REVIEWS OF

MODERN PHYSICS

VOLUME 9 JANUARY, 1937 NUMBER 1

The Physics of the Ionosphere

HARRY ROWE MIMNO
Harvard University, Cambridge, Massachusetts

TABLE OF CONTENTS

A. INTRODUCTION

ON October 22, 1924, an eminent British engineer delivered a lecture on "Unsolved Problems of Wireless" before the Radio Society of Great Britain. In his address, R. H. Barfield proposed the following problems:

1. Why is long distance wireless transmission possible at all?
2. Why are signals stronger by night than by day?
3. Why do direction finding stations experience large errors by night, while the errors by day are practically negligible?
4. Why does the phenomenon known as fading occur?

During the next decade, these questions and a number of allied problems were completely solved from the engineering point of view. During the course of the investigation some additional engineering problems were disclosed, which still await solution. These will be described below in their proper setting.

However, the most interesting results, derived from intensive experimental research during the past twelve years, belong properly in the domain of pure physics. New and interesting experimental and theoretical problems confront the physicist. The development of new experimental tools and methods permits exact measurements in a field of atmospheric physics which is not too complex for useful theoretical analysis. The theory is not complete, and the experimental methods require extension, but a satisfactory beginning has been made in a young and active branch of physical research. The new ionosphere problems are intimately connected with a number of other geophysical, lunar and solar investigations which have attracted the attention of physicists for a number of years. Such related experiments include investigations of cosmic rays, terrestrial magnetism, solar activity, auroral displays, ozone distribution, thunderstorm activity, atmospheric temperatures, luminous clouds, meteor trails, air-mass movements, earth currents, stratosphere meteorology, and elastic deformations of the earth's crust.

After a brief historical survey of the early radio theories and experiments, I shall attempt to present a unified account of a number of researches carried on during the last twelve years. As a complete collection of the original publications of this active period would fill a number of large volumes, it is necessary for me to exclude many valuable papers from this review. Therefore the papers mentioned in the bibliography have been selected from a much larger group, as they appear to illustrate adequately various points in the presentation. The references are intended to be representative, though by no means complete. Supplementary references may be found in nearly all of the papers which have been quoted. I have not adhered to a strict chronological sequence, and have not attempted to solve innumerable problems of priority.

In every active branch of physics numerous highly controversial questions arise, and a number of them may be found in the ionosphere field. When discussing such matter I believe that the reviewer should not straddle the argument, but should formulate an unbiased personal opinion whenever sufficient evidence appears to be available. While attempting to do this, I have tried to avoid arbitrary statements, and have included representative references covering both sides of the dispute. The field is new and extensive and no progress can be expected without occasional justifiable errors in mathematical theory, experimental technique, and interpretation. I shall welcome correction and reproof.

B. Brief Historical Survey

When Marconi first transmitted radio signals across the Atlantic, on December 12, 1901, Lord Rayleigh pointed out that plane wave transmission could not account for the observed facts in a simple manner, and suggested that some sort of diffraction theory must be employed in order to explain how the wave followed the curvature of the earth.

The diffraction theory was investigated by Rayleigh,[1] MacDonald,[2] Poincaré,[3] Zenneck,[4] Sommerfeld,[5] Nicholson,[6] March,[7] von Rybczynski,[8] Love,[9] Van der Pol,[10] and numerous others. In the twenty-fourth Kelvin lecture Sir Frank E. Smith[11] summarized these investigations succinctly: "Many years' work by some of the most distinguished of the world's mathematicians did not suffice to bring this apparently innocent problem beyond a stage at which Nicholson could say that it was one in the whole field of mathematical analysis about which most divergent views are held." However it is generally conceded that Watson[12] presented an adequate survey of the problem in 1919, and exposed the errors which were responsible for such wide disagreement. The diffraction theory fails by an enormous factor to account for the field intensities observed at points far below the optical horizon. As pointed out by Larmor[13] in 1924, a 100-meter wave on the earth corresponds in scale to a wave of visible light on a sphere of 6 cm radius, and it is not natural to expect a sensible bending of the ray as a diffraction effect alone. The elaborate mathematical analysis merely confirms this expectation. Some divergent views have been expressed in comparatively recent years by Meissner[14] and by Kiebitz[15] but the computations offered by Kiebitz have been attacked by Mesny[16] with apparent success.

Since 1924, interest in the diffraction problem has declined to a considerable extent, since

[1] Lord Rayleigh, Proc. Roy. Soc. **72**, 40 (1904).

[2] H. M. MacDonald, Proc. Roy. Soc. **71**, 251 (1903); **72**, 59 (1904); Roy. Soc., Phil. Trans. **A210**, 113 (1911); Proc. Roy. Soc. **A90**, 50 (1914).

[3] H. Poincaré, Proc. Roy. Soc. **72**, 42 (1904); Rendiconti Circolo Matematico Palermo **29**, 169 (1910); Comptes rendus **154**, 795 (1912).

[4] J. Zenneck, Ann. d. Physik **23**, 846 (1907); *Lehrbuch der drahtlosen Telegraphie*, 2te. Auffl. (Stuttgart, 1913); *Wireless Telegraphy*, tr. by Seelig (New York, 1915).

[5] A. Sommerfeld, Ann. d. Physik **28**, 665 (1909).

[6] J. W. Nicholson, Phil. Mag. **19**, 516 and 757 (1910); Phil. Mag. **20**, 157, 1910; Phil. Mag. **21**, 62 and 281 (1911).

[7] H. W. March, Ann. d. Physik **37**, 29 (1912).

[8] W. von Rybczynski, Ann. d. Physik **41**, 191 (1913).

[9] A. E. H. Love, Roy. Soc., Phil. Trans. **215**, 105 (1915).

[10] B. Van der Pol, Phil. Mag. **38**, 365 (1919).

[11] F. E. Smith, J. Inst. Elec. Eng. **73**, 574 (1933); wireless section **9**, 22 (1934).

[12] G. N. Watson, Proc. Roy. Soc. **95**, 83 (1919).

[13] J. Larmor, Phil. Mag. **48**, 1025 (1924).

[14] A. Meissner, *Jahrbuch der drahtlosen Telegraphie* **24**, 85 (1924).

[15] F. Kiebitz, E. N. T. **3**, 376 (1926).

[16] R. Mesny, Onde Élec. **5**, 650 (1926); Onde Élec. **6**, 127 (1927).

numerous new experimental tests have proven conclusively that long-distance transmission is mainly dependent on entirely different factors. However, diffraction undoubtedly does play an important role in certain types of short-distance transmission, and I shall have occasion to refer to the matter again in Section C.

While the diffraction problem was being studied mathematically, various investigators were also considering the several possible methods of propagation in the earth's atmosphere. In 1878 Stewart[17] had suggested that conducting layers in the upper atmosphere might account for certain types of cyclic variations in terrestrial magnetism. This hypothesis was developed further by Schuster[18] in 1889. In March, 1902, Kennelly[19] published a short article, suggesting that "There is well-known evidence that the waves of wireless telegraphy, propagated through the ether and atmosphere over the surface of the ocean, are reflected by that electrically conducting surface." Three months later Heaviside[20] wrote a similar brief comment, which was published in December: "There may possibly be a sufficiently conducting layer in the upper air. If so, the waves will, so to speak, catch on it more or less. Then the guidance will be by the sea on one side and the upper layer on the other." The original contributions of Kennelly and Heaviside have been republished recently.[21, 22]

By the year 1912, G. W. Pierce and L. De Forest were discussing in private correspondence the probable explanation of radio signal "fading" in terms of interference between a ground wave and a sky wave. It is evident that the general nature of the phenomenon was well understood at that time, although there was some difference of opinion in regard to the exact path of the indirect ray.

In his early publication, Kennelly did not attempt to discuss the mechanism of atmospheric conduction, but justified his assumptions by referring to Thomson's[23] measurements of the conductivity of air in discharge tubes. Two important advances were therefore made by Eccles[24, 25] in 1912. At that time Eccles discussed the ionizing effect of solar radiation, and also presented the fundamental theory of ionic refraction. This phase of the theory was extended by Salpeter[26] and Van der Pol.[27] The next major advance was made in 1924, when Larmor[13, 28] reexamined the entire problem and ascribed the major part of the ionic refraction to the presence of free electrons in large numbers.

The possibility of refraction in the *lower* atmosphere, as a result of barometric gradient, water-vapor gradient, or temperature inversion, was also examined by several investigators. These effects were computed, and it soon became evident that the maximum amount of bending which could be produced by the lower atmosphere would be insufficient to account for long distance radio transmission. This phase of the matter has been summarized by Fleming[29] and by Larmor.[28]

Meanwhile, a simple empirical equation, known as the "Austin-Cohen formula," had obtained universal recognition as a basis for engineering design. It resulted from the analysis of a series of experiments on over-water transmission carried on under the supervision of Austin[30] in 1909 and 1910. On the cruisers Birmingham and Salem, numerous quantitative observations of signal strength *vs.* wave-length were made in cooperation with a powerful land station at Brant Rock, Massachusetts, while the ships moved away from the fixed station. The extreme distances were of the order of 1200 miles. Most of the original measurements were made on wave-lengths of 1000 meters and 3750 meters.

---

[17] B. Stewart, *Encyclopedia Britannica*, ninth edition, p. 181, 1878.

[18] A. Schuster, Roy. Soc., Phil. Trans. **A180**, 467 (1889); Roy. Soc., Phil. Trans. **A208**, 163 (1907).

[19] A. E. Kennelly, Elec. World **39**, 473 (1902).

[20] O. Heaviside, *Encyclopedia Britannica*, tenth edition, Vol. 33.

[21] K. W. Wagner, E. N. T. **8**, 515 (1931).

[22] F. E. Smith, J. Inst. Elec. Eng. wireless section **9**, 38 (1934).

[23] J. J. Thomson, *Recent Researches in Electricity and Magnetism* (Oxford, 1893), p. 101.

[24] W. H. Eccles, Proc. Roy. Soc. **A87**, 79 (1912).

[25] W. H. Eccles, Electrician **79**, 1015 (1912); *Jahrbuch der drahtlosen Telegraphie* **8**, 253 and 282 (1914).

[26] J. Salpeter, Physik Zeits. **14**, 201 (1913).

[27] B. Van der Pol, Dissertation (Utrecht, 1920).

[28] J. Larmor, *Jahrbuch der drahtlosen Telegraphie* **25**, 141 (1925).

[29] J. A. Fleming, *Principles of Electric Wave Telegraphy and Telephony* (1919), p. 660.

[30] L. W. Austin, National Bureau of Standards, Bulletin **7**, 315 (1911).

Austin's experimental results were adequately described by the empirical relation:

$$I_R = 4.25 I_S (h_1 h_2 / \lambda d) e^{-\alpha d \lambda^{-\frac{1}{2}}},$$

where I_R = current in a 25 ohm receiving antenna of height h_2 kilometers,
I_S = current in a transmitting antenna of height h_1 kilometers,
d = distance in kilometers,
λ = wave-length in kilometers,
$\alpha = 0.0015$ for transmission over sea water.

Fleming[31] shows that the first part of this equation is entirely consistent with the original equations of H. Hertz, and represents a simple inverse-square-law decrease of energy in the wave front. Austin added the exponential factor in order to fit the experimental curves, and he ascribed this to "atmospheric absorption." The quantity $\lambda^{-\frac{1}{2}}$, which occurs in the exponent, was introduced by L. Cohen, after a careful analysis of Austin's original data.

The theoretical formula deduced from diffraction theory[3, 6, 8] contained a somewhat similar exponential term. However, in addition to a large discrepancy in the numerical value of α, the diffraction formula involved the term $\lambda^{-\frac{1}{3}}$ in place of $\lambda^{-\frac{1}{2}}$. The experimental data seemed precise enough to preclude this substitution. Hence the "reflecting layer" hypothesis received considerable additional support in 1919 when Watson[32] derived the Austin-Cohen formula by solving a difficult mathematical problem based on wave propagation in a medium bounded by concentric conducting spheres. At a later date, Kenrick[33] indicated that Watson's result would not be invalidated if the boundaries of the conducting spheres were not sharply defined.

For a number of years, all new experiments, involving additional frequencies, longer distances, and various types of overland transmission, seemed merely to give greater support to the Austin-Cohen formula and to extend its range of usefulness. Immense sums were spent in constructing powerful long wave stations with multiple lines of antenna towers ranging up to 800 feet in height All wave-lengths below 200 meters were considered practically worthless for long distance communication. Consequently this entire range of wave-lengths was allotted to the amateurs.

Under such circumstances it is not surprising that the amateur operators were influenced by conventional ideas at first and therefore kept their transmitters very close to the 200-meter wave which represented their legal upper boundary. Amateur activities were severely restricted in many countries, and were entirely forbidden in others, but liberal governmental cooperation was offered in the United States by the Department of Commerce.

The subsequent astonishing progress of amateur radio represents an important and most unusual chapter in scientific history. The American Radio Relay League, a strictly noncommercial organization founded in 1914 by the late H. P. Maxim, became the nucleus of a flourishing international society which has an active membership list drawn from almost every remote corner of the world. It includes men and women from all age groups and has representatives in nearly every trade and profession. The League maintains a competent technical staff and its technical journal (QST), and handbook have become practically indispensable in professional engineering laboratories.

This remarkable progress accompanied and resulted from equally unexpected developments in the study of the short wave region, originally assigned to amateur use. Under favorable conditions erratic 200-meter communication was obtained at distances far beyond the limit predicted by the Austin-Cohen formula. Systematic tests over longer and longer distances were arranged. The latest improvements in vacuum tube apparatus were immediately incorporated in amateur circuits, while commercial progress was frequently hampered by patent restrictions and conservative economic policies. Finally, in 1921, the American amateurs sent an expedition, equipped with the latest type of receiving device, to the coast of Scotland, and carried on successful one-way short wave transatlantic tests on a prearranged schedule. The tests were repeated during the following winter and occasional two-way communication was established. It became increasingly evident that the orthodox formula was not a dependable guide in this short wave region.

[31] J. A. Fleming, *Principles of Electric Wave Telegraphy and Telephony* (1919), p. 650.
[32] G. N. Watson, Proc. Roy. Soc. **95**, 546 (1919).
[33] G. W. Kenrick, Phys. Rev. **16**, 1040 (1928).

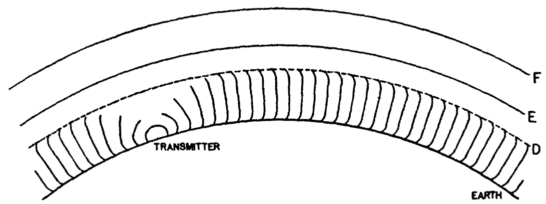


FIG. 1. Long wave daylight transmission.

Stimulated by natural curiosity, and by the overcrowded conditions near 200 meters, some of the more venturesome amateurs began to investigate still shorter waves. To everyone's surprise the transmission became stronger and less erratic. Naturally the downward movement was accelerated, and was limited only by the design of vacuum tube circuits for increasingly higher frequencies.

Below 50 meters another entirely new effect, of outstanding importance, was observed and partially explained. Apparently the signal strength decreased to zero at points relatively close to the transmitter (say 50 miles) but exceptionally efficient transmission could be maintained, under favorable conditions, between low power stations thousands of miles apart. The outer boundary of the "zone of silence" or "skip region" appeared to be sharply marked.[34]

A similar "zone of silence" had been noted previously in connection with ordinary sound wave transmission of loud noises (produced by artillery fire, or by heavy blasting). The occurrence of such acoustic mirages may be explained in a satisfactory quantitative manner by considering the refraction produced by temperature inversion in the stratosphere. However, the refraction of electromagnetic waves in this region is insufficient to account for the observed radio mirages. Numerous "skip zone" observations were carried on in 1924 by an amateur group headed by Reinartz[35] who correctly ascribed this new phenomenon to the effect of the ionized region. These amateur investigations were immediately given scientific verification, and were extended by Taylor and Hulbert.[36] Rukop[37] has given an interesting survey of the early experiments, and has pointed out that the research departments of various commercial organizations were equipped with suitable short wave transmitters and receivers at an early date, but had never suspected the possibility of communication over great distances with such existing apparatus.

When the great economic importance of these long-neglected wave-lengths was realized, revised frequency allocations were soon made by international agreement, and the construction of new long wave stations was practically abandoned overnight. In the ensuing struggle between rival countries and rival commercial interests the amateurs quickly lost about 90 percent of their former territory, though a few channels were preserved through the friendly efforts of the American governmental representatives.

With these new experimental discoveries, and with the theory of electronic refraction offered by Larmor in 1924, the long period of random exploration came to a definite end, and a new field of atmospheric physics began to open up. In previous years physicists had given casual attention to results obtained as a by-product of radio communication. During the succeeding decade most of the advances were made by direct physical measurements, in which the radio apparatus served merely as an incidental research tool. Unfortunately this altered situation has received no recognition whatever in the monumental structure of American governmental regulations. Academic scientific institutions are badly handicapped by inflexible rules which became obsolete twelve years ago. As this situation constitutes a major obstacle, without parallel in other branches of experimental physics, it is necessary to examine the matter in greater detail in Section Q. Supervisory officials are aware of the present difficulties and have offered to cooperate, but the cumbersome machinery can be accelerated by a more general knowledge of the facts.

C. BASIC EXPERIMENTAL FACTS

Under this heading I shall attempt to present a simple analysis of the more prominent features which distinguish the various parts of the radio spectrum, reserving for later treatment the detailed examination of particular experiments

[34] A. H. Taylor, Proc. I. R. E. **13**, 677 (1925).
[35] J. L. Reinartz, Q. S. T. **9**, 9 (1925).
[36] A. H. Taylor and E. O. Hulburt, Q. S. T. **9**, 12 (1925).
[37] H. Rukop, Zeits. f. Hochfrequenztechnik **28**, 41 (1926).

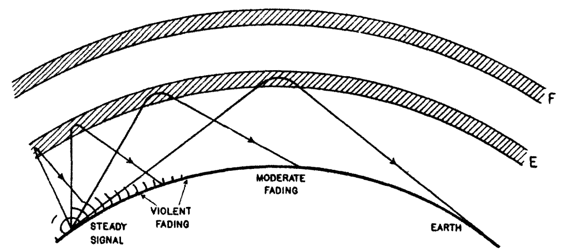


FIG. 2. Broadcast wave transmission at night.

designed to test some special aspect of the ionosphere theory. In sketching this general background the picture may be clarified by dividing the radio spectrum into definite sections which exhibit characteristic types of behavior. A simple quantitative classification of this sort may be useful for illustrative purposes even though it necessarily involves somewhat arbitrary allocations of boundaries between different regions of the spectrum. With this reservation concerning the numerical data we may adopt the following terminology:

		Wave-Length in Meters	Frequency in Kilocycles
"Long waves"		30,000–600	10–500
"Broadcast waves"		600–200	500–1500
"Short waves"	A	200–100	1500–3000
	B	100–50	3000–6000
	C	50–10	6000–30,000
"Quasioptical waves"		10–1	30,000–300,000
"Microwaves"		1–0.1	300,000–3,000,000

We shall also need to consider several different concentric regions or "layers" in the ionosphere which affect various parts of the radio spectrum in somewhat different fashions. Adopting Appleton's alphabetical notation, and neglecting all "fine structure" for the moment, we may distinguish three main regions which govern most of the observed effects:

F region—strongly ionized. . . . approximate height, 240 km
E region—moderately ionized . approximate height, 100 km
D region—weakly ionized . . . . approximate height, 50 km

After sunset the ionization slowly decreases in each region, and the weak *D* region appears to be relatively ineffective at night. The normal diurnal cycle of ionization and recombination is often modified by sudden and erratic increases in ionization which may occur in any region at any hour of day or night. Such changes are especially common in the *D* and *E* layers. The 11-year sunspot cycle apparently affects the average ionization in all parts of the atmosphere. The stated heights merely denote representative values chosen to illustrate the order of magnitude. A more detailed discussion of the measured "heights" will be given in Section K. Free electrons in the *F* region, set in motion by the radio waves, have a comparatively long mean free path, and lose little energy in collisional friction. In the *D* region, collisional friction is the controlling factor.

"Long waves" obey the Austin-Cohen formula rather well and the concentric-conducting-sphere type of transmission offers a reasonable and adequate explanation of their behavior. In the absence of conclusive experimental evidence, we may tentatively assume that the *D* region serves as the outer conductor in the case of long distance grazing incidence daylight transmission. In comparison with the wave-length it is probable that the lower boundary of the layer is fairly sharply defined. Consequently the longer waves do not penetrate the ionized layer appreciably and are not absorbed by the high attenuation which would accompany low frequency transmission in an ionized region. The layer acts like a simple metallic reflector, though considerable absorption doubtless results from the slight residual ionization in the troposphere and lower stratosphere. A moderate decrease in attenuation is noted when the transmission path lies on the dark side of the earth. Under such conditions it is natural to suppose that the upper boundary shifts to the *E* region.

Fig. 1 indicates continuous wave fronts, extending from the ionized layer to the earth. This distance is comparatively small when measured in wave-lengths, and there is no complete separation into a "ground wave" and a "sky wave." Since most of the energy is carried to the receiver by a single ray, the long waves are particularly suitable for "radio compass" applications, being comparatively free from the distortion commonly associated with plural

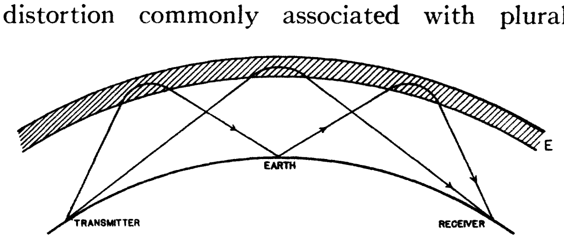


FIG. 3. Plural path sky wave transmission.

path transmission. The electric vector is approximately vertical, having a slight forward tilt which depends upon the amount of energy absorption in the soil or water traversed. At these low frequencies the ground attenuation is relatively small. The effect of ground conductivity has been discussed recently by Dean.[38] In the case of long wave transmission over sea water Yokoyama and Tanimura[39] find some evidence which suggests a zigzag ray, reflected successively by the ionosphere and by the sea.

The "long wave" region provides a limited number of high grade transoceanic telegraph channels, free from service interruptions except during severe magnetic storms or abnormally heavy "static" disturbances. The use of "long waves" for telephony is not common, as the wider frequency bands occupied by telephone channels would cause excessive crowding. Most of the telegraph channels are occupied by expensive high power stations which have been in steady use for many years. In planning *new* construction the relative stability and reliability of "long wave" transmission does not ordinarily outweigh the decreased cost of a comparable "short wave" circuit.

"Broadcast waves" provide reliable high quality service in a limited area in the neighborhood of the transmitting station. The dimensions and shape of the area depend upon the antenna design, wave-length, geological conditions, and sunspot cycle, but a circle of 50 mile radius may be mentioned as a representative example. Outside of this area there is a narrow zone characterized by violent fading, which is particularly evident at night. Beyond 150 miles stations of intermediate power are not ordinarily received during daylight hours, though thousands of miles are frequently covered during the night and transoceanic reception is not unusual. Nocturnal distant transmission is unreliable, however, and subject to moderate fading.

During the daytime nearly all of the energy which arrives at the "broadcast wave" receiver is carried by a "ground wave." The exact nature of this "ground wave" has been the subject of extensive examination. According to the

[38] S. W. Dean, Proc. I. R. E. **17**, 1440 (1929).
[39] E. Yokoyama and I. Tanimura, Proc. I. R. E. **21**, 263 (1933).

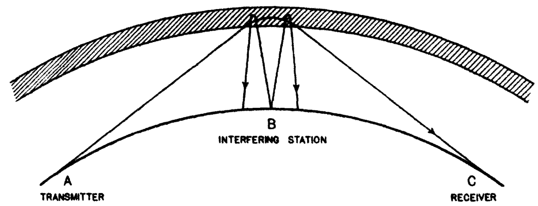


FIG. 4. "Interaction."

majority viewpoint the "ground wave" is a true guided wave, similar in nature to the high frequency waves which can be propagated along a single copper wire. On the basis of Zenneck's[4] diffraction equations, Sommerfeld[5, 40] developed a formula describing the transmission of a wave along the interface between the atmosphere and the semiconducting earth. Numerical and graphical evaluations of this formula have been given by Hoerschelmann,[41] Ratcliffe and Barnett,[42] Rolf,[43] Van der Pol,[44] Wise,[45] Eckersley,[46] Niessen,[47] and Nunier.[48] A separate treatment by Murray[49] was corrected by Niessen[50] and shown to be consistent with the Sommerfeld formula. Eckersley considers the formula valid for computing direct ray field strengths produced by waves from 60 meters to 2000 meters at distances up to 2000 miles.

Barfield[51] has applied the formula in determining geological differences, but the short cut which he uses has been criticised by Englund.[52] Noether[53] admits that the surface waves treated by Zenneck and Sommerfeld are theoretically possible, but contends that existing methods do not produce them. In this connection it is interesting to observe that, in dealing with seismic waves, Muskat[54] has recently found it necessary

[40] A. Sommerfeld, *Jahrbuch der drahtlosen Telegraphie* **4**, 157 (1911); Ann. d. Physik **81**, 1135 (1926).
[41] Hoerschelmann, *Jahrbuch der drahtlosen Telegraphie* **5**, 14 (1912); **5**, 188 (1912).
[42] J. A. Ratcliffe and M. Barnett, Proc. Camb. Phil. Soc. **23**, 288 (1926).
[43] B. Rolf, Proc. I. R. E. **18**, 391 (1930); Ingeniors Vetensk. Acad. No. 96 (1929).
[44] B. Van der Pol, Tijds. Nederland Radiogenootschap **4**, 105 (1930).
[45] W. H. Wise, Proc. I. R. E. **18**, 1971 (1930).
[46] T. L. Eckersley, Proc. I. R. E. **20**, 1555 (1932).
[47] B. Van der Pol and K. F. Niessen, Ann. d. Physik **6**, 273 (1930).
[48] W. Nunier, Ann. d. Physik **20**, 513 (1934).
[49] F. H. Murray, Proc. Camb. Phil. Soc. **28**, 433 (1932).
[50] K. F. Niessen, Ann. d. Physik **16**, 810 (1933).
[51] R. H. Barfield, J. Inst. Elec. Eng. **66**, 204 (1928).
[52] C. R. Englund, J. Inst. Elec. Eng. **67**, 931 (1929).
[53] F. Noether, E. N. T. **10**, 160 (1933).
[54] M. Muskat, Physics **4**, 14 (1933).

to introduce a modification in the theory of propagation of elastic waves along an interface between two homogeneous elastic media.

We could regard the guided wave hypothesis as proven experimentally were it not for the fact that a direct wave, propagated through the troposphere by the shortest route from the transmitting antenna, would be bent into valleys and around obstacles by ordinary diffraction, and might arrive at nearby receiving points without the help of ground conductivity. This effect should be computable, but simple knife-edge diffraction is not a sufficiently good approximation, and the complete problem apparently has not been solved in a satisfactory manner as yet. For our present purpose it is sufficient to describe the "ground wave" as a stable reliable wave which does not show appreciable diurnal variation. Its intensity decreases exponentially with distance. Its range increases with wave-length, and with the power available, and may be increased by designing an antenna which concentrates the emitted energy in low angle radiation.[55]

During the daytime practically all high angle radiation from "broadcasting" antennas is absorbed by the ionosphere. The frequency is high enough to permit the wave to penetrate into the *D* region where it is rapidly attenuated by collisional friction.

After nightfall the *D* ionization decreases, and the "broadcast waves" are strongly "reflected" from the *E* layer, as indicated in Fig. 2. The decreased attenuation results from the longer mean free path of the electrons at the higher altitude. (The refraction and polarization effects will be discussed later.) At points close to the transmitter the energy of the sky wave is much smaller than that of the "ground wave," but the contribution of the "sky wave" does not decrease rapidly with distance. As a consequence of the exponential decrease in the "ground wave" we soon arrive at a zone in which the two waves produce approximately equal fields. The slightest change in atmospheric conditions will cause violent fading as the phase relation of the two waves varies. The commonly observed periodic fading is an indication of a progressive change in the equivalent path of the "sky wave."

At distant points the "ground wave" is entirely ineffective, but the sky wave does appear at great distances under favorable conditions. Namba and Hiraga[56] have observed an improvement in propagation across the Pacific during the years near the sun spot minimum. Such transmission is better in autumn than in midwinter and is worst in midsummer. Berkner[57] has reported on the reception of numerous American broadcasting stations in the south polar district, at distances exceeding 12,000 km. Beyond the ground wave zone the signal is often comparatively steady, though not dependable. Slow alterations in field strength occur as a result of changes in absorption. On other occasions periodic fading results from the type of plural path sky wave transmission indicated in Fig. 3.

The ionized region is not a completely linear transmitting medium. As indicated in Fig. 4, transmission from point *A* to point *C* may be noticeably distorted by cross modulation from a powerful interfering station located near the halfway point on the great circle which connects the transmitter and receiver. These "interaction" effects will be discussed in Section *S*. The disturbance is unimportant except at the low frequency end of the broadcast band.

For convenience in describing the "short wave" region, I have divided it into three parts. The wave-lengths from 100 to 200 meters are notoriously erratic and unsatisfactory for distant communication. The ground wave is attenuated so rapidly that the local service area is too small for efficient broadcast utilization. The sky wave is also subject to excessive attenuation, and this part of the spectrum may be considered as an actual atmospheric "absorption band." Though other theories will be mentioned later, it is almost certain that these anomalous atmospheric effects are due to the resonance frequency of the free electrons set in motion by the wave. The electrons travel in circular orbits about the lines of force of the earth's magnetic field and have a natural frequency in this general range.

[55] H. E. Gihring and G. H. Brown, Proc. I. R. E. **23**, 311 (1935).

[56] S. Namba and D. Hiraga, Radio Research Japan, Report **2**, 9 (1932).

[57] L. V. Berkner, Proc. I. R. E. **20**, 1324 (1932).

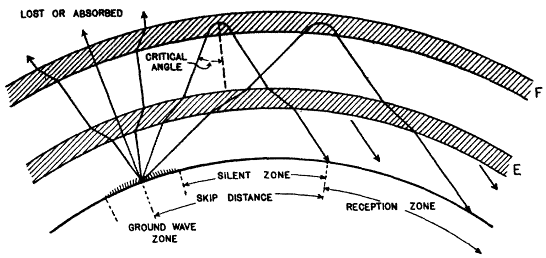


FIG. 5. "Skip distance."

The erratic properties of the 100–200 meter waves arise in part from an additional complication. This is a transition range of frequencies, and reflections may be expected from the E layer or the F layer or simultaneously from both layers. The shorter waves in this band are often able to penetrate the E region, but their frequencies are not high enough to prevent appreciable partial reflection, attenuation, polarization and reduction in group velocity.

In "short wave" group B, we may conveniently include the range extending from 50 to 100 meters. These waves are especially valuable for overland transmission within the confines of a single continent and they are widely used by aviation interests and by the military forces. Except at points within 30 miles of the transmitter, where the ground wave produces an appreciable effect, reception depends upon sky wave propagation. In general the dependable sky wave comes from the F layer, though strong E layer "reflections" are not infrequent. Even in the extreme case of signals received at nearby points after "reflection" from the ionized region at nearly normal incidence, E layer "reflections" often occur. These transient "reflections" frequently appear to be caused by relatively small dense ionic clouds, drifting at random in the E region. On other occasions there is a general increase in ionization which shifts the path from the F to the E region for a number of hours. As the shift often occurs very quickly, transitions of this sort frequently occur without severe fading or interruption of service. Though group B is sometimes used in transoceanic service the grazing incidence absorption is somewhat greater than in group C.

Group C, extending from 10 to 50 meters, is chiefly characterized by remarkable efficiency in long distance intercontinental transmission, and by the prominence of the skip distance effect, previously mentioned in the historical survey.

In order to understand this "skip distance" phenomenon let us first fix our attention on the nocturnal F layer "sky wave" or echo, which returns to earth at a point near the transmitter after "reflection" at nearly normal incidence. As the ionization density slowly decreases during the night the "sky wave" gradually penetrates to greater and greater heights in the ionized region. The corresponding time lag of the echo signal is gradually increased. A similar effect may be produced artificially by increasing the transmitter frequency while the ionization density remains substantially constant. In either case a critical condition may be attained at which the slow increase of "effective height" (measured as time lag of echo) is broken by a sudden upward surge. At the same time the strength of the "reflected" signal decreases rapidly, and it suddenly vanishes completely.

The sudden increase of time lag is readily accounted for as evidence of a corresponding decrease of group velocity in the wave, as it penetrates into and passes through a considerable thickness of ionized F layer. Similar effects are frequently produced by the E region, when the ionization density is near a critical value which will just suffice to produce "reflection" at the given frequency. Such phenomena are predicted by the mathematical theory, and it is natural to expect an associated decrease in signal strength due to attenuation.

By analogy with the behavior observed under similar circumstances in the E region, it is possible to assume that the final complete disappearance of the signal corresponds to complete penetration of the F layer and resultant loss of the signal in interstellar space. This is the customary assumption and it has been regarded as a self-evident fact by many writers. However, Eckersley favors the alternative idea of complete absorption within the F layer. The matter will be referred to again in Section K.

According to either hypothesis a signal which has been lost in this manner can be restored by increasing the angle of incidence beyond a definite critical value. Consequently we should expect the ray pattern represented in Fig. 5.

At points within a short distance (say 15 miles) of the transmitter a "ground wave" signal appears, but this is completely ineffective at distant points on account of excessive attenuation. At points beyond this small "ground wave" zone no dependable signal is observed until we reach a sharply marked boundary curve where the usual downcoming sky wave appears. The main reception zone lies beyond this boundary.

By careful examination these "skip-distance" phenomena may be detected at night in the 50–100 meter region, but in general such skip distances are short and the effect is largely masked by ground wave propagation. In the 10–50 meter band, however, the skip distances may be measured in thousands of miles. Here the effect becomes the controlling factor in determining the best frequency for use on a definite long distance circuit at a particular hour, season, and year.

The "silent zone" is not completely dead. Strong signals produce a peculiar reverberant echo, easily recognized by an experienced ear in voice or code transmission. These "scattered" signals are not yet completely explained, but the matter will be referred to in greater detail in Section R.

As the frequency of the wave is increased the skip zone boundary moves outward until the "silent zone" finally embraces the entire earth, and transmission by means of ionosphere "reflection" is no longer possible. Under average conditions this limitation is reached in the vicinity of 10 meters. Consequently, wave-lengths somewhat longer than 10 meters are of great value in daylight transmission over very long paths, while wave-lengths slightly shorter than the critical value are quite useless for this purpose. However, the ionization density of the *F* layer varies considerably from day to day. Departures from the average value not infrequently permit freak transmission on wave-lengths as short as 8.5 meters and it is quite certain that 5 meter signals can be received at great distances when the ionization is exceptionally high.

In general, however, absolutely no dependence can be placed on ionosphere transmission or guided ground wave transmission in the "quasi-optical" range, extending from 1 to 10 meters.

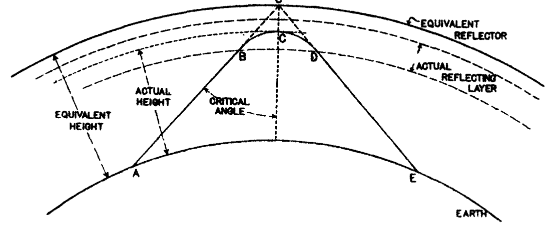


FIG. 6. Relation between actual path and "equivalent" path of ray.

Though apparently narrow when measured in terms of wave-length, this classification evidently covers an enormous range of frequencies, and is therefore particularly attractive for future television applications. These waves are also especially applicable to two-way communication with police patrol cars, since a short vertical rod, easily carried on a moving car, serves as a decidedly efficient quarter-wave transmitting antenna. The dependable service area is practically limited to the territory included within the optical horizon, and these waves are chiefly useful for communication within a single metropolitan district.

Unfortunately, though, the interference area of quasioptical transmission is much greater in extent than the dependable service area. Even with comparatively low power used in present experimental installations, very strong 5 meter signals are often received far below the optical horizon at points more than 100 miles from the transmitter. This effect is definitely associated with weather conditions in the lower troposphere and it appears to be a simple mirage phenomenon caused by the temperature inversions which frequently exist in the atmosphere a few thousand feet above the ground level. In fact it appears to be possible to use this type of transmission as a new meteorological tool in studying the distribution and movement of air masses. Preliminary studies, conducted by the American Radio Relay League in cooperation with Harvard University, have indicated a remarkably close correlation between quasioptical transmission over the Boston-Hartford path and the corresponding meteorological temperature gradients ("lapse rates") measured at the Boston Airport and at Mitchell Field, near New York City. By means of simultaneous field experiments on wave-lengths of 1.25, 2.50 and 5.00 meters we are now investigating the possible effect of hilltop diffraction on quasioptical transmission. On account of excessive attenuation the simple guided wave, following the surface of the ground, appears to be quite ineffective at such high frequencies. W. W. Mumford considers that the large dipole moment of water vapor molecules may contribute greatly to the bending of quasioptical waves. Water vapor gradients usually accompany temperature inversions.

Presumably the "microwaves," extending from 0.1 to 1.0 meter, will exhibit characteristics resembling the "quasioptical" waves. Research in this region requires the construction of special types of transmitting tubes. The power outputs and efficiencies, thus far obtained, are very low, and little information is yet available on transmission. However, a successful two-way 17 cm channel has been established across the English Channel and other experiments are in progress.

D. Elementary Theory

The detailed study of the action of the ionosphere requires an intricate mathematical formulation, but the basic facts may be explained in a very simple way.

Consider the action of a single electron placed between the plates of a condenser. If the frequency of the alternating voltage applied to the condenser plates is low, and if the motion of the electron is restrained by elastic forces, the displacement of the electron is substantially in phase with the applied voltage, and the vibrating charge constitutes an alternating current, in phase with the Maxwell displacement current through the otherwise empty space. In other words, the negative electron approaches the positive plate of the condenser at the peak of the positive half-cycle. It therefore neutralizes a portion of the charge on the condenser plates and permits the available voltage to drive a larger charging current through the external circuit. Under such circumstances we are accustomed to say that the presence of the bound electron has *increased* the dielectric constant of the region permeated by the electric field.

On the other hand, if the elastic restraining force acting on the electron is zero, or if the frequency is so high that the inertial force predominates, the phase relations ***reverse***, and the vibrating charge may ***reduce*** the effective dielectric constant of the medium below unity. We are familiar with this effect in optics, since it provides the conventional elementary "resonance" explanation of the anomalous dispersion observed on the high frequency side of an absorption line in the optical spectrum. If present, collisional friction introduces a resistance term which decreases the magnitude of the change in dielectric constant but cannot alter the sign.

Precisely the same facts determine the basic conditions governing the propagation of a radio wave in the ionosphere. Due to the action of free electrons, set in motion by the electromagnetic field of the wave, the dielectric constant and the corresponding index of refraction of the ionized region are less than unity. Consequently the layer is a medium which is "less dense" in the optical sense than the nonionized strata of air beneath. If the boundary between the media is reasonably sharp (in comparison with the wave-length of the electromagnetic radiation) total internal reflection takes place at the boundary, and rays which strike the layer at angles exceeding the critical angle will be strongly reflected back to the earth.

If the boundary between the two media is ***not*** sharp the incident ray is gradually *refracted* farther and farther away from the normal, describing a curve which depends upon the rate of increase of the free-electron density. If the total increase of electron density suffices, the ray will eventually attain a horizontal direction and will then follow a symmetrical downward path which brings it back to the earth. If the maximum density is insufficient the ray will penetrate the layer. For a layer of given maximum density (and a wave of given frequency), there is evidently a definite critical angle of original incidence, which determines whether or not a given ray will return to earth. The study of this type of refraction may be regarded as a detailed examination of the mechanism which is responsible for the "reflection" of an electromagnetic wave from a conducting surface. When the boundary of the conducting "layer" is not sharply defined, it is sometimes advantageous to consider the actual curved path of the ray. For many purposes, however, it is sufficient to replace the curved path $ABCDE$ by the angular "equivalent path," $ABC'DE$, represented in Fig. 6, and to speak of the "equivalent height" of a fictitious "reflector." We may use the term "actual height" to refer to the height at which maximum free-electron density occurs. In the special case of a ray which strikes the layer at the "critical angle," this height coincides with the maximum height of the curved ray. All other "totally reflected" rays will fall beneath this height.

Breit and Tuve[58] have called attention to the fact that the time required for a definite radio signal to traverse the route $ABCDE$, through the actual ionized medium, is exactly equal to the time which would be required for the signal to follow the route $ABC'DE$ in empty space. Consequently the "equivalent height," defined geometrically by Fig. 6, is properly measured by direct observations on the time lag of echoes, without correction for the reduction in group velocity which delays the signal in the ionized layer. Equivalent heights measured in this fashion are therefore consistent with equivalent heights determined geometrically by a measurement of the angle of arrival of the downcoming ray.

Consequently the "actual height" of a layer and its ionization gradient cannot be measured directly by a single observation of any sort. These quantities must be inferred indirectly from a series of observations of "equivalent height" using different frequencies or different paths or both. There is reason to believe that the lower boundary of the E layer is relatively sharply defined and that the measured "equivalent height" exceeds the "actual height" by only a few percent. The F region is more diffuse and the "actual heights" are not well known as yet. Preliminary data suggest that the "equivalent heights" may exceed the "actual heights" by at least 25 percent. When not more explicitly defined, statements in the literature in regard to layer "height" nearly always refer to the "equivalent height" determined by direct experimental observation.

[58] G. Breit and M. A. Tuve, Phys. Rev. **28**, 554 (1926).

Disregarding, for the moment, modifications produced by the earth's magnetic field and by collisional friction, we may readily deduce a simple formula for the index of refraction of an ionized medium. A more elaborate analysis will be given in Section H. Pedersen[59] has suggested the following simplified treatment, based on elementary considerations, which yields the familiar equation of Eccles[24] and Larmor.[13]

Consider a condenser with plates of unit area, separated in vacuum by a distance of 1 cm. When provided with suitable guard rings, the unit condenser has a capacity of $1/4\pi$ absolute units.

Now introduce N electrons of charge e and mass m in the cubic centimeter included between condenser plates and apply an alternating voltage of instantaneous value v, of angular frequency ω, and amplitude V. The electrons will be set in motion by the electric field, and their velocity u will lag a quarter period behind the applied voltage and will have a maximum value

$$U=(e/\omega m)V.$$

But N charges of magnitude e, moving with the common velocity u, are equivalent to a current element of 1 cm length and with the strength

$$i_e=Ne\cdot u.$$

The amplitude of this lagging alternating current is therefore

$$I_e=NeU=\frac{Ne^2}{\omega m}V=\frac{V}{\omega(m/Ne^2)}$$

and the vibrating electrons produce the same effect as an equivalent inductance of m/Ne^2 absolute units, shunted across the condenser.

But so far as the total current in the external circuit is concerned this combination is equivalent to a condenser with the capacity

$$C'=1/4\pi-(Ne^2/\omega^2 m)$$

and therefore to the unit condenser with the dielectric constant

$$\epsilon=1-4\pi Ne^2/\omega^2 m.$$

[59] P. O. Pedersen, Wireless Engineer **7**, 16 (1930).

In this equation ω and N may be regarded as independent variables, though only ω is under experimental control.

It is evident that the theory also predicts an additional reduction of dielectric constant due to the presence of heavier ions of either sign, and we could readily introduce additional terms of similar algebraic form to take account of this effect. However, as the mass of the ion occurs in the denominator of the subtractive term, a single electron is more effective than a vast number of heavy ions. Nevertheless, some physicists believe that hydrogen ions may produce a detectable effect in the ionosphere. This possibility will be referred to in a later section. It is also evident that the increase in dielectric constant which occurs during the night may result mainly from neutralization of electrons by positive ions, or mainly from mere attachment of electrons to heavier neutral atoms or molecules.

Since N may become very large it is evidently possible for the dielectric constant to be reduced to zero and even to be reversed in sign. Such action commonly occurs in the ionosphere in the frequency ranges used in practical communication. It is therefore important to examine the effect on the index of refraction of this reversal of the sign of the dielectric constant.

Pedersen points out that the usual approximate expression, $n=\sqrt{\epsilon}$, is valid only for positive values of ϵ. This is an important restriction as a number of writers have assumed that a negative dielectric constant necessarily implies an imaginary refractive index. In a nonmagnetic conducting medium the correct relation is

$$n=[\epsilon/2+\{\epsilon^2/4+(2\pi c^2\sigma/\omega)^2\}^{\frac{1}{2}}]^{\frac{1}{2}}.$$

Where the conductivity may be neglected this reduces to

$$n=[\epsilon/2+(\epsilon^2/4)^{\frac{1}{2}}]^{\frac{1}{2}}$$

which gives $n=\sqrt{\epsilon}$ for $\epsilon\geqq 0$,
$n=0$ for $e\leqq 0$.

We may also note that the phase velocity, V_p, of the wave in the ionosphere is given by the expression

$$V_p=c/(1-4\pi Ne^2/\omega^2 m)^{\frac{1}{2}},$$

where c is the velocity of light.

The velocity V_p (which exceeds the velocity of light) is not a directly measurable quantity. The speed of travel of the observed radio signal is measured by the group velocity, V_g, where

$$V_g=V_p\Big/\left(1-\frac{\omega}{V_p}\cdot\frac{dV_p}{d\omega}\right)$$
$$=c(1-4\pi Ne^2/\omega^2 m)^{\frac{1}{2}}.$$

The signal is therefore retarded by the presence of free ions in the medium.

When applied to the ionosphere these considerations have direct application to the most prominent experimental facts encountered in ordinary communication, though we must later consider the modifications caused by the magnetic field of the earth and by collisional friction.

As the index of refraction of an ionized layer approaches zero the critical angle also approaches zero. In the limit we might expect that total internal reflection would occur even at normal incidence. We may express this prediction in other language by saying that the "skip zone" should shrink to zero for a sufficiently high electron density in the layer or a sufficiently low frequency at the transmitter. These expectations are amply verified by experiment.

It is therefore customary to say that a radio signal, directed vertically upward, penetrates into a region of increasing free-electron density and travels with a slower and slower group velocity, until it finally reaches an altitude where the electron density is just sufficient to reduce the index of refraction to zero. At this point the direction of the ray is reversed by total internal reflection and the signal is propagated downward at an increasing speed, attaining the velocity of light as it emerges from the ionized layer. If the concentration of free electrons at the most dense portion of the layer is not quite sufficient to reduce the index of refraction to zero, the vertical ray passes completely through the region and the signal emerges above the layer with the velocity of light. At some greater height it may then encounter a denser layer which will return it toward the earth, and it passes through the lower region a second time on its way to the receiving antenna. When the maximum electron density in the lower layer is only slightly less than the critical value, the signal is considerably retarded by this round trip through the lower stratum and the measured "equivalent height" of the upper layer may be abnormally increased by at least 100 percent. Such effects are readily recognized and interpreted, however, when they occur in a series of experimental observations.

Though this elementary explanation of vertical incidence "reflection" in a diffuse medium is strongly suggestive and does agree well with the experimental observation, it cannot be regarded as a rigorous and complete theory. We have no apparent right to generalize the simple laws of total internal reflection and apply them so confidently in a region which has continuously variable optical properties. Various attempts to resolve this difficulty will be referred to in Section G, but the complete theoretical problem is a difficult one which apparently has not yet been solved in a manner which is thoroughly satisfactory.

E. The Nature of the Accelerating Field

Though the elementary theory, presented in Section D does correlate a number of interesting experimental facts, we shall need to invoke a more elaborate mathematical treatment in order to explain the double-refraction effects which are frequently observed. As a preliminary step in considering the more advanced theory we shall first examine several fundamental postulates in regard to the forces which determine the motion of the electrons.

Fabry[60] has called attention to the fact that free electrons are quite effective in scattering radio waves, since phase addition is to be expected at these comparatively low frequencies. In the case of free electrons set in motion by a light wave, only the intensities add. We should therefore expect little resultant modification of the light beam, though it is possible that slight astrophysical effects may be detected.

However, when we attempt to compute scattering of radio waves resulting from electron motion we immediately become involved in a serious controversy in regard to the nature of the accelerating field. In the ionosphere, in addition to the free electrons, positive ions and neutral molecules are present in large numbers. In the elementary treatment presented above, and in the theoretical papers which first appeared, the simple space-average value of the electric field was employed, without critical examination, in computing the motion of the individual electrons. When investigating the alteration of dielectric consant due to the presence of ordinary neutral atoms between condenser plates, we know that this "pipe force" must be corrected by the Lorentz "polarization" term which takes account of the actual corpuscular distribution of matter. We often visualize the individual atom or molecule as an object enclosed in an approximately spherical cavity, influenced by the field due to induced charges on the walls. This physical picture may be used in order to compute the Lorentz "polarization" term, though several other methods of computation lead to the same mathematical expression. The resulting mathematical formula has been tested by experiments upon the dielectric constant of ordinary materials, and good quantitative agreement has been obtained.

Consequently, Hartree[61–63] included this Lorentz correction term in several mathematical papers on the scattering of radio waves in stratified media, and emphasized the fact that Goldstein[64] and previous writers had omitted it. The corrected expression has been employed by Appleton[65–67] in several papers, and the resulting "Appleton-Hartree formula," for the dielectric constants of a doubly-refracting medium, has been applied extensively in numerical calculations by Taylor[68] and numerous others. The numerical correction is a large one, which alters by fifty percent the value of the electron density computed from measurements of penetration frequency. The actual equations will be given later.

[60] C. Fabry, Comptes rendus **187**, 777 (1928).

[61] D. R. Hartree, Proc. Camb. Phil. Soc. **25**, 97 (1929).
[62] D. R. Hartree, Proc. Camb. Phil. Soc. **27**, 143 (1931)
[63] D. R. Hartree, Proc. Roy. Soc. **A131**, 428 (1931).
[64] S. Goldstein, Proc. Roy. Soc. **A121**, 260 (1928).
[65] E. V. Appleton, J. Inst. Elec. Eng. **71**, 642 (1932).
[66] E. V. Appleton and R. Naismith, Proc. Roy. Soc. **137**, 36 (1932).
[67] E. V. Appleton and G. B. Builder, Proc. Phys. Soc. **45**, 208 (1933).
[68] Mary Taylor, Proc. Phys. Soc. **46**, 408 (1934).

In 1933, Tonks[69, 70] questioned the validity of Hartree's derivation and contended that the polarization term should be omitted except where "there is some detailed arrangement of the negative with respect to the positive charges." In this objection he was supported by Norton[71] who maintained that the space-average field is applicable, since the electron moves over a distance which is large, with respect to the electron spacing, in a time which is short, in comparison with the period of the wave.

Though the papers by Tonks and Norton started a lively theoretical argument there was no immediate agreement in regard to the merit of the various viewpoints which had been presented. Hartree[72] pointed out two unjustified assumptions in the arguments offered by Tonks, and one such assumption in his own previous derivations. The important question at stake remained undecided. In self-defence several experimental physicists formed the habit of publishing two sets of numerical computations, allowing the reader to take his own choice.

In 1934 Darwin published a preliminary paper,[73] indicating that the Lorentz term should *not* be included. In December, 1934, he presented an extensive and thorough reexamination[74] of the entire problem which appears to settle the matter decisively. In his recent paper Darwin notes that the problem is a perfectly definite one and it should be possible to solve it completely without resort to experiment. The main work on the subject was done more than fifty years ago. A purely classical treatment is sufficient, though quantum mechanics methods may be employed without altering the result. Though superficially simple, the problem is exceedingly treacherous. Since the average field is very small in comparison with the fields which exist at local points in the medium the computed value may be greatly affected by a slight alteration in the method used for evaluating the average. Darwin first considers the various methods which have been employed previously. By means of arguments "quite as convincing as many of those always accepted in theoretical physics" it is possible to support either of two entirely contradictory formulae. Though the reexamination of these older viewpoints does not provide a clear-cut answer to the question, the general weight of evidence favors the omission of the Lorentz term when computing the force which acts on a free electron in the presence of positive ions or neutral atoms and molecules. Since atoms cannot interpenetrate, an individual *atom* is affected only by the *external* fields of other atoms. This restriction does not apply to the free *electron* and the simple space average should be used in determining the force.

However, Darwin obtains a more decisive formulation of the problem by an entirely new method which avoids the troublesome ambiguities involved in the close analysis of internal electric fields in matter, and he supplies conditions for discriminating between substances requiring the two types of formula. By the application of Hamiltonian dynamics it is possible to formulate the electric moment of a small volume of the medium, the size of the region being such that the retardation of the waves is negligible. The conditions for the omission of the polarization term are fulfilled in the ionosphere and in metals. When applied to the determination of the properties of metals, the theory is supported by quantitative experimental evidence, and it appears to afford a reliable basis for the quantitative computation of the properties of the ionosphere.

F. Other Forces Acting on the Electron

Having discussed the nature of the electric field produced by the incident wave, we may next consider several additional factors which may affect the motion of the scattering electrons. The frictional force due to collisions and the deflecting force produced by the earth's magnetic field will be treated mathematically in Section G. Perturbing forces caused by interfering signals will be described in Section S. In addition to these effects we must examine a hypothetical "quasielastic" or "relaxation" force which has come into the literature as a result of laboratory experiments designed to test the properties of ionized gases.

---

[69] L. Tonks, Nature **132**, 101 (1933).
[70] L. Tonks, Nature **132**, 710 (1933).
[71] K. A. Norton, Nature **132**, 676 (1933).
[72] D. R. Hartree, Nature **132**, 929 (1933).
[73] C. G. Darwin, Nature **133**, 62 (1934).
[74] C. G. Darwin, Proc. Roy. Soc. **146**, 17 (1934).

In 1913, Salpeter[75, 26] developed equations for the refraction of electric waves in an ionized gas. Though similar to the treatment given by Eccles[24] in the previous year, Salpeter's derivation included a more detailed consideration of collisional friction. In 1920, Van der Pol[27] performed an experiment on an ion plasma between the plates of a condenser and attempted to verify the theory of Eccles and Salpeter by direct test. He sought to measure a change in condenser capacity resulting from the formation of ions, and for this purpose he employed the Lecher wire technique suggested in 1897 by Drude.[76] For certain adjustments of the apparatus, Van der Pol obtained a qualitative indication of a reduction in the dielectric constant.

In 1927, H. Gutton and G. Clément[77–79] repeated this experiment with slightly different experimental technique and tried to obtain improved quantitative data. As predicted by the theory, they did obtain a decrease in the dielectric constant when ions were formed. However, when they increased the ionization beyond a certain critical value, the dielectric constant apparently suddenly *increased*, in contradiction to the expected behavior. This critical value of the ion density could be reduced by performing the experiment with a resonant circuit of lower natural frequency.

Gutton and Clément attempted to explain this anomalous behavior by assuming that the vibrating ions were subjected to a "quasielastic" restoring force produced by the mutual action of the ions. The ion plasma would therefore have a natural period of oscillation which would increase with the density of the ionized gas. By extrapolating similar laboratory measurements C. Gutton[80] sought to show that such gaseous resonance would occur in the ionosphere at a frequency in the neighborhood of 1500 kc and he considered this to be the explanation of the poor transmission obtained with wavelengths of the order of 200 meters. For frequencies above resonance the dielectric constant would be reduced and refraction could occur. For lower frequencies transmission could be ascribed to conductivity only.

Though this ionosphere extension of the theory was never widely accepted by other investigators it was generally admitted that Gutton and Clément had undoubtedly made accurate observations of the alterations in the apparent capacity of their laboratory circuit, and many persons believed that "quasielastic" forces might account for some of the ultra-high-frequency oscillations which occur in gaseous electron tubes. H. Gutton[81] believed that such forces produced the "plasmoidal oscillations," treated theoretically and experimentally by Tonks and Langmuir[82] though he concluded that the theory was not precise enough to provide quantitative verification.

In 1927 and 1928, Pedersen[83] and Rybner[84] claimed that the apparent anomaly was entirely due to the apparatus which Gutton and Clément had used. In 1929, Bergmann and Düring[85] repeated Van der Pol's original Lecher wire experiment but substituted electrons liberated from a hot cathode in place of the ionized gas used by previous investigators. Consequently the mean free path of the electrons was large. Under these conditions they obtained complete quantitative agreement with the simple theory within the limits of accuracy of the measurements. Using a magnetron tube, Benner[86] likewise verified the simple magneto-ionic theory regarding resonant changes in dielectric constant and decrement of an ionized region.

In a separate paper Benner[87] offered an improved formula for the Bergmann-Düring experiment, which better suited the special experimental conditions involved. In his original derivation Salpeter had assumed oscillations interrupted by collisions, whereas Bergmann and Düring had made use of a mean free path much larger than the amplitude of oscillation.

[75] J. Salpeter, *Jahrbuch der drahtlosen Telegraphie* **8**, 247 (1914).
[76] P. Drude, Ann. d. Physik **61**, 466 (1897).
[77] H. Gutton and J. Clément, Comptes rendus **184**, 441 (1927).
[78] H. Gutton and J. Clément, Comptes rendus **184**, 676 (1927).
[79] H. Gutton and J. Clément, Onde Élec. **6**, 137 (1927).
[80] C. Gutton, Ann. de physique **14**, 5 (1930).
[81] H. Gutton, Ann. de physique **13**, 98 (1930).
[82] L. Tonks and I. Langmuir, Phys. Rev. **33**, 195 (1929).
[83] P. O. Pedersen, *The Propagation of Radio Waves* (Copenhagen, 1927).
[84] J. Rybner, Onde Élec. **7**, 428 (1928).
[85] L. Bergmann and W. Düring, Ann. d. Physik **1**, 1041 (1929).
[86] S. Benner, Naturwiss. **17**, 120 (1929).
[87] S. Benner, Ann. d. Physik **3**, 993 (1929).

In a general survey article on the internal action of thermionic systems, published in 1931, Benham[88] criticized this revised formula, but Benner[89] promptly pointed out that the equation had been misused by Benham.

In 1932, Niessen[90] published a theoretical paper indicating that, in the ionosphere, the attenuating effect of a "relaxation force" could be neglected in comparison with collisional friction. "Relaxation resonance" would be impossible. *If* present, however, the "relaxation force" would always cause absorption and might also exert an apparent binding action on the free electrons.

In 1932, after completing experiments begun by Appleton and Childs,[91] Appleton and Chapman[92] announced a thorough experimental reexamination of the Gutton-Clément investigation. A supplementary report of this test was presented by Appleton[93] at the 1934 Congress of the U. R. S. I. After duplicating the apparatus used by Gutton and Clément, Appleton and Chapman checked the original experimental observations but denied all of the conclusions. There were no "quasielastic" forces. All "quasiresonance" effects were due to the fact that the dielectric constant assumes negative values. The explanations previously offered by Pedersen and Rybner were verified quantitatively. The condenser containing the ionized gas is not a simple structure with a uniform dielectric. In part of the space permeated by the electric field the dielectric constant is positive. The stray capacities of the coil and wiring must also be considered. When describing the actual distributed circuit by an equivalent simple circuit with lumped constants, it is useful to consider *two* condensers in series with the coil. In *one* of these condensers the dielectric constant can assume negative values, but the *other* condenser contains a normal positive dielectric which is not affected by the ionization. As the ionization density increases the *resultant* series capacity assumes a large negative value and then suddenly changes to a large positive value. Though this critical condition is purely a property of the electric circuit itself, the effect produces a striking imitation of a resonant change in the properties of the ionized gas.

[88] W. E. Benham, Phil. Mag. **11**, 457 (1931).
[89] S. Benner, Phil. Mag. **11**, 1252 (1931).
[90] K. F. Niessen, Physik Zeits. **33**, 705 (1932).
[91] E. V. Appleton and E. C. Childs, Phil. Mag. **10**, 969 (1930).
[92] E. V. Appleton and F. W. Chapman, Proc. Phys. Soc. **44**, 246 (1932).
[93] E. V. Appleton, U. R. S. I., London Congress (1934).

Though Appleton considers the Gutton "quasiresonance" entirely spurious, he does verify certain experimental observations made by Tonks[94] which indicate the presence of a *second* "resonance" point, obtained with low values of tube current. Appleton considers that Tonks and Langmuir are not justified in describing this effect as "plasma-resonance," however. He prefers to regard it as "sheath-resonance," which takes place at the *boundary* of the plasma, since "it can be shown that an electron vibrating from the main discharge into a sheath will be subjected to a restoring force and thus have a natural vibration frequency." Though important in the theory of discharge tubes, sheath-resonance may evidently be disregarded in the ionosphere.

In France, the concept of "quasielastic" forces still receives strong support at the Laboratoire Nationale de Radioélectricité. C. Gutton,[95] director of the laboratory, recently presented an interesting paper summarizing the French experiments, but apparently he has not refuted Appleton's quantitative data as yet.

If we therefore reject the idea of "quasiresonance" in a plasma, we may still conduct a search for real resonance frequencies due to the internal vibrations of large molecules. In 1934, Ziegler[96] summarized the theories and measurements of dielectric constant and concluded that there is no satisfactory evidence of any such effect in the ultra-high-frequency radio spectrum extending from 100 cm down to 10 cm. At a wave-length of 1.1 cm however, Cleeton and Williams[97] have demonstrated an absorption effect due to the "turning-inside-out frequency" of the ammonia molecule. This type of research will undoubtedly be extended but molecular resonance may be disregarded so far as ionosphere refraction is concerned.

[94] L. Tonks, Phys. Rev. **37**, 1458 (1931).
[95] C. Gutton, U. R. S. I., London Congress (1934).
[96] W. Ziegler, Physik Zeits. **35**, 476 (1934).
[97] C. E. Cleeton and N. H. Williams, Phys. Rev. **45**, 234 (1934).

G. The Equations of Motion of the Electrons

Theoretical investigations, verified by direct experimental tests, indicate that the magnetic field of the earth is responsible for the double-refraction effects frequently encountered in commercial radio transmission and in ionosphere research. The experiments are precise enough to verify the known quantitative facts in regard to the variation of magnetic field intensity with latitude, and to provide means for determining the variation with height within the earth's atmosphere.

In the special cases of transmission along the magnetic field and transmission perpendicular to the magnetic field this effect was studied in 1926 by Appleton and Barnett[98] and independently by Nichols and Schelleng.[99] Below 70–80 km electrons collide so frequently with gas molecules that the magnetic field is relatively ineffective. Above this critical height the collision frequency is lower than the rotational frequency about the earth's field. Consequently the inductivity in the direction of the field is appreciably different from the inductivity at right angles to the earth's field, and the plane of polarization is rotated. The Faraday effect and the Kerr effect are produced according to the wave-length. So long as the investigation was limited to the two special cases mentioned above it was possible to make use of the equations previously employed by Voigt,[100] Drude,[101] and Lorentz[102] in the study of the propagation of optical waves in the field of a magnet.

Appleton[103] and a number of others soon extended the equations of magneto-ionic double refraction in order to include the general case of propagation at an arbitrary angle with respect to the direction of H. The appearance of the resulting expressions for the refractive indices depends greatly on the selection of the directions of coordinate axes and the nature of the abbreviations used in reducing the cumbersome expressions to a compact and useful form. Although the analysis is already available in the scientific literature of several different countries, the importance of the effect justifies the repetition of the derivation in outline form at this point. As Appleton's[104] adaptation of the Lorentz equations has been widely used in numerical computation, I shall adopt his notation and choice of axes, but shall eliminate the Lorentz correction term for the reasons given in Section E. The reader may also wish to study the derivations, presented in somewhat different form, by Försterling and Lassen,[105] Baker and Green,[106] and Pierce.[107]

Considering plane wave propagation along the x axis, and choosing our coordinate system so that there is no component of the earth's magnetic field along y, we denote the x and z components of this field by the symbols H_L and H_T, respectively.

For Maxwell's field equations (expressed in Lorentz units), we may write:

$$\begin{cases} \text{Curl } H = \dfrac{1}{c}\left(\dfrac{\partial E}{\partial t} + j\right), & (1) \\[2ex] \text{Curl } E = -\dfrac{1}{c}\dfrac{\partial H}{\partial t}, & (2) \end{cases}$$

where E and H are the electric force and magnetic force of the plane wave.

j is the convection current density. We therefore have:

$$j = dP/dt, \tag{3}$$

where P is the polarization at any point.

When expanded, Maxwell's equations become:

$$\partial E_x/\partial t + j_x = 0, \tag{4}$$

$$-\partial H_z/\partial x = (1/c)(\partial E_y/\partial t + j_y), \tag{5}$$

$$\partial H_y/\partial x = (1/c)(\partial E_z/\partial t + j_z), \tag{6}$$

$$\partial H_x/\partial t = 0, \tag{7}$$

$$-\partial E_z/\partial x = -(1/c)(\partial H_y/\partial t), \tag{8}$$

$$\partial E_y/\partial x = -(1/c)(\partial H_z/\partial t). \tag{9}$$

[98] E. V. Appleton and M. A. F. Barnett, Nature **115**, 333 (1925); Proc. Roy. Soc. **109**, 621 (1925).
[99] H. W. Nichols and J. C. Schelleng, Bell Sys. Tech. J. **4**, 215 (1925).
[100] W. Voigt, *Magneto und Elektro-optik* (Leipsig, 1908).
[101] P. Drude, *Lehrbuch der Optik* (Leipsig, 1900).
[102] H. A. Lorentz, *Theory of Electrons* (Leipsig, 1909).
[103] E. V. Appleton, U. R. S. I., Washington Assembly (1927).
[104] E. V. Appleton, J. Inst. Elec. Eng., wireless section, **7**, 257 (1932).
[105] K. Försterling and H. Lassen, Ann. d. Physik **18**, 26 (1933).
[106] W. G. Baker and A. L. Green, Radio Research Board Report No. 3 (Melbourne, 1932).
[107] G. W. Pierce, Cruft Laboratory lectures.

Eq. (4) indicates that the waves are not entirely transverse since there will ordinarily be a longitudinal component of electrical force. Eq. (7) indicates that there is no longitudinal component of H.

The remaining four equations determine four wave equations. Assuming that the field vectors contain the factor $e^{ip(t-qx)}$, these wave equations simplify to:

$$\left.\begin{aligned}(c^2q^2-1)E_y-P_y&=0,\\(c^2q^2-1)E_z-P_z&=0,\\(c^2q^2-1)H_z-cqP_y&=0,\\(c^2q^2-1)H_y+cqP_z&=0,\end{aligned}\right\}\qquad(10)$$

whence

$$E_y/E_z=P_y/P_z \quad\text{and}\quad H_y/H_z=-E_z/E_y. \qquad (11)$$

Consequently the resultant components of magnetic force and electric force are at right angles to each other in the wave front.

If we now study the motion of an electron by considering its displacement components, ξ, η, ζ, along the x, y, and z axes, we may write the equations of motion:

$$m\frac{d^2\xi}{dt^2}=-eE_x-g\frac{d\xi}{dt}\frac{eH_T}{c}\frac{d\eta}{dt},$$

$$m\frac{d^2\eta}{dt^2}=-eE_y-g\frac{d\eta}{dt}-\frac{eH_L}{c}\frac{d\zeta}{dt}+\frac{eH_T}{c}\frac{d\xi}{dt},$$

$$m\frac{d^2\zeta}{dt^2}=-eE_z-g\frac{d\zeta}{dt}+\frac{eH_L}{c}\frac{d\xi}{dt},$$

or in terms of the electric polarization:

$$\frac{m}{Ne}\frac{d^2P_x}{dt^2}=-eE_x-\frac{g}{Ne}\frac{dP_x}{dt}-\frac{H_T}{Nc}\frac{dP_y}{dt},$$

$$\frac{m}{Ne}\frac{d^2P_y}{dt^2}=-eE_y-\frac{g}{Ne}\frac{dP_y}{dt}-\frac{H_L}{Nc}\frac{dP_z}{dt}+\frac{H_T}{Nc}\frac{dP_x}{dt},$$

$$\frac{m}{Ne}\frac{d^2P_z}{dt^2}=-eE_z-g\frac{dP_z}{dt}+\frac{H_L}{Nc}\frac{dP_z}{dt}.$$

Since $P_x=Ne\xi$; $P_y=Ne\eta$; $P_z=Ne\zeta$; where N is the number of electrons per cm^3 and e is the magnitude of the negative charge of the electron, and g is a friction factor.

Assuming that all dependent variables contain time only in the factor ϵ^{ipt}, these equations become:

$$E_x=(\alpha+i\beta)P_x+i\gamma_TP_y, \qquad (12)$$
$$E_y=(\alpha+i\beta)P_y-i\gamma_TP_x+i\gamma_LP_z, \qquad (13)$$
$$E_z=(\alpha+i\beta)P_z-i\gamma_LP_y, \qquad (14)$$

where

$$\alpha=-mp^2/Ne^2, \qquad (15)$$
$$\beta=pg/Ne^2, \qquad (16)$$
$$\gamma_L=\frac{mpeH_L/mc}{Ne^2}, \qquad (17)$$
$$\gamma_T=\frac{mpeH_T/mc}{Ne^2}. \qquad (18)$$

In these expressions e, H_L, and H_T are expressed in Lorentz units. If e and H are to be expressed in electrostatic units (for convenience in computing numerical values for the coefficients) each expression should be divided by 4π.

Lorentz suggested the rough approximation, $g=2m\nu$, where m would be the mass of the electron in this case, and ν would be the average number of collisions per second which the electron makes with the air molecules. The expression has been accepted without comment by several writers, though its validity is doubtful.[102] For the present it seems better to leave g in the expression as an empirical constant.

From (4) and (12) we obtain:

$$P_x=-i\gamma_TP_y/(1+\alpha+i\beta). \qquad (19)$$

From (19) and (13) we have:

$$E_y=(\alpha+i\beta)P_y-(\gamma_T^2/(1+\alpha+i\beta))P_y+i\gamma_LP_z. \qquad (20)$$

Combining (20) with (14), and eliminating the E_y, E_z, P_y, and P_z:

$$\gamma_L^2=\left\{(\alpha+i\beta)-\frac{1}{c^2q^2-1}\right\}\times\left\{(\alpha+i\beta)-\frac{\gamma_T^2}{1+\alpha+i\beta}-\frac{1}{c^2q^2-1}\right\}, \qquad (21)$$

whence

$$c^2q^2=1+\frac{2}{2(\alpha+i\beta)-\gamma_T^2/(1+\alpha+i\beta)\pm[\gamma_T^4/(1+\alpha+i\beta)^2+4\gamma_L^2]^{\frac{1}{2}}}. \qquad (22)$$

For convenience in analyzing this expression let us write:

$$p_0^2=4\pi Ne^2/m,\quad p_L=eH_L/mc,\quad p_T=eH_T/mc,$$

where e, H_L, and H_T will now be expressed in electrostatic units. Whence

$$\alpha=-p^2/p_0^2,\qquad \beta=pg/p_0^2m,$$
$$\gamma_L=pp_L/p_0^2,\qquad \gamma_T=pp_T/p_0^2.$$

Substituting, we obtain:

$$c^2q^2=(\mu-ick/p)^2$$
$$=1-\frac{2p_0^2}{2(p^2-ipg/m)-(p^2p_T^2/(p^2-p_0^2-ipg/m))\mp[p^4p_T^4/(p^2-p_0^2-ipg/m)^2+4p^2p_L^2]^{\frac{1}{2}}}. \quad (23)$$

In general, therefore, the ray treatment of the problem indicates that we are to expect two different indices of refraction, and two corresponding modes of propagation through the medium. To find the polarization belonging to either of these modes of propagation we may use the auxiliary equation:

$$\frac{H_y}{H_z}=\frac{i\gamma_L}{1/(c^2q^2-1)-(\alpha+i\beta)} \quad (24)$$

obtained by substituting (10) in (14).

If we had included the Lorentz correction term, we would have obtained the following equation in place of Eq. (23):

$$c^2q^2=(\mu-ick/p)^2$$
$$=1-\frac{2p_0^2}{2(p^2+\frac{1}{3}p_0^2-ipg/m)-p^2p_T^2/(p^2-\frac{2}{3}p_0^2-ipg/m)\mp[p^4p_T^4/(p^2-\frac{2}{3}p_0^2-ipg/m)^2+4p^2p_L^2]^{\frac{1}{2}}}. \quad (23')$$

Disregarding the slight additional modification produced by evaluating the friction factor, g, in terms of the collision rate, ν, we may note that Eq. (23′) is essentially the Appleton-Hartree formula, which has been used extensively in numerical computations by Taylor[108] and a number of others. For the reasons mentioned in Section E, however, we now prefer Eq. (23) which is equivalent to Appleton's[104] simpler formula.

Equations, similar to those derived above, have been widely used in ionosphere theory and have been extensively supported by experiments. We will consider their significance in further detail in several of the following sections. However, it would be incorrect to assume that the magneto-ionic theory of the ionosphere has received universal acceptance or that it represents the only possible attack on the problem. Before beginning an analysis of the conventional theory, I will therefore insert several references representing viewpoints not covered by the derivations above.

A rival theory offered by Eckersley[109, 110] has received a considerable amount of attention. Eckersley contends that magneto-optical effects are only of major importance at night and perhaps on extremely long waves in the daytime. He considers that daylight transmission is primarily controlled by absorption phenomena which in turn depend upon whether the frequency of the radio wave is higher or lower than the average collision rate of the electrons. At an early date Mesny[111] contrasted Eckersley's hypothesis with the conventional theory and offered interesting criticisms of the theories of Lassen, Nagoaka, Pedersen, Breit, and Chapman, and the experiments of Pickard.

The validity of the ray method of electric wave analysis has been questioned by a number of writers. Eckersley[112–114] wishes to substitute an approximate phase integral treatment, which presents a formal analogy with quantum theory and uses the methods of Bohr and Sommerfeld. He considers the ray method inadequate for long waves. Hartree[63] finds that the conditions for use of the ray method are not completely satisfied, but the correction introduced by the

[108] Mary Taylor, Proc. Phys. Soc. **45**, 245 (1933).
[109] T. L. Eckersley, J. Inst. Elec. Eng. **71**, 405 (1932).
[110] T. L. Eckersley, Phil. Mag. **9**, 225 (1930).
[111] R. Mesny, Onde Elec. **7**, 130 (1928).
[112] T. L. Eckersley, Proc. Roy. Soc. **A132**, 83 (1931).
[113] T. L. Eckersley, Proc. Roy. Soc. **A136**, 499 (1932).
[114] T. L. Eckersley, Proc. Roy. Soc. **A137**, 158 (1932).

wave method is unimportant. Epstein[115, 116] justifies the use of geometrical optics since there is no appreciable reflection unless the conditions approximate those of *total* reflection. He considers that the refracted ray is properly traced by the ray method and this refracted ray carries the main part of the energy.

Finally, we may refer very briefly to a theory based on "subelectrons," offered by Bagchi[117] and to a theory involving diffusion by little cloudlets of electrons, advanced by Ponte and Rocard.[118] Apparently neither of these theories has attracted much attention in the general literature.

H. Analysis of Magneto-Ionic Double Refraction

The elementary theory, presented in Section *D*, indicates that we may expect that a radio wave, directed vertically upward, will experience total internal reflection when it arrives at a level where the electron density is just sufficient to reduce the refractive index to zero. We developed a simple expression relating the refractive index to the electron density and to the angular frequency of the wave.

In Section G we undertook a more thorough investigation of the problem, taking account of the modifications produced by terrestrial magnetism. As a result we obtained a more elaborate formula (Eq. (23)) for the refractive index, involving the longitudinal and tangential components of the earth's magnetic field in addition to the quantities previously mentioned. Our improved equation indicates that there are *two* different numerical values of the complex refractive index corresponding to a given electron density, wave frequency, and direction of propagation. We find, therefore, that the ionosphere is a doubly refractive medium, and we are to expect two modes of propagation of radio waves in the ionized regions.

In general we find that a radio wave, emitted as a single ray from a sending antenna, will be resolved into an ordinary and an extraordinary ray during its passage through the ionized layers. These rays will follow different paths, will penetrate to different levels, and will return to earth at different times. If both waves happen to return to the same point at approximately the same time, they recombine to produce a resultant signal at the receiving antenna. Under such conditions the resultant wave frequently undergoes rapid change in intensity and polarization as a result of minor alterations in the path of either ray. Under other conditions one of the component rays may be lost by absorption or by penetration, or it may be delayed so long that we may clearly recognize it as a separate echo when it finally returns.

Until very recently, practically all direct measurements of the properties of the ionosphere have been carried on by means of transmitting and receiving stations placed relatively near together, so that the waves have been "reflected" at normal incidence by the layers. Two main types of experiments are often carried on under such conditions. In "variable frequency" experiments the angular frequency of the emitted wave is altered over a wide range while the corresponding echo sequence, obtained at the receiving point, is recorded automatically or noted by an observer who watches a cathode-ray oscillograph screen. In this way it is frequently possible to obtain a *pair* of "critical frequencies." *One* of these "critical frequencies" is the highest frequency at which the *extraordinary* ray will be "reflected" by the electron concentration prevailing in a given layer at the time of the experiment. The *other* "critical frequency" is the highest frequency which will permit the return of the *ordinary* ray under the same conditions. Evidently each "critical frequency," measured accurately by direct experiment, is just high enough to reduce to zero the refractive index for the corresponding mode of propagation, at the point in the layer where the electrons are most dense. For higher frequencies the corresponding refractive indices do not decrease to zero at any point in the layer, and one or both of the component rays will penetrate to higher layers or will escape to interstellar space. If we may assume the validity of the analysis presented in Section G, it should be possible to compute the electron density and the strength of the earth's field from

---

[115] P. S. Epstein, Proc. Nat. Acad. Sci. **16**, 37 (1930).
[116] P. S. Epstein, Proc. Nat. Acad. Sci. **16**, 627 (1930).
[117] S. C. Bagchi, Nature **134**, 701 (1934).
[118] Ponte and Rocard, Comptes rendus **187**, 942 (1928).

our measurements of critical frequency. As the complete experiment occupies but a short time, the electron density in the layer remains sensibly constant during the test.

"Constant frequency" experiments, on the other hand, are usually continuous experiments designed to study the normal and abnormal variations in electron density. In studying the F region a transmitter frequency is selected which will normally show no vertical incidence reflections from that layer during the early morning hours (say 3492.5 kc). During the sunrise period the electron density increases as a result of the absorption of ultraviolet light. Eventually the electron density becomes large enough to reduce to zero the refractive index for the extraordinary ray (in the numerical case just mentioned). The exact time of the beginning of this type of echo may be recorded automatically to the nearest minute. Some thirty minutes later the further increase of electron density suddenly permits the "reflection" of the ordinary ray. Similar effects occur in reverse order as the electron density slowly decreases during the late evening hours. Considerable variations occur from day to day but by the use of automatic apparatus it is possible to secure sufficient statistical data for the study of ionization rates and decay rates. Continuous measurements of this sort are also extremely useful in accumulating precise data on the effect on radio propagation of magnetic storms, sunspots, auroral displays, meteors, weather conditions, and other geophysical phenomena. The two types of experiments are complementary.

In interpreting the data of both types of experiments it is evident that we shall be interested in studying the variation of refractive index with angular frequency and with electron density, and will wish to know the type of polarization corresponding to each of the two refractive indices. We shall be especially interested in the conditions which just suffice to reduce the refractive indices to zero. The equations derived in Section G will therefore be analyzed from this point of view. Such analyses have been given by Goldstein,[64] Ratcliffe,[119] Taylor,[108] and others. The results obtained by the various writers are equivalent except for the quantitative differences which result from the inclusion or omission of the Lorentz correction term previously discussed. Ratcliffe's treatment is particularly good, as he gives special attention to the progressive changes which accompany alterations in the relative directions of the wave normal and the earth's magnetic field. The following outline is suggested by his article, though the notation is altered to some extent and the discussion is necessarily greatly abbreviated.

As differences in terminology have caused some misunderstanding[120, 121] in regard to the actual polarization of downcoming radio waves, it is necessary for us to examine the reference axes and sign conventions with considerable care.

As in Section G, let us consider a radio wave propagated in the direction of the positive x axis (Fig. 7), with its own magnetic field entirely in the yz plane. The magnetic field of the earth is represented by a vector in the xz plane, with a longitudinal component, H_L, along the positive x axis, and a tangential component, H_T, along the positive z axis. In the ionized medium the wave may be resolved into two parts represented by the two ellipses in the yz plane. The dotted ellipse represents a ray with "right-hand" polarization since its magnetic component, H_y, attains its maximum positive value 90 electrical degrees in advance of the component H_z. The magnetic vector of this ray therefore rotates in a clockwise sense when viewed in the direction of propagation. As the opposite conditions obtain in the case of the solid ellipse, this represents a "left-hand" polarization.

In the present section we will neglect collisional friction and drop all terms containing the friction factor, g. With this simplification we may rewrite Eq. (23) in the form:

$$c^2q^2=(\mu-jck/\omega)^2=1-\frac{2(\omega_0^2/\omega^2)(1-\omega_0^2/\omega^2)}{2(1-\omega_0^2/\omega^2)-\omega_T^2/\omega^2\mp[\omega_T^4/\omega^4+4\omega_L^2/\omega^2(1-\omega_0^2/\omega^2)^2]^{\frac{1}{2}}},$$

[119] J. A. Ratcliffe, Wireless Engineer **10**, 354 (1933).
[120] T. L. Eckersley, Nature **130**, 398 (1932).
[121] T. L. Eckersley, Nature **130**, 472 (1932).

where $\omega_0{}^2 = 4\pi Ne^2/m,$
$\omega_L = H_L e/mc,$
$\omega_T = H_T e/mc.$

We will also use:

$\omega_1 = He/mc,$

$\left.\begin{array}{l} Y = \lambda/\lambda_1, \\ X = \omega_0{}^2/\omega^2, \end{array}\right\}$ Note that X and Y are ratios, not identified with the space variables, x and y.

$\lambda_1 = 2\pi c/\omega_1,$

θ = orientation of H, measured from the direction of propagation.

At a point where the earth's magnetic field is approximately 0.5 e.m.u., the numerical value of λ_1 is approximately 200 meters. The corresponding frequency is a resonance frequency produced by the rotation of the free electrons about the magnetic lines of force. By analogy with optical problems it is natural to expect a considerable difference in the properties of the medium above and below this critical region. The resonance is not sharply marked since the strength of the magnetic field varies appreciably with altitude and since there will be frictional damping in the practical case. Let us first consider the special case, $\lambda = 100$ meters, since this will illustrate the typical conditions encountered in the transmission of high frequency radio waves. The corresponding curves of μ^2 *vs.* X have been computed and plotted by Ratcliffe (Fig. 8) for three different orientations of the direction of propagation with respect to the magnetic field of the earth.

Notice that the abscissa, X, is directly proportional to the ionization density N. The shaded areas enclose a family of curves which might be drawn for any of the intermediate orientations included between the longitudinal and transverse directions of propagation. We may arbitrarily select the 45 degree curve as a typical representative of this group. By using the lower sign in Eq. (23) we obtain the continuous curve which passes through the (0, 1) point. Though the shape of this curve does depend upon the strength of the magnetic field, we may note that H has *no* effect upon the location of the point where μ^2 becomes zero. Consequently the name "ordinary ray" is assigned to the corresponding mode of propagation. By the use of the upper

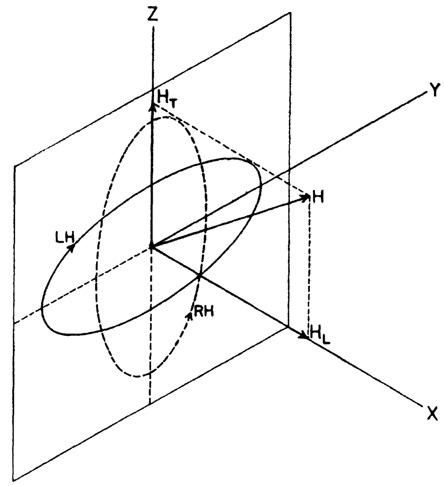


FIG. 7. Polarization of radio waves.*

sign in Eq. (23) we obtain the curve for the "extraordinary ray." This curve has one infinite point, and crosses the horizontal axis at the two points, $X = 1 - Y$ and $X = 1 + Y$. The location of these zero points evidently depends upon the magnitude of H but not upon its direction.

By examining Eq. (23), it is possible to show that the ordinary ray has left-hand polarization ($H_z/H_y < 0$), for values of $X < 1$; plane polarization for $X = 1$; and right-hand polarization for $X > 1$. The extraordinary ray has right-hand polarization for $X < 1$; plane polarization for $X = 1$; and left-hand polarization for $X > 1$. (Note that these statements are based on the assumption that H has a positive component in the direction of propagation, as in the case of downcoming waves in the Northern Hemisphere. The polarizations are reversed in the Southern Hemisphere. Both waves are plane polarized at right angles in the limiting case of completely transverse propagation.)

Let us now apply these results to the practical problems of short wave radio transmission and experimental ionosphere research. A radio wave, transmitted vertically upward, is divided into an extraordinary and an ordinary component when it reaches the ionosphere. These two components continue to move upward through a region of increasing electronic density. Eventually they may reach a level where there are enough electrons to reduce to zero the refractive index of

* Fig. 1, Wireless Engineer **10**, 355 (1933).

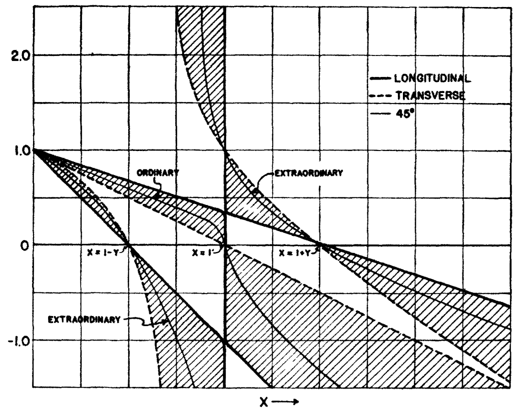


FIG. 8. Variation of refractive index with ionic density ($Y=\frac{1}{2}$).†

the extraordinary ray ($X=1-Y$). This ray then experiences total internal reflection, but the ordinary ray must rise to a greater height in search of a region of appreciably greater density, such that $X=1$. If the layer is sharply defined the difference in path may be too small to measure by ordinary means, and the reflected signals overlap closely. During the sunrise period and the late evening period, however, the difference in path is frequently large and the two echoes are easily resolved. It is not unusual for an 86 meter vertical incidence extraordinary ray reflection to persist throughout the night when no trace of the ordinary ray can be found.

The second zero point of the extraordinary ray, at $X=1+Y$, probably cannot be detected experimentally, since practically all of the energy of the ray has been lost in the reflection at $X=1-Y$, and in collisional friction. Theoretically, even in the case of "total" internal reflection, a small amount of energy does penetrate into the outer region, but this becomes inappreciable in a distance of the order of a wave-length. Consequently the zero points at $X=1-Y$ and at $X=1$ are the points which may be correlated most directly with experiment.

In the case of nonvertical incidence the results are qualitatively similar. The extraordinary and ordinary rays rise to different heights, as their refractive indices must become small, though they need not be reduced to zero. The electron density may be sufficient to return one or both components to earth at distant points while insufficient to produce reflections of either type at small angles of incidence.

Let us next consider the behavior of waves which are appreciably longer than 200 meters. Fig. 9 represents Ratcliffe's study of the propagation of 400 meter waves ($Y=2$). Boundary curves have again been drawn for the limiting cases of transverse and longitudinal propagation, with a representative intermediate curve for $\theta=30°$, and shaded areas to show the transition region.

It is apparent that the ordinary and extraordinary rays have exchanged roles, the extraordinary ray now producing the echo which travels over the longer path and is reflected from the more dense region. Although the extraordinary ray evidently reverses its polarization as it passes through the level where $X=1$, the extraordinary ray again has right-hand polarization in the lower levels of the earth's atmosphere where it may be examined experimentally. The ordinary ray retains its left-hand polarization. (Both statements apply to downcoming waves in the Northern Hemisphere.)

In conclusion we may refer briefly to a few representative experiments which illustrate the methods used in the study of ionospheric double refraction. The construction of a typical polarimeter receiver for studying the nature of downcoming waves is described by Ratcliffe and White.[122, 123] Martyn and Green[124, 125] have used a three-aerial system of measurement for determining the polarization of downcoming waves in the Southern Hemisphere. Their results differ from the results obtained by similar measurements in the Northern Hemisphere and the differences are correctly predicted by the usual magneto-ionic theory.

Berkner and Wells[126, 127] have compared critical frequencies measured in Washington, D. C., and Huancayo, Peru, and conclude that the results are substantially in agreement with the known

† Fig. 9, Wireless Engineer **10**, 359 (1933).

[122] J. A. Ratcliffe and E. L. C. White, Phil. Mag. **16**, 125 (1933).
[123] J. A. Ratcliffe and E. L. C. White, Phil. Mag. **16**, 423 (1933).
[124] D. F. Martyn and A. L. Green, Proc. Roy. Soc. **148**, 104 (1935).
[125] A. L. Green, Proc. I. R. E. **22**, 324 (1934).
[126] L. V. Berkner and H. W. Wells, Proc. I. R. E. **22**, 680 (1934).
[127] L. V. Berkner and H. W. Wells, Proc. I. R. E. **22**, 1102 (1934).

difference in magnetic intensity. Appleton[128] and Chapman[129] consider that it is now possible to make an accurate determination of magnetic intensity as a function of altitude by means of the precise measurement of critical frequencies. Preliminary results are in accord with theory.

Though the high frequency echo component of lesser delay is *normally* the extraordinary ray, Appleton and Builder[67] have found that group retardation in a lower layer can cause a reversal of position when the experimental frequency is just above the penetration frequency of the low layer. This reversal is caused by the fact that the ordinary ray has the higher group velocity.

By employing exceptionally good experimental technique in oscillograph measurements on discrete pulses, von Handel and Plendl[130] made a thorough test of the distortion of a radio signal (selective side-band interference) produced by double refraction in the ionosphere.

Green and Builder[131] have explained and interpreted observations of Hollingsworth, Naismith, and Namba on the rotation of the plane of polarization of long radio waves.

I. Collisional Friction

The effect of collisional friction has been considered at some length by a number of writers, but no general agreement has been reached as yet in support of any single definite and quantitative mathematical treatment. I shall therefore limit this discussion to a brief qualitative survey with references to some of the typical exploratory computations undertaken by various research groups. The problem may be attacked from a number of different angles. We may be interested in the nature of the molecular and atomic collisions, or in the frictional loss which might limit the useful transmitting range, or in the selective absorption which alters the polarization of a "reflected" wave.

Hulburt[132, 133] has contributed a number of interesting papers on the absorption of radio waves in the upper atmosphere. By treating collisions between electrons and molecules he develops a simple attenuation formula and considers that the measured values may give data on electronic and molecular densities at high altitudes.

Yokoyama and Nakai[134] find that the observed east-west attenuation is decidedly greater than the north-south attenuation in the case of long wave transmission during daylight hours in fairly high latitudes. However, the agreement between experimental measurements and the various theories which have been offered is none too good.

G. Kreutzer[135] examines the absorption of a finite wave train in a dielectric and finds that the absorption constant depends on depth of penetration into the medium, on the frequency, and on the length of the train. The absorption is smaller than that given by classical optics theory for frequencies near resonance. The theory has been tested by experiments on ethyl alcohol.

By assuming that the directed momentum of the electron is destroyed at each impact, and that the velocity acquired between collisions is small compared with the random velocity of thermal agitation, Childs[136, 137] computes the theoretical conductivity of a gas and finds that it is of the same order of magnitude as the ob-

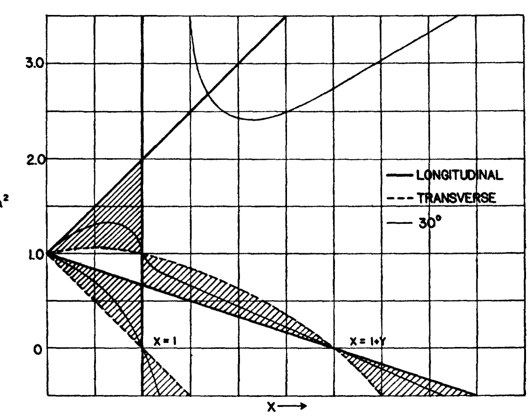


FIG. 9. Variation of refractive index with ionic density ($Y=2$).‡

[128] E. V. Appleton, Nature **133**, 793 (1934).
[129] S. Chapman, Nature **133**, 908 (1934).
[130] P. von Handel and H. Plendl, E. N. T. **10**, 76 (1933).
[131] A. L. Green and G. Builder, Proc. Roy. Soc. **145**, 145 (1934).
[132] E. O. Hulburt, Phys. Rev. **29**, 365 (1927).
[133] E. O. Hulburt, Phys. Rev. **29**, 706 (1927).
[134] E. Yokoyama and T. Nakai, Proc. I. R. E. **17**, 1240 (1929).
[135] G. Kreutzer, Zeits. f. Physik **60**, 825 (1930).
[136] E. C. Childs, Proc. Phys. Soc. **44**, 246 (1932).
[137] E. C. Childs, Phil. Mag. **13**, 873 (1932).
‡ Fig. 10, Wireless Engineer **10**, 360 (1933).

served conductivity. He concludes that the kinetic theory is valid for ionosphere computations.

Using the Appleton-Hartree formula in a form similar to that derived in Section G (but with explicit use of the Lorentz polarization term and the Lorentz evaluation of the friction factor in terms of collision frequency), Taylor[68] computes dispersion curves for four radiofrequencies and four collision frequencies. The insertion of the frictional term removes the infinities from the dispersion curves presented in Section H, but does not affect the general shape of these curves or the conclusions derived from them. However, an important additional piece of information now appears. (Similar but less complete discussions have been presented previously by other writers.) At broadcast frequencies and in the Northern Hemisphere the extraordinary ray is greatly weakened by attenuation and absorption. Though the ordinary ray penetrates deeper into the ionosphere, and experiences a longer delay, it eventually returns with a greater amplitude than the extraordinary ray. In the case of downcoming broadcast waves in the Northern Hemisphere we therefore find that left-hand polarization normally predominates. However, if the electron density drops through a critical value and the weak extraordinary ray alone remains, the polarization changes suddenly to the right-hand type. These changes have been observed experimentally by Appleton and Builder,[67] White[138] and others, and this portion of the collisional friction theory appears to be in entire accord with experiment.

From our present viewpoint it is somewhat unfortunate that the Lorentz polarization term was used in Taylor's elaborate and arduous numerical computations. Though most of her results are undoubtedly qualitatively correct, the importance of the subject would seem to justify a repetition of the quantitative work in accordance with Darwin's reexamination of the problem. In addition to the polarization effects mentioned above, Taylor finds a transition from quasitransverse to quasilongitudinal transmission at a critical collision frequency. By considering indices of attenuation in addition to indices of refraction, she also concludes that the lower boundary of the lower layer must be sharp in the optical sense.

In *short-distance* observations on long wave phenomena, Naismith[139, 140] finds that the strongest downcoming wave is obtained in the north-south direction, and he therefore believes that there will be correspondingly *less* energy available for *long distance* transmission in this direction. He also finds that magnetic storms *increase* the long wave field at short distances but *decrease* it at great distances.

J. Complete Analysis by Conformal Representation

Bailey[141] and Martyn[142] have recently developed an interesting graphical method for obtaining quantitative solutions of the Appleton-Hartree formula (or similar equations for the complex refractive index of the ionosphere), taking full account of the damping produced by collisional friction as well as the effect of the earth's magnetic field.

Though this new attack on the problem deserves special mention in an individual section of this report, it does not seem necessary to introduce any extensive description of the specific geometrical processes involved. The new procedure is a straightforward application of the methods of conformal mapping,[143] and it is fully described and illustrated in two recent articles[141, 142] which are available in practically all scientific libraries. "The two charts needed for the determination respectively of the polarization and the refractive index are easily drawn, as they involve only circles, ellipses, and parabolas."

Martyn[142] has carried out the graphical analysis for "five typical wave-lengths (100 m to 20,000 m), three collisional frequencies (10^4, 10^5, 10^6, per sec.) likely to cover the range of practical importance for radio propagation, and for directions making the three angles 0°, 40°, and 90° with the magnetic field." In general his results

[138] F. W. G. White, Proc. Phys. Soc. **46**, 91 (1934).

[139] R. Naismith, J. Inst. Elec. Eng. **69**, 875 (1931).
[140] R. Naismith, Wireless Engineer **8**, 254 (1931).
[141] J. A. Bailey, Phil. Mag. **18**, 516 (1934).
[142] D. F. Martyn, Phil. Mag. **19**, 376 (1935).
[143] J. H. Jeans, *Electricity and Magnetism* (Cambridge, 1908).

verify and supplement the conclusions presedent in Sections H and I of this report. In the detailed analysis of the limiting values of polarization obtained with low electron densities, Martyn's results differ essentially from those of Taylor[68] though they are in accordance with the conclusions of Baker and Green.[106] As Taylor and Martyn have studied the same fundamental equation (the Appleton-Hartree formula, including the Lorentz polarization term), the difference is evidently due to a minor analytical discrepancy which can undoubtedly be removed. As Martyn refers, in a footnote, to the debatable character of the Lorentz polarization term, we may hope that he will soon discuss the quantitative modifications which would result from its omission.

K. "Fine Structure" of the Ionosphere

When reviewing the basic experimental facts (Section C), we noted the existence of two main "reflecting" regions, commonly designated by the symbols E and F. By an analysis of the effect of the earth's magnetic field (Section H) we were led to expect two modes of propagation in each region. The echo pattern obtained at the receiver is further complicated by the common occurrence of "multiple" reflections, representing waves which have been "reflected" repeatedly from one or more of the layers and from the earth. (In our own measurements we have observed as many as 18 round trips of this character at times of low attenuation.)

In addition to these well-known effects, however, a number of observers have noted the presence of echoes which can only be accounted for by assuming that the F layer commonly divides into two parts during the daytime; and by assuming that additional "reflections" may sometimes be produced by minor concentrations of electrons above, between, or below the main E and F regions. Thus echo analysis becomes almost as complicated as spectrum analysis and special symbols are obviously needed in order to describe the "fine structure" of the echo pattern. As the best measurements have been made by a relatively small number of experimental groups, the effect of geographic differences is not completely understood, and the literature in regard to "fine structure" is still somewhat contradictory. Further confusion arises from the various terminologies and symbols employed in different countries. Nevertheless, references to a few representative experimental reports in regard to stratification may be of interest. Wherever applicable I shall use the symbols adopted by international agreement at the London convention of the U. R. S. I. in 1934, though other notations are often used in current periodicals.

As early as 1916, while considering the propagation of radio waves in the atmosphere, Lowenstein[144] stated, "Measurements of the intensity of light taken at sunset show three distinct discontinuities, when the last rays become tangent to the layers of air at the height of 11 km, 75 km, and 220 km." As yet we cannot know whether or not Lowenstein's 75 km and 220 km "discontinuities" are closely related to the E and F ionic regions, but additional observations of this character might prove to be valuable.

Using his original experimental methods (studying natural "fading," and producing artificial interference changes at the receiving point by slowly varying the frequency of the transmitter), Appleton and his co-workers[145–148] noted the existence of two "reflecting" regions, and assigned to them the symbols E and F. Apparently these letters were chosen in order to provide designations which could later be extended to lower or higher "layers." Appleton also found evidence of an absorbing layer below the E layer.

During the same period, Breit, Tuve and Dahl[58, 149] developed the "pulse" method of observation (transmission of a very short radio signal and measurement of the time lag of the echoes), and obtained layer "heights" ranging from 85 to 220 km.

In honor of Appleton's numerous contributions to ionosphere research, other English writers

[144] F. Lowenstein, Proc. I. R. E. **4**, 271 (1916).
[145] E. V. Appleton, Tijds. Nederland Radiogenootschap 2, 115 (1925).
[146] E. V. Appleton and M. Barnett, Proc. Roy. Soc. **113**, 450 (1926).
[147] E. V. Appleton, Nature **120**, 330 (1927).
[148] E. V. Appleton and A. L. Green, Proc. Roy. Soc. **128** 159 (1930).
[149] G. Breit, M. A. Tuve, and O. Dahl, Proc. I. R. E. **16**, 1236 (1929).

have often referred to the F layer as the "Appleton layer," reserving the name "Kennelly-Heaviside layer" for the E layer, which presumably played a greater part in the transmission of the long waves used in Marconi's first transatlantic experiments. (Kennelly originally postulated more than one reflecting level, while Heaviside referred to a single layer.) In general, however, the term "Kennelly-Heaviside region" has been used as a synonym for the word "ionosphere," since the use of proper names as designations of the individual parts of the ionosphere seems likely to produce additional confusion.

After the existence of two main layers was apparently well established, Goubau and Zenneck[150] published a report ascribing all reflections to a single low layer. Their paper was summarized in English by Howe[151] and it was soon followed by contributions from Eckersley,[152] Schafer and Goodall,[153] Gilliland, Kenrick, and Norton,[154] pointing out the overwhelming experimental evidence indicative of at least two layers.

Let us now consider some of the evidence in regard to the minor layers previously mentioned. In order to interpret transmission data taken in China in 1927–28, N. H. Edes[155] suggested the existence of a low lying region of small ionization. He estimated the height of this layer as 10 km. Recent experiments made by Colwell[278] tend to verify the existence of such a "C layer."

In 1927 and 1928, Appleton,[147] Heising,[156] and Goldstein[64] found indications of an absorbing region somewhat below the E layer. This (partly hypothetical) region is now referred to as the "ozone layer" or "D layer." In general its existence is inferred from indirect evidence, though Appleton,[157] Goubau[158] and several others have reported weak reflections. Estimates of its height range from 30 km to 65 km. Goldstein believes that it occurs at the Lindemann-Dobson temperature inversion height. Lugeon[159, 160] reports a 50 km layer over France. Additional observations have been reported by Bontch-Bruewitch[161] and Sillitoe,[162] but Kirby and Judson[163] maintain "that if the absence of refractions above 5000 kilocycles during summer midday is due to absorption, the absorption is not mainly below the E layer."

In 1933, Shafer and Goodall[164] reported an "intermediate" layer at 150 km and tentatively assigned to it the letter "M." Their observation was soon verified by Appleton[165] and by Ratcliffe and White.[166, 167] This use of the letter "M" appears somewhat unfortunate, since it breaks the established alphabetical sequence, and since the same letter has been used in England as a graphic description of an entirely different type of ray path, which presumably follows the zigzag route illustrated in Fig. 10.

If my interpretation of the recent U. R. S. I. recommendation is correct, the main E layer should now be designated as E_1 while the "intermediate" layer is to be referred to as E_2.

However, there are additional complications in the E region. In addition to the normal daytime ionization (presumably caused by simple ultraviolet absorption), practically all observers have noted the frequent occurrence of "abnormal," "sporadic," or "nocturnal" reflections from approximately the same height. To distinguish this random, intermittent reflection from the regular daytime effect, Ratcliffe and White[167] suggest the use of the lower case letter, "e," as a description of the sporadic phenomenon. In the present review I shall use the designation "abnormal E" to describe this nocturnal reflection.

The critical frequency which is just high enough to penetrate the E_1 layer is denoted by

[150] G. Goubau and J. Zenneck, Zeits. f. Hochfrequenztechnik **37**, 207 (1931).
[151] G. W. O. Howe, Wireless Engineer **8**, 463 (1931).
[152] T. L. Eckersley, Marconi Review **31**, 1 (1931).
[153] J. P. Schafer and W. M. Goodall, Proc. I. R. E. **19**, 1434 (1931).
[154] T. R. Gilliland, G. W. Kenrick, and K. A. Norton, Proc. I. R. E. **20**, 286 (1932).
[155] N. H. Edes and J. C. Coe, Proc. I. R. E. **20**, 740 (1932).
[156] R. A. Heising, Proc. I. R. E. **16**, 75 (1928).
[157] E. V. Appleton, Proc. Roy. Soc. **126**, 542 (1930).
[158] G. Goubau, E. N. T. **10**, 72 (1933).
[159] J. Lugeon, Comptes rendus **191**, 525 (1930).
[160] J. Lugeon, Comptes rendus **191**, 676 (1930).
[161] M. A. Bontch-Bruewitch, Nature **133**, 175 (1934).
[162] S. Sillitoe, Can. J. Research **11**, 163 (1934).
[163] S. S. Kirby and E. B. Judson, Proc. I. R. E. **23**, 733 (1935).
[164] J. P. Schafer and W. M. Goodall, Nature **131**, 804 (1933).
[165] E. V. Appleton, Nature **131**, 872 (1933).
[166] J. A. Ratcliffe and E. L. C. White, Nature **131**, 873 (1933).
[167] J. A. Ratcliffe and E. L. C. White, Proc. Phys. Soc. **46**, 107 (1934).

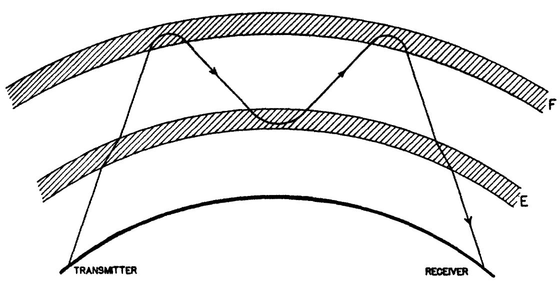


FIG. 10. "M" reflection.

the symbol f_{E_1}. If the resolving power of the experimental apparatus is sufficiently high, the two component critical frequencies, $f^0{}_{E_1}$ and $f^x{}_{E_1}$ corresponding to the ordinary and extraordinary rays may be separated. (Magneto-ionic splitting of the E layer has been observed in our own experiments, though the effect is much more conspicuous in the F region on account of the smaller electron-density gradients encountered at the higher levels.)

Similar indications of substratification have been found in the F region. Kirby, Berkner, and Stuart[168] note the existence of an F_1 layer somewhat below the main F_2 layer. The boundary becomes indistinct in winter and at night, but is frequently distinguishable during summer daylight hours. Appleton[169] confirms this observation. Having in mind the graph of ion density as a function of altitude, he considers that F_1 is a sort of ledge or protuberance on the lower side of the main F_2 region. It is quite possible for the two types of reflection to appear simultaneously. In fact Henderson[170] calls attention to the occurrence of anomalous echoes from the high levels, which occur at times when Region E is much more intensely ionized than region F.

As the experimental frequency is progressively increased, the F_1 region becomes completely penetrable, and eventually the F_2 reflections also disappear. In the latter case, however, we do not have an entirely clear indication of the sudden decrease in group velocity which definitely marks the penetration frequencies of the lower layers. The reflections often fade out somewhat gradually without a marked increase in the equivalent height. Furthermore, as F_2 is presumably the outermost strongly reflecting shell, we cannot demonstrate its penetration by exhibiting powerful echoes from points beyond. The indirect evidence afforded by skip zone observations appears to be entirely consistent with the common assumption that electron limitation determines the short wave limit of ionosphere transmission and that waves of very high frequency therefore escape into interstellar space. However, several writers[168, 171] have suggested that the very short waves merely pass so far into the F_2 region that they are completely absorbed. Our own measurements support the "complete penetration" hypothesis, but the question cannot be regarded as definitely settled.

Magneto-ionic double refraction effects may be observed without difficulty in the F region. Consequently the symbols $F^0{}_1$, $F^x{}_1$, $F^0{}_2$, and $F^x{}_2$ are needed in order to describe the complete set of first order reflections. The experimental results are in entire agreement with the magneto-ionic theory presented in Sections H and I. Several types of polarimeters have been employed. Some statistical results will be presented in Section O.

Although no strong and steady reflections from points more distant than F_2 are commonly observed, there are many indications of echoes which are weak, or intermittent, or diffuse, or occasional. In our earliest measurements, made with simple mechanical oscillographs, strong echoes were frequently received from points more that 1500 km distant, but these echoes ordinarily vanished in a fraction of a second. Similar effects may be readily demonstrated by projecting the echo pattern on a cathode-ray oscillograph screen. Taylor and Young[172, 173] and Quäck and Mögel[174, 175] noticed anomalous echoes at points within the normal "skip zone," and it is probable that these are due to the "scattering" phenomena mentioned in Section R. By operating our own continuous automatic re-

[168] S. S. Kirby, L. V. Berkner, and D. M. Stuart, Proc. I. R. E. **21**, 757 (1933).
[169] E. V. Appleton, Electrician **110**, 857 (1933).
[170] J. T. Henderson, Proc. I. R. E. **22**, 679 (1934).
[171] T. L. Eckersley, Wireless Engineer **4**, 213 (1927).
[172] A. H. Taylor and L. C. Young, Proc. I. R. E. **16**, 561 (1928).
[173] A. H. Taylor and L. C. Young, Proc. I. R. E. **17**, 1491 (1929).
[174] E. Quäck and H. Mögel, Proc. I. R. E. **17**, 824 (1929).
[175] E. Quäck and H. Mögel, E. N. T. **6**, 74 (1929).

corders[176–178] under maximum sensitivity conditions we have found two different types of extremely weak but systematic reflections[179] coming from points distant 600–1800 km from the transmitter. These echoes sometimes remain comparatively steady in signal strength and position for several hours. The direction of arrival of such echoes is not yet known.

Hollingworth[180, 181] finds indirect evidence suggesting that 30 meter waves are occasionally trapped between the *E* and *F* regions and transmitted long distances by this mode of propagation. Similar ideas have been advanced by Janco.[182]

In order to determine the relation between "apparent height" and "true height" in the *F* region we shall probably need improved simultaneous measurements employing different ray paths. Suggestions in regard to the "true height" have been made by Schelleng,[183] Kenrick and Jen,[184] Ranzi[185] and numerous others.

In 1927 M. Hals observed strong echoes which were apparently reflected from regions far outside of the earth's atmosphere, since the returning signal could be identified by ear, the time lag measured by an ordinary stopwatch. His observations were repeated and confirmed by Störmer[186–189] and Van der Pol.[190] Störmer considers that these reflections are due to distant clouds of moving electrons, which form a converging mirror of toroidal shape as a result of their motion in the presence of the earth's magnetic field. Van der Pol claims that the enormous time lag (of 30 seconds or more) merely results from an abnormally low group velocity and believes that the reflection takes place in the ionosphere. Pedersen[191] disputes this theory, since abnormally low group velocities should be accompanied by excessive attenuation. Though the original observations seem entirely trustworthy, recent attempts to reproduce the echoes have been unsuccessful. Probably they can only be heard during an especially favorable part of the sunspot cycle.

L. Why Does Stratification Exist?

From a knowledge of the nature and intensity of the solar radiation, and a knowledge of the composition, the absorption coefficients, and the motion of the earth's atmosphere, it would seem possible to give a complete interpretation of the stratification observed in the ionosphere. Unfortunately, as Chapman[192, 193] has recently pointed out, the present tables of atmospheric pressure, temperature, density, and composition, for heights much above 30 km, are largely speculative. At best they merely illustrate the conclusions to be drawn from alternative hypotheses.

Consequently, the extensive new quantitative data derived from ionosphere measurements must play an important part in extending our knowledge of the atmosphere and the incident radiation. Do atmospheric winds exist at great heights? What is the degree of dissociation of the atmospheric constituents? Does hydrogen predominate at great heights or is it almost completely absent? To what extent is ozone carried by winds? A complete analysis of such questions would extend through the entire field of meteorology. In the present space it is, therefore, impossible to do more than to mention a limited number of references which may be used as starting points in investigating the theoretical background of ionospheric stratification. Current papers are largely contradictory, though often useful in suggesting new experiments.

Chapman[194] has offered a theoretical analysis which covers:

---

176 H. R. Mimno and P. H. Wang, Phys. Rev. **41**, 395 (1932).
177 H. R. Mimno and P. H. Wang, Proc. I. R. E. **21**, 529 (1933).
178 H. R. Mimno and P. H. Wang, Proc. Fifth Pacific Science Congress (1934), p. 2195.
179 H. R. Mimno, Nature **134**, 63 (1934).
180 J. Hollingworth, J. Inst. Elec. Eng. **72**, 229 (1933).
181 J. Hollingworth, Wireless Engineer **10**, 89 (1933).
182 N. Janco, Proc. I. R. E. **22**, 923 (1934).
183 J. C. Schelleng, Proc. I. R. E. **16**, 1471 (1928).
184 G. W. Kenrick and C. K. Jen, Proc. I. R. E. **17**, 711 (1929).
185 I. Ranzi, N. Cimento **10**, 21 (1933).
186 C. Störmer, Naturwiss. **17**, 643 (1929).
187 C. Störmer, Comptes rendus **189**, 365 (1929).
188 C. Störmer, Comptes rendus **190**, 106 (1930).
189 C. Störmer, Proc. Roy. Soc. **50**, 187 (1930).
190 B. Van der Pol, Nature **122**, 878 (1928).
191 P. O. Pedersen, Proc. I. R. E. **17**, 1750 (1929).
192 S. Chapman, J. Roy. Meteorolog. Soc. **60**, 127 (1934).
193 S. Chapman, U. R. S. I., London Congress (1934).
194 S. Chapman, Proc. Phys. Soc. **43**, 26 (1931).

1. Absorption of ionizing radiation
2. Absorption of nonionizing radiation such as that which produces ozone.
3. Treatment of dissociating radiation where products of dissociation recombine according to the simple law:

$$dn/dt = I - \alpha n^2.$$

He derives values for the density of the dissociation products as a function of height, time of day, latitude, and season. Appleton[195] finds that the measured values of ionization density show diurnal variations resembling those which Chapman deduced in his treatment of ionization caused by monochromatic radiation.

In the course of a theoretical survey Hulburt[196, 197] compares radio determinations of electron density with values obtained from magnetic theories and other available information. He also discusses the significance of the measured values of the "recombination" rate in the E and F regions, maintaining that ionic recombination takes place in the E region. He believes that the decrease of free electron density in the F_1 region may be ascribed to the attachment of electrons to oxygen molecules. He considers that the F_2 region is due to ionic winds flowing outward from the heated regions near the earth's equator.

However, Appleton[198] claims that Hulburt has confused the main F_2 region which was discovered in 1926, with the subsidiary shelf, F_1, which was discovered in 1933. He further maintains that Hulburt is incorrect in considering that the E region consists mainly of ions of molecular mass. Chapman[199] finds that true recombination is more important than the attachment of electrons to neutral particles (in the F region as well as in the E region). Eckersley[200] also presents numerical values supporting this viewpoint.

Lassen[201] states that the effect of free electrons is unimportant in comparison with hydrogen ions, while Conroy[202] finds no hydrogen at any height. He coordinates the data obtained from observations of meteors, auroral spectra, and the refraction of sound waves in the stratosphere. The dissociation of oxygen has been treated extensively by Krütschkow.[203] Nagaoka[204] believes that the upper layer is due to the ionization of helium.

Nagaoka also ascribes certain differences in east-west and west-east transmission to an asymmetry in the layer. The matter has also been investigated by Nakai.[205] Namba[206–209] offers a general theory of ionosphere propagation and attempts to compute the effect of the altitude of the sun on the properties of the layer.

Various hypotheses regarding the nature of the incident radiation have been examined by Chapman[210] in his extensive theoretical treatment of magnetic storms. Elias[211] supposes a permanent ionization produced by corpuscular rays from the sun, plus a temporary daytime ionization due to solar wave radiation. Swann[212] notes that the earth's electric charge (as indicated by measurements up to a height of 10 km) would decrease 90 percent in ten minutes if not replenished in some unknown manner. He postulates a possible slow death of positive charge on the earth and discusses difficulties with the solar electron-stream hypothesis. He proposes a new scheme of electrodynamics. However, various other writers believe that the earth's electric charge is maintained in equilibrium by constant thunderstorm activity in the great static-producing areas of the earth's surface.

M. Tidal Effects in the Ionosphere

By measuring the signal strength produced by distant radio stations, several observers have obtained indirect evidence of a lunar tidal effect

[195] E. V. Appleton, Nature **127**, 197 (1931).
[196] E. O. Hulburt, Proc. I. R. E. **18**, 1231 (1930).
[197] E. O. Hulburt, Phys. Rev. **46**, 822 (1934).
[198] E. V. Appleton, Phys. Rev. **47**, 89 (1935).
[199] S. Chapman, Proc. Roy. Soc. **141**, 697 (1933).
[200] T. L. Eckersley, Proc. Roy. Soc. **141**, 697 (1933).
[201] H. Lassen, E. N. T. **4**, 174 (1927).
[202] C. C. Conroy, Science **74**, 113 (1931).

[203] S. Krütschkow, J. Applied Physics, Moscow **7**, 61 (1930).
[204] H. Nagaoka, Radio Research Japan, Report **1**, 1 (1931).
[205] T. Nakai, Electrot. Laborat. Tokyo, Japan, Researches, No. 241 (1928).
[206] S. Namba, Proc. I. R. E. **21**, 238 (1933).
[207] S. Namba, J. Inst. Elec. Eng., Japan **52**, 103 (1933).
[208] S. Namba, Electrot. Laborat. Tokyo, Japan, Researches, No. 336 (1932).
[209] S. Namba, Radio Research Japan, Report **2**, 303 (1932).
[210] S. Chapman, Terr. Mag. **36**, 77 (1931).
[211] G. Elias, Zeits. f. Hochfrequenztechnik **27**, 66 (1926).
[212] W. F. G. Swann, J. Am. Inst. Elec. Eng. **47**, 209 (1928).

acting on the ionosphere. Stetson[213, 214] has recently reviewed several experiments and described his own observations. Apparently the signal intensity is decreased as the moon passes over the observer's meridian. The cause of the tidal effect and the nature of the ionosphere disturbance remain somewhat speculative. The simple gravitational atmospheric tide is very small.

Using Gilliland's[215] experimental data, Vreeland[216] finds an indication of a reduction in layer height at the time of new moon. However, the layer height observations do not appear to be extensive enough to allow us to discriminate with certainty between the lunar cycle and the period of sunspot rotation, with its associated magnetic changes.

Breckel[217] suggests that Stetson's conclusions apply only to relatively long waves, and considers that the effect is reversed in the 3500–4000 kc region of the radio spectrum. Various additional observations have been contributed by Pickard,[218] Stoye,[219] Vincent[220] and Shannon,[221] but no general agreement has yet been reached, and it is obvious that direct layer-height measurements extending over a long period are needed.

Stetson[214] has also studied certain systematic discrepancies between radio time signals transmitted from the Greenwich observatory and the time signals transmitted from the Naval Observatory at Washington. He finds that these discrepancies depend upon the hour angle of the moon and attain values as large as ± 0.03 sec. This alteration is provisionally assumed to be due to a tidal distortion of the earth's crust which may alter the distance between the two observatories, by as much as ± 32 ft. This distortion is somewhat unexpectedly large. From the same numerical data he computes the time of transmission of a radio signal across the Atlantic and finds that it is approximately 0.04 sec. This time of transmission also shows systematic variations of the order of ± 0.01 sec., which are tentatively ascribed to tidal effects in the ionosphere. Though the apparent correlation of time signal discrepancies with the lunar hour angle is of undoubted interest, I feel that the provisional numerical analysis is somewhat weakened by the basic assumption "that the radio wave passes either way across the Atlantic in equal time intervals." As the results imply that the effective ground-level speed of the signal is less than half of the velocity of light, it is evident that the properties of the transmitting medium must receive thorough consideration, and it is possible that the west-east and east-west asymmetries, previously mentioned, may invalidate the basic assumption. Furthermore, the paths of the useful rays would probably depend in part upon the vertical-plane directional characteristics of the transmitting and receiving antennas. As the time of transmission is apparently of the same order of magnitude as the observed discrepancies, it seems somewhat hazardous to treat the symmetry condition as a self-evident fact.

N. Sunspots, Magnetic Indices, and Auroral Displays

The existence of a close connection between sunspots, magnetic disturbances, auroral displays and radio reception is universally admitted, though the details of the connecting mechanism are little understood. The technical literature covering this phase of radio transmission is particularly extensive.

A large amount of data on the field strength of distant radio stations has been compiled by Austin,[222, 223] Pickard[224–227], and many others. When plotted on a yearly, monthly, or weekly basis the correlation of radio reception with Wolfer sunspot numbers is obvious. In general, transmission becomes worse as the sunspot numbers increase, but the magnitude and even the direction of the change depends to some

[213] H. T. Stetson, Phys. Rev. **37**, 1021 (1931).
[214] H. T. Stetson, *Earth, Radio, and the Stars* (McGraw-Hill, 1934).
[215] T. R. Gilliland, Proc. I. R. E. **19**, 114 (1931).
[216] F. K. Vreeland, Proc. I. R. E. **19**, 1500 (1931).
[217] H. F. Breckel, Radio Engineering **11**, 19 (1931).
[218] G. W. Pickard, Bull. National Research Council (1931), p. 125.
[219] K. Stoye, Funktechnische Monatshefte (April, 1933), p. 152.
[220] P. Vincent, Onde Élec. **5**, 554 (1926).
[221] D. Shannon, Wireless Engineer **3**, 429 (1926).
[222] L. W. Austin, Proc. I. R. E. **15**, 825 (1927).
[223] L. W. Austin and I. Wymore, Proc. I. R. E. **16**, 166 (1928).
[224] G. W. Pickard, Proc. I. R. E. **15**, 83 (1927).
[225] G. W. Pickard, Proc. I. R. E. **15**, 749 (1927).
[226] G. W. Pickard, Proc. I. R. E. **15**, 1004 (1927).
[227] G. W. Pickard, Proc. I. R. E. **19**, 353 (1931).

extent on the wave-lengths employed. In short wave transmission between definite geographic points, Plendl[228] and Mögel[229] find that the optimum wave-length may be increased as much as 30 percent by the decrease of average ionization density apparently occurring at a sunspot minimum. Austin,[230] Yokoyama, and Nakai[231] find a closer solar correlation for short waves than for long waves, though Abbot[232] believes that seven periodicities in the values of the solar constant also appear in long wave propagation data.

However, there does *not* appear to be a good *day-to-day* correlation of radio field strength with the passage of definite sunspot groups across the central portion of the sun, though the radio data does follow the daily variations in the magnetic field of the earth. It is entirely possible that the sunspots are not the direct cause of terrestrial disturbances but merely symptoms of some underlying solar perturbation. If corpuscular rays cause the ionization changes it is also possible that these rays are deviated from the simple radial path after emission from the sun.

Surveys of the earlier experiments on sunspot correlations have been given by Mesny[233] and Stetson.[234] Recent observations on transatlantic signals have been presented by Judson.[235] Gunn[236] and Larmor[237] have considered the magnetic fields of sunspots, while Dauvillier[238] has examined the possibility of a deformation of the ionosphere under pressure of solar radiation.

Though easily obtainable with a minimum of apparatus, radio field strength measurements present no direct indication of the ionization density at various points in the ionosphere. Depending upon the wave-lengths employed in the test and the geographic locations of the transmitter and receiver, some signals become stronger and others weaker as the magnetic conditions and the sunspot numbers change.

During recent years, therefore, we have begun to make continuous daily records of the actual reflections received from various levels of the ionosphere. Such an experiment requires elaborate automatic apparatus, but it yields direct information in regard to ionization densities. We have now obtained enough reflection records to permit a preliminary comparison with available magnetic data. An 86 meter signal, transmitted vertically upward, usually penetrates the E layer without producing a reflection, but at times the ionization in this region exceeds the critical value and the corresponding echo is photographically recorded. On days when the average ionization of the E region is high we should expect to record such echoes during a relatively large fraction of the 24-hour period.

In Fig. 11 we have plotted the total duration of these E reflections in hours per day. In drawing these curves we have arbitrarily designated 86-meter E layer reflections as "abnormal" when they commence between 9 P.M. and ground level sunrise. Evidently these "abnormal" reflections are *not* caused by simple uniform ionization arising from the absorption of ultraviolet light. All other E layer reflections are described as "normal." The "normal" reflection curve shows a decided agreement with the magnetic index (which has been plotted downward since the correlation is inverse). The numerical value of the correlation coefficient is -37. Large values of the magnetic index indicate days characterized by extensive fluctuations of the earth's magnetic field. These indices represent average values obtained from the data of 48 magnetic observatories by G. van Dijk. They are published in *Caractère Magnétique des Jours* and in *Terrestrial Magnetism and Atmospheric Electricity*. The "abnormal" nighttime reflections do not show such a pronounced correlation with the magnetic conditions. (The numerical coefficient is -15.)

In general it appears that there is a reduced probability of daytime E layer reflections when the magnetic conditions are disturbed. This indicates a decreased average ionization under such conditions or an increased absorption. The first hypothesis is strongly supported by the F_1 layer observations presented in Fig. 12.

---

228 H. Plendl, Proc. I. R. E. **20**, 520 (1932).
229 H. Mögel, Telefunken Zeits. **13**, 32 (1932).
230 L. W. Austin, Proc. I. R. E. **20**, 280 (1932).
231 E. Yokoyama and T. Nakai, Proc. I. R. E. **19**, 882 (1931).
232 C. G. Abbot, Science **75**, 607 (1932).
233 R. Mesny, Onde Élec. **8**, 103 (1929).
234 H. T. Stetson, J. Frank. Inst. **210**, 403 (1930).
235 E. B. Judson, Proc. I. R. E. **21**, 1354 (1933).
236 R. Gunn, Terr. Mag. **34**, 154 (1929).
237 J. Larmor, Nat. Roy. Astron. Soc. Monthly **94**, 469 (1934).
238 A. Dauvillier, U. R. S. I., London Congress (1934).

Whenever the average ionization is low we should expect that morning sunlight would have to act upon the F region for a relatively long period before producing an ion density great enough to reflect 86 meter waves at normal incidence. The commencement of reflections during the sunrise period should therefore be delayed. Under similar conditions the steady decrease of ionization after sunset should terminate the F_1 layer reflections at a relatively early hour in the evening. The "critical times" at which F_1 reflections commence and cease may usually be determined precisely for the extraordinary ray and for the ordinary ray. During these critical periods rapid changes of group velocity produce corresponding alterations in the observed equivalent height of the reflector and it is therefore convenient to speak of a "rising" F_1 layer in the evening hours and a "falling" F_1 layer during the sunrise period. The critical times correspond to the attainment of definite critical densities in the ionized region.

With the method of graph plotting adopted in Fig. 12 all four upper curves should rise whenever the mean ionization is low. Apparently this occurs when the magnetic condition is disturbed. To show the general trend we have plotted the ionosphere data and the magnetic index on the basis of ten-day averages. Similar correlations may be obtained by plotting the daily values directly, but some confusion is caused by excessive detail unless the horizontal axis of the graph is greatly extended.

From Fig. 11 and Fig. 12 we have two separate statistical indications of a *decrease* in average atmospheric ionization on days when magnetic fluctuations occur. This raises interesting theoretical problems. Evidently we cannot accept the naive idea of great clouds of electrons, ejected violently from the sun during sunspot eruptions, which might cause auroral displays, magnetic fluctuations, and *immediate increases* of atmospheric ionization. Furthermore this detailed analysis of day-to-day alterations must eventually be reconciled with accepted individual observations[169, 239] which indicate that the yearly average ionization is greatest at the maximum of the sunspot cycle. In order to obtain further experimental information, it would seem desirable to continue the registration of echo patterns over an extended period. Automatic records totaling approximately 10,000 hours are available at present in our files, but these cover a limited portion of the solar cycle.

[239] E. O. Hulburt, Physics **4**, 196 (1933).

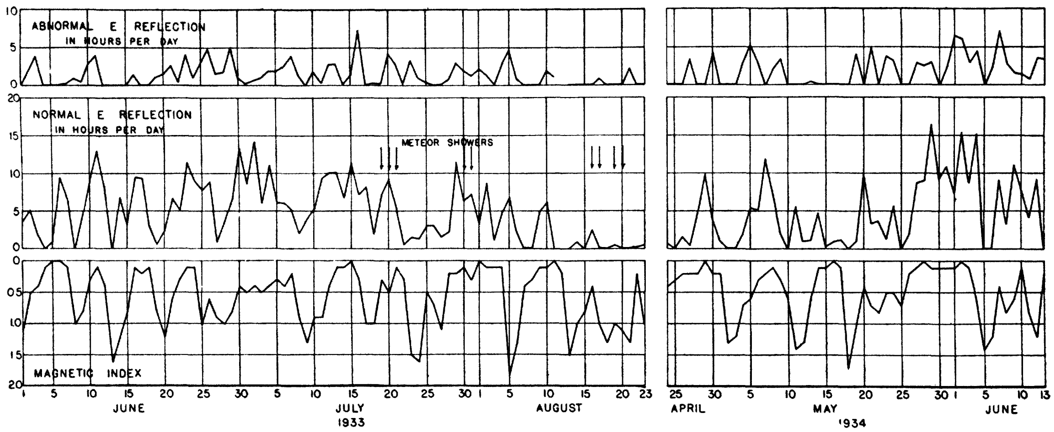
There appears to be general agreement among experimentalists in regard to the effect of polar aurora on radio transmission in Northern latitudes. Essentially similar reports have been made by Wagner,[240] Helbronner,[241] Ogilvie,[242] Sutton[243] and many others. A marked decrease of signal strength occurs simultaneously with the onset of a nearby visible aurora. This effect is particularly marked in the case of short wave transmission, and is frequently severe enough to shut off the signals completely. Echoes disappear in a manner which suggests complete absorption. At more distant points less violent effects of the same general nature occur, though there is some indication of an appreciable time lag between the aurora and the resultant weakening of signals. The visible aurora is accompanied by simultaneous rapid fluctuations in the magnetic field of the earth. Düll[244] states that it is preceded by rapid variation in the direction of arrival and the intensity of radio signals.

Dauvillier[245, 246] divides auroral phenomena into two distinct stages. He states that an initial cosmic effect, which is frequently of short duration, is followed by a relatively slow spread of phosphorescence due to excitation, ionization, and the production of ozone. He observes luminous clouds of ions in rapid motion at an approximate height of 200 km and believes that this ionic "wind" is due to strong electromagnetic forces. "Wind velocities" of the order of tens of kilometers per second occur.

The nature of the fundamental cosmic effect remains somewhat obscure, though Störmer[247, 248] has shown that the spectacular multiple folds of

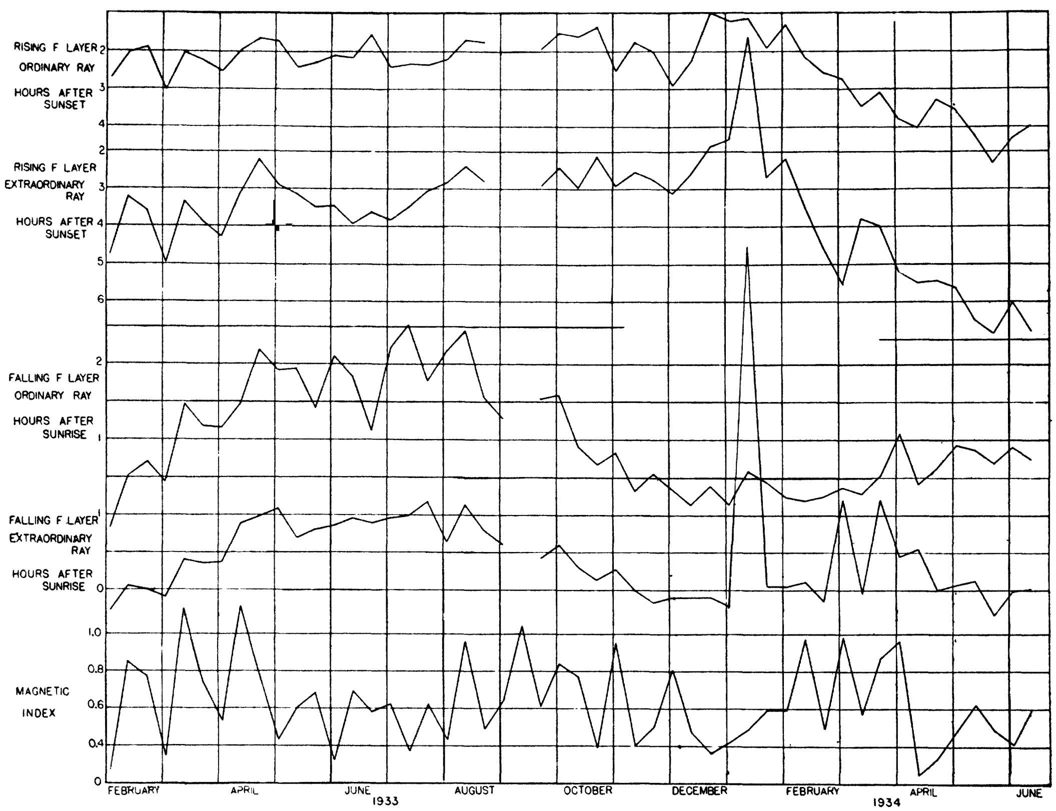
[240] K. W. Wagner, E. N. T. **11**, 37 (1934).
[241] P. Helbronner, Comptes rendus **191**, 536 (1930).
[242] N. J. Ogilvie, Am. Geophys. Union, Trans., 14th Meeting (April, 1933), p. **44**.
[243] W. Sutton, Q. S. T. **10**, p. 23, October, 1926.
[244] D. Düll, Electronics **5**, 268 (1932).
[245] A. Dauvillier, J. de phys. et rad. **5**, 398 (1934).
[246] A. Dauvillier, Comptes rendus **194**, 192 (1932).
[247] C. Störmer, Terr. Mag. **35**, 193 (1930).
[248] C. Störmer, Terr. Mag. **36**, 133 (1931).

FIG. 11.



E LAYER ACTIVITY AND MAGNETIC INDEX

FIG. 12.



CRITICAL F LAYER TIMES AND MAGNETIC ACTIVITY

the auroral "draperies" may be reasonably explained by considering families of trajectories of incoming electrons. Brüche[249–251] has imitated some of the auroral phenomena by demonstrating Störmer's electron ray hypotheses on a small scale model of the earth. His experimental technique is especially interesting. On the basis of Störmer's auroral theory and the associated theory of long delay radio echoes. Dostal[252] has computed the density of the electronic space charge in the incident electron ray, and main-

[249] E. Brüche, Zeits. f. Astrophys. 2, 30 (1931).
[250] E. Brüche, Zeits. f. techn. Physik 13, 336 (1932).
[251] E. Brüche, Zeits. f. Physik **64**, 186 (1930).
[252] H. Dostal, Ann. d. Physik **14**, 971 (1932).

tains that it is insufficient to cause direct reflection of 30-meter radio waves.

Vegard[253] has recently reviewed extensive investigations of the auroral spectrum. He concludes that the electric rays producing the aurora may possibly be regarded as a mixture of electrons and ordinary matter which largely exists in the form of positive ions, precipitated towards the earth from the corona of the sun. It has also been suggested[254] that coronal matter to a large extent consists of oxygen. If the average charge of the ray bundle is negative and numerically small, "we may explain the great angular distance between the auroral zone and the magnetic axis point."

Larmor[255] has also presented speculations on the cause of the aurora and its effect upon terrestrial ionization. He observes that the "fact that a quite small density of ions entirely upsets the optical elasticity of space as regards long waves, provides a cause preserving ionized gaseous clouds of astronomical size, for example in the interior of stars, from rapid dissipation or dispersal in bulk."

O. Magnetic Storms and Meteoric Showers

Most of the day-to-day variations in the magnetic index represent alterations in the character and duration of long continued minor fluctuations in the magnetic field of the earth. Occasionally, however, the earth experiences a sudden disturbance of such violet intensity that it deserves the name of "magnetic storm." In extreme cases, magnetic storms completely disrupt wire communication as well as radio communication. Fortunately these intense disturbances are usually brief.

Magnetic storms have been known to recur, after approximate 27-day intervals, on as many as eight successive occasions.[256] Though it is commonly believed that this is due to the rotation of the sun, it is not certain that a storm is associated with a definite visible sunspot, or other visible form of solar abnormality. In certain individual instances storms of extreme violence *have* accompanied definite sunspot groups of unusual size. Skellett[257] offers preliminary statistical evidence suggestive that ". . . the presence of an area whose activity may be seen with a spectrohelioscope is a necessary though not a sufficient condition for a radio disturbance. . . ." There is also some tentative indication of a definite time lag, of the order of one day, between the transit of the active solar area and the appearance of concomitant terrestrial effects. More complete solar data is needed.

Dellinger[258] has recently described a "cosmic phenomenon" which caused brief interruptions of long distance, short wave communication on four occasions in 1935. Such interruptions were effective over the entire illuminated half of the globe and were spaced approximately 54 days apart (twice the period of rotation of the central portion of the sun). We were fortunate enough to obtain actual layer height records[259] during extremely turbulent periods associated with three successive minor magnetic storms. It appears quite certain[260] that such turbulence would be sufficient to produce the effects described by Dellinger. Until the apparent double period of 54 days is further substantiated, however, it seems preferable to assume that brief intervening disturbances at the 27-day points may have escaped observation. It would seem that this might easily occur if the most active or most susceptible radio channels happened to lie on the dark side of the earth at the crucial moment.

A number of observers have reported moderate indications of ionization changes associated with meteoric showers. However, the existing statistical data are not sufficiently conclusive and some of the apparent results of individual experiments might be ascribed to chance. Some astronomers believe that the energy of the heaviest known showers is insufficient to produce an observable ionization change unless we may assume that the visible meteors are accompanied by very large amounts of meteoric dust. In Fig. 11

[253] L. Vegard, Geofysiske Publ. **9**, No. 11, Oslo (1932).
[254] T. L. De Bruin, Naturwiss. **20**, 269 (1932).
[255] J. Larmor, Nature **133**, 221 (1934).
[256] G. Angenheister and J. Bartels, *Handbuch der Experimentel Physik*, Vol. 25, Part 1 (Wien-Harms, 1928), p. 674.
[257] A. M. Skellett, Proc. I. R. E. **23**, 1361 (1935).
[258] J. H. Dellinger, Science **82**, 351 (1935).
[259] H. R. Mimno and P. H. Wang, Phys. Rev. **43**, 769 (1933).
[260] H. R. Mimno, Science **82**, 516 (1935).

we have indicated the time of occurrence of some of the periodic meteor showers, but no evident correlation appears as yet in our data.

Tentative correlations with commercial transmission records have been reported by Nagoaka,[261] Quäck,[262] and Pickard.[263] Skellett[264, 265] finds a reduction in the height of the *E* layer and offers ionization computations to support his viewpoint. Minohara and Ito[266] claim that meteors greatly increase the number of echoes. Schafer and Goodall[267] are somewhat more conservative, but they also find some reason to believe that meteors cause an observable increase in ionization. Interesting theoretical papers have been contributed by Lindemann,[268] Maris,[269] Mascart,[270] Millman[271] and Malzer.[272]

Visual observation of the motion of meteor trails have indicated the existence of winds at great heights in the upper atmosphere. This phase of the subject has been discussed by Hulburt.[273]

P. Thunderstorms and Barometric Effects

Most of the meteorological readjustments which determine our daily weather conditions take place in the lower 10 kilometers of the earth's atmosphere. Even under extreme conditions the observed ionization in this region is relatively small and one would not expect significant absorption or ionic refraction of radio waves. Nevertheless there is a considerable amount of reliable evidence indicating that ground level weather observations are related in some way to radio propagation.

In the "quasioptical" region below 10 meters ordinary optical mirage effects frequently occur under suitable conditions of temperature gradient and moisture gradient. As mentioned in Section C, distant transmitting stations, one or two degrees below our optical horizon, are occasionally received in Cambridge with phenomenal intensity. The transmission path is presumably confined to the troposphere, and the slight amount of atmospheric refraction involved does not depend upon the existence of ionization.

In the regions above 10 meters, however, there is no general agreement upon the explanation of the effects which are observed. Bureau[274] suggests that "the ionized layers of the upper atmosphere play the principal role; but in certain conditions . . . a slight modification can decide between two possible and different paths along these layers. This modification may be the act of the phenomena of the troposphere which would then become the arbiters of the propagation and would decide the fate of the wave." Though this viewpoint might explain changes of signal intensity sometimes noted at sea after crossing air mass boundaries, it would scarcely account for the numerous reports of actual alterations in *E* layer height and electron density accompanying barometric changes. Ranzi[275] finds great increases in *E* ionization after sunset when barometrical depressions occur at the place of observation or *north* of it. Ratcliffe[276] agrees with Ranzi's observations and maintains that a large *rain* cloud may affect the *E* layer. Martyn[277] reports a close correlation between *E* layer ionization density and ground-level barometric pressure 12 to 36 hours *later*. Colwell[278] claims an accuracy of 85 to 90 percent in weather predictions based on reception of broadcasting stations. Fuchs[279] finds that variations in signal strength depend on atmospheric pressure along the transmission path.

Thunderstorms seem to produce effects which are even more remarkable. In Fig. 13 we have represented reflection conditions before, during, and after local thunderstorms which occurred in Cambridge at the point of transmission and reception of an 86 meter signal. The "*F* layer

[261] H. Nagaoka, Radio Research Japan, Report **2**, 49 (1932).
[262] E. Quäck, E. N. T. **8**, 46 (1931).
[263] G. W. Pickard, Proc. I. R. E. **19**, 1166 (1931).
[264] A. M. Skellett, Phys. Rev. **37**, 1668 (1931).
[265] A. M. Skellett, Proc. I. R. E. **20**, 1933 (1932).
[266] T. Minohara and Y. Ito, Radio Research Japan, Report **3**, 115 (1933).
[267] J. P. Schafer and W. M. Goodall, Proc. I. R. E. **20**, 1941 (1932).
[268] F. A. Lindemann, Nature **118**, 195 (1926).
[269] H. B. Maris, Proc. I. R. E. **16**, 177 (1928).
[270] J. Mascart, Comptes rendus **198**, 544 (1934).
[271] P. M. Millman, Proc. Nat. Acad. Sci. **19**, 34 (1933).
[272] V. Malzer, Nature **132**, 137 (1933).
[273] E. O. Hulburt, Am. Geophys. Union, Trans., 13th Meeting (1932), p. 124.

[274] R. Bureau, Comptes rendus **188**, 455 (1929).
[275] I. Ranzi, Nature **130**, 545 (1932).
[276] J. A. Ratcliffe, Proc. Roy. Soc. **45**, 399 (1933).
[277] D. F. Martyn, Nature **133**, 294 (1934).
[278] R. C. Colwell, Nature **130**, 627 (1932); R. C. Colwell and A. W. Friend, Phys. Rev. **50**, 632 (1936).
[279] J. Fuchs, Funkmagazin **2**, 1021 (1929).

height" curve is an average obtained from the records of 22 thunderstorms. Ten additional storms took place during the test period but these have been omitted from the average on account of insufficient data, or because the storm occurred at a time when the *F* layer was seriously affected by the normal diurnal cycle, or by magnetic storms. The apparent increase in the "height" of the *F* layer, at the outbreak of the storm, might be attributed to decreased group velocity in the underlying regions which the signal must traverse.

Since vertical incidence *E* layer reflections are less common than *F* layer reflections at 86 meters, the curve of *E* layer height represents an average obtained from the three storms which happened to occur while adequate *E* layer reflections were present. In complete agreement with the *F* layer curve, our *E* layer observations indicate a maximum echo delay at the time of the outbreak of the storm. The curves of Fig. 13 also indicate that the break of the storm decreases the probability of occurrence of sporadic *E* reflections. However, when such reflections *do* occur at this time, an abnormally high intensity is indicated by an increase in the length of the train of echoes (i.e., in the number of multiple reflections). Such action might be the result of a general increase in the ionization of the upper atmosphere, which frequently extends downward into the "absorbing" *D* region, but occasionally does *not* appreciably affect the levels below the *E* layer. C. T. R. Wilson and others[280] have suggested that the enormous electric fields of thunder clouds may produce a penetrating radiation which could cause ionization at a considerable distance from the storm center. Additional experimental data is needed in order to decide whether this hypothetical radiation actually exists and to determine its nature.

Although our few individual observations, made in New England, consistently indicate an increase in *E* layer height at the break of the storm, it is possible that other effects may be observed under different meteorological conditions. Ratcliffe[281] reports a storm which apparently produced a marked temporary lowering of the *E* layer.

[280] R. A. Watson-Watt, Nature **132**, 13 (1933).
[281] J. A. Ratcliffe, Science **80**, 86 (1934).

Q. Local Ionospheric Clouds

In Section K reference was made to the fact that the ion density in the *E* region does not decrease steadily and smoothly after sunset as a simple recombination hypothesis might suggest. Numerous observers have noted high night ionization[282] and the very sudden appearance of strong nocturnal *E* layer reflections[283] which do not appear to be associated with local storms or general magnetic disturbances.[284]

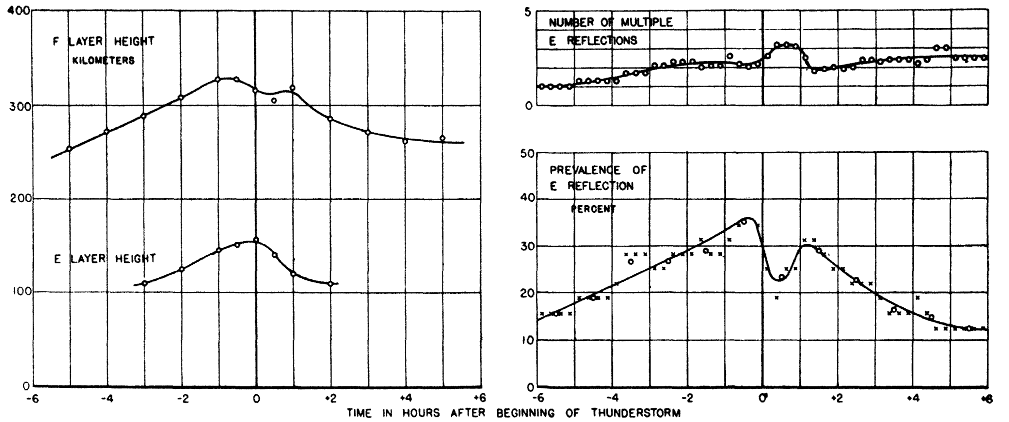
Frequently the ion densities suddenly attained on such occasions exceed the greatest densities reached during the day.[285]

Our detailed analysis of the prevalence of 86 meter *E* layer reflections in New England is presented in Fig. 14. An ordinate of 10 percent indicates that *E* reflections were present on one-tenth of the occasions observed. It will be noted that the daytime values are low in winter and high in summer. The depression during midday hours is presumably due to absorption in the underlying *D* region. (In agreement with this hypothesis it may be noted that F_1 layer reflections are also less prevalent at noon than in the forenoon and afternoon.) The probability of observing 86 meter *E* layer reflections is a maximum in midsummer, but there is also a subsidiary maximum in midwinter. This winter maximum is contributed largely by nocturnal ionization. There is a small hump in the daily prevalence curve shortly before sunrise. This occurs both in summer and in winter, and has been observed during two successive years. No adequate explanation of this presunrise effect has been brought to our attention.

The sudden and frequent appearance of very strong *E* layer reflections of short duration might be due to a violent momentary increase in the activity of some corpuscular ionizing agent capable of acting over a wide area on the dark side of the earth. It might equally well be due to a dense moving cloud of ions, with relatively sharp boundaries of limited extent, which happens to drift over the local area where the

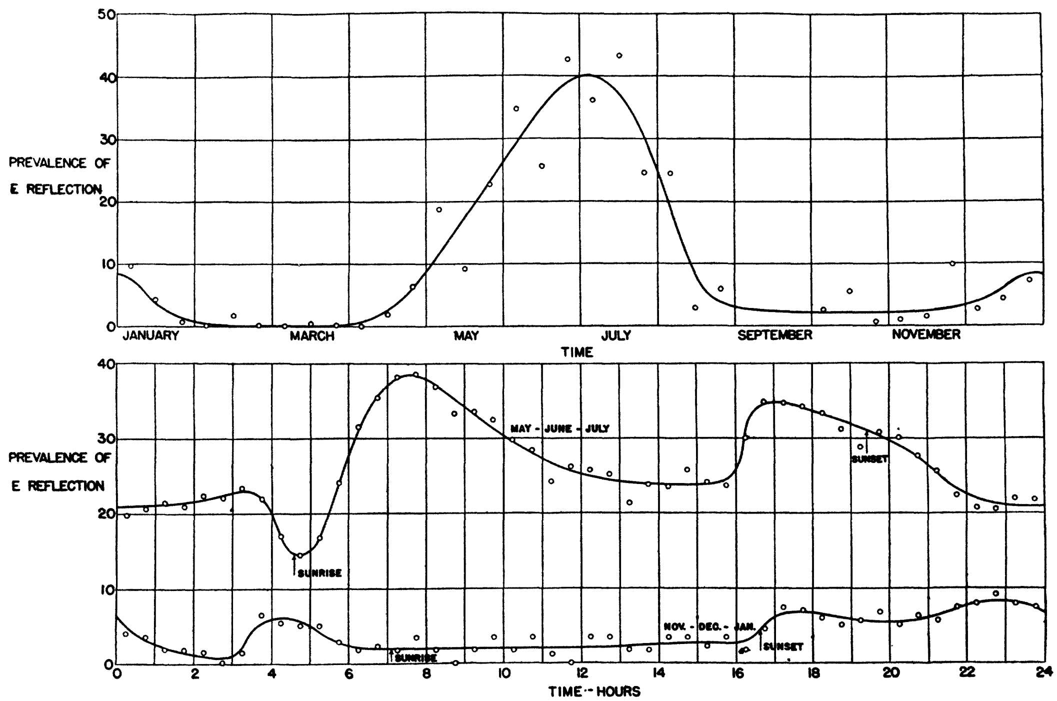
[282] T. R. Gilliland, Nat. Bur. Stand. J. Research **11**, 141 (1933).
[283] H. R. Mimno and P. H. Wang, Phys. Rev. **45**, 291 (1934).
[284] H. Mögel, Telefunken Zeits. **14**, 21 (1933).
[285] E. V. Appleton, R. Naismith, and G. Builder, Nature **132**, 340 (1933).

FIG. 13.



IONIZED LAYER ACTIVITY AND THUNDERSTORMS

FIG. 14.



experiments are being performed. Our own preliminary experiments, conducted simultaneously at Cambridge and at Worcester, tend to confirm the second hypothesis. There appear to be measurable time differences, of the order of one minute, consistently indicating a progressive motion of such a cloud over the 40 mile intervening distance.

If these preliminary indications could be confirmed they might prove to be of considerable importance in connection with protection of life and property. The original aircraft radio beacons were subject to serious errors caused by the reflection of an undesired "sky wave." This condition was greatly improved by a modification in the design of the antenna system, which decreased the energy radiated upward. Nevertheless, even the improved antenna system must radiate upward to a moderate extent, and it is at least conceivable that sudden reflections from small dense local clouds of ions might occasionally cause fatal local deviations of short duration at a particular spot on the beacon range. Aircraft pilots mention such "dead spots." The reports of recent disastrous crashes indicate that experienced pilots sometimes lose their way, without apparent reason, at times when their beacon is producing a satisfactory signal at distant ground stations, and is properly guiding aircraft over other portions of the same route.

Unfortunately this experiment has been completely interrupted for a period of nearly three years by a remarkable type of quasilegal controversy unexpectedly initiated by the Federal Communications Commission.[286] In order to terminate a long protracted argument involving merely the precise meaning of the phraseology of the Radio Act, it has now become necessary for us to seek a special Act of Congress specifically authorizing use of the automatic devices appropriate to our research No scientific objection to their use has ever been offered, since it is admitted that our methods conform to reasonable engineering practice in all respects.

Further indications of an unfortunate lack of governmental cooperation in fundamental research appear in the frequency assignments offered to experimental stations. In the range of wave-lengths extending from 15 to 30,000 meters, less than one-half of one percent of the spectrum is available for general experimental use. Such assignments are thoroughly disproportionate and inadequate. The present practice allows legitimate scientific research to be crowded out by adjacent services having greater numerical strength. Such a policy is obviously completely short-sighted and against public interest, but it cannot be adequately combated by individual physicists.

R. Scattering of Radio Waves

In Section C we have referred to a "silent zone" or "skip region" which surrounds transmitting stations operating on a sufficiently high frequency. Points in this area lie beyond the range of the ground wave but are not distant enough to receive the usual sky wave from the E or F regions. This is not always a zone of complete silence, however, for steady signals of moderate intensity sometimes appear at points which should normally be inaccessible.

An experienced radio telegraph operator often realizes that these signals possess some unusual character. The dots and dashes have a hollow ringing quality reminiscent of footsteps echoing from a stone staircase. By transmitting ground wave signals simultaneously on a longer wavelength, Taylor and Young[172] were able to show that these freak echoes are abnormally delayed. On arrival at a point only 420 km from the transmitter the echoes had various time lags corresponding to 2000 to 10,000 km of path length. At first they ascribed these echoes to signals scattered backward into the "silent zone" by distant mountain ranges. On the basis of directional observations they later decided[173] that the distant scattering object might be the sea.

Eckersley[287–290] suggests that the ionosphere has a cloudlike structure which might produce a complex scattering effect. He also suggests[109] that scattering may be due to multiple splitting. Howe[291] has collected a number of observations on this subject, including Mögel's opinion that directive beams produce scattering only at the top of the ray path. Our own observations[179] definitely indicate the presence of nonuniform, cloud-like ion formations, and tend to prove that the scattering centers are in the ionosphere and not on the earth. The ion formations observed at great distances from our transmitter seem capable of producing the type of echo observed by Taylor and Young at times when the E and F layers are readily penetrable.

S. Interaction of Radio Waves

In the usual mathematical treatment of radio propagation it is tacitly assumed that the ionosphere is a linear transmitting medium, and that the behavior of a designated wave is therefore unaffected by other waves which happen to be passing through the same region of space at the same time. Considerable interest

[286] H. R. Mimno, Science **83**, 54 (1936).

[287] T. L. Eckersley, Nature **121**, 245 (1928).
[288] T. L. Eckersley, J. Inst. Elec. Eng. **67**, 992 (1929).
[289] T. L. Eckersley, Electrician **102**, 468 (1929).
[290] T. L. Eckersley, Wireless Engineer **6**, 255 (1929).
[291] G. W. O. Howe, Wireless Engineer **8**, 579 (1931).

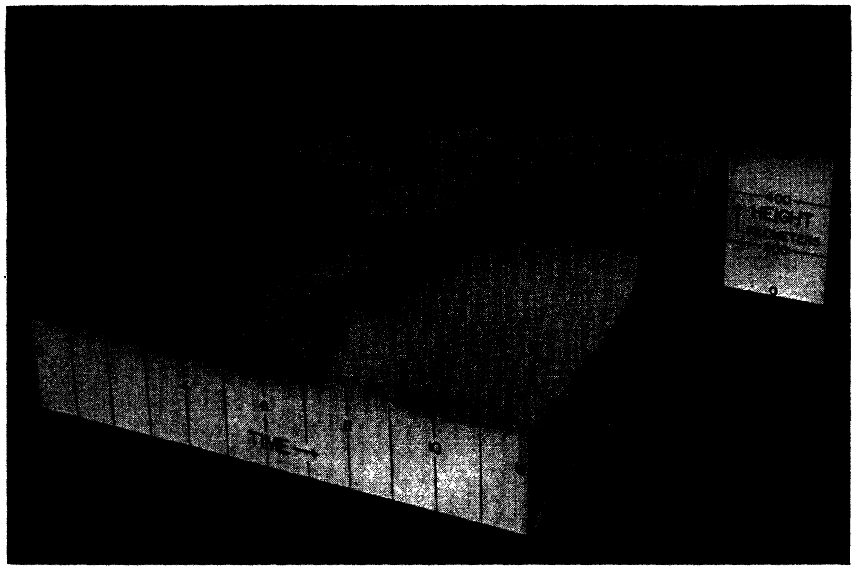


FIG. 15. Reflection conditions during solar eclipse of 1936.

was therefore aroused in Europe in 1933 when it was found that the powerful broadcasting station at Luxembourg was producing "cross modulation" which could be heard faintly on the waves emitted by other stations. When first reported by Tellegen[292] and various broadcast listeners[293–295] some skepticism was shown, as it was obvious that a spurious "cross modulation" might readily take place in the broadcast receiving set.

The reality of the effect was carefully confirmed by Van der Pol[296, 297] and it was found that waves passing directly over the powerful transmitter as indicated in Fig. 4 experienced a maximum distortion.

A possible explanation has been proposed by Bailey and Martyn[298–300] who find that a station of 200 kw operating on a wave-length of 1190 meters can produce an appreciable change in the mean velocity of agitation of the electrons in the ionosphere at points nearly over the transmitter. This will, in turn, produce a change in the frequency of collision of the electrons with molecules and hence in the absorbing power of that portion of the ionosphere. The absorbing power therefore varies in accordance with the modulation frequency of the station, and so the modulation will be impressed on any other carrier wave which may traverse the region. A quantitative treatment has been presented.

Although there is probably no relation between the two effects, it may be interesting to refer to an experiment performed by Ferrier and Donder[301] who were investigating telephony over an ultraviolet and an infrared beam of light. They state that an absorber placed near the transmitter where the rays are concentrated, has a different effect from the same absorber placed in the weaker radiation field near the receiver.

T. Eclipse Observations

By means of observations made at the time of a total solar eclipse it is possible to obtain additional experimental evidence in regard to the nature of the incident solar radiation and the stratification of the ionosphere.

[292] B. D. H. Tellegen, Nature **131**, 840 (1933).
[293] B. D. H. Tellegen, World Radio **17**, 165 (1933).
[294] B. D. H. Tellegen, World Radio **18**, 353 (1934).
[295] U. R. S. I. Discussion, Wireless World **35**, 263 (1934).
[296] B. Van der Pol, Science **79**, 11 (1934).
[297] B. Van der Pol, U. R. S. I., London Congress (1934).
[298] V. A. Bailey and D. F. Martyn, Nature **133**, 218 (1934).
[299] V. A. Bailey, Nature **133**, 869 (1934).
[300] V. A. Bailey, Phil. Mag. **18**, 369 (1934).
[301] R. Ferrier and Th. de Donder, R. G. E. **27**, 125D (1930); **27**, 133D (1930).

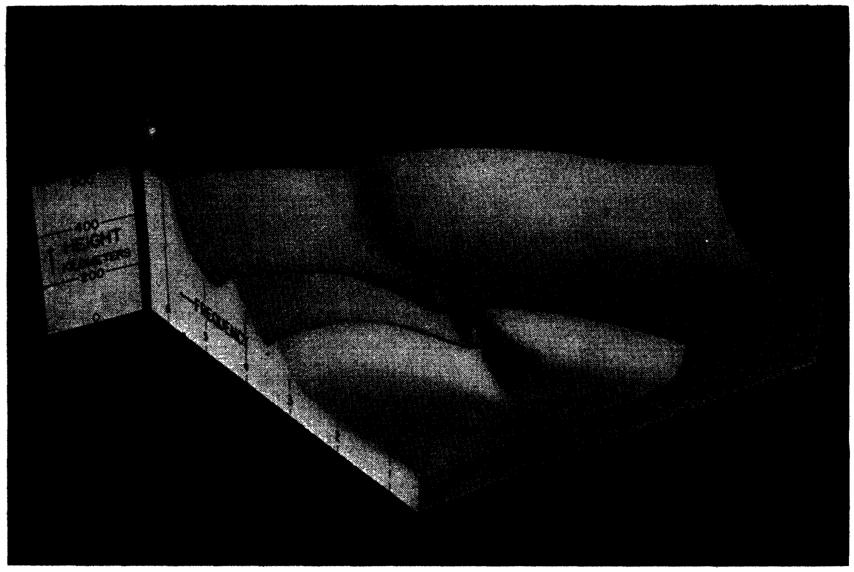


FIG. 16. Reflection conditions during solar eclipse of 1932.

The space model shown in Fig. 15 illustrates the general nature of the eclipse effect. In this space model, which will soon be described in greater detail elsewhere, the X coordinate is the hour of the day (from midnight to noon), the Y coordinate is transmitter frequency (from 2500 to 8500 kc), and the vertical Z coordinate is layer height. The sharp ridge, which occupies the central portion of the model, is the boundary which separates the F_1 reflection from the F_2 reflection. The hatched area represents penetration. We may pass from one type of reflection to the other by varying the frequency or the time of observation. On a normal morning this boundary ridge would rise steadily. The large V shaped dip results from the passage of the moon's shadow across the ionosphere. This depression centers at optical totality, within the limits of accuracy of the observation.

Fig. 15 has been prepared by Mr. J. A. Pierce and is based upon extensive data obtained by our group during the June 19, 1936, Russian eclipse by the simultaneous use of the variable frequency and constant frequency methods of observation referred to in Section H. The measurements were made at Ak-Bulak in western Turkestan. Observations of previous eclipses[177, 302–309] though individually less complete, are in good agreement with Fig. 15 when fitted together. Fig. 16 is a composite space model of the 1932 eclipse obtained by combining our own observations with the data reported by other groups. As this eclipse occurred in New England in the late afternoon, the time scale runs from noon to midnight, the frequency scale from 500 kc to 6500 kc, and the general trend of the F_1F_2 boundary ridge is downward. The smaller ridge below the main F_1F_2 boundary represents the transition from E to F_1 reflections. It was omitted from Fig. 15 as the field equipment taken to Russia was primarily designed to investigate the F region. The progress of the sunspot cycle, rather than the geographical difference, is chiefly responsible for the fact that

[302] K. Maeda, Radio Research Japan, Report **4**, 89 (1934).

[303] T. Minohara and Y. Ito, Radio Research Japan, Report **4**, 109 (1934).

[304] E. V. Appleton and S. Chapman, U. R. S. I., London Congress, September (1934).

[305] S. S. Kirby, L. V. Berkner, T. R. Gilliland, and K. A. Norton, Proc. I. R. E. **22**, 247 (1934).

[306] J. T. Henderson and D. C. Rose, Can. J. Research **8**, 29 (1933).

[307] J. P. Schafer and W. M. Goodall, Science **76**, 444 (1932).

[308] G. W. Kenrick and G. W. Pickard, Proc. I. R. E. **21**, 546 (1933).

[309] (General Survey), Nature **130**, 385 (1932).

the F_1F_2 boundary has been shifted to a higher frequency in the 1936 eclipse.

The general simularity of Fig. 15 and Fig. 16 is of considerable interest since it may be shown that hypothetical corpuscular radiation, accompanying the light of the sun, should have produced markedly different behavior in the two cases. The general simplicity of the eclipse effect seems to indicate that ultraviolet light is responsible for all of the observed phenomena.

The large magnitude of the alteration, and the fact that it is observable with considerable accuracy, indicates that eclipses provide useful quantitative data, available as a numerical test of the accuracy of any ionization and recombination theories which may be presented in order to account for the observed stratification of the ionosphere. These quantitative observations should be continued and it is possible that partial eclipses will also be of interest.[275]

U. Conclusion

In this relatively new branch of physics it is possible to perform a variety of experiments, some of which yield precise numerical data. The physical system under examination is exceedingly complex and must be studied by statistical methods. Classical magneto-ionic theory offers satisfactory quantitative interpretations for many individual experimental facts, though there are some anomalies which require further attention. No thorough quantitative study of molecular or atomic ionization and recombination processes, designed to explain the observed stratification of the ionosphere, has yet produced any generally accepted comprehensive theory. However, in view of the detailed experimental knowledge already available this major theoretical problem is worthy of serious attention at the present time.

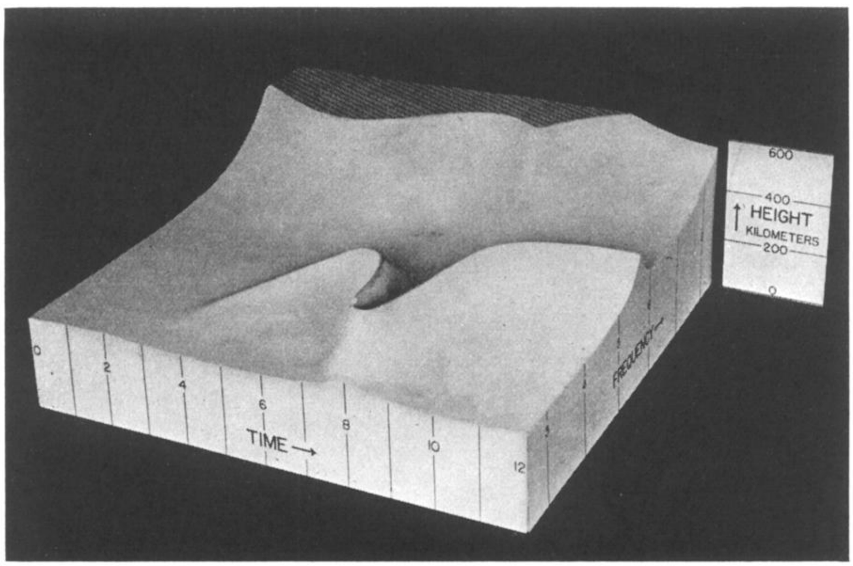


FIG. 15. Reflection conditions during solar eclipse of 1936.

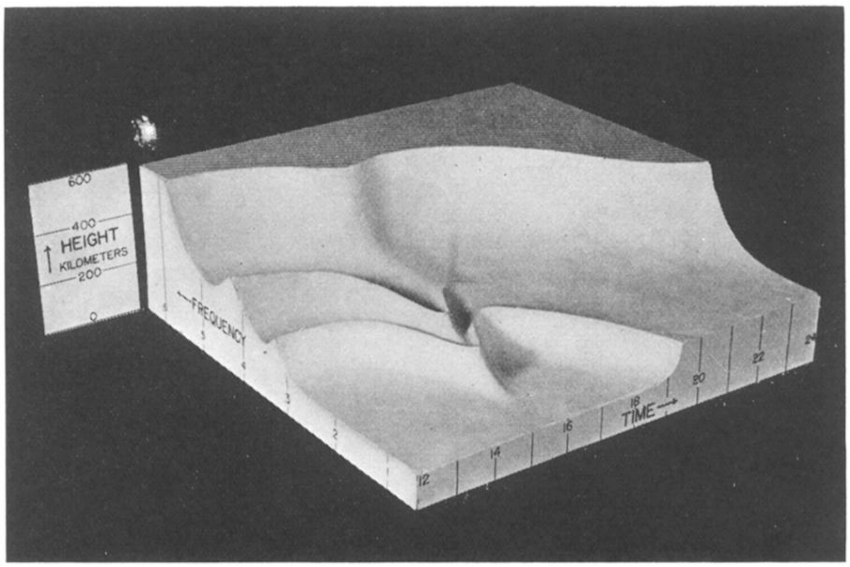


FIG. 16. Reflection conditions during solar eclipse of 1932.