several aurorae extending as low as 70 km were photographed. The results from the measurement of 200 simultaneous photographs, taken at Chesterfield during the polar year also have been plotted, using the number of points for each 2 km interval.[15] The ray and nonray distributions are shown. The height distribution is similar to that found over Norway, with a pronounced maximum for the quiet and homogeneous forms occurring between

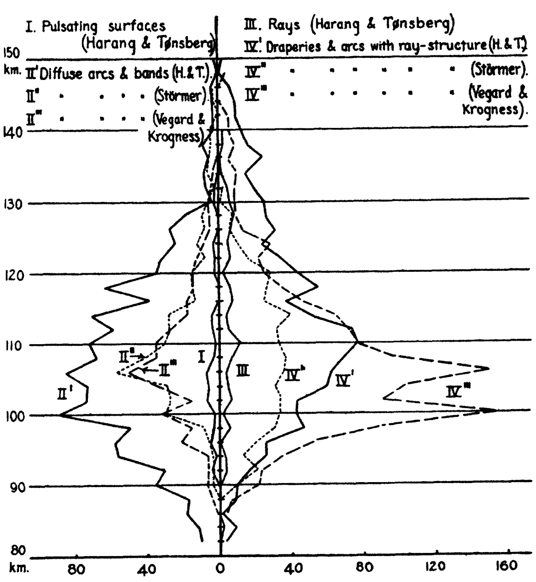

FIG. 10. Heights of various forms of aurora.

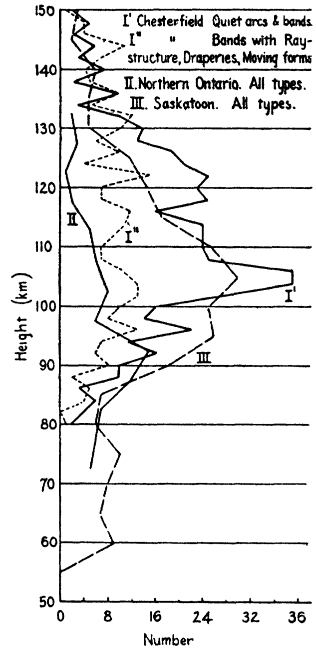

FIG. 11. Auroral heights as found from several Canadian stations.

104 and 106 km. On the same graph is shown the height distribution found at Saskatoon by the polar year group situated there, from 220 auroral points, taken in 5 km groups. Störmer has stated that the lower limit of height in Norway is 80 km.[58] On two occasions at Saskatoon bands were photographed, the lower limits of which were at a height of only 60 km. Thirty-seven measurements were made of auroral features, the heights of which were less than 80 km, and eighteen of these showed heights of less than 70 km. The lowest recorded was at 59 km. In these photographs the intensity was good and the edge of the aurora clearly defined.

The photographs taken at both Chesterfield and Saskatoon have been analyzed by the "artificial star" method of Harang and Tønsberg which has already been described, while Störmer's earlier method was used for the Onakawana ones.

Since Störmer's statement that the lower limit in Norway was 80 km was published, Harang and Bauer in March 1932, working at the Tromsö Observatory photographed an aurora with its lower edge at a height of only 65 km.[24] It was a green arc with a dark red lower edge.

This red edge could only be found in the lowest position of the arc. Thus comparatively low aurorae occur both in Canada and Norway.

It should also be noted that the curves for Canada do not show the two maxima at 100 and 106 km which the curves for Norway so clearly indicate. It is possible that they would appear if a larger number of determinations were made.

The explanation of these two maxima is uncertain. In an interesting paper, Egedal has attempted to explain this by the lunar tides of the upper atmosphere.[20] He suggests that at the time of the ebb of these tides corpuscular rays coming from outer space, which might cause the aurora, penetrate more deeply than at the time of the flood, for in the first case they reach the wave valleys and in the second are stopped by the wave crests. Egedal investigated approximately 1700 heights of aurorae with the corresponding times of flood and ebb, and found suggestive indications to justify his explanation. Harang and Tønsberg tried to trace the same effect, using Egedal's method of calculation, but neither the frequency-curves of the arcs and bands, nor that of the draperies gave any certain confirmation to this theory.[23]

It has often been stated by observers that they have seen aurorae near the surface of the earth, so low that the aurora can be distinguished against some distant background, such as a mountain range. It has also been claimed that aurorae are sometimes accompanied by sounds, such as swishing or crackling. Simpson has recently strongly discounted the possibility of either occurring.[48] He gives several reasons for the great improbability that aurorae could extend to so near the earth's surface, and also why they are not likely to be accompanied by sound.

B. Sunlit aurorae

One of the most spectacular aurorae seen in Norway in the last twenty years occurred on the night of March 22–23, 1920. The different forms were observed photographically the whole night through by seven of Störmer's stations, and more than 600 photographs taken. The measurement and calculation of these brought out the striking fact that after sunset and before sunrise the auroral rays had a considerably higher position than in the middle of the night. At that time no explanation could be found.

It was a remarkable violet-gray aurora appearing on September 8, 1926, which suggested to Störmer the conditions under which these high rays occurred.[54] This aurora extended from the horizon in the west up to a certain height in the sky, and then disappeared. Calculations showed it was situated over the Shetland Islands, and was at the unusual height of 300–500 km above the earth. The great height, unusual color and situation near the region where the sun had set suggested that it lay in the upper atmosphere in full sunshine. Computation showed that most of these high rays were in sunshine, and that some few of them attained the altitude of 1000 km. A few rays seemed to extend into the earth's shadow, but were separated from the higher rays by dark intervals.

Further analysis showed that the high rays of the brilliant aurora of 1920, seen before sunrise and after sunset, had been illuminated by the sun.

On the night of March 15–16, 1929, an especially interesting series of photographs was taken both of sunlit auroral rays and of ordinary aurorae in the earth's shadow. Fig. 12, representing a vertical section of the atmosphere with the boundary between sunlit and dark atmosphere marked by the tangent, shows the positions of the aurorae observed that night. For the first time rays consisting of two parts were observed, the upper one in sunshine and the lower in shadow, the two being connected by an invisible part. Two such rays can be seen in the figure approximately 1800 km from the tangent point.

The reality of this phenomenon has been confirmed by photographs taken on several occasions during the past few years.[57, 23] A recent photograph of sunlit aurora is shown in Fig. 13. This figure shows sunlit aurora rays near the stars of the Great Bear, photographed by Störmer and his assistants at 3 hr. 19 min. 55 sec. G.M.T. on 17th October, 1936, simultaneously from Oslo and Lillehammer. The length of the base line was about 159 km and direction about S–N. The height of the foot of the ray nearest to the stars in the Great Bear was 319 km, and that of the summit of the same ray was 661 km.

Some attempts have been made to try to explain how and why these sunlit aurorae occur as they do.

Krogness has advanced the hypothesis that the radiation pressure of the sun pushes the upper parts of the atmosphere tangentially backwards like a little comet's tail, and that the high sunlit aurorae occur in this tail.[28]

The auroral forms measured by Störmer on the night of September 8, 1926, do lend weight to this idea, but observations at other times indicate rather that the whole atmosphere is raised by the action of the sun's rays, and that it drops down again quite quickly as soon as the earth's shadow reaches a height of approximately 300 km.

This elevation of the upper atmosphere by the sunlight has also been suggested by Vegard.[68] By his theory of a photoelectric action of the radiation from the sun on the upper atmosphere, Vegard has been led to the hypothesis that the upper atmosphere must be electrically charged and so driven upward by the resulting electrical force.

That the sun's rays and especially the ultraviolet ones have a direct influence upon the action by which light is produced, is evident first from the fact that the bottom points have a distinct tendency to follow the boundary line between sunlight and darkness, and secondly because of rays of the type seen on March 15–16, 1929, by Störmer, which extend from great heights in sunlight down into the darkness, with a short break at the boundary between sunlight and shadow.

Some of the facts could be explained if it were assumed that the energy of an exciting agency from outer space was strongly absorbed by the sunlit atmosphere. Perhaps all the energy would be absorbed by the sunlit atmosphere if it extended down to 200 km, and so no ordinary aurorae would be seen in the earth's shadow. When the strongly absorbing sunlit region extended down to 300 km, then only part of the energy would be lost, and the auroral ray could reappear when it had penetrated some distance into the earth's shadow. Two examples of this type are clearly shown in Fig. 12. Finally when only the layers higher than 400 km are illuminated, very little absorption would occur in traversing this space, and the exciting agency could penetrate further into the shadow region, causing aurorae down to a level of 100 km. All three stages can be seen in Fig. 12.

C. Comparison of ordinary and sunlit aurorae

The most outstanding difference between these two types of aurora is the height at which they

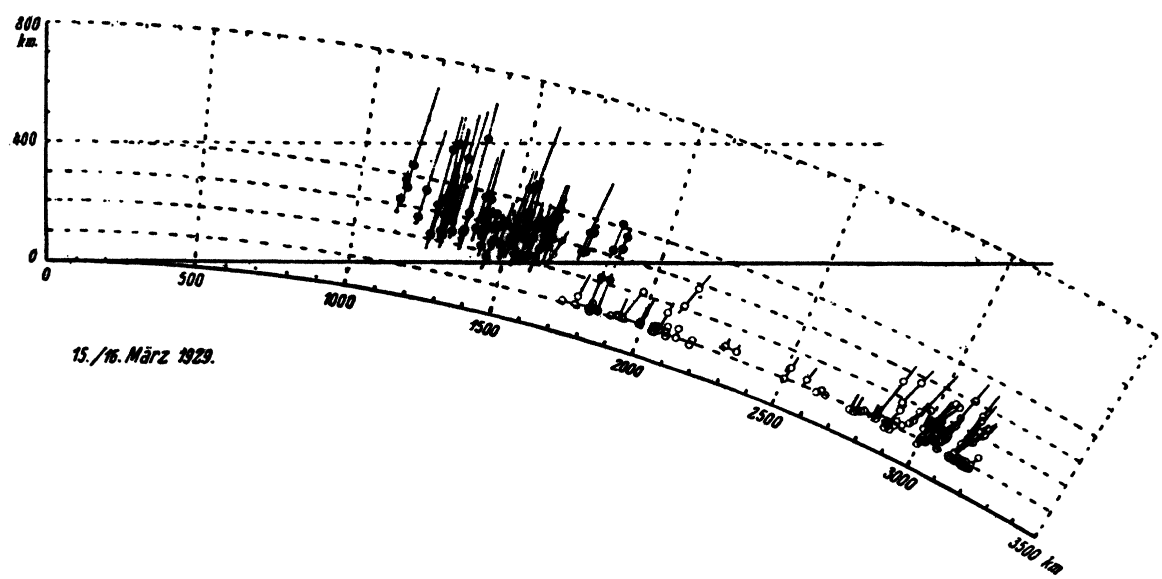

FIG. 12. Sunlit aurorae (after Störmer).

occur. The ordinary aurora rarely extends to a height greater than 400 km, while the sunlit type is usually considerably higher than this, and has been observed at heights greater than 1000 km.

The grayish-violet color, even blue on some occasions, is a distinguishing feature of the sunlit aurora. A yellowish-green color characterizes the ordinary aurora, but on rare occasions a deep red is the predominant color.

As would be expected from the color distribution of each, the sunlit aurora spectrum contains a large number of blue and violet lines not found in the ordinary aurora spectrum. Of this, more will be said later.

V. The Auroral Spectrum

A. Early investigations

The earliest study of the spectrum of the aurora appears to have been made in 1869 by Angstrom.[2] He supposed that he would obtain the same spectrum as that of the electrical discharge in air, but found only a green line at 5567A, and three very weak lines in the blue region, all of which were unknown to him.

From Angstrom's time until 1912, a considerable number of observations of the auroral spectrum were made by various observers. Most of these were visual observations, but attempts were also made by Paulsen[42] and Westmann[81] to employ a photographic method. The small spectrographs and low dispersion gave results which were far from satisfactory. H. Kayser, in his *Handbuch der Spektroscopie* has summarized the observations made during this early period. The inaccuracy in the wave-length measurements was so great that identification was nearly impossible.

In the winter of 1912–13, Vegard, realizing the importance of a detailed study of the auroral spectrum, undertook an expedition to Bossekop, mainly to investigate it.[65] Here he photographed the spectrum of the aurora during the last three months of 1912. His calculations showed that the three strongest lines were at wave-lengths of 4708, 4278 and 3914A, and that fainter lines were at 5571, 4647, 4234 and 4200A. He decided that all these lines except the yellowish green one at 5571 corresponded to heads of nitrogen bands. This suggestion was later confirmed by Lord Rayleigh.[44]

B. The auroral green line

Measurements of the green line were continued at Oslo with a spectroscope of considerably higher dispersion and a good micrometer arrangement. The most reliable series of measurements gave a wave-length of 5577.6A.

A green line in the spectrum of the night sky had been noted by several observers, and in 1919 Slipher[49] measured its wave-length, and found it to be 5578.05A. Comparison of the values of the wave-lengths of these two lines led to the conclusion that the lines were identical. In 1922 Babcock[4] on Mount Wilson and at Pasadena, accurately measured this line with a Fabry and Perot etalon. A series of careful determinations gave the value 5577.350 ± 0.005A for the wave-length, and indicated that the width of the line was not greater than 0.035A.

In order to attempt to discover the origin of the auroral green line, Vegard started more systematic investigations of the spectrum of the aurora in 1922–23. Suitable spectrographs were constructed and mounted on the roof of the Geophysical Institute at Tromsö. The first winter's work gave a number of good spectrograms, and 35 lines and bands were measured in the visible and ultraviolet parts of the spectrum.[69] With the exception of the green line and three faint lines or bands, the lines were identified with known nitrogen ones.

Although the stronger lines in the auroral spectrum were greatly over exposed, no hydrogen nor helium lines were found. This fact to a large extent invalidated the widely prevalent view that the atmosphere above 100 km consisted mainly of light gases.

This, combined with the fact that nearly the whole spectrum consisted of nitrogen lines, supported Vegard's view, first expressed by him in 1910, that the green line also belonged to nitrogen, and that it was emitted under the special conditions existing in the upper atmosphere. He assumed that the nitrogen was in the form of a crystalline dust.

In order to test this, Vegard, working in the Cryogenic Laboratory at Leyden bombarded solid nitrogen with cathode rays, and got the

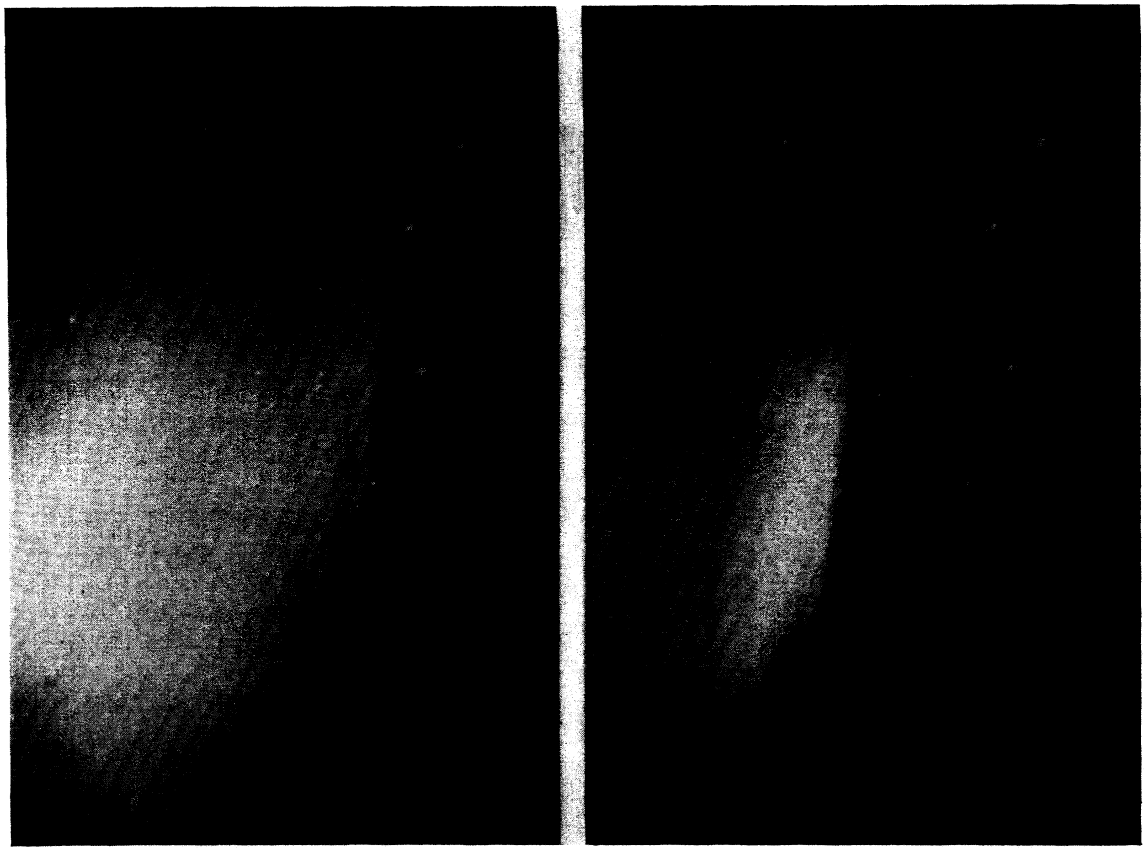

FIG. 13. Sunlit aurora, photographed simultaneously from Oslo and Lillehammer.

type of light effect he had expected.[70] In addition to a broad line or narrow band at about 5577A, the spectrogram showed as well a line at 5230A, which is about the wave-length of a fairly conspicuous line in the auroral spectrum. Vegard considered that this was sufficient to show that the typical auroral spectrum is emitted by solid nitrogen.

In Canada, McLennan and Shrum[31] also investigated the spectrum of solidified nitrogen made luminous by cathode rays, and instead of a single spectral line at 5577A, found the spectrum to consist of three broad lines or narrow bands, with the mean wave-lengths 5556, 5617 and 5654A. On the basis of their findings, they concluded that the auroral green line did not originate in solid nitrogen.

The search for the origin of this line took on a new aspect when in 1925 these two investigators, McLennan and Shrum, announced that they had obtained a spectral line from an electric discharge in a mixture of oxygen and one or other of the rare gases helium and neon, with the rare gas in excess, that had all the characteristics of the auroral green line.[32] They also obtained it faintly in pure oxygen at low pressures. Careful measurements showed that its wave-length was 5577.35 ± 0.15A.

In 1926 a more extensive investigation on the nature and occurrence of the line was undertaken by McLennan, McLeod and McQuarrie.[34] This further work showed that this 5577.35A spectral line can never be observed in the spectrum of any electrical discharge in the absence of oxygen. From observations on the Zeeman resolution of the line, they were able to show that it was one that could be provided for in the scheme of spectral terms worked out for atomic oxygen by McLennan, McLay and Smith.[33]

Precision measurements of the wave-length of this oxygen line made in 1927 by McLennan and McLeod,[35] using a Fabry and Perot interferometer gave as a mean 5577.341±0.004A. Recently, the auroral green line was measured interferometrically by Vegard and Harang.[78] A number of observations with a 5 mm Fabry and Perot etalon yielded the mean value 5577.3445±0.0027A. Babcock's value of the night sky line was 5577.350±0.005A. The close agreement between the wave-lengths found for the auroral line, night sky line and oxygen line seems to leave no doubt that they are identical.

Further study of the Zeeman resolution of the line by McLennan, McLeod and Ruedy[36] indicated that it originates in transitions of oxygen atoms from a $^1S_0$ metastable energy state to another metastable one designated as $^1D_2$. Frerichs,[38] using new lines observed by him in the oxygen arc spectrum, was enabled to calculate the frequency difference for the $(^1S_0-{}^1D_2)$ transition, and this frequency was found to be identical with that for the auroral green line. This confirmed the classification suggested by McLennan.

By analogy with the oxygen line found through laboratory experiments, McLennan and Ireton[38] have suggested that molecules of nitrogen in an excited metastable energy state of approximately 10.5 to 11.8 volts play an important part in the production of the auroral green line in the upper atmosphere.

In an attempt to obtain information from another direction about the origin of the excitation of this line Lord Rayleigh[45] organized a survey to try to determine whether the intensity of the green line radiation of the night sky varies with the time and place of observation. Results from various stations throughout the world indicate an annual periodicity with a maximum intensity in October in the northern hemisphere and in April in the southern hemisphere. Observations in England alone showed an intensity maximum at midnight. Independently, McLennan found from observations at Flagstaff,[37] in the southern United States, and near Toronto, Canada, a maximum about one and a half hours after midnight.

Working in France, Garrique[21] studied the variation of intensity with direction. By the use of a special arrangement whereby exposure was reduced to one hour or less and the spectra of two different parts of the sky photographed simultaneously, he found that the intensity of the night sky line at the northern horizon was approximately double that at the zenith, and about 1.2 times that at the southern horizon. The maximum intensity occurred about 1 A.M.

C. The auroral red line

In the 1870's a large number of deep red auroral displays were observed from many parts of Europe. At that time no spectroscopic studies were made of them, and unfortunately so few similar aurorae have appeared since then, that it has been impossible to accurately determine the lines emitted during such auroral displays. Vegard[71] reports having seen one on the night of January 26, 1926. The striking feature of this aurora was its intense red color, which spectroscopic analysis showed was due mainly to a single sharp line in the red, at about a wave-length of 6323A. Its intensity was nearly as great as that of the green line. Spectrograms taken at Oslo and Tromsö showed the characteristic auroral spectrum, consisting of the green line 5577A, and the negative bands of nitrogen. This prominent red line has also been recorded on two previous occasions at Tromsö, and its mean wave-length deduced from the three observations is 6322.4A. In the ordinary auroral spectrum, a very weak and diffuse line has been observed a number of times by Vegard[74] at about this same wave-length. Doubtless this is the same line as the one which is emitted with such great strength from red aurorae. Seven measurements of the line gave a mean of about 6318A. Neither nitrogen nor oxygen lines seem to correspond with this wave-length. Further investigation is necessary before the origin of this line can be ascertained.

D. The spectrum of the sunlit aurora

On several occasions, when Störmer[53] has looked at sunlit auroral rays through a pocket spectroscope, he has noticed that the auroral green line was very faint as compared with strong lines in the blue and violet. To investigate this matter more fully, Störmer in the winter of 1928–29[56] had two small spectrographs constructed. Two successful spectrograms of the aurora of March 15–16, 1929, illustrated in

Fig. 12, were obtained. The first, exposed for about five hours, recorded spectra from all auroral forms during the first part of the night, and the second, with an exposure of an hour, photographed exclusively the spectra of all the high sunlit aurora rays just before dawn. About a month later, similar comparison spectrograms were obtained.

These observations indicate that the green line 5577A, which is very strong for the common aurora in the earth's shadow, is very much fainter for the high sunlit aurora rays as compared with the lines of ionized nitrogen 4728A and 3914A. It was interesting to note that no hydrogen nor helium lines appear in the spectrum of these high rays. These results were only preliminary, and the investigation is being continued by Störmer.

Doubt has been expressed by Vegard[72] that this is a real effect. He has suggested that the two spectra given by Störmer do not fulfill the conditions which might make them comparable. First, the two spectra are taken on different plates, and the second and more serious defect is that one of the spectra is very strongly exposed while the other spectrum is very weak. Also, even if it had been possible from Störmer's material to conclude that the sunlit aurora gave a relatively weaker green line, this effect might be due merely to the altitude effect discovered in 1923 by Vegard.[69a]

An enhancement of red lines and bands in the auroral spectrum from a sunlit atmosphere has also been noted by several workers. Vegard and Tønsberg found from spectrograms of auroral arcs, some exposed to sunlight and some not exposed, taken on October 27, 1935, that the red line 6300A relative to the green line 5577A is 4–5 times stronger from the sunlit atmosphere than from the dark one.[79] They also noted that on this occasion the bands of the first positive group of nitrogen were enhanced to an even greater degree. This enhancement of the red line 6300A was also observed by Currie and Edwards at Chesterfield during the polar year,[16] by Kosirev and Eropkin at Kirowsk (March 1935)[27] and by Störmer during two nights in April, 1936 and during a night in October, 1936.[62] Vegard also observed the enhancement on the latter night.[80]

E. The infrared section

The importance of obtaining auroral spectra in the infrared has long been recognized, and recently considerable progress has been made in this direction. Bauer photographed the aurora borealis arc, using a filter which removed the visible and ultraviolet radiation.[5] By carefully considering both the limit of sensitivity of the plate and the permeability of the filter, it was found that the auroral spectrum possesses a considerable amount of infrared radiation between 7500 and 8400A.

Investigations of the infrared have formed part of Vegard's spectroscopic program for the new Auroral Observatory at Tromsö, and in a preliminary report, Vegard[73] announced the first few measurements. The spectrum in the infrared had the appearance of one strong and one weak narrow band, with sharp edges towards the long wave-lengths. Rough measurement of these lines gave values of 7883A (strong) and 8095A (weak). They had the appearance of band sequences belonging to the first positive group of nitrogen.

Further infrared spectrograms by Vegard and Harang[77] have shown eight infrared bands extending as far as 8150A. The measurements suggest that the whole of the observed infrared spectrum forms part of the first positive group of nitrogen.

F. The ultraviolet section

A number of ultraviolet lines have also been observed and measured by Vegard.[74] Assuming that the visible spectrum ends at the strong line 3914A, his spectrograms show 17 lines in the ultraviolet, extending from 3904A down to 3136A. Most of these are fairly strong, and can be identified as lines belonging to the second positive group of nitrogen. In addition to these 17, a few faint lines were found.

G. Summary

A summary of the first three years' work at the Auroral Observatory at Tromsö has been given by Vegard.[76]

Also, a summary of the results of investigations on the auroral spectrum from 1912–33 by Vegard and his collaborators has recently been published by Vegard.[75] The gist of his summary is given below.

Interferometer measurements of the auroral green line, with the collaboration of Harang, have recently been made. The agreement with Babcock's value for the night sky line, shows conclusively that these two lines are identical. Interferometer pictures taken both parallel and perpendicular to the auroral ray-streamers show no Doppler shift.

In the region between 3000 and 9000A, the auroral spectrum has been quite thoroughly explored. Up to the present, 85 lines and bands have been recorded photographically and measured.

The following are the more important results obtained:

(1) Apart from the green line, the spectrum of an ordinary aurora is dominated by negative and positive nitrogen bands.

(2) All the eight infrared bands yet observed, and a large number of red bands, belong to the first positive group of nitrogen.

(3) The second green line or band has a wavelength of about 5240A, and cannot be identified with oxygen lines or with bands of the negative nitrogen group. It may be referred to the first positive group of nitrogen, or the band N_2 of solid nitrogen.

(4) In the ultraviolet, a diffuse band series coinciding with one emitted from solid nitrogen has been observed.

(5) In the red region, two lines, 6302A and 6365A appear, which may be identical with the O_1 lines $({}^1D_2-{}^3P_2)$ and $({}^1D_2-{}^3P_1)$, respectively. No O_1 lines are found in the infrared, and it is doubtful whether O_1 lines, other than those attached to the levels ${}^1S_0$, ${}^1D_2$ and ${}^3P_{012}$ appear in the auroral spectrum. This curious fact may be explained by assuming that oxygen is mainly excited indirectly by collisions of the second kind with active nitrogen. The energy of active nitrogen available for excitation (9.55 volts) and the energy required for dissociation according to the equation $O_2=O_1+O({}^1S_0)$ (9.3 volts) are nearly equal, and the probability for the excitation process therefore becomes large because of a kind of resonance. Kaplan has shown experimentally that the green O_1 line can actually be excited by active nitrogen.

(6) Lines of hydrogen and helium are either not present or extremely weak, indicating that there is no dominating layer of light gases floating at the top of the atmosphere.

The average energy of the more prominent bands and lines have been measured. Considerable changes in the intensity distribution occur. The following changes have been detected and studied:

(a) Variations with the auroral type.

(b) Lines in the 6300A region are enhanced and produce the red colored aurora. The enhancement is a universal phenomenon, and seems to follow the sunspot cycle. In another type of aurora, where the red color is confined to a band near the lower limit, this red coloring can be explained by the enhancement of the first positive group of nitrogen.

(c) The enhancement with increase of altitude of the negative nitrogen bands relative to the green line. This altitude effect, detected in 1923 has recently been verified by Störmer and Harang.

VI. Theory of Aurora

A. Corpuscular theories

(a) *The history of the problem*

It appears that the German physicist Goldstein[22] was the first to publish the idea that the sun sends into space electrical rays analogous to cathode rays, and so explain the connection between solar activity and magnetic and electric disturbances on the earth. Some years later, the Danish meteorologist Paulsen[41] brought forth the hypothesis that the aurora is due to cathode rays. He believed that these cathode rays originated in the upper atmosphere.

In the year 1896, Birkeland[6] experimented with cathode rays in a magnetic field. He found that a magnetic pole tended to make a beam of cathode rays, originally moving parallel, converge towards a point. This suggested to him that the aurora was due to a similar effect of the earth's magnetic field on cathode rays coming from the sun. To test this hypothesis, Birkeland[7] exposed a small spherical electromagnet to a stream of cathode rays, and found close analogies to the shape and nature of the aurora. The auroral belts were particularly well shown.

Arrhenius, in 1900, published his hypothesis[3] that the sun continually sends out small electrified particles, varying in size from 0.0001 to 0.001 mm in diameter. He suggested that these particles are repelled from the sun by the pressure of light, and cause aurorae on reaching the earth's atmosphere.

In 1903, Störmer became interested in Birkeland's theory of the aurora, and particularly in the problem of finding the trajectories of electric corpuscles moving in the magnetic field around a magnetized sphere, such as the one with which Birkeland had been experimenting. This was similar to the problem solved just before by Poincaré[48] for the case where the magnetic field was that due to a single magnetic pole. Störmer's first results seemed promising.[50]

(b) *Mathematical treatment of the problem*

Störmer has published various aspects of his mathematical study of aurorae.[51]

Simplifying hypotheses.—In order that the mathematics might not be too difficult, Störmer made a number of simplifying assumptions.

Considering the Gaussian series[26] for the magnetic field in space, due to magnetic masses in the interior of the earth, the approximation to the first term becomes greater and greater with increase of distance from the earth. Störmer therefore chose the magnetic field due to the first term, which is the magnetic field around a homogeneous magnetic sphere; in other words, the magnetic field of an elementary magnet at the center of the earth. The magnetic axis is then a diameter cutting the surface of the earth near Smith's Sound in northern Greenland, halfway between the magnetic pole and the geographic north pole.

The next simplification was to assume that the time required for the corpuscles to travel from the sun to the earth was so small that the position of the earth relative to the sun does not change appreciably during that time. This requires that the corpuscles attain velocities comparable to those of cathode rays.

A third assumption was that long thin streams of corpuscles may occur, similar to those observed in cathode ray experiments, and that they do not dissolve because of electrostatic repulsion. Some auroral forms seem to justify this assumption.

Finally, it was assumed that the only force acting on the corpuscles was that due to the earth's magnetic field. Thus, solar magnetism, electrostatic or electromagnetic effects from the earth and sun, gravitation and light pressure were neglected.

It seemed advisable to attempt to solve the problem using these assumptions, and then to disregard each in turn in order to approach actual

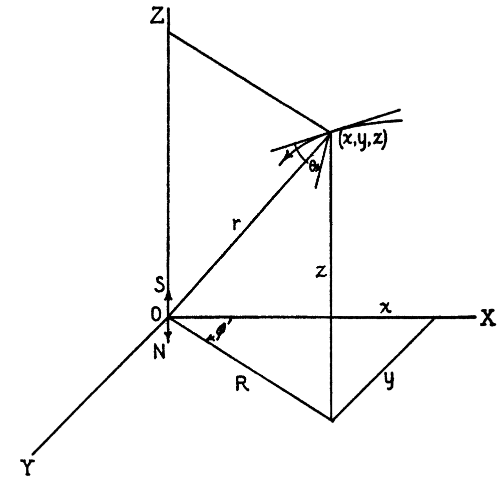

FIG. 14.

conditions more closely. It is surprising to find how many peculiarities in auroral phenomena can be explained by the solution of this simplified problem. However, even this solution is very difficult, and is still far from complete.

This problem may be defined precisely as the determination by mathematical analysis of all the possible trajectories of an electrically charged corpuscle in the magnetic field of an elementary dipole.

Equations of motion.—In a system of rectangular Cartesian coordinates (Fig. 14) let the elementary magnetic dipole be placed at the origin with its south pole pointing in the direction of increasing z. The equations of motion of a negatively charged particle moving in the field of the elementary dipole are then:

$$m\frac{d^2x}{dt^2}=e\left[H_z\frac{dy}{dt}-H_y\frac{dz}{dt}\right],$$

$$m\frac{d^2y}{dt^2}=e\left[H_x\frac{dz}{dt}-H_z\frac{dx}{dt}\right],$$

$$m\frac{d^2z}{dt^2}=e\left[H_y\frac{dx}{dt}-H_x\frac{dy}{dt}\right],$$

where the letters have the following meanings: x, y and z are the coordinates of the electrons, t the time, H_x, H_y and H_z the components of the magnetic vector, m and e are the mass and charge of the electron, expressed in the c.g.s. system and in the electromagnetic system. By changing the sign on the right-hand side of these equations, they hold for positively charged particles. It follows by multiplying each of the equations respectively by dx/dt, dy/dt, and dz/dt, and adding that the velocity of each electron is constant. This is a well-known fact which follows because the force acting on the electrons is always perpendicular to their paths. Let this constant velocity be v cm/sec. Furthermore, for the elementary magnet, there are three equations,

$$H_x = -M(3xz/r^5), \quad H_y = -M(3yz/r^5), \quad H_z = -M(3z^2 - r^2)/r^5,$$

where M is the magnetic moment of the magnet, and $r = (x^2+y^2+z^2)^{\frac{1}{2}}$ is the radius vector.

Using the relation that $ds = vdt$, the equations of motion pass over into the form

$$\frac{d^2x}{ds^2} = \frac{C^2}{r^5}\left[3yz\frac{dz}{ds} - (3z^2 - r^2)\frac{dy}{ds}\right],$$

$$\frac{d^2y}{ds^2} = \frac{C^2}{r^5}\left[(3z^2 - r^2)\frac{dx}{ds} - 3xz\frac{dz}{ds}\right],$$

$$\frac{d^2z}{ds^2} = \frac{C^2}{r^5}\left[3xz\frac{dy}{ds} - 3yz\frac{dx}{ds}\right], \quad \text{where} \quad C = (Me/mv)^{\frac{1}{2}}.$$

If C centimeters is selected as a new unit of length, the equations take the form

$$\left.\begin{aligned} \frac{d^2x}{ds^2} &= \frac{3yz}{r^5}\frac{dz}{ds} - \frac{3z^2 - r^2}{r^5}\frac{dy}{ds}, \\ \frac{d^2y}{ds^2} &= \frac{3z^2 - r^2}{r^5}\frac{dx}{ds} - \frac{3xz}{r^5}\frac{dz}{ds}, \\ \frac{d^2z}{ds^2} &= \frac{3xz}{r^5}\frac{dy}{ds} - \frac{3yz}{r^5}\frac{dx}{ds}. \end{aligned}\right\} \quad \text{(I)}$$

By introducing polar coordinates R and φ given by

$$x = R\cos\varphi, \quad y = R\sin\varphi$$

and integrating, the equation

$$R^2 d\varphi/ds = 2\gamma + R^2/r^3 \quad \text{(II)}$$

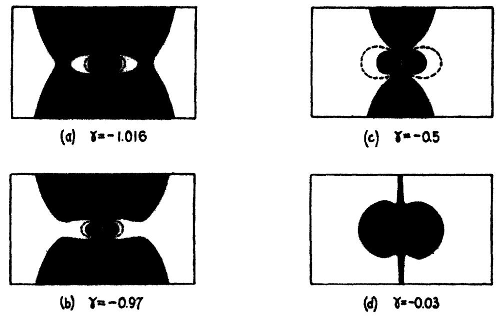

FIG. 15. Sections through the Z axis of four of the more characteristic Q_γ regions to which the trajectory must be confined. The regions are represented by the white parts when rotated about the Z axis.

is obtained for φ (Fig. 14). R and z are given by the system

$$\left.\begin{aligned} d^2R/ds^2 = \tfrac{1}{2}\partial Q/\partial R, \quad d^2z/ds^2 = \tfrac{1}{2}\partial Q/\partial z, \\ (dR/ds)^2 + (dz/ds)^2 = Q. \end{aligned}\right\} \quad \text{(III)}$$

γ is a constant of integration, and Q is a function of R and z given by the equation

$$Q = 1 - [2\gamma/R + R/r^3]^2.$$

The problem of finding the trajectories is thus reduced to the integration of the system (III), which can be done by the integration of a second order differential equation and a quadrature. Then φ may be found from Eq. (II) and a new quadrature.

But even without integrating the equations, it is possible, as will be shown, to draw important conclusions directly from the equations as they now stand.

Regions of space beyond which the trajectories cannot go.—Eq. (II) will give the first important result for the general discussion of the trajectories. Calling the angle between the tangent in the direction of motion and the plane through its point of contact with the trajectory and the Z axis, θ (Fig. 14) it follows that

$$\sin\theta = Rd\varphi/ds.$$

The equation (II) therefore gives the formula

$$\sin\theta = 2\gamma/R + R/r^3.$$

Now $\sin\theta$ can only have values between $+1$ and -1. The trajectory must then be confined to the region of space where

$$-1 \leqq 2\gamma/R + R/r^3 \leqq 1.$$

Let this region be called Q_γ.

To each value of the constant of integration γ there is a corresponding region Q_γ, and no trajectory corresponding to the same value of γ can get beyond this region.

The value of γ for a given trajectory can be found immediately from the equation

$$2\gamma = R\sin\theta - R^2/r^3$$

by substituting on the right-hand side the values R, r and θ for an arbitrarily selected point on the trajectory.

A complete discussion of these Q_γ spaces has been published by Störmer.[50]

Figure 15 shows sections through the Z axis of four of the more characteristic Q_γ regions. These regions are described by the white parts when they are rotated about the Z axis.

A full investigation of the Q_γ spaces shows that only those corresponding to

$$-1 < \gamma \leqq 0$$

are open from the origin to infinity. All others are either closed or do not reach the origin of the system of coordinates. In Fig. 15, (a) and (d) show Q_γ spaces which are not open from the origin to infinity.

A mechanical interpretation of the equations for R and z.—Still more useful information concerning the trajectories may be obtained by giving a suitable mechanical interpretation to system (III).

If the arc s be considered as the time, and R and z as the Cartesian coordinates of a material point p in a plane, then system (III) defines the

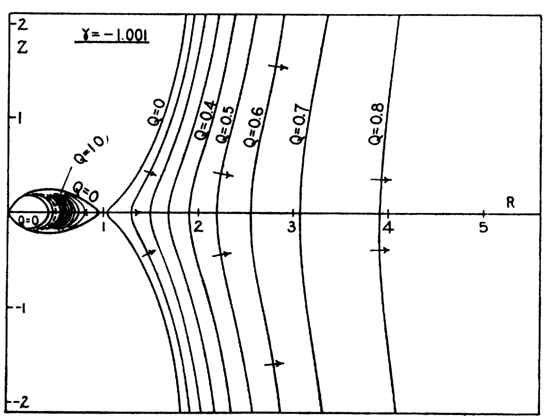

FIG. 16. Level lines for $\gamma = -1.001$ (after Störmer).

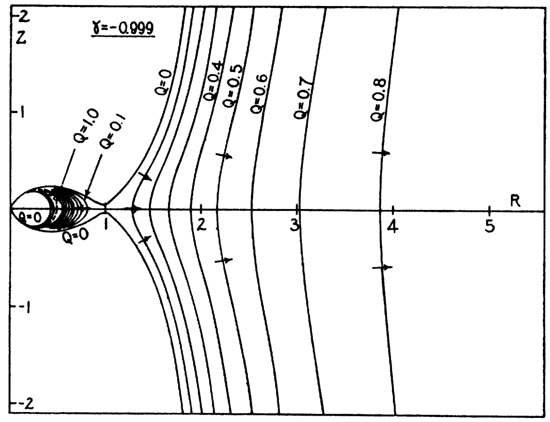

FIG. 17. Level lines for $\gamma = -0.999$ (after Störmer).

motion of that point under the action of a force derived from the force function $\frac{1}{2}Q$.

Consider an arbitrary plane ROZ through the Z axis (Fig. 14) as a plane of this kind. Let P be the corpuscle moving along the trajectory T. Draw a circle through P parallel to the $X-Y$ plane, with its center on the Z axis. Then p will be the point of intersection of this circle and the plane ROZ, and when the corpuscle P moves with constant velocity along T, the corresponding point p moves in the plane ROZ according to the above mentioned mechanical law, and will describe a certain plane curve K. Conversely, if the shape of the curve K is known, the form of the corresponding trajectory T in space may be easily found by the formula for $\sin\theta$. To each curve K there are in general two corresponding sets of trajectories, each comprising all trajectories that can be obtained from one of them by rotation about the Z axis. The first set corresponds to a motion along K in one direction; the second to a motion in the opposite direction, and the two sets are symmetrical to each other, referred to the fixed plane ROZ.

The study of the curves K is comparatively easy when the level lines Q are drawn, with Q taking equidistant values ranging from $Q=0$ to $Q=1$. These lines are situated exclusively in that part of the plane lying within the region previously called Q_γ, and in particular $Q=0$ marks the boundary of this region.

To facilitate the understanding of the trajectories, level lines corresponding to $Q=0$, 0.1, 0.2, 0.3$\cdots$0.9 and 1 for three values of the constant of integration are shown in Figs. 16, 17 and 18. The force acting on the point p will always be directed normally to the level lines, and towards the regions of higher values of Q, as

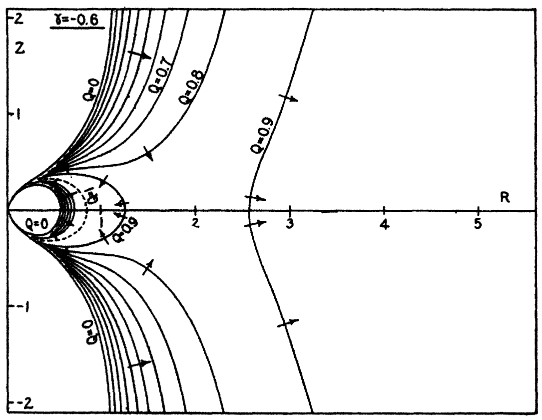

FIG. 18. Level lines for $\gamma=-0.6$ (after Störmer).

indicated by the arrows. The magnitude of this force will vary approximately in inverse proportion to the breadth of the spaces between two consecutive level lines. The motion of the point p will be very similar to that of a small sphere rolling without friction on an uneven surface, where the level lines indicate the shape of the surface, just as contour lines on a geographical chart show the topography of a region. The valleys are marked by level lines $Q=1$, and the highest regions by $Q=0$. The level lines $Q=0$ represent barriers over which the small sphere may not pass.

Graphical and numerical integration.—The above methods are very useful for a qualitative discussion of the trajectories. However, they are insufficient for a quantitative investigation. Yet without rigorously integrating the differential equations of motion, it is possible to find the trajectories with any desired degree of accuracy. This can be done by using graphical and numerical methods to integrate the differential equations. The former should be employed when a preliminary and not very exact view is required; the latter, which is quite laborious, when the greatest possible accuracy is desired.

Störmer has described the method of graphical integration which he has employed.[51f] It is based on a further development of a method given by Lord Kelvin, and makes use of a simple construction of the radius of curvature. The earlier method of numerical integration has also been discussed in detail by Störmer.[51g] He found it relatively simple to integrate Eq. (II) by means of Simpson's formula, when the curve K had been calculated.

In 1930, Störmer commenced further research on trajectories, using a more convenient form of numerical integration.[51p] Instead of directly integrating the equations of system (III) he transformed these to more tractable forms, and then integrated the transformed system. The steps by which this was done will be indicated very briefly.

Two new dependent variables, and an independent variable were introduced into system (III), by means of a Goursat transformation. By employing Hamilton's principle, the equations of (III) were brought into a form which permitted numerical integration. R and z were found by working back through the transforming equations. Simpson's formula gave the arc length s. When the electron is far from the magnet, the transformed equations have the form of a Liouville type system, which is completely integrable by means of elliptic functions.[51j] Details of approximate formulae for electron paths, found by means of elliptic functions and integrals, have recently been published by Störmer.[51r]

A general view of the trajectories.—A short summary will be given of the outstanding features of the trajectories found by numerical integration.

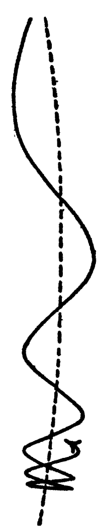

FIG. 19.

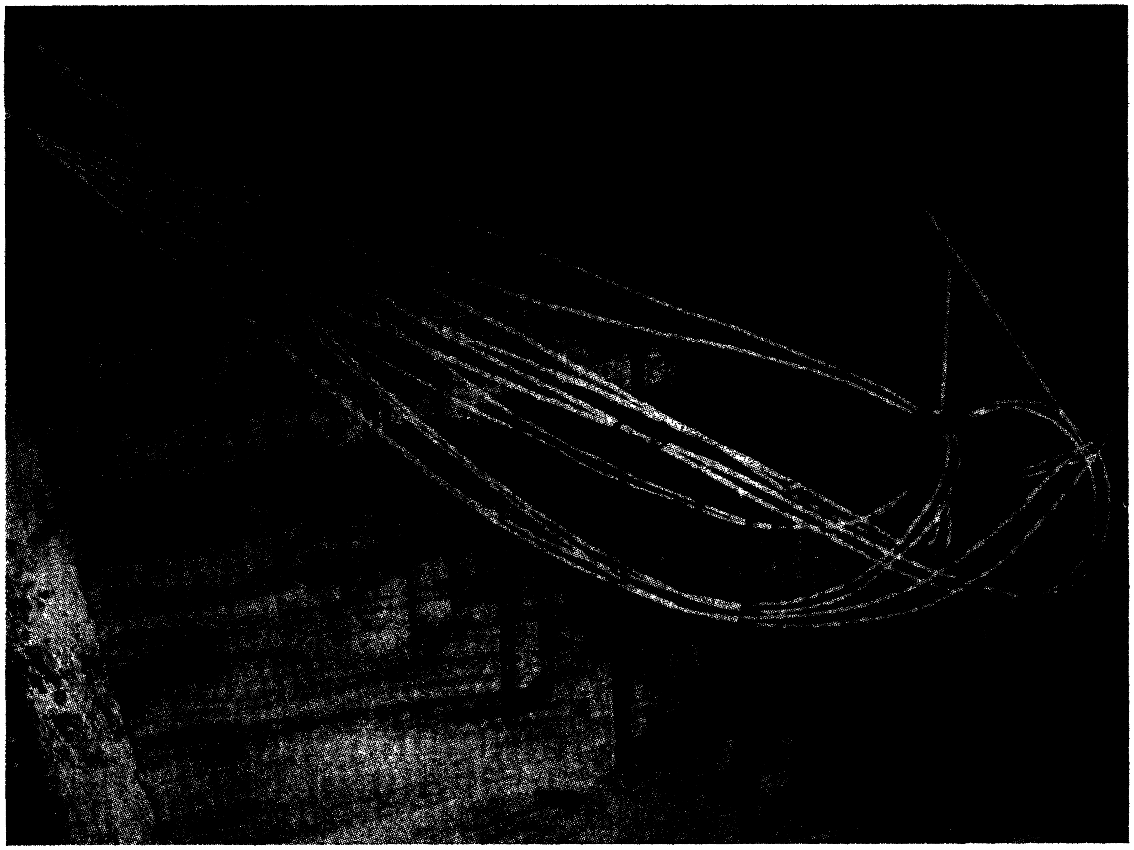

FIG. 20. Electron orbits toward the earth (after Störmer).

Because of possible applications to the aurora, the first calculations were made with the object of finding the trajectories of corpuscles coming from infinity and reaching the origin. Several series of such trajectories, issuing from a distant point, were computed, but without much success. After coming into the neighborhood of the origin, the trajectories, for the most part, receded. Only in a few cases were trajectories found that approached near enough to verify theoretical conclusions as to the shape of the orbits near the elementary magnets. The latter were similar to the well-known geodetic lines of a cone of revolution, with a very small opening, as seen in Fig. 19.

A model of the various trajectories in their relation to the earth was constructed by Störmer, with the aid of the graphical method of integration.[511] This model, shown in Fig. 20, illustrates certain characteristic features. The first of interest is the large group of trajectories turning around the nocturnal (evening and night) hemisphere of the globe, these being supposed to come from the sun. In the neighborhood of this group there are trajectories encircling the globe in undulating fashion. On the morning side, the model shows another series of paths bending abruptly away. Near the early afternoon side there is a whirl, in the center of which, trajectories approach the globe in spirals, as shown in Fig. 19.

Trajectories passing through the origin.—For application to the aurora, the problem of finding the paths of corpuscles coming from infinity and reaching the origin is of fundamental importance.

The distribution of the points of intersection of the trajectories with the surface of the small sphere is shown on a large scale in Fig. 21. The computed paths are only the simplest ones corresponding to the values of γ written by the side of each point. For γ between -0.93 and -1, there is an immense number of remarkable trajectories, which have curious forms. For instance, a corpuscle following a trajectory of this kind may make a number of revolutions around the globe passing over and under the $X-Y$ plane many times before reaching the origin. Others may approach in spirals and then recede before finally reaching the origin.

(c) *Comparison of theory with experiment*

The foregoing results from theoretical considerations agree remarkably well with Birkeland's experiments.

An experiment done by him,[8] in which a mass of cathode rays were sent towards a magnetized globe, showed that there was a toroidal space about the sphere which was free from electrons similar to those spaces which can be noticed in Fig. 15, (b) and (c).

Also, some of Birkeland's other work showed the positions where the electron beams struck the magnetized sphere. These positions agreed very well with those shown by Störmer's wire model.[8]

Recently, Brüche has experimentally investigated the paths of electron rays in the field of a small magnetized sphere.[9] The paths as found by him show remarkably fine agreement with those calculated by Störmer from theoretical considerations.

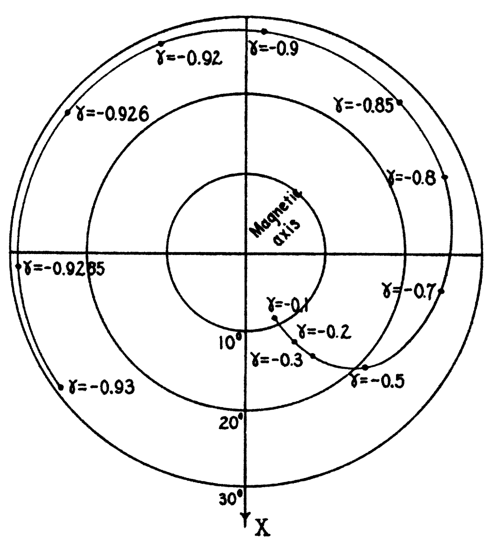

FIG. 21. Distribution of the points of intersection of the trajectories with the surface of a small sphere.

(d) *Application to the aurora*

Agreement between theory and observation.—It should be remembered, in applying these results to the aurora, that the earth's magnetic field is considered identical with the field due to a magnetic dipole with its middle point at the center of the earth, and its axis cutting the earth's surface near Smith's Sound in Greenland. The magnetic moment of this dipole is about

$$M=8.4\times10^{25} \text{ c.g.s. units.}$$

The standard of length is given by

$$C=(Me/mv)^{\frac{1}{2}}.$$

Now a charged particle moving perpendicular to the lines of force of a magnetic field of strength H will describe an arc of a circle of radius ρ, given by the equation

$$(m/e)v=H\rho.$$

The following table gives the values of $H\rho$ and C for cathode rays, β-rays and α-rays.

		$H\rho$	C IN KM
Cathode rays	from	108	8,900,000
	to	543	4,000,000
β-rays	from	1801	2,200,000
	to	4524	1,400,000
α-rays	from	291,000	170,000
	to	398,000	146,000

It can be seen that the standard of length chosen is very large compared to the dimensions of the earth.

Figure 22 is an enlarged diagram of the toroidal space into which corpuscles cannot penetrate which theory indicates should be about the earth, similar to the ones that can be seen in Fig. 15. If in the equation

$$\sin\theta=2\gamma/R+R/r^3,$$

$R=r\cos\psi$ be substituted, and $\gamma=-1$ be considered near enough to the actual values of γ for corpuscles reaching the earth, an equation

$$r=C\frac{\cos^2\psi}{1+(1+\cos^3\psi)^{\frac{1}{2}}} \qquad \text{(IV)}$$

in polar coordinates may be obtained for the cross section of this toroidal space.

Figure 22 shows that corpuscles coming to the earth from a distance greater than C, can only meet the earth in two zones around the magnetic axis, limited by the intersection of this toroidal space with the earth's atmosphere.

This space cuts the magnetic equatorial plane of the earth in a circular trace of radius

$$r=(2^{\frac{1}{2}}-1)C.$$

Call Ω the angular distance from the magnetic axis to a point on the intersection of the toroidal space with the upper atmosphere, so that $\Omega=90°-\psi$, and call A the radius of the upper atmosphere. Then if A is small compared with C, it follows from Eq. (IV) that

$$\sin\Omega=(2A/C)^{\frac{1}{2}}.$$

This formula determines theoretically the southern boundary of the aurora borealis, and the northern boundary of the aurora australis.

For cathode rays, Ω lies between 2° and 4°, and for β-rays between 4° and 6°. However, in the case of α-rays, Ω is between 16° and 19°. Now the zone of maximum frequency of aurorae has an angular radius of about 23° (Fig. 2).

Here is met the first objection to the theory that the aurora is due to negatively charged particles. The resulting radius of the aurora zone is far too small. For the positively charged particles however, theory gives a zone much closer to the observed one. Störmer has suggested an explanation for this, which would also provide a reasonable hypothesis for the fact that, during magnetic storms, the aurora borealis extends much farther south than at ordinary times.[51c and d] It has been mentioned that the theory leads to a large number of corpuscular rays bending around the earth on the afternoon and night side. It seemed probable that so many corpuscular rays, and even more so a closed corpuscular ring, might have an influence on the corpuscles coming from without, and effect the situation of the aurora belts. Störmer believes that the wireless echoes heard in 1928 from the Dutch station Eindhoven[55] support his suggestion. Brüche found experimentally that the presence of a circular current of electricity about his magnetized sphere brought the precipitation zone of the cathode rays considerably nearer the equator of the sphere.[9]

Chapman, however, thinks that the hypothesis of a ring-current with as large a radius as

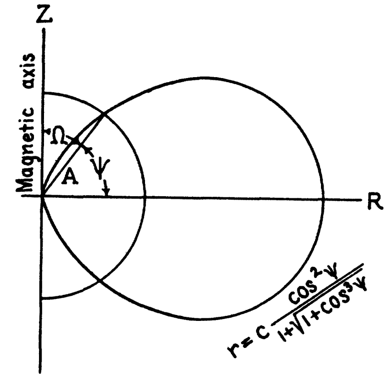

FIG. 22. Section of earth and toroidal space.

Störmer proposed, creates more difficulties than it solves.[14] The main objection to it is that such a ring-current could not hold together against the mutual electrostatic repulsion of its parts if composed of charges of one sign only, especially if, as Störmer suggested, these are electrons. Besides, a ring of such a size would far more than nullify the earth's field in the region where it is set up, thus destroying and reversing the forces that are assumed to create it. Chapman and Ferraro[12] have considered the problem, and they conclude that the ring, if it exists, is likely to be far smaller than Störmer supposed.

Another limitation of the auroral zones appears when the results concerning the trajectories through the origin are applied. If it be assumed that $H\rho$ lies between 100 and 400,000, only those corpuscles with trajectories situated in the vicinity of those reaching the origin can reach the earth. The others return into space. Now the angle between the plane normal to the earth's magnetic axis and the line from the earth to the sun, varies from $-35°$ to $+35°$ and consequently, trajectories coming from points situated outside this interval of direction must be excluded.

This means that regions Q_γ where γ lies between 0 and -0.2 must be discarded. Thus an inner limitation of the auroral zones may be accounted for.

The breadth of the auroral belts is reduced if only the aurorae seen during the night are considered. This would exclude all paths for γ from approximately -0.2 to -0.5.

Vegard has drawn attention to a fact which cannot be accounted for by the simple theory.

How the aurora borealis can appear in winter, when the earth-sun line makes such an angle with the magnetic equatorial plane, has not yet been explained. The present theory indicates that corpuscles coming from that direction could not cause aurora. Working through the theory again with the assumption of a magnetic sun might clear this difficulty up.

A more detailed study of the trajectories which come down to the earth's atmosphere shows that bundles of rays may sometimes be spread out by the magnetic field of the earth into long thin curtains, many thousand times longer than they are thick, corresponding to one of the most striking features of the auroral phenomena.[51b] Trajectories corresponding to γ- values of about -0.939 give a qualitative explanation of the horseshoe-formed curtains.[51q]

The theory also explains very well the rapidly moving forms of aurorae. Because of the fact that the magnetic axis of the earth swings daily about its spin axis and that emanation centers on the sun are in all probability continually in motion relative to the earth-sun line, it would be expected that aurorae would not stay for long in the same position, but would show considerable motion.

The auroral bows which often stay in one position for many hours are not so easily explained. Simultaneous observations all along the maximal zone such as were taken during the polar year, should help to solve this problem.

The nature of the corpuscles.—Störmer considers that some of the ray forms consist of thin corpuscle bundles, in which the individual bands form spirals about the lines of force of the earth's field. By assuming that the corpuscles undergo a reversal of direction at the bottom of the luminous portion of a ray streamer, as shown in Fig. 19, he has calculated from the width of these streamers, that the product $H\rho$ is probably less than 15,000 c.g.s. units. Vegard and Krogness,[67] after visual observation of ray thicknesses, conclude that $H\rho < 10{,}000$ c.g.s. units. On this basis, the corpuscles which cause the streamers, corresponding to cathode or β-rays, must be negative, for if they were α-particles, the thickness of the streamers would be considerably greater than the actually observed thickness.

The above discussion indicates that negatively charged particles cause the auroral forms with ray structure. The theory shows that the corpuscles may be electrons if a corpuscular ring about the earth can be assumed, as mentioned previously.

On the basis of these two results, the opinion at present is that the ray structure forms are caused by negatively charged particles.

However, the shapes taken up by the aurorae without ray structure, especially the pulsating surfaces, suggest that they may be due to positively charged particles.

Objections to this corpuscular theory.—One of the most serious objections to the corpuscular theory is the one brought forward by Schuster, that such beams of charged particles as required by the theory would dissolve because of electrostatic repulsion, long before reaching the earth.[47] Schuster's objection has been amplified and strengthened by Lindemann,[29] Swann,[63] and Chapman and Ferraro.[12] The latter do not believe that by the time the stream reached the earth, a residue of charge sufficient to produce the aurora would remain.

However, Schuster considered that his objection concerned only big streams whose magnetic fields are comparable to those shown during magnetic storms, and he regarded Birkeland's auroral theory as still tenable if the electron stream was suitably rare. Störmer's view is similar, believing that if very *thin* bundles are shot out from the sun and maintained for a time, the influence of electrostatic repulsion is not very great.

(e) *Recent mathematical treatment*

Störmer has also investigated the equations of motion of a charged particle when the particle, besides being influenced by the field of a magnetized sphere is also moving under the effect of a force which varies as the inverse square of the distance from the charged particle to the center of the sphere.[51p]

A great many different types of spaces have been studied. The spaces instead of depending on just one constant, as in the simple case, are determined by two constants. The number of these spaces is very great, and they assume widely differing shapes.

This more complex case has not been treated fully yet, but it is expected that further research on the problem will bring new understanding of the aurora.

(f) *A further corpuscular stream theory*

Chapman and Ferraro have studied in detail the question of corpuscular streams from the sun, and to avoid the above mentioned difficulties about electrostatic repulsion in the stream, they postulate a neutral stream of ions, electrons and perhaps neutral atoms.[10, 12]

Their study of the motion of a neutral ionized stream in a magnetic field indicates that within the stream the ions and electrons can move together almost rectilinearly, without spiralling, and with only a slight deflection by the field. The magnetic field of the stream nearing the earth opposes the advance of the stream into the stronger earth's field. In this manner a hollow region around the earth is formed in the stream, open at the back of the earth, as viewed from the sun. This hollow gradually shrinks, at a diminishing rate as the surface advances. If the stream were directed towards the earth for a considerable period of time, the hollow region would eventually close up on the earth. However, because the sun is rotating, the stream would probably pass on before this happens. The part of the stream which has collected near the earth remains for some time, probably as a ring around the equator, but gradually disappears by the passage of the ions and electrons along the lines of force of the earth's field, into the atmosphere in high latitudes.

Milne has developed an interesting theory about the emission of corpuscles from the sun.[40] According to this, positive ions are projected outwards from the sun by selective radiation pressure, while the radiative acceleration of the electrons is much smaller. However, it is clear that the electrons would be drawn after the ions by electrostatic forces. Neutral atoms would be emitted in the same manner as ions. Thus this theory provides for the emission from the sun of just such a stream as Chapman and Ferraro considered, a neutral stream of ions, electrons and neutral atoms.

But Milne computed that the velocity of calcium ions emitted from the sun would be of the order of 1.6×10^8 cm/sec. He has indicated that for this velocity the equivalent air range, that is, the range in air of the density existing at ground level in the earth's atmosphere, of calcium atoms is only about 0.15 cm, while the equivalent air range in the earth's atmosphere vertically down to 80 km, which is approximately the lower limit of the aurora, is 5.6 cm.

Chapman has suggested that 5.6 cm is too low a value, since there is little doubt that the calculated densities on which this estimate was based are too small for heights above 80 km. He believes that 10 cm is a preferable value.[10]

Milne's theory gives an equivalent air range of 0.15 cm, while auroral height observations seem to demand an equivalent air range of at least 10 cm. This is a serious discrepancy, and suggests that the theory is of little value in explaining auroral phenomena.

B. Secondary corpuscle theory

(a) *The ultraviolet light theory*

A somewhat different theory of the aurora has been put forward by Maris and Hulbert.[30] Their theory discusses the aurora under two conditions:

1. When the sun shows no unusual activity.
2. When the sun is exceptionally active.

Quiet sun.—Maris and Hulbert postulate that, in the region above 450 km, where molecular free paths are very long, a portion of the highly absorbed ultraviolet radiation from the sun is converted into kinetic energy by processes of atomic excitation and ionic recombination, and dissociates about 10^6 atoms per square centimeter per second. Further, these atoms travel away from the earth with velocities as high as 10 km/sec., reach heights of 30,000 to 50,000 km in three hours, and are then ionized by the ultraviolet sunlight. The ion pairs thus formed spiral about the lines of force of the earth's magnetic field, and the majority, if they were originally in the equatorial regions, reach the polar latitudes after about nine hours. They fall into a zone roughly 25° from the magnetic poles, and give rise to the aurorae there. Maris and Hulbert quote the fact that short wireless waves traverse polar regions in support of the idea that the ionization is due to the ion influx from lower latitudes.

Active sun.—It was assumed that the sun, when active, emits a sudden (one half-hour) blast of ultraviolet radiation. During a blast like this, the ultraviolet energy between the wave-lengths 500 to 1000A would be 10^5 times greater than ordinarily. This ultraviolet energy, completely absorbed by the upper atmosphere at about 200 km, blasts some of the gases out to great distances. These atoms may produce ions up to 40,000 km. The first effect of these ions traveling at such great heights is to produce a sudden current of the order of 10^6 amperes, in planes parallel to the equator, which causes a magnetic field of 10^{-3} gauss simultaneously over the whole earth. This is in agreement with what has been observed when magnetic storms occur. Numbers of ions descend to the zones 23° from the magnetic poles, and give rise to the aurorae there.

Some consequences of the theory.—Hulbert[25] has drawn attention to the following consequences of the ultraviolet theory of aurorae.

(a) There may be auroral displays in polar regions with no magnetic storms in temperate latitudes.

(b) Auroral displays in temperate latitudes should as a rule be accompanied by magnetic storms.

(c) Polar aurorae should follow several hours or a day after strong magnetic perturbations observed in temperate latitudes.

He finds that these inferences are in agreement with the data from magnetic observatories and auroral observations in temperate and polar latitudes.

Some difficulties.—Chapman, although finding this theory ingenious and interesting, has pointed out several serious objections to it.[11] Chapman shows that the velocities of 10 km/sec. are too low to permit the corpuscles to penetrate to 80 km. Furthermore, he considers that this theory does not explain a ıroral forms as well as Störmer's does.

VII. Conclusion

It may be seen from this review of the more important facts which are known about aurorae, that there are still a great number of problems which remain to be solved.

Only a beginning has as yet been made towards the solution of the complete problem. It is a field in which a great deal of profitable research may yet be done.

In conclusion, the writer wishes to acknowledge his indebtedness to Professor C. Störmer, who has read in detail the manuscript of this paper, and has made a number of valuable suggestions. In addition, he has been kind enough to supply the photographs.

The writer also wishes to express his thanks to Mr. J. Patterson, Controller of the Meteorological Service of Canada; to Mr. A. Thomson, the Chief Physicist; to Mr. W. E. K. Middleton of the Meteorological Service; and to Professor E. F. Burton and Dr. P. Millman of the University of Toronto, for helpful criticism and advice.

Bibliography

1. T. Alty and F. J. Wilson, Nature **133**, 687–688 (1934).
2. A. J. Angstrom, Pogg. Ann. **137**, 161–163 (1869).
3. S. A. Arrhenius, Öfv. Kong. Vet.-Akad. Handl. (Stockholm, 1900).
4. H. D. Babcock, Astrophys. J. **57**, 209–221 (1923).
5. W. Bauer, Naturwiss. **20**, 287–288 (1932).
6. K. Birkeland, Arch. Sci. Phys. **1**, 497–512 (1896).
7. K. Birkeland, Vid. selsk. Skr. No. 1 (1901).
8. K. Birkeland, *The Norwegian Aurora-Polaris Expedition*, Vol. I, Sect. 2 (1902–1903), p. 712, 598.
9. E. Brüche, Terr. Mag. and Atmos. Elec. **36**, 41–52 (1931).
10. S. Chapman, Mon. Not. R. Astr. Soc. **89**, 456–470 (1929).
11. S. Chapman, Mon. Not. R. Astr. Soc. Geophys. Supp. **2**, 296–300 (1930).
12. S. Chapman and V. C. A. Ferraro, Terr. Mag. and Atmos. Elec. **36**, 77–97 (1931); **36**, 171–186 (1931); **37**, 147–156 (1932); **37**, 269–272 (1932); **37**, 421–429 (1932); **38**, 79–96 (1933).
13. S. Chapman, Terr. Mag. and Atmos. Elec. **39**, 299–303 (1934).
14. S. Chapman, Terr. Mag. and Atmos. Elec. **40**, 353 (1935).
15. B. W. Currie, Terr. Mag. and Atmos. Elec. **39**, 293–297 (1934).
16. B. W. Currie and H. W. Edwards, Terr. Mag. and Atmos. Elec. **41**, 265–278 (1936).
17. M. A. Dauvillier, Comptes rendus, Acad. Sc. **197**, 780–781 (1933).
18. M. A. Dauvillier, (a) J. de phys. et rad. **5**, 398–412 (1934). (b) The Electrician **112**, 787–789 (1934).
19. F. T. Davies. Terr. Mag. and Atmos. Elec. **36**, 199–230 (1931).
20. J. Egedal, Pub. Danske Meteorolog. Inst., Copenhagen. No. 10. 1930.

21. H. Garrique, Comptes rendus Acad. Sc. **198**, 1398–1400 (1934).
22. Goldstein, Wiedemanns Ann. **12**, 266 (1881).
23. L. Harang and E. Tønsberg, Geofysiske Pub. 9, No. 5 (1932).
24. L. Harang and W. Bauer, Gerlands Beitr. z. Geophys. **37**, 109–115 (1932).
25. E. O. Hulbert, Phys. Rev. **34**, 344–351 (1929).
26. J. H. Jeans, *Electricity and Magnetism*, fifth edition, (Cambridge Univ. Press,(1927), p. 403.
27. Kosirev and Eropkin, Poulkovo Astr. Circular, No. 18 (1936).
28. O. Krogness, reported by C. Störmer, Nature **123**, 82–83 (1929).
29. Lindemann, Phil. Mag. **38**, 669–684 (1919).
30. H. B. Maris and E. O. Hulbert, Phys. Rev. **33**, 412–431 (1929).
31. J. C. McLennan and G. M. Shrum, Proc. Roy. Soc. **A106**, 138–149 (1924).
32. J. C. McLennan and G. M. Shrum, Proc. Roy. Soc. **A108**, 501–512 (1925).
33. J. C. McLennan, A. B. McLay and H. G. Smith, Proc. Roy. Soc. **A112**, 76–94 (1926).
34. J. C. McLennan, J. H. McLeod and W. C. McQuarrie, Proc. Roy. Soc. **A114**, 1–22 (1927).
35. J. C. McLennan and J. H. McLeod, Proc. Roy. Soc. **A115**, 515–527 (1927).
36. J. C. McLennan, J. H. McLeod and R. Ruedy, Proc. Roy. Soc. **A120**, 327–357 (1928).
37. J. C. McLennan, Roy. Soc. Can. **22**, 397 (1928).
38. J. C. McLennan and H. J. C. Ireton, Proc. Roy. Soc. **A129**, 31–43 (1930).
39. J. C. McLennan, H. S. Wynne-Edwards and H. J. C. Ireton, Can. J. Research **5**, 285–296 (1931).
40. E. A. Milne, Mon. Not. R. Astr. Soc. **86**, 459–473 (1926).
41. A. Paulsen, *Sur la nature et l'origine de l'aurore boreale.* (Kjøbenhavn, 1894.)
42. A. Paulsen, Comptes rendus Acad. Sc. **130**, 655–656 (1900).
43. H. Poincaré, Comptes rendus Acad. Sc. **123**, 530–533 (1896).
44. Lord Rayleigh, Proc. Roy. Soc. **A101**, 114–124 (1922).
45. Lord Rayleigh, Proc. Roy. Soc. **A119**, 11–33 (1928).
46. W. J. Rooney, Terr. Mag. and Atmos. Elec. **39**, 103–109 (1934).
47. A. Schuster, Proc. Roy. Soc. **A85**, 45–50 (1911).
48. G. C. Simpson, Quarterly J. Roy. Meteorolog. Soc. **59**, 185–190 (1933).
49. V. M. Slipher, Astrophys. J. **49**, 266–275 (1919).
50. C. Störmer, Vid. selsk. Skr. No. 3 (1904).
51. C. Störmer, (a) Arch. Math. Naturv. **28**, No. 2 (1906). (b) Arch. Sci. Phys. **24** (1907). (c) **32** (1911). (d) **33** (1912). (e) Vid. selsk. Skr. (1907). (f) ibid. (1908). (g) (I, II, III) (1913), (1914). (h) No. 7 (1912). (i) No. 5 (1916). (j) Comptes rendus Acad. Sc. **162**, 829–832 (1916). (k) Terr. Mag. and Atmos. Elec. **22**, 23–34, 97–112 (1917). (l) *The Book of the Rice Inst.*, Vol. 3 (Houston, Texas, 1918). (m) Rep. Internat. Math. Cong., Toronto (1924). **1**, 139–147 (1928). (n) Zeits. Astrophys. **1**, 237–274 (1930). (p) Gerlands Beitr. Geophys. Supplement **1**, 37–86 (1931). (q) Terr. Mag. and Atmos. Elec. **37**, 375–388 (1932). (r) Zeits. Astrophys. **6**, 333–344 (1933).
52. C. Störmer, (a) Vid. selsk. Skr. No. 17 (1911) (b). Terr. Mag. and Atmos. Elec. **18**, 133–135 (1913). (c) **20**, 1–12, 159–174 (1915). (d) **21**, 45–56, 157–168 (1916). (e) Geofysiske Pub. 1: No. 5 (1921).
53. C. Störmer, Geofysiske Pub. **4**, 104 (1924).
54. C. Störmer, Gerlands Beitr. z. Geophys. **17**, H. **2**, 254–269 (1927).
55. C. Störmer, Nature **122**, 681 (1928).
56. C. Störmer, Nature **124**, 263–264 (1929).
57. C. Störmer, Zeits. Geophysik **5**, 177–194 (1929).
58. C. Störmer, *Atlas of Aurora*, International Geodetic and Geophysical Union. (Oslo, A. W. Brøgger, 1930.)
59. C. Störmer, Gerlands Beitr. z. Geophys. Supplementband **1**, 10–13 (1931).
60. C. Störmer, Terr. Mag. and Atmos. Elec. **37**, 475–477 (1932).
61. C. Störmer, Elek. Nachrichten-technik **10**, 60–68 (1933).
62. C. Störmer, Nature **139**, 584 (1937).
63. Swann, Phil. Mag. **47**, 306 (1924).
64. L. Vegard, (a) Arch. Math. Naturv. (1911). (b) Phil. Mag. **23**, 211–237 (1912). (c) Vid. selsk. Skr. (1916).
65. L. Vegard, Physik. Zeits. **14**, 677–681 (1913).
66. L. Vegard, Ann. d. Physik **50**, 855 (1916).
67. L. Vegard and O. Krogness, Geofysiske Pub. **1**, No. 1 (1920).
68. L. Vegard, Zeits. f. Physik **16**, 367–390 (1923).
69. L. Vegard, (a) Phil. Mag. **46**, 193, 577 (1923). (b) Nature **113**, 716–717 (1924). (c) Ann. d. Physik **79**, 377–441 (1926).
70. L. Vegard, Koninkl. Akad. Wet. Amsterdam, **27**, Nos. 1, 2 (1924).
71. L. Vegard, Nature **117**, 356 (1926).
72. L. Vegard, Nature **124**, 947–948 (1929).
73. L. Vegard, Nature **129**, 468 (1932).
74. L. Vegard, Geofysiske Pub. **9**, No. 11 (1932).
75. L. Vegard, Nature **132**, 682 (1933).
76. L. Vegard, Geofysiske Pub. **10**, No. 4 (1933).
77. L. Vegard and L. Harang, Geofysiske Pub. **10**, No. 5 (1933).
78. L. Vegard and L. Harang, Geofysiske Pub. **11**, No. 1 (1934).
79. L. Vegard and E. Tønsberg, Nature **137**, 778–779 (1936).
80. L. Vegard, Nature **138**, 930–931 (1936).
81. J. Westmann, *Mission Sc. pour la mesure d'un Arc de Méridien au Spitzberg*, Vol. 2 (Stockholm, 1904).

FIG. 1. Draperies with ray structure, Kongsberg, Norway, January 24, 1936. Phot. Busengdal.

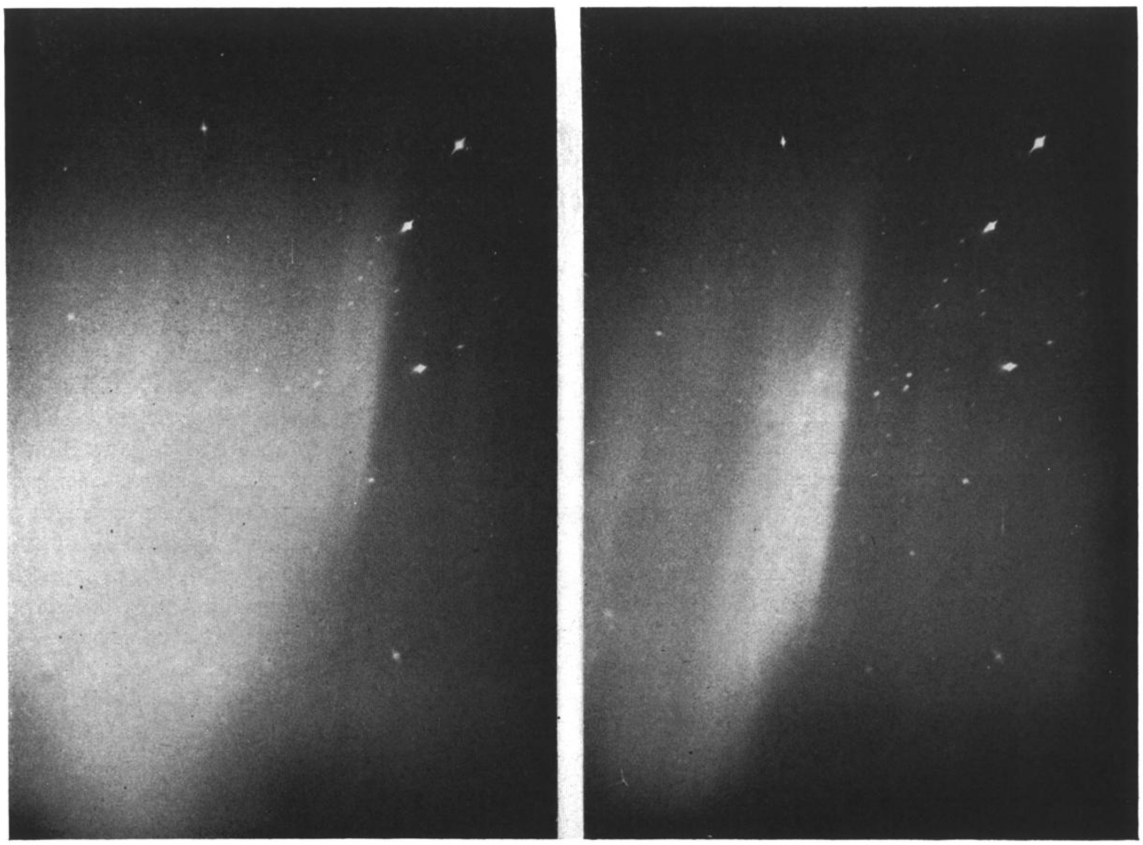

FIG. 13. Sunlit aurora, photographed simultaneously from Oslo and Lillehammer.

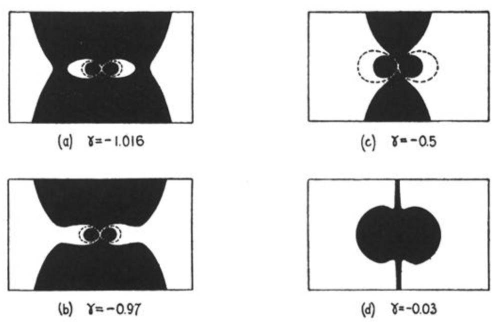

FIG. 15. Sections through the Z axis of four of the more characteristic Q_γ regions to which the trajectory must be confined. The regions are represented by the white parts when rotated about the Z axis.

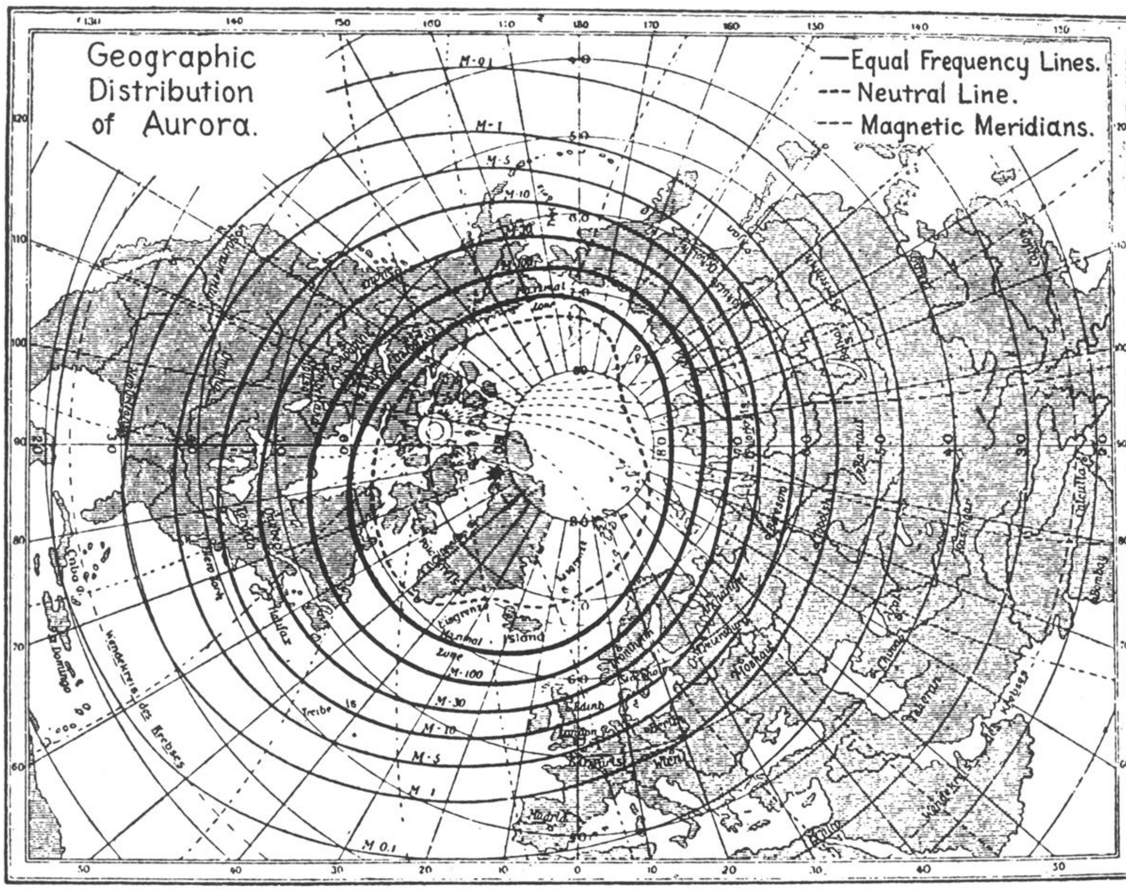

Fig. 2. Map showing the geographic distribution of aurora. (The star indicates approximately the point of intersection in 1900 of the earth's magnetic axis with the Northern Hemisphere.)

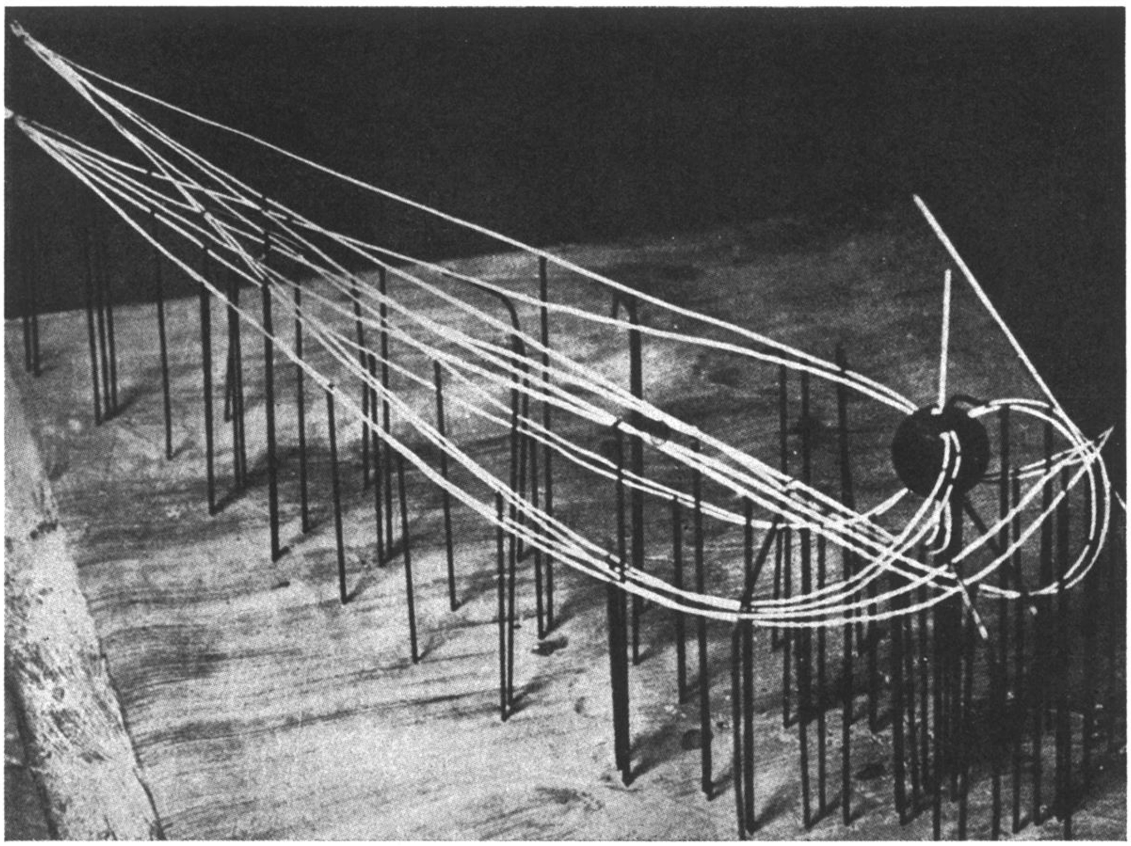

FIG. 20. Electron orbits toward the earth (after Störmer).