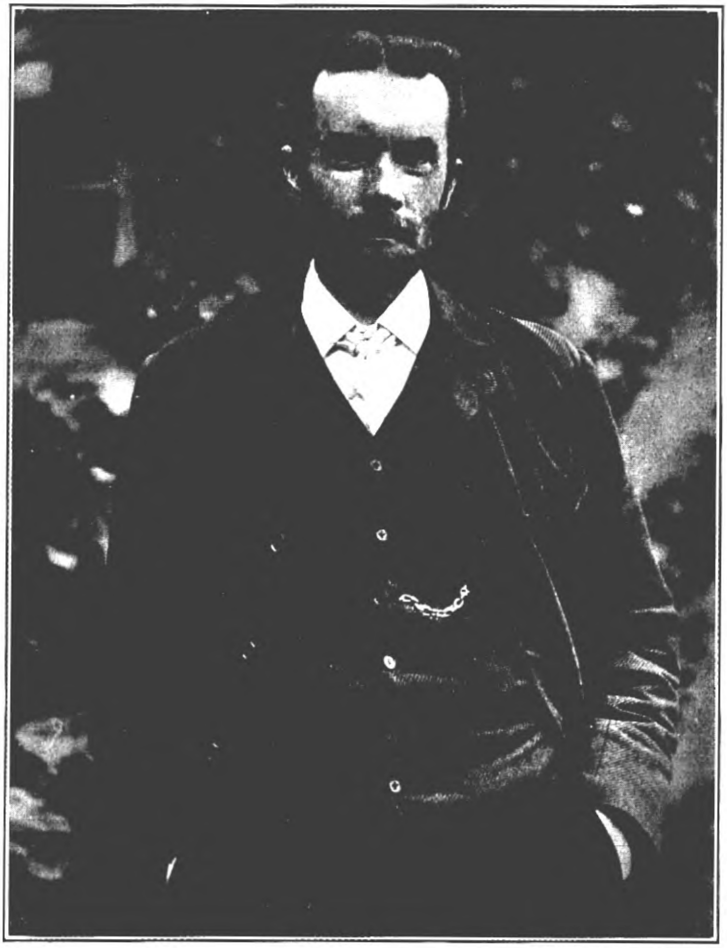


Courtesy of "Electrician," London

THE LATE OLIVER HEAVISIDE, F.R.S.

The Bell System Technical Journal

July, 1925

Oliver Heaviside

By F. GILL

ALTHOUGH abler pens[1] have expressed appreciation of the late Oliver Heaviside, it is perhaps permissible for an English telephone engineer to present a note regarding him. Of his life-history not very much is known; but he may have been influenced in his choice of a career by the fact that he was a nephew of the famous telegraph engineer Sir Charles Wheatstone. Heaviside was born in London on May 13, 1850; he entered the service of the GreatNorthern Telegraph Company, operating submarine cables, and he remained in that service, at Newcastle-on-Tyne, until 1874. While he was with the Telegraph Company, he published in 1873 a paper showing the possibility of quadruplex telegraphy.

At the age of about 24, owing, it is suggested, to increasing deafness, he left the service of that Company and took up mathematical research work. How he acquired his mathematical training does not seem to be known;[2] perhaps he was self-taught,—in some of his Papers he implies it. By whatever means he mastered the principles, it is evident that he was an ardent student of Maxwell, for constantly in Heaviside's own writing runs a vein of appreciation of Maxwell. For some time he lived in London, then he moved to Paignton in Devonshire; his Electrical Papers are written from there, and he died at the neighboring town of Torquay on February 4, 1925, in his 75th year.

That is about all the personal history at present available, and yet it gives a clue to a dominant note in his character, viz., reluctance to come into prominence, originating, perhaps, in a kind of shyness, which ultimately led to the recluse state. It is strange that so remarkable an investigator should, in his earlier manhood, have convinced so few, notwithstanding the fact that his voluminous writings made his name well known. It must, however, be remembered that his articles were very difficult, even for advanced mathematicians to follow, for he used a system of mathematics which, at that time

[1] *The Electrician*, Vol. XCIV, p. 174, by Sir Oliver Lodge, F.R.S., O.M. *Nature*, Vol. 115, p. 237, by Dr. Alex. Russell, F.R.S.

[2] Was he the youth with the frown in the library? He says he "then died," but also says "he was eaten up by lions." (E.M.T., Vol. III, pp. 1 and 135.)

was unusual. Whatever the cause, the fact remains that until about the year 1900 few engineers understood him.

Coming to his work, what was it that Heaviside did, and upon what does his fame rest? That is too large a subject for a telephone engineer to answer fully, but as regards communication engineering something may be said. His great achievement was the discovery of the laws governing the propagation of energy in circuits. He recognized the relationship between frequency and distortion; he illustrated it by numerical examples, and he showed what was required to make a "distortionless circuit." Further, he showed the effects of "attenuation" and the result of "inductance" (these words were his own coinage) in improving telephony. He also explained how the inductance of circuits could be increased; he suggested the use of continuous loading, of lumped inductance in the form of coils, and he pointed out the difficulty of obtaining sufficiently low resistance in such coils. He investigated the effect of sea and land and the upper atmosphere on the propagation of radio energy and how it was that this energy could be transmitted over the mountain of earth intervening between two distant places.

His activity in these matters can best be illustrated by extracts from his writings, as follows:

In his "Electrical Papers," Vol. II, written in 1887, p. 164, he gives numerical examples of frequency distortion and of its correction, and says:

> "It is the very essence of good long distance telephony that inductance should *not* be negligible."

In his "Electromagnetic Theory," Vol. I, published in 1893, he considers in Section 218, p. 441.

> "various ways, good and bad, of increasing the inductance of circuits"

He suggests, page 445, the use of

> ". . . inductance in isolated lumps. This means the insertion of inductance coils at intervals in the main circuit. That is to say just as the effect of uniform leakage may be imitated by leakage concentrated at distinct points, so we should try to imitate the inertial effects of uniform inductance by concentrating the inductance at distinct points. The more points the better, of course . . . The Electrical difficulty here is that inductance coils have resistance as well, and if this is too great the remedy is worse than the disease.

. . . To get large inductance with small resistance, or, more generally, to make coils having large time constants, requires the use of plenty of copper to get the conductance, and plenty of iron to get the inductance, employing a properly closed magnetic circuit properly divided to prevent extra resistance and cancellation of the increased inductance . . . This plan . . . is a straightforward way of increasing the L largely without too much increasing the resistance and may be worth working out and development. But I should add that there is, so far, no direct evidence of the beneficial action of inductance brought about in this way."

In "Electrical Papers," Vol. II, p. 311, he deals with reflected waves, and on page 347 he says:

" . . . but the transmitter and the receiving telephone distort the proper signals themselves. The distortion due to the electrical part of the receiver may, however, be minimized by a suitable choice of its impedance.

"Electromagnetic Theory," Vol. I, p. 404:—

"We have seen that there are four distinct quantities which fundamentally control the propagation of 'signals' or disturbances along a circuit, symbolized by R, K, L, and S, the resistance, external conductance, inductance, and permittance;"

"Electromagnetic Theory," Vol. I, p. 411:—

"It is not merely enough that signals should arrive without being distorted too much; but they must also be big enough to be useful . . . Nor can we fix any limiting distance by consideration of distortion alone. And even if we could magnify very weak currents, say a thousandfold, at the receiving end, we should simultaneously magnify the foreign interferences. In a normal state of things interferences should be only a small fraction of the principal or working current. But if the latter be too much attenuated, the interferences become relatively important, and a source of very serious distortion. We are, therefore, led to examine the influence of the different circuit constants on the attenuation, as compared with their influence on the distortion."

"Electrical Papers," Vol. II, p. 402:—

"I was led to it (the distortionless circuits), by an examination of the effect of telephones bridged across a common circuit (the proper place for intermediate apparatus, removing their impedance) on waves transmitted along the circuit."

With regard to Radio Communication, one extract must suffice writing on The Electric Telegraph in June, 1902, for the Encyclopedia Britannica, he says,—"Electromagnetic Theory," Vol. III, p. 335:—

"There is something similar in 'wireless' telegraphy. Sea water, though transparent to light, has quite enough conductivity to make it behave as a conductor for Hertzian waves, and the same is true in a more imperfect manner of the earth. Hence the waves accommodate themselves to the surface of the sea in the same way as waves follow wires. The irregularities make confusion, no doubt, but the main waves are pulled round by the curvature of the earth, and do not jump off. There is another consideration. There may possibly be a sufficiently conducting layer in the upper air. If so, the waves will, so to speak, catch on to it more or less. Then the guidance will be by the sea on one side and the upper layer on the other. But obstructions, on land especially, may not be conducting enough to make waves go round them fairly. The waves will go partly through them."

Probably due to his long seclusion, his approach to certain subjects was rather critical. At one time I tried to get a portrait of him for the Institution of Electrical Engineers, but failed;—he did not wish to have his photograph exhibited, he thought that "one of the worst results (of such exhibition) was that it makes the public characters think they really are very important people, and that it is therefore a principle of their lives to stand upon doorsteps to be photographed."

On another occasion when I sent him a copy of an article by a distinguished telephone engineer on "The Heaviside Operational Calculus," he replied that he had "looked through the paper . . . with much interest, to see what progress is being made with the academical lot, whom I have usually found to be very stubborn and sometimes wilfully blind."

Some have held that Heaviside was not recognized as he ought to have been. This was probably the case some time ago, but not in recent years. The same is true of many very great men who were much in advance of their time, for the English have the national characteristic that they do not make much fuss about their great men. So if Heaviside suffered, he shared this experience in common with other pioneers who deserved higher recognition. See, for example, what Heaviside himself said about one of these, in a footnote in "Electromagnetic Theory," Vol. III, p. 89:

"George Francis Fitzgerald is dead. The premature loss of a man of such striking original genius and such wide sympathies will be considered by those who knew him and his work to be a national misfortune. Of course, the 'nation' knows nothing about it, or why it should be so."

During the last 20 years or more, the significance and luminous quality of the work of Heaviside has been increasing by acknowledged mathematicians and by practical telephone, telegraph and radio engineers. To other electrical engineers his treatment of wave-transmission has not yet appealed quite so strongly.

Probably his first recognition came from his contribution to the problem—"Electromagnetic Induction and its Propagation" in the *Electrician.* It appeared as a series of articles between January, 1885 and December, 1887. His "Electrical Papers" were written at various times and were published in two volumes in 1892. Then followed his three volumes on "Electromagnetic Theory"—on the basis of the *Electrician* articles—published in 1893, 1899 and 1912. He also wrote, in 1902, the article on the "Theory of the Electric Telegraph" in the "Encyclopedia Britannica."

In 1891, the Royal Society made him a Fellow. In 1899, the American Academy of Arts and Sciences elected him an Honorary Member. In 1908 the Institution of Electrical Engineers did the same, followed by the American Institute of Electrical Engineers in 1917. The Literary and Philosophical Society of Manchester also elected him an Honorary Member. He was an Hon. Ph.D. of the University of Gottingen, and in 1921, the Institution of Electrical Engineers conferred upon him the highest award in their gift—the Faraday Medal. He was the first recipient of this Medal which was established to commemorate the 50th anniversary of the founding of the original Society of Telegraph Engineers and of Electricians, and since then the medal has been bestowed upon Sir Charles Parsons, Dr. S. Z. de Ferranti, and Sir J. J. Thomson.

From time to time there were reports of his living in great poverty, and attempts were made to help him. These reports lacked proportion, but it is true he had not much money and perhaps still less comfort; he was a difficult man to help. Towards the end of his life he received from the British Government a Civil Pension. His independent character rendered it necessary that offers of assistance should be tactfully made and apparently this was not always the case, as I believe help was sometimes refused; but there were those who succeeded. Another difficulty was his unconventional mode of living which caused him, in his last years, to live as a recluse, cooking and looking after his house alone.

Just what other work Heaviside did, in addition to his published writings, is not at present known to me. I believe he left a good deal of manuscript, but whether it is in such a state that it could be completed by another, I do not know. Let me conclude this note by an extract from his last chapter of his last book, "Electromagnetic Theory," Vol. III, page 519:—

"As the universe is boundless one way, towards the great, so it is equally boundless the other way, towards the small; and important events may arise from what is going on in the inside of atoms, and again, in the inside of electrons. There is no energetic difficulty. Large amounts of energy may be very condensed by reason of great forces at small distances. How electrons are made has not yet been discovered. From the atom to the electron is a great step, but is not finality.

"Living matter is sometimes, perhaps generally, left out of consideration when asserting the well-known proposition that the course of events in the physical world is determined by its present state, and by the laws followed. But I do not see how living matter can be fairly left out. For we do not know where life begins, if it has a beginning. There may be and probably is no ultimate distinction between the living and the dead."

The Loaded Submarine Telegraph Cable[1]

By OLIVER E. BUCKLEY

SYNOPSIS: With an increase of traffic carrying capacity of 300% over that of corresponding cables of the previous art, the New York-Azores permalloy-loaded cable marks a revolution in submarine cable practice. This cable represents the first practical application of inductive loading to transoceanic cables. The copper conductor of the cable is surrounded by a thin layer of the new magnetic material, permalloy, which serves to increase its inductance and consequently its ability to transmit a rapid succession of telegraph signals.

This paper explains the part played by loading in the operation of a cable of the new type and discusses some of the problems which were involved in the development leading up to the first commercial installation. Particular attention is given to those features of the transmission problem wherein a practical cable differs from the ideal cable of previous theoretical discussions.

Brief mention is made of means of operating loaded cables and the possible trend of future development.

Permalloy Loading

THE announcement on September 24, 1924, that an operating speed of over 1,500 letters per minute had been obtained with the new 2,300 mile New York-Azores permalloy-loaded cable of the Western Union Telegraph Company, brought to the attention of the public a development which promises to revolutionize the art of submarine cable telegraphy. This announcement was based on the result of the first test of the operation of the new cable. A few weeks later, with an improved adjustment of the terminal apparatus, a speed of over 1,900 letters per minute was obtained. Since this speed represents about four times the traffic capacity of an ordinary cable of the same size and length, it is clear that the permalloy-loaded cable marks a new era in transoceanic communication.

The New York-Azores cable represents the first practical attempt to secure increased speed of a long submarine telegraph cable by inductive loading and it is the large distributed inductance of this cable which is principally responsible for its remarkable performance. This inductance is secured by surrounding the conductor of the cable with a thin layer of permalloy. Fig. 1 shows the construction of the deep sea section of the cable. In appearance it differs from the ordinary type of cable principally in having a permalloy tape 0.005 inch thick and 0.125 inch wide, wrapped in a close helix around the stranded copper conductor.

Permalloy, which has been described by Arnold and Elmen,[1] is an alloy consisting principally of nickel and iron, characterized by very

[1] Presented before the A. I. E. E., June 26, 1925.

[2] *Jour. Franklin Inst.*, Vol. 195, pp. 621–632, May 1923; *B. S. T. J.*, Vol. II, No. 3, p. 101.

high permeability at low magnetizing forces. The relative proportion of nickel and iron in permalloy may be varied through a wide range of additional elements as, for example, chromium may be added to secure high resistivity or other desirable properties. On account

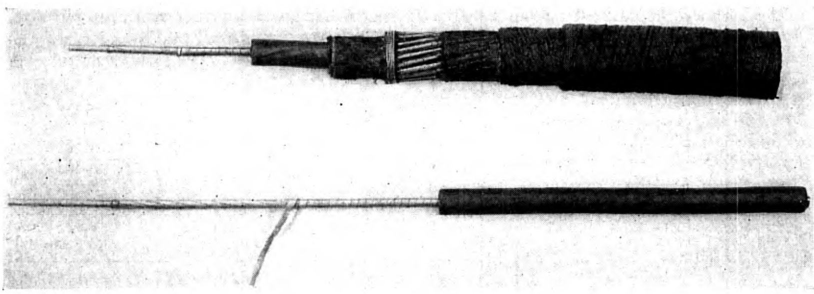


Fig. 1—Permalloy-Loaded Cable. Above, section of deep sea type showing construction. Below, section of core showing permalloy tape partly unwound.

of its extremely high initial permeability a thin layer of permalloy wrapped around the copper conductor of a cable greatly increases its inductance even for the smallest currents.

In the case of the New York-Azores cable the permalloy tape is composed of approximately 78½% nickel and 21½% iron and gives the cable an inductance of about 54 millihenries per nautical mile. An approximate value of the initial permeability of the permalloy in that cable may be got by assuming the helical tape replaced by a continuous cylinder of magnetic material of the same thickness.[3] This material would have to have a permeability of about 2,300 to give the observed inductance. A better appreciation of the extraordinary properties of the new loading material may be obtained by comparing this permeability with that which has previously been obtained with iron as the loading material. The Key West-Havana telephone cables are loaded with 0.008 inch diameter soft iron wire. The permeability of this wire, which was the best which could be obtained commercially when that cable was made, is only about 115,

[3] The true initial permeability is slightly higher. To compute it, account must be taken of the fact that, contrary to what has been sometimes assumed, the magnetic lines of induction in the tape do not form closed loops around the wire but tend to follow the tape in a helical path. The pitch of the helical path of the lines of induction is slightly less than that of the permalloy tape with the result that a line of induction takes a number of turns around the conductor, then crosses an airgap between two adjacent turns of tape and continues along the tape to a point where it again slips back across an airgap. O. E. Buckley, British Patent No. 206,104, March 27, 1924, also K. W. Wagner, E.N.T., Vol. I, No. 5, p. 157, 1924.

or approximately one-twentieth that of the permalloy tape of the New York-Azores cable.

Problems Encountered

The proposal to use permalloy loading to increase the speed of long telegraph cables was one outcome of an investigation undertaken by the author soon after the war to determine whether some of the new methods and materials developed primarily for telephony might not find important application to submarine telegraphy. In the subsequent development of the permalloy loaded cable a large number of new problems, both theoretical and practical, had to be solved before the manufacture of a cable for a commercial project could be undertaken with reasonable assurance of success. The problems encountered were of three principal kinds. First was that of the transmission of signals over a cable having the characteristics of the trial conductors made in the laboratory. Although the theory of transmission over a loaded cable had been previously treated by others, the problem considered had been that of an ideal loaded cable with simple assumptions as to its electrical constants and without regard to the practical limitations of a real cable. The second class of problems had to do with the practical aspects of design, manufacture and installation. In this connection an extensive series of experiments was conducted to determine the means required to secure at the ocean bottom the characteristics of the laboratory samples on which the transmission studies were based. Among the numerous problems which arose in this connection were those concerned with protecting the copper conductor from any possible damage in the heat-treating operation which was necessary to secure the desired magnetic characteristics, and those concerned with protecting the strain-sensitive permalloy tape from being damaged by submerging the cable to a great depth. The third class of problem had to do with terminal apparatus and methods of operation. The prospective speed of the new cable was quite beyond the capabilities of standard cable equipment and accordingly new apparatus and operating methods suited to the loaded cable had to be worked out. In particular it was necessary to develop and construct instruments which could be used to demonstrate that the speed which had been predicted could actually be secured. The success of the investigations along all three lines is attested by the results which were obtained with the New York-Azores cable. Fig. 2 shows a section of cable recorder slip, the easily legible message of which was sent from

Horta, Fayal, and received at New York at a speed of 1,920 letters per minute.

It is principally with regard to the first of these classes of problems, that of the transmission of signals, that the following discussion is concerned. No attempt will be made here to discuss the details of

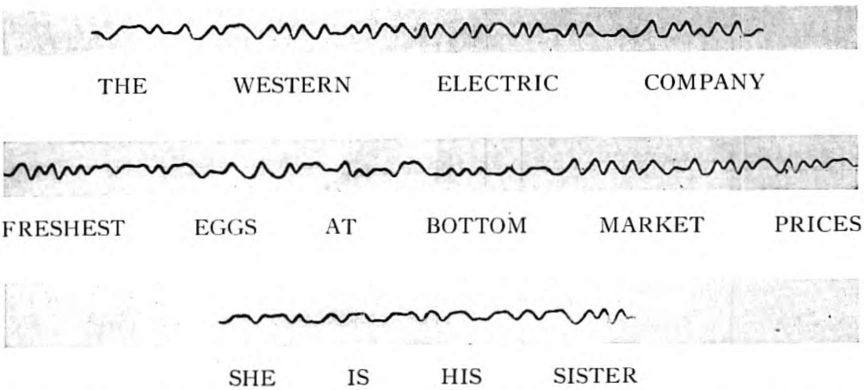


Fig. 2—Test Message. Western Union New York-Azores Permalloy-Loaded Cable. Sent from Horta (Azores) and received at New York, November 14, 1924. Speed—1920 letters per minute. Recorded with special high speed siphon recorder

design and development of the physical structure of the cable, nor will there be given a detailed description of the operating results or how they were obtained. These subjects must be reserved for later publication. It is desired in what follows to explain how inductive loading improves the operation of a submarine cable and to point out some of the problems concerned with the transmission of signals which had to be considered in engineering the first long loaded cable.

Factors Limiting Speed of Non-Loaded Cable

In order to understand the part played by loading in the transmission of signals it is desirable first to review briefly the status of the cable art prior to the introduction of loading and to consider the factors then limiting cable speed and the possible means of overcoming them. A cable of the ordinary type, without loading, is essentially, so far as its electrical properties are concerned, a resistance with a capacity to earth distributed along its length. Although it does have some inductance, this is too small to affect transmission at ordinary speeds of operation except on cables with extremely heavy conductors. The operating speed of a non-loaded cable is approximately inversely proportional to the product of the total resistance by the total capacity: that is,

$$S = \frac{k}{CRl^2},$$

where C is capacity and R resistance per unit length, and l is the length of the cable. The coefficient k is generally referred to as the speed constant. It is, of course, not a constant since it depends on such factors as terminal interference and method of operation, but is a convenient basis for comparing the efficiency of operation of cables of different electrical dimensions. As the technique of operating cables has improved the accepted value of k has increased, its value at any time being dependent on the factor then limiting the maximum speed obtainable. This factor has at times been the sensitiveness of the receiving apparatus, at other times the distortion of signals, and in recent years interference. During a great part of the history of submarine cable telegraphy distortion was considered the factor which limited the speed of operation of long cables and on this account most of the previous discussions of submarine cable transmission have been concerned principally with distortion and means for correcting it. As terminal apparatus was gradually improved means of correcting distortion were developed which practically eliminated distortion as an important factor in the operation of long cables. With distortion thus eliminated the speed was found to be limited principally by the sensitiveness of the receiving apparatus. This limit was, however, eliminated in turn by the development of signal magnifiers. During recent years, in which numerous cable signal magnifiers have been available and methods of correcting distortion have been understood, the only factor limiting cable speed has been the mutilation of the feeble received signals by interference. Most cables are operated duplex, and in these the speed is usually limited by interference between the outgoing and incoming signals. In cables operated simplex, and also in cables operated duplex where terminal conditions are unfavorable, speed is limited by extraneous interference which may be from natural or man-made sources and which varies greatly in different locations. The strength of the received current must in either case be great enough to make the signals legible through the superposed interference current. Owing to the rapidity with which the received signal amplitude is decreased as the speed of sending is increased, the limiting speed is quite sharply defined by the interference to which the cable is subject.

Means of Increasing Speed

With the speed of operation thus limited there were two ways in which the limiting speed could be increased: the interference could be reduced, or the strength of signals made greater. No great reduction in interference due to lack of perfect duplex balance could be expected, as balancing networks had already been greatly refined. Extraneous interference in certain cases could be reduced by the use of long, properly terminated sea-earths. The signal strength could be increased either by increasing the sending voltage or by decreasing the attenuation of the cable. However, with duplex operation nothing at all is gained by increasing the voltage in cases where lack of perfect duplex balance limits the speed, and with simplex operation any gain from raising the voltage is obtained at the cost of increased risk to the cable, the sending voltage being usually limited to about 50 volts by considerations of safety. The attenuation of the cable could be reduced and the strength of the signal increased by use of a larger copper conductor or by using thicker or better insulating material. None of these possible improvements, however, seemed to offer prospect of very radical advance in the art.

In telephony, both on land and submarine lines, an advantage had been obtained by adding inductance [3] in either of two ways, by coils inserted in series with the line or by wrapping the conductor with a layer of iron. The insertion of coils in a long deep-sea cable was practically prohibited by difficulties of installation and maintenance. Accordingly, only the second method of adding inductance, commonly known as Krarup or continuous loading, could be considered

[3] The idea of improving the transmission of signals over a line by adding distributed inductance to it originated with Oliver Heaviside in 1887 (Electrician, Vol XIX, p. 79, and Electromagnetic Theory, Vol. I, p. 441, 1893), who was the first to call attention to the part played by inductance in the transmission of current impulses over the cable. He suggested as a means for obtaining increased inductance the use of iron as a part of the conductor or of iron dust embedded in the gutta percha insulation. He also proposed inserting inductance coils at intervals in a long line. Other types of coil loading were proposed by S. P. Thompson (British Patent 22,304—1891, and U. S. Patents 571,706 and 571,707—1896), and by C. J. Reed (U. S. Patents 510,612 and 510,613—1893). M. I. Pupin (A. I. E. E. Trans., Vol. XVI, p. 93, 1899, and Vol. XVII, p. 445, 1900) was the first to formulate the criterion on the basis of which coil loaded telephone cables could be designed. Continuous loading by means of a longitudinally discontinuous layer of iron covering the conductor was proposed by J. S. Stone in 1897 (U. S. Patent 578,275). Breisig (E. T. Z., Nov. 30, 1899) suggested the use of an open helix of iron wire wound around the conductor and Krarup (E. T. Z., April 17, 1902) proposed using a closed spiral so that the adjacent turns were in contact. J. H. Cuntz (U. S. Patent 977,713 filed March 29, 1901) proposed another form of continuous loading. Recent general discussions of loaded telegraph cable problems have been given by Malcolm (Theory of Submarine Telegraph and Telephone Cable, London, 1917) and by K. W. Wagner (Elektr. Nachtr. Tech., Oct., 1924).

for a transoceanic telegraph cable and it is primarily with regard to continuous loading that the following discussion is concerned.

Effects of Loading

Most of the proposals to load telegraph cables have had the object of reducing or eliminating distortion, and accordingly most of the mathematical treatments of loading have been from that point of view. The reduction of distortion is, however, not the only benefit to be obtained from loading and, in fact, may not always be secured in the high speed operation of a loaded cable. The principal benefit of loading from the practical standpoint is to decrease the attenuation of the signals so that for a given frequency more current will be received or so that the minimum permissible current may be received with a greater speed of signalling. From the mathematical standpoint there are two ways of treating the problem of the loaded cable, first with regard to the transmission of a transient impulse, and second with regard to setting up steady alternating currents of definite frequency. In the ultimate analysis the solution of either problem can be got from the other. However, for practical purposes they are two distinct means of attack. Which should be used depends on the object to be secured. If one is concerned primarily with the effect of the cable on the wave shape of the signal transmitted over it, it is fairly obvious that the transient treatment has advantages. If, however, one is concerned only with the strength of the received signal, as is the case if there is assurance that the signal shape can in any event be corrected by terminal networks, then the steady state treatment is sufficient and much more convenient to apply. In the case of the real loaded cable the complete transient solution is extremely complex and the steady state treatment relatively simple. The solution of the transient problem of an ideal loaded cable is, however, very valuable to give a physical picture of how inductive loading aids the high speed transmission of signals.

The transient solution of the problem of an ideal heavily loaded cable has been worked out by Malcolm [4] and more rigorously by Carson [5], who have determined the curve showing the change of current with time at one end of the cable if a steady e.m.f. is applied at zero time between the cable and earth at the distant end. Such a curve is called an "arrival curve" and for an ideal loaded cable comprising only constant distributed resistance, capacity and inductance may have a form like that shown in Curve b of Fig. 3, which is to be

[4] Theory of the Submarine Telegraph and Telephone Cable, London, 1917.
[5] Trans. A. I. E. E., Vol. 38, p. 345, 1919.

compared with Curve a which is the arrival curve of a non-loaded cable. The straight vertical part of Curve b represents the "head" of the signal wave which has travelled over the cable at a definite speed and with diminishing amplitude. The definite head of the arrival curve is the most striking characteristic difference between the ideal loaded and the non-loaded cable. In the latter, as is evident from Fig. 3, the current at the receiving end starts to rise slowly almost as soon as the key is closed at the transmitting end. When an e.m.f. is applied to the sending end of the non-loaded cable a charge spreads out rapidly over the whole length, the receiving end charging up much more slowly than the sending end on account of the resistance of the intervening conductor. Hence, if a signal train consisting of rapidly alternating positive and negative impulses is applied to the sending end, the effect at the receiving end of charging the cable positively is wiped out by the succeeding negative charge before there has been time to build up a considerable positive potential and the successive alternating impulses thus tend to annul each other. In the loaded cable the effect of inductance is to oppose the setting up of a current and to maintain it once it has been established, and thus to maintain a definite wave front as the signal impulse travels over the cable. Hence, with inductive loading the strength and individuality of the signal impulses are retained and a much higher speed of signalling is possible. It should be noted that by speed of signalling is meant the rapidity with which successive impulses are sent and not the rate at which they travel over the cable. This speed of travel is actually decreased by the addition of inductance, about one-third of a second being required for an impulse to traverse the New York-Azores cable from end to end.

It should be noted that Curve b of Fig. 3 is for an ideal loaded cable in which the factors of resistance, capacity and inductance are constant. In a real loaded cable none of these factors are constant and the arrival curve cannot be simply and accurately computed. Even the capacity which is usually assumed as constant for real cables varies appreciably with frequencies in the telegraph range, and owing to the fact that gutta percha is not a perfect dielectric material, its conductance, which is also variable with frequency, must be taken into account. Although the inductance of the cable is substantially constant for small currents of low frequency, it is greater for the high currents at the sending end of the cable on account of the increase of magnetic permeability of the loading material with field strength and is less at high frequencies than at low on account of the shielding effect due to eddy currents. The resistance is highly variable since it comprises, in addition to the resistance of the copper conductor, effective resistance due to eddy currents and hysteresis in the loading material, both of which vary with frequency and current amplitude. Furthermore, there is variable inductance and resistance in the return circuit outside the insulated conductor which must be

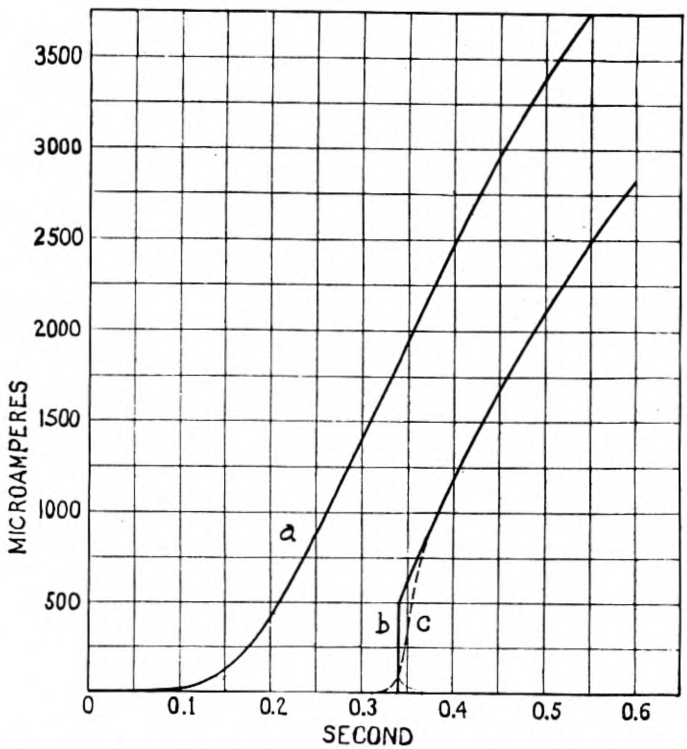


Fig. 3—Arrival Curves. a. Non-loaded cable. b. Ideal loaded cable. c. Real loaded cable (approximate)

taken into account. Although it is very difficult to compute the exact arrival curve of a cable subject to all of these variable factors, an approximate calculation in a specific case like that of the New York-Azores cable shows that the arrival curve has the general shape of Curve c of Fig. 3. It will be noticed that although this arrival curve lacks the sharp definite head, characteristic of the ideal loaded cable, it still has a relatively sharp rise and that the time required for the impulse to traverse the cable is not greatly different from that of the ideal loaded cable.

Although it is difficult to take exact account of the variable characteristics of the loaded cable in the solution of the transient problem, it is easy to take account of them in the steady state or periodic analysis by means of well-known methods. If a steady sinusoidal voltage, V_s, is applied at one end of the cable the resulting voltage, V_r, at the distant end will be given by the equation

$$V_r = k\,V_s\,\epsilon^{-Pl},$$

where l is the length, P, the propagation constant of the cable and k, a constant which depends on the terminal impedance and which is unity in case the cable is terminated at the receiving end in its so-called characteristic impedance. The propagation constant is given by the formula,

$$P = \sqrt{(R+ipL)(G+ipC)} = \alpha + i\beta,$$

where R is the resistance, L, the inductance, G, the leakance and C, the capacity per unit length and p is 2π times the frequency. The real part of the propagation constant, α, is called the attenuation constant and the imaginary part, β, the wave length constant. By separating α and β the amplitude and phase displacement of the received voltage relative to the sent voltage may be computed for any particular frequency and the behavior of a complex signal train may be worked out by analyzing it into its Fourier components and treating them separately. The phase shift is, however, of importance mainly as regards the shape of the received signals and their amplitude may, in general, be obtained from the attenuation constant alone. Thus if it is known that the signal shape can in any case be corrected by terminal networks there is no need to be concerned with more than the attenuation constant to compute the speed of the cable.

In the case of a cable of the permalloy loaded type, α is given with an approximation [6] sufficiently close for the purposes of this discussion by the equation,

$$\alpha = \frac{1}{2}\sqrt{\frac{C}{L}}\left(R + \frac{G}{C}L\right).$$

For the purpose of computing R it is convenient to separate it into its components, giving

$$\alpha = \frac{1}{2}\sqrt{\frac{C}{L}}\left(R_c + R_e + R_s + R_h + \frac{G}{C}L\right),$$

[6] For accurate computation of attenuation the complete formula for α must be used.

where R_c = copper resistance per unit length

R_e = eddy current resistance per unit length

R_s = sea return resistance per unit length

R_h = hysteresis resistance per unit length

The copper resistance R_c is that determined by a direct current measurement of the loaded conductor since the resistance of the loading tape is so high and its length is so great that the current flowing longitudinally through it may be safely neglected.

The eddy current resistance R_e is given approximately by the formula,

$$R_e = \frac{m\mu^2 t^3 f^2}{\rho\,(d-t)},$$

where t is the thickness or diameter of the loading tape or wire, d, the outside diameter of the loaded conductor, f, the frequency, ρ, the resistivity of the loading material, μ, its magnetic permeability and m, a constant which depends on the form of the loading material and is in general greater for tape than for wire loading. Although it is possible to compute a value of m, the value found in practice is always larger than the theoretical value which is necessarily based on simple assumptions and does not take into account such a factor as variation of permeability through the cross-section or length of the loading material. Accordingly it is necessary to determine m experimentally for any particular type of loaded conductor.

The sea-return resistance may be safely neglected in the computation of slow speed non-loaded cables, but it is a factor of great consequence in the behavior of a loaded cable. By sea-return resistance is meant the resistance of the return circuit including the effect of the armor wire and sea water surrounding the core of the cable. Although the exact calculation [7] of this resistance factor is too complex to be discussed here, the need for taking it into account may be quite simply explained. Since the cable has a ground return, current must flow outside the core in the same amount as in the conductor. The distribution of the return current is, however, dependent on the structure of the cable as well as on the frequencies involved in signalling. If a direct current is sent through a long cable with the earth as return conductor the return current spreads out through such a great volume of earth and sea water that the resistance of the return path is negligible. On the other hand if an alternating current is sent through the cable the return current tends to concentrate

[7] See Carson and Gilbert, *Jour. Franklin Inst.*, Vol. 192, p. 705, 1921; *Electrician*, Vol. 88, p. 499, 1922; *B. S. T. J.*, Vol. I, No. 1, p. 88.

around it, the degree of concentration increasing with the frequency. With the return current thus concentrated the resistance of the sea water is of considerable consequence. It is further augmented by a resistance factor contributed by the cable sheath. This may be better understood by considering the cable as a transformer of which the conductor is the primary and the armor wire and sea water are each closed secondary circuits. Obviously the resistances of the secondary circuits of armor wire and sea water enter into the primary circuit and hence serve to increase the attenuation. The presence of the armor wires may thus be an actual detriment to the transmission of signals.

To take account of the hysteresis resistance, R_h, and also of the increased inductance and eddy current resistance at the sending end of the cable it is most convenient to compute the attenuation of the cable for currents so small that R_h may be safely neglected. The attenuation thus computed is that which would be obtained over the whole cable if a very small sending voltage were used. The additional attenuation at the sending end for the desired sending voltage may then be approximated by computing successively from the sending end the attenuation of short lengths of cable over which the current amplitude may be considered constant, the attenuations of separate lengths being added together to give the attenuation of that part of the cable in which hysteresis cannot be neglected. In this computation account must, of course, be taken of the increased inductance and eddy current resistance accompanying the higher currents at the sending end.

Having calculated or obtained by measurement the several resistance factors and knowing the capacity, leakance and inductance, the whole attenuation of a cable for any desired frequency may be computed and a curve drawn showing the variation of received current with frequency for a given sending voltage. This relation for a particular case is shown in Curve c of Fig. 4. Curve a shows for comparison the relation between frequency and received current of a non-loaded cable of the same size, that is, a cable having a conductor diameter the same as that of the loaded conductor and having the same weight of gutta percha. Curve b shows the behavior of an ideal loaded cable having the same inductance, capacity and d.c. resistance as the real loaded cable of Curve c, but in which the leakance and alternating current increments of resistance are assumed to be zero.

Now, if the level of interference through which the current must be received is known, the maximum speed of signalling for the loaded cable may be obtained from Curve c. It is that speed at which the highest frequency necessary to make the signals legible is received with sufficient amplitude to safely override the superposed interference. Just what the relation of that frequency is to the speed of signalling cannot be definitely stated, since it depends on the method of operation and code employed as well as on the desired perfection

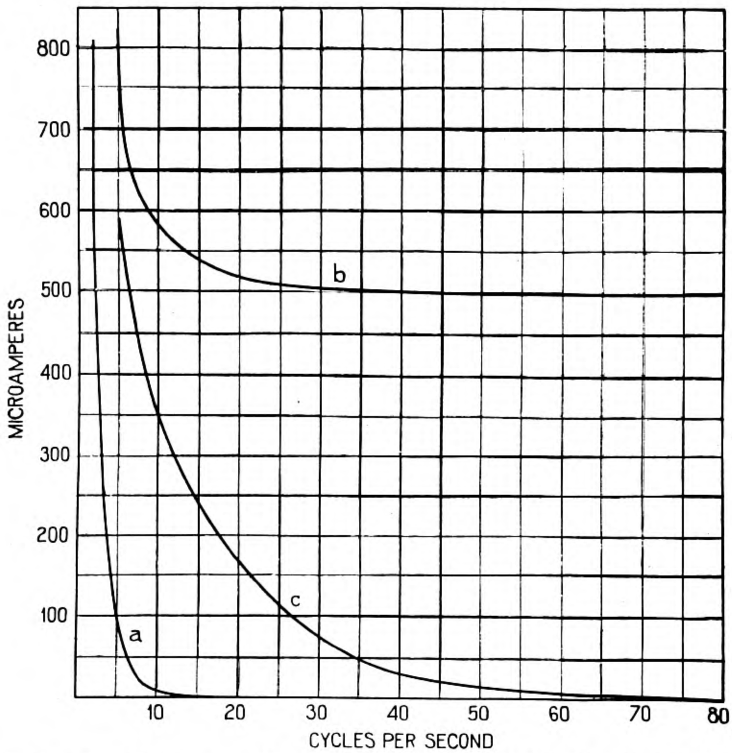


Fig. 4—Received Current vs. Frequency. a. Non-loaded cable. b. Ideal loaded cable. c. Real loaded cable

of signal shape. J. W. Milnor [8] has suggested that for cable code operation and siphon recorder reception a fair value is about 1.5 times the fundamental frequency of the signals, that is, the fundamental frequency when a series of alternate dots and dashes is being sent.

Remarks on the Design of Loaded Cables

By referring again to the equation for α, above, it can now be explained why high permeability is a necessary characteristic of the

[8] *Journal A. I. E. E.*, Vol. 41, p. 118, 1922. Transactions A. I. E. E., Vol. 41, p. 20, 1922.

loading material if a benefit is to be obtained from continuous loading. The addition of the loading material has two oppositely directed effects; on the one hand it tends to improve transmission by increasing the inductance and consequently decreasing the attenuation, and on the other hand it tends to increase the attenuation by increasing the effect of leakance and by the addition of resistance. Not only are the hysteresis and eddy-current factors of resistance added by the loading material but it must also be looked upon as increasing either the copper resistance or the capacity on account of the space it occupies. Generally it is more convenient to look upon the loading material as replacing some of the copper conductor in the non-loaded cable with which comparison is made, since by so doing all of the factors outside of the loaded conductor are unchanged. Now, if the loading material is to be of any benefit, the decrease in attenuation due to added inductance must more than offset the increase due to added resistance, including the added copper resistance due to the substitution of loading material for copper. In the limiting case the lowest permeability material which will show a theoretical advantage from this point of view is that which, as applied in a vanishingly thin layer, gives more gain than loss. For any particular size and length of cable there is a limiting value of permeability which will satisfy this condition, this limiting value being greater the longer the cable and the smaller the diameter of its conductor.[9] For transatlantic cables of sizes laid prior to 1923 the minimum initial permeability required to show an advantage is higher than that of any material known prior to the invention of permalloy. Actually a considerably higher permeability than this theoretical minimum was, of course, required to make loading an economic advantage since there are practical limits to the thickness of loading material and since the cost of applying it has also to be taken into account. Further, there are limits on methods of operation imposed by loading which necessitate still higher permeability to make loading worth while.

Since the addition of loading has two opposite tendencies in its effect on attenuation, the practical design of the cable must be based on a compromise between them. Thus, to secure the maximum gain from loading a cable of a given size, the loading material should be chosen of such a thickness that the gain due to increased inductance from a slight increase of thickness just offsets the loss due to increased resistance and dielectric leakance. In practice, of course, economic considerations of the cost of various thicknesses of loading must also be taken into account.

[9] See British Patent No. 184,774—1923, to O. E. Buckley.

In designing the New York-Azores cable some assumption had to be made as to the extraneous interference which would be encountered. Theoretical considerations led us to believe that the loaded cable would be no more subject to external interference than non-loaded cables. It even appeared that it would be less affected by some types of interference, for, owing to the shorter wave-length for a given frequency, a disturbance which affects a great many miles of cable simultaneously is less cumulative in its effect at the terminal of a loaded than a non-loaded cable. A reasonable assumption seemed to be that the total overall attenuation which could be tolerated for the loaded cable was at least as great as that which experience had shown to be permissible for simplex operation of non-loaded cables. This maximum permissible attenuation depends, of course, on conditions of terminal interference and no fixed value can be given as applicable to all cables. However, for average conditions of terminal interference in locations free from power line disturbances and where the cable lies in relatively deep water near to its terminal landing, a reasonable value of total attenuation constant for the fundamental frequency of cable code is about 10 (86.9 T.U.) for recorder operation and about 9 (78.2 T.U.) for relay operation. These were the approximate values assumed for the New York-Azores cable and later experience has demonstrated that they were well justified.

Distortion in Loaded Cables

Throughout all of the preceding discussion it has been assumed that the relation between attenuation and terminal interference would limit the speed of simplex operation rather than that distortion of signal shape would be the limiting factor. Although this is, in fact,[10] the case with non-loaded cables it was not self-evident as regards the loaded cable, and to make reasonably certain that the speed could be determined from the attenuation-frequency relation required a demonstration that the signal distortion of a real loaded cable could be corrected by suitable terminal apparatus. One of the merits long claimed for loading was that it would reduce distortion and, indeed, an ideal loaded cable with constant inductance and without magnetic hysteresis, eddy current loss, dielectric leakance and sea return resistance would have very little distortion and would give a speed limited only by terminal apparatus. However,

[10] Recent work of J. R. Carson (U. S. Patent 1,315,539—1919) and R. C. Mathes (U. S. Patent 1,311,283—1919) has shown that with the combined use of vacuum tube amplifiers and distortion correcting networks, distortion in non-loaded cables can be compensated to any desired degree.

a real loaded cable, the inductance of which varies with both current and frequency and in which all the above noted resistance factors are present, may give, and in general will give when operated at its maximum speed, greater distortion of signals than a non-loaded cable.

To solve the question of distortion on a purely theoretical basis required consideration of the transmission of a transient over the loaded cable. This was made extremely difficult by the existence of numerous possible causes of signal distortion, the effects of which could only be approximated in the solution of the transient problem. In addition to the distortion resulting from the rapid increase of attenuation with frequency due to the various sources of alternating current losses, distortion peculiar to the magnetic characteristics of the loading material had also to be taken into account. There are several types of magnetic distortion to be concerned about. First, there is the production of harmonics as a result of the non-linear magnetization curve of the loading material; second, there is a possible asymmetrical distortion due to hysteresis, and third, there is a possible modulation resulting from the superposition of signals on each other, that is, in effect, a modulation of the head of the wave of one impulse by the tail of the wave of a preceding impulse. The first two of these are effective at the sending end of the cable and the third near the receiving end.

A computation of distortion, including the peculiar magnetic effects, by a steady state a.c. method based on measurements of short loaded conductors indicated that the cable should operate satisfactorily with ordinary sending voltages. Further evidence that none of these various types of distortion would be of serious consequence and that the distortion of a loaded cable could be corrected by terminal apparatus, was obtained by experiments with an artificial line constructed to simulate closely, with regard to electrical characteristics, the type of loaded conductor with which we were then experimenting. This artificial line was loaded with iron dust core coils which served the purpose admirably, not only as regards inductance and alternating current resistance but also as regards magnetic distortion. Iron dust is, of course, very different in its magnetic characteristics from permalloy. However, owing to the large number of turns on a coil, it is operated at much higher field strengths and on a part of the magnetization curve corresponding approximately to that at which permalloy is operated on the cable. The case for magnetic distortion was in fact a little worse with the artificial line than with the then proposed cable. Fig. 5 shows a photograph of the artificial line, the coils of which are in the large iron pots and the resistance and paper condenser capacity units of which are in the steel cases. This line was equivalent to a 1,700 nautical mile cable loaded with 30 millihenries per n.m. and over it legible

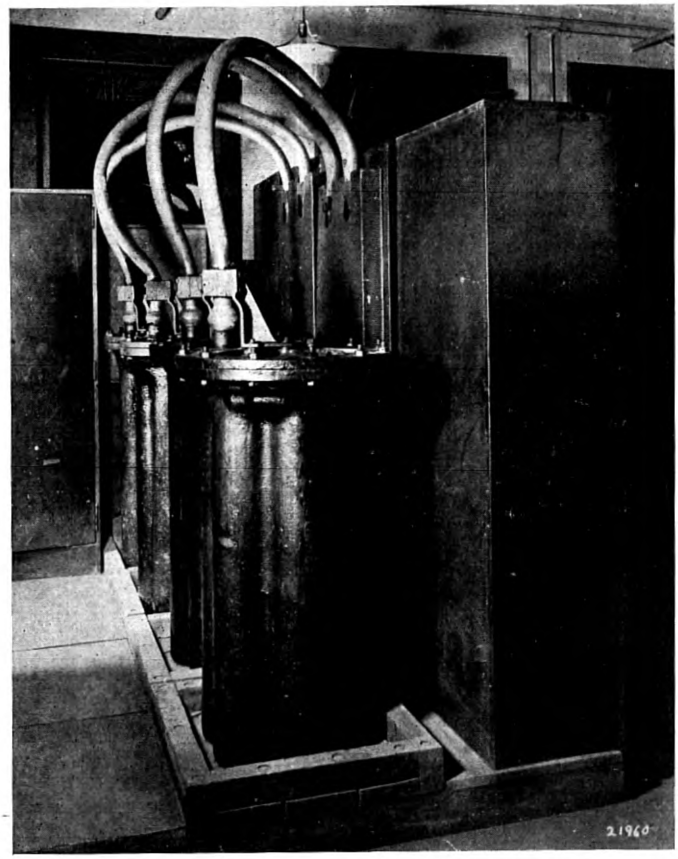


Fig. 5—Loaded Artificial Line

signals were secured at speeds up to more than 2,600 letters per minute. Such a speed of operation was quite beyond the range of the then available telegraph instruments, and accordingly special transmitting and receiving instruments were required. The multiplex distributor of the Western Electric printing telegraph system proved an excellent transmitter for experimental purposes and, for receiving, use was made of a combined vacuum tube amplifier and signal shaping network, the signals being recorded on a string oscillograph. Fig. 6 shows part of a test message received over the loaded artificial cable at a speed of 2,240 letters per minute.

The results of the tests with the artificial loaded cable were entirely in agreement with our calculations and showed that it was

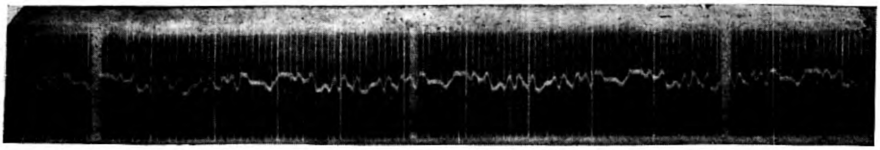


Fig. 6—Test Message. Signals received April 16, 1920, over coil-loaded artificial line equivalent to a 1700 n.m. cable with 30 m.h./n.m. Speed 2240 letters per minute

possible to obtain satisfactory signal shape with a coil-loaded cable having alternating current resistance and distortion factors approximating those of the permalloy-loaded cable. The exact behavior of the proposed cable, including such factors as sea-return resistance and a somewhat variable distributed inductance, could not, of course, be duplicated without prohibitive expense. The approximation was considered, however, to be sufficiently good to justify proceeding with a loaded cable installation so far as questions of signal shaping were concerned. It is interesting to note that the factor which limited the operating speed of the artificial loaded cable was one which is not present in a continuously loaded cable but which would possibly be a serious factor in the operation of a coil loaded cable, namely the oscillations [11] resulting from the finite size and separation of the inductance units.

Operation of Loaded Cables

With the completion of the artificial loaded cable tests there was still one principal question of transmission which had to remain unanswered until a cable had been installed. This was the question of balancing the cable for duplex operation. Ordinary submarine cables are generally operated duplex, the total speed in the two directions being usually from about 1.3 to 2 times the maximum simplex or one-way speed. Except in cases where the external interference is very bad, the limiting speed of duplex operation is determined by the accuracy with which an artificial line can be made the electrical equivalent of the cable. Ordinarily the artificial line is

[11] Carson, Trans. A. I. E. E., Vol. 38, p. 345, 1919.

made up only of units of resistance and capacity arranged to approximate the distributed resistance and capacity of the cable. Sometimes inductance units are added to balance the small inductance which even a non-loaded cable has. In the actual operation of cables, artificial lines are adjusted with the greatest care and a remarkable precision of balance is obtained. This is necessary because of the great difference in current amplitude of the outgoing and incoming signals, the former being of the order of 10,000 times the latter. It is quite obvious that it will be much more difficult to secure duplex operation with a loaded than with an ordinary cable, since not only do the copper resistance and the dielectric capacity have to be balanced, but the artificial line must also be provided with inductance and alternating current resistance. Also the sea-return resistance and inductance which vary with frequency must be balanced.

In view of these difficulties it will probably be impossible to get as great a proportionate gain from duplex operation of loaded cables as is secured with ordinary cables. However, it is quite evident that it will be possible to secure duplex operation at some speed, since, with loaded as with non-loaded cables, the ratio of received to sent current increases rapidly as the speed is reduced and on this account it is much easier to duplex the cable at low speeds than at high. To make duplexing worth while on a cable with approximately equal traffic loads in both directions it is in general only necessary to get a one-way duplex speed half as great as the simplex speed. In fact in some cases the operating advantages of duplex would warrant even a slower duplex speed. On the other hand, there are cables on which the traffic is largely undirectional through most of the day and which would accordingly require a one-way duplex speed somewhat higher than half the simplex speed to justify duplex operation. Whether a sufficiently great speed of duplexing could be secured to justify designing a cable on the basis of duplex operation could not be judged in advance of laying the first cable, and accordingly it was decided to engineer that cable on the basis of simplex operation.

Although it was expected that the new cable might at first have to be operated simplex it should not be supposed that any great difficulty or loss of operating efficiency was anticipated on this account. The speed of the New York-Azores cable is so great that to realize its full commercial advantage practically requires working it on a multi-channel basis as, for example, with a Baudot code, multiplex system, similar to that used on land lines. Such a system may be conveniently adapted to automatic direction reversal and with this modification most of the common objections to simplex operation are removed. Indeed, simplex operation may in this case possess a real advantage over duplex from the commercial point of view since it permits dividing the carrying capacity of the cable most efficiently to handle the excess of traffic in one direction.

Although means have been made available for making efficient use of the loaded cable it should be recognized that the method of operation best suited to satisfy commercial demands must be determined from future experience with cables of the new type. This is especially true with regard to relatively short cables. The discussion of the loaded cable problem in this paper has been confined wholly to the realm of long ocean cables where the limitations of the cable rather than terminal equipment or operating requirements determine the best design. This is the simplest case and the one which at present seems to show the greatest gain from loading. Where traffic requirements are limited and where there is no prospect of ever requiring higher speed than can be obtained with a non-loaded cable of reasonable weight, the advantage of loading is less and becomes smaller as the weight of non-loaded cable which will accomplish the desired result decreases. It should not be concluded, however, that loading will not find important application to short cables. Many short cables are parts of great systems and must be worked in conjunction with long cables. In such cases it may pay to load short sections where otherwise loading would not be justified. Permalloy loading also offers great possibilities for multiple-channel carrier-telegraph operation on both long and short cables and with this type of operation in prospect it is too early, now, to suggest limits to the future applications of permalloy to cables or to predict what will be its ultimate effect on transoceanic communication.

Useful Numerical Constants of Speech and Hearing

By HARVEY FLETCHER

NOTE: The material given in this paper was prepared in a more condensed form for publication in the International Critical Tables. In order to make it available in convenient form for the use of telephone engineers it was deemed advisable to publish it in this journal. The author is indebted to Dr. J. C. Steinberg for able assistance in collecting and arranging the material.

I. Bibliography

A BIBLIOGRAPH of papers on Pitch Discrimination, Intensity Discrimination, Absolute Sensitivity of the Ear, Upper Limit of Audibility, Lower Limit of Audibility, Theories of Hearing and other miscellaneous works on Speech and Hearing are given in a paper by H. Fletcher, *Bell Tech. Jour.*, Vol. II, 4, pp. 178–180, Oct., 1923.

II. Absolute Sensitivity of the Ear

The sensitivity is the minimum audible rms pressure in dynes cm^{-2} in ear canal. The values below are the average of the results of Wien (*Arch. f. ges. Physiol.* 97, p. 1, 1903), Fletcher and Wegel (*Phys. Rev.*, 19, p. 553, June, 1922), and Kranz (*Phys. Rev.*, 21, p. 573, May, 1923) weighted 3, 72, and 14, respectively according to number of ears tested

TABLE I

Frequency (dv)[1]	64	128	256	512	1024	2048	4096
Sensitivity (dynes)	.12	.021	.0039	.001	.00052	.00041	.00042

III. Minimum Audible Power for a Normal Ear

The power in microwatts passing through each square centimeter in the wave front of a free progressive wave in air under average conditions is related to the rms pressure in dynes by the formula

$$p=20.5\sqrt{J}.$$

The figures of Table I may be converted by this formula to minimum audible powers. It is thus seen that the minimum audible accoustical power is at frequencies between 2,000 and 4,000 vibrations per second and is equal to 4×10^{-10} microwatts per square centimeter

[1] The symbol dv is used to denote "double" or complete vibrations.

IV. Range of Audition in Frequency and Intensity

In Fig. 1 the lower curve is a plot of the average sensitivity values given in Table I. The upper curve gives the pressures that produce a sensation of feeling and serves as a practical limit to the range of auditory sensation. (Wegel, *Bell Tech. Jour.*, 1, p. 56,

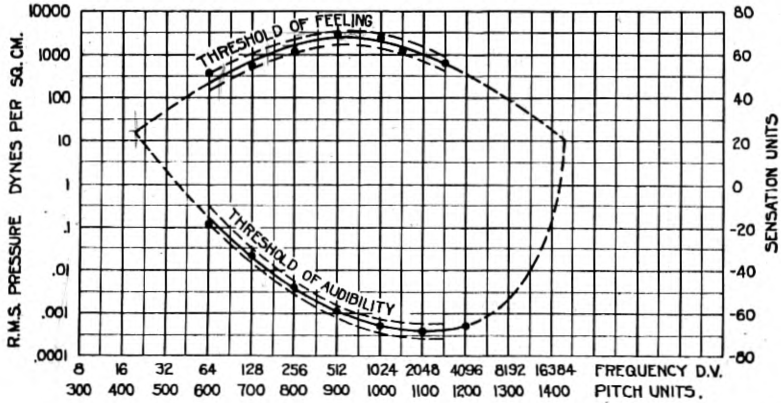


Fig. 1

November, 1922.) Investigators vary from about 8 to 40 dv for the lower pitch limit and from about 12,000 to 35,000 dv for the upper limit. (See I.) The values of 20 and 20,000 dv shown on the chart were taken as being most representative. Half of the observations lie within the dotted curves. The pitch is equal to 100 $\log_2 N$ and the sensation units equal to 20 $\log P$ where N is the frequency and P is the pressure. (Fletcher, *Jour. Frank. Inst.*, 194,

V. Minimum Perceptible Increase in Intensity and Frequency

(Knudsen, *Phys. Rev.* 21, p. 84, Jan., 1923)

Sensation Level in Sensation Units or TU's	Per Cent Increase in Intensity to be Just Perceptible
10	23
20	14
30	12
40	11
50	10.6
60 to 100	10

Frequency	Per Cent Increase in Frequency to be Just Perceptible
64	.93
128	.59
256	.40
512	.32
768 to 4096	.30

p. 289, Sept., 1923.) The sensation level S of a sound is defined by $S = 20 \log \frac{P}{P_o}$ where P_o is the threshold pressure, or it is the number of sensation units above the threshold of audibility. These sensation units are the same as the transmission units used in telephone engineering.

The per cent increase in frequency to be just perceptible varies with sensation level in about the same way as does the per cent increase in intensity to be just perceptible. The values are for monaural reception the tones being heard successively.

VI. The Number of Double Vibrations Necessary to Determine Pitch

(Bode, *Psychol. Stud.*, 2, p. 293, 1907)

TABLE II

	Weak Tones		Medium Tones	
Freq. dv	Time (sec.)	No. of dv	Time (sec.)	No. of dv
128	0.0496	12.1		
256			0.06908	17.6
384	.0672	24.08	0.0445	17.1
512	.0579	29.64	0.04274	21.8

VII. The Masking Effect of One Sound Upon the Audibility of Another Sound

(Wegel and Lane, *Phys. Rev.*, 23, p. 266, Feb., 1924)

If the ear is stimulated by a pure tone of frequency N_1, it is in general rendered less sensitive to other pure tones. The tone that constantly stimulates the ear is called the masking tone. The tone that is heard in the presence of this stimulating tone is called the masked tone. The masking is measured in sensation units or TU's. It is equal to $20 \times \log_{10}$ of the ratio of the pressures necessary to perceive the masked tone with and without the presence of the masking tone. In other words it is equal to the number of units that the threshold has been shifted. Fig. 2 shows the amount of masking (ordinate) of tones of various frequencies as a function of the sensa-

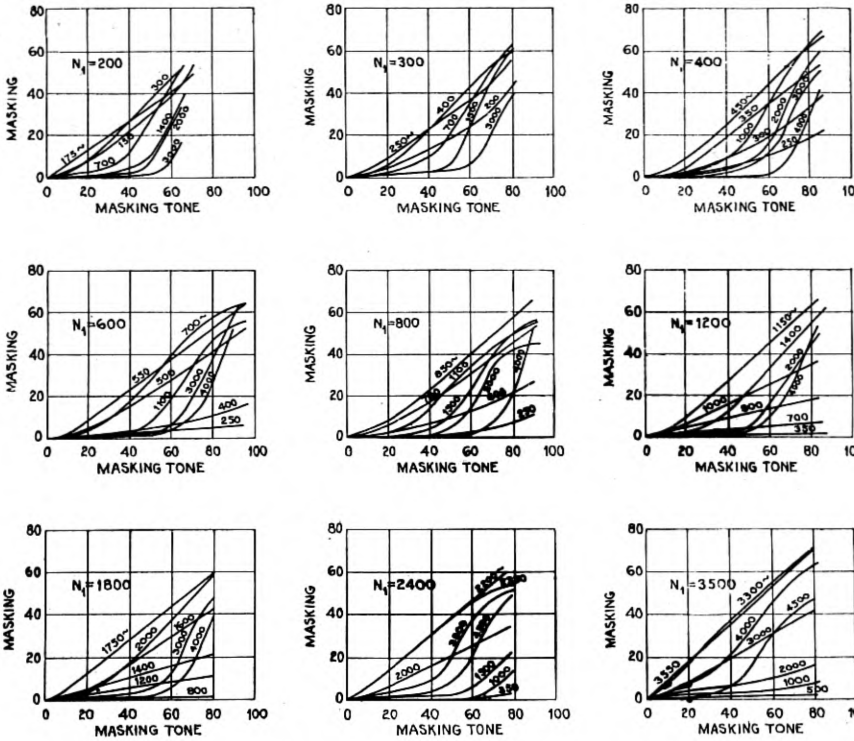


Fig. 2—Masking for Tones in Same Ear

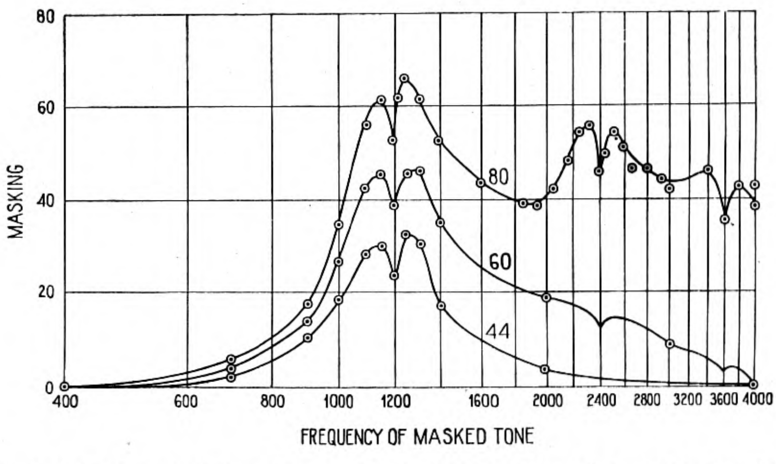


Fig. 3—Masking of Various Frequencies by 1,200 Cycles at Sensation Levels of 80, 60, and 44 Units, Respectively

tion level (abscissa) and frequency N_1 of the masking tone. In Fig. 3 data for a masking tone of 1,200 dv is plotted in which the frequencies of the masked tones are plotted on the abscissa. In order to get satisfactory curves of this kind it is necessary to take more

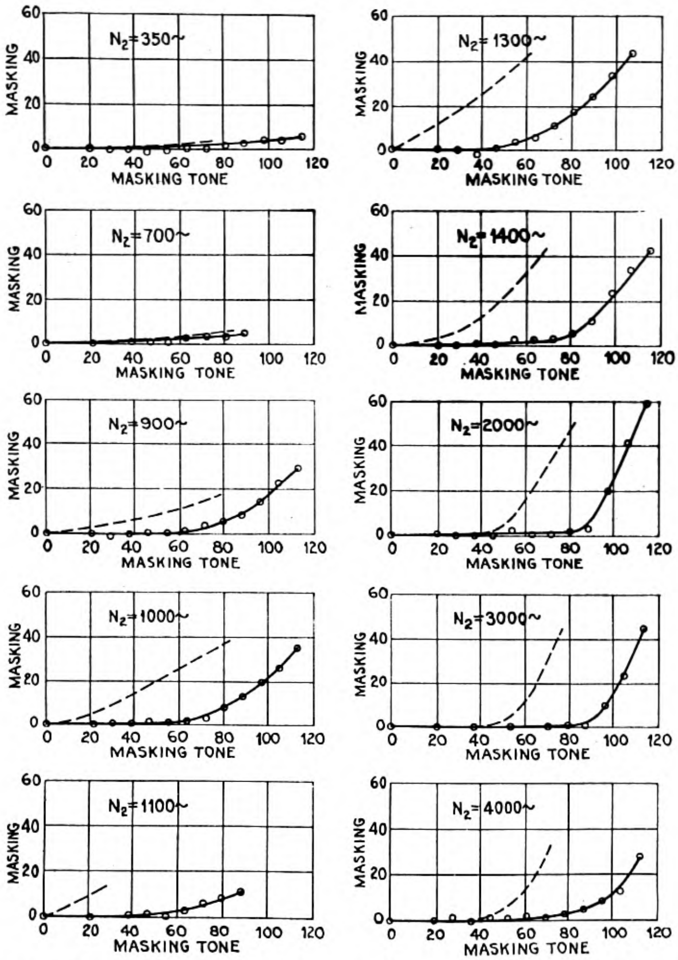


Fig. 4—Masking Data. Tones in Opposite Ears. Masking Tone 1,200 Cycles

comprehensive data than that shown in Fig. 2. The solid curves of Fig. 4 show the masking when the masked and masking tones are introduced into opposite ears. The dotted curves were taken from Fig. 2.

VIII. Conduction of Skull Between the Two Ears

A comparison of the two curves in Fig. 4 shows that the attenuation introduced by the skull from one ear to the other when the tone is introduced by a telephone receiver is between 40 and 50 sensation units corresponding to an intensity ratio of from 10^4 to 10^5. This becomes 7 TU greater when rubber caps are interposed between the head and the receiver cap.

IX. Localization of Pure Tones as a Function of the Phase Difference at the Two Ears

(G. W. Stewart, *Phys. Rev.*, 25, p. 425, May, 1920)

The experimental results can be represented by the formula

$$\frac{\Phi}{\Theta}=0.0034N+.8 \text{ (approx.)}$$

Φ is the phase difference in degrees of the tones at the two ears.

Θ is the number of degrees to the right or left of the median plane that an observer locates the source of sound. The direction of location is toward the ear leading in phase.

N is the frequency of the tone in dv. The relation applies only for frequencies of 100 to 1,000 dv., inclusive.

X. Constants Used in the Computation of the Loudness of a Complex Sound

(Fletcher and Steinberg, *Phys. Rev.*, 24, p. 306, Sept., 1924)
(Steinberg, *Phys. Rev.* To be published soon)

If L be the loudness as judged by an average normal ear, then

$$L=3.33 \log_{10}\left[\sum_{n=1}^{n=k}(W_n p_n)^{\frac{2}{r}}\right]^{\frac{r}{2}}$$

where

p_n = rms pressure of the n^{th} component,

W_n = a weight factor for the n^{th} component (Fig. 5)

r = a root factor (Fig. 5)

The sensation levels (See IV) given in the chart are for the complex tone.

XI. Dynamical Constants of the Hearing Mechanism

(Howell, W. H., "A Textbook of Physiology"

(Wrightson, Sir Thomas, "Analytical Mechanism of the Internal Ear")

(*a*) Ear Canal

Length, 2.1–2.6 cm.
Volume, 1 cm^3.
Area at Opening, .33 to .50 cm^2.

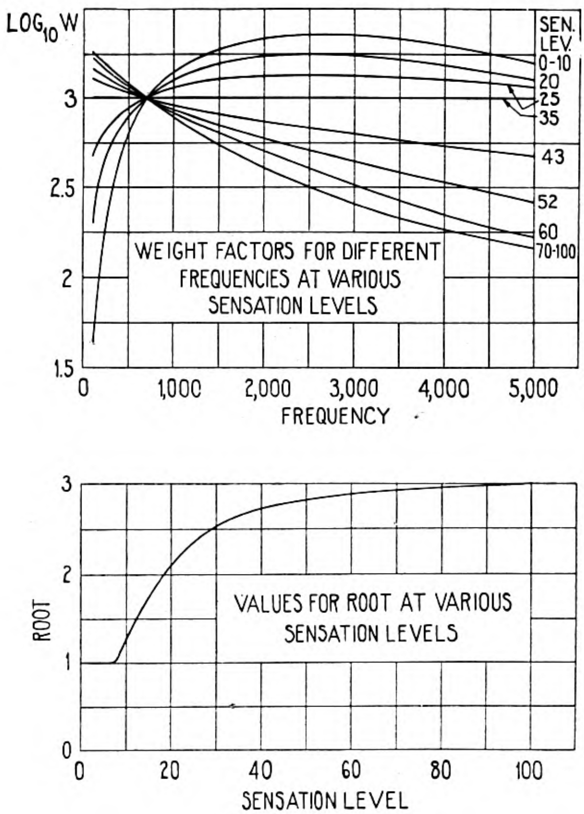


Fig. 5

(*b*) Drum

Vertical Diameter, .85 cm.
Horizontal Diameter, 1.00 cm.
Area, .65 cm^2.

(*c*) Hammer
Length, .8 to .9 cm.
Weight, 23 mg.

(*d*) Anvil
Weight, 25 mg.

(*e*) Stirrup
Weight, 3 mg.

(*f*) Mechanical Impedance of the Ear Drum
(Data by Wegel and Lane, Bell Telephone Laboratories)

The order of magnitude is 20 to 30 mechanical ohms (cgs units) over the frequency range from 200 to 4,000 dv.

XII. Speech Energy

A. Speech Power

(Data furnished by C. F. Sacia and L. J. Sivian, Bell Telephone Laboratories)

1. The average speech power delivered by an average speaker is about 10 microwatts. In the process of obtaining the average the silent intervals were included. If they are excluded the average increases about 50%. The peak power frequently rises to 2,000 microwatts.

2. Variation of average speech power delivered by different persons during conversation. (Fig. 6.)

B. Energy Frequency Distribution of Average Speech

(Crandall and MacKenzie, *Phys. Rev.*, 19, p. 221, March, 1922)
(Fig. 7)

C. Acoustic Power in Vowel Sounds

(Data furnished by C. F. Sacia of the Bell Telephone Laboratories.

This data together with a description of the apparatus and methods used in obtaining it will be given in a paper soon to be published.)

Table III contains data on the power of individual vowels obtained from analyzing the vowel portions of the syllables shown in the key-word. The first two columns give the average power in microwatts of 8 males and 8 females during the particular cycle of the fundamental containing the maximum energy for unaccented vowels. A rough estimate of the corresponding figures of typical accented

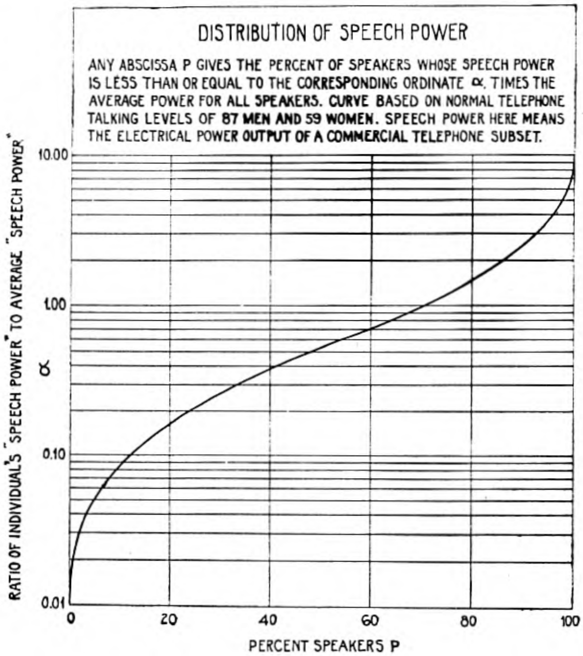


Fig. 6

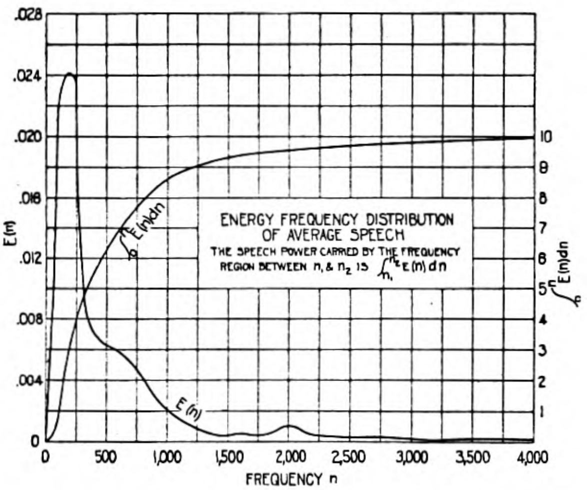


Fig. 7

vowels may be obtained by multiplying these values by a factor of 3. The third and fourth columns give peak factors which convert the power figures of the first two columns into maximum instantaneous powers. Columns 5 and 6 give the maximum values of these peak factors found among the male and female voices, respectively.

TABLE III

Acoustic Power in Microwatts of the Vowel Sounds

Vowel	Key	(1) P_m 8 males	(2) P_m 8 fem.	(3) Av. Peak Factor 8 males	(4) Av. Peak Factor 8 fem.	(5) Max. Peak Factor 8 males	(6) Max. Peak Factor 8 fem.
ū	tool	27	41	2.6	2.8	3.8	3.4
u	took	32	49	4.0	3.1	4.9	3.4
ō	tone	33	44	4.1	3.4	6.4	4.9
o′	talk	37	49	4.5	3.3	5.7	3.6
o	ton	29	38	4.6	3.9	6.8	5.7
a	top	50	48	4.2	3.6	4.2	4.7
a′	tap	43	39	5.4	4.7	7.4	5.2
e	ten	25	30	5.6	3.8	6.3	4.6
ā	tape	21	30	5.3	4.5	6.0	5.1
i	tip	25	31	4.1	3.8	5.8	5.7
ē	team	32	23	4.7	2.6	5.8	3.6

XIII. Frequency of Occurrence of English Speech Sounds

(Table IV contains data from a book by Godfrey Dewey, "The Relative Frequency of English Speech Sounds," Harvard University Press)

TABLE IV

Relative Frequency of Occurrence of English Speech Sounds

Speech Sound	Key	Rel. Freq.	Speech Sound	Key	Rel. Freq.
a	top	3.3	g		0.74
ā	tape	1.84	h		1.81
a′	tap	3.95	j		0.44
e	ten	3.44	k		2.71
ē	eat	2.12	l		3.74
er	term	0.63	m		2.78
i	tip	8.53	n		7.24
ī	dike	1.59	ng	hang	0.96
o	ton	6.33	p		2.04
ō	tone	1.63	r		6.88
o′	talk	1.35	s		4.55
u	took	0.71	sh	shell	0.87
ū	tool	1.89	th	(thin)	.37
ou	our	0.59	th	then	3.43
b		1.81	t		7.13
ch	chalk	0.52	v		2.28
d		4.31	w		2.08
f		1.84	y		0.60
			z		2.97

XIV. Interpretation of Speech

(Fletcher, H., *Jour. Frank. Inst.*, 193, 6, June, 1922)

A measure of the interpretation of speech was obtained by means of articulation tests. Meaningless syllables were pronounced and observers were required to record the syllables. The articulation is

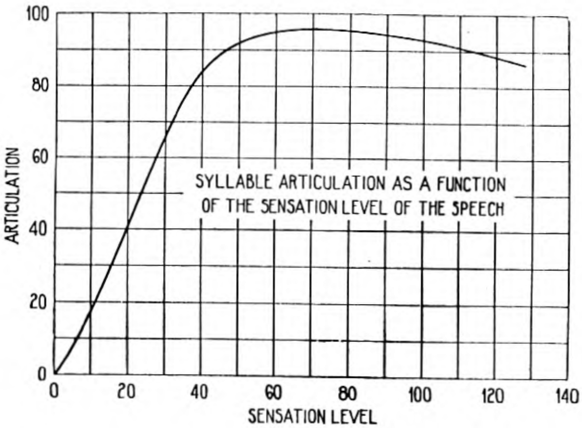


Fig. 8

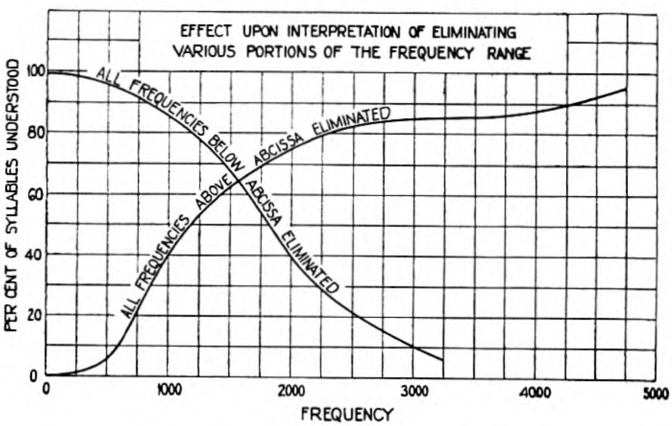


Fig. 9

the per cent of syllables that were correctly recorded. The articulation depends upon the sensation level of the speech (Fig. 8), and upon the width of the frequency band transmitted (Fig. 9).

The syllables that were recorded in these tests were analyzed to show the articulation of the fundamental speech sounds. Fig. 10

shows these articulations as functions of the sensation level of the speech. In Fig. 11 they are shown as functions of the width of the transmitted frequency band. It should be noted that the term articulation as here employed denotes only the correct interpretation of unrelated speech sounds and is not a measure of voice naturalness which is also an important factor in the telephonic transmission of speech.

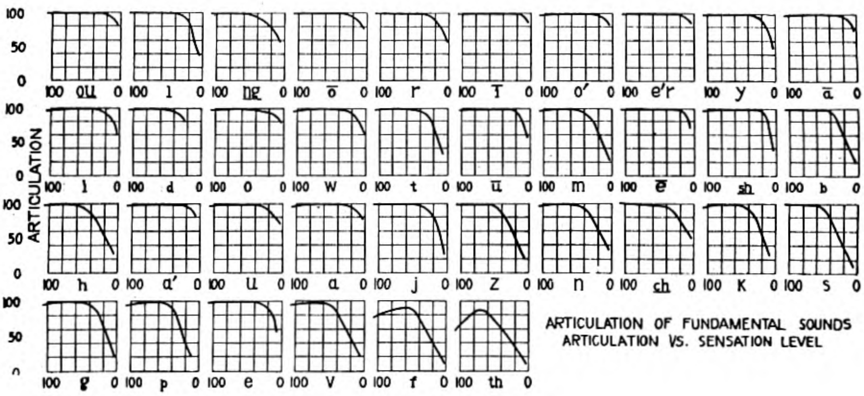


Fig. 10

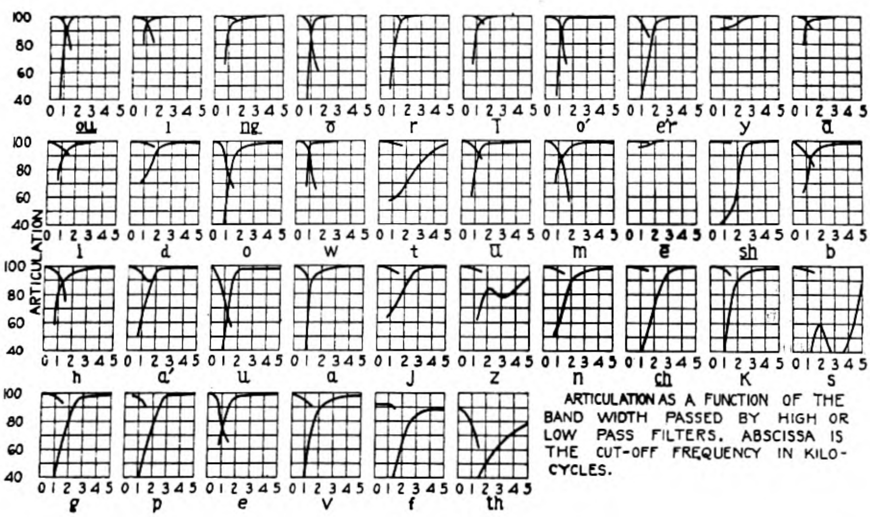


Fig. 11

Graphic Representation of the Impedance of Networks Containing Resistances and Two Reactances

By CHARLES W. CARTER, Jr.

ABSTRACT: The driving-point impedance of an electrical network composed of any number of resistances, arranged in any way, and two pure reactances, of any degree of complication within themselves but not related to each other by mutual reactance, inserted at any two points in the resistance network, is limited to an eccentric annular region in the complex plane which is determined by the resistance network alone.

The boundaries of this region are non-intersecting circles centered on the axis of reals. The diameter of the exterior boundary extends from the value of the impedance when both reactances are short-circuited to its value when both are open-circuited. The diameter of the interior boundary extends from the value of the impedance when one reactance is short-circuited and the other open-circuited to its value when the first reactance is open-circuited and the second short-circuited.

When either reactance is fixed and the other varies over its complete range, the locus of the driving-point impedance is a circle tangent to both boundaries. By means of this grid of intersecting circles the locus of the driving-point impedance may be shown over any frequency range or over any variation of elements of the reactances. This is most conveniently done on a doubly-sheeted surface.

The paper is illustrated by numerical examples.

INTRODUCTION

SUPPOSE that any number of resistances are combined into a network of any sort and provided with three pairs of terminals, numbered (1) to (3) as in Fig. 1. The problem set in this paper is to investigate the driving-point impedance[1] of such a network at

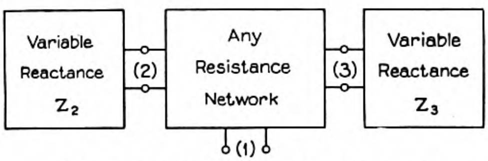


Fig. 1—The Network to be Discussed

terminals (1) when variable pure reactances, Z_2 and Z_3, are connected to terminals (2) and (3), respectively. Z_2 and Z_3 are formed of capacities, self and mutual inductances. They are not connected to each other by mutual reactance, but they may be of any degree of complication within themselves.

The problem is dealt with in terms of the complex plane: that is, the resistance components of the impedance, S, measured at terminals

[1] The driving-point impedance of a network is the ratio of an impressed electromotive force at a point in a branch of the network to the resulting current at the same point.

(1) are plotted as abscissas and the reactance components as ordinates. To every value of the impedance, then, there is a corresponding point, and to the values of the impedance over a range of variation of some element, or over a frequency range, there corresponds a locus, in the complex plane. This locus may be labelled at suitable points with the corresponding value of the variable. So labelled, it combines into one the curves which are usually plotted to show separately the variation of the reactance and resistance components or to show separately the variation of absolute value and angle.

The use of the complex plane is not new: it is the basis of most of the vector diagrams for electrical machinery. The characteristics of both smooth and loaded transmission lines have also been displayed by its means. Its application to electrical networks, however, is not common, and it is a subsidiary purpose of this paper to illustrate the fact that the properties of certain networks, which have complicated characteristics if exhibited in the usual way, may be shown quite simply in the complex plane. This simplicity, combined with generality, is attained by application of theorems concerning functions of a complex variable which are immediately available.

The Fundamental Equations

The impedance measured in branch 1 of any network is

$$S=R+iX=\frac{\Delta}{\Delta_{11}} \tag{1}$$

where Δ is the discriminant of the network, either in terms of branches or n independent meshes.[2]

Assigning the reactances Z_2 and Z_3 to meshes 2 and 3

$$\Delta=\begin{vmatrix} R_{11} & R_{12} & R_{13} & . & . & R_{1n} \\ R_{21} & R_{22}+Z_2 & R_{23} & . & . & R_{2n} \\ R_{31} & R_{32} & R_{33}+Z_3 & . & . & R_{3n} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ R_{n1} & R_{n2} & R_{n3} & . & . & R_{nn} \end{vmatrix} \tag{2}$$

where R_{jj} is the resistance in mesh j and $R_{jk}(=R_{kj})$ that common to meshes j and k.

[2] See: G. A. Campbell, Transactions of the A. I. E. E., 30, 1911, pages 873–909, for a complete discussion of the solution of networks by means of determinants.

Therefore $$S=\frac{A+A_{22}Z_2+A_{33}Z_3+A_{22\cdot 33}Z_2Z_3}{A_{11}+A_{11\cdot 22}Z_2+A_{11\cdot 33}Z_3+A_{11\cdot 22\cdot 33}Z_2Z_3} \tag{3}$$

where A is the discriminant of the resistance network alone and $A_{jj\cdot kk\cdot ll}$ denotes the cofactor of the product of the elements of A located at the intersections of rows j, k and l with columns j, k and l, respectively.

For convenience this is written as

$$S=\frac{a+bZ_2+cZ_3+dZ_2Z_3}{a_1+b_1Z_2+c_1Z_3+d_1Z_2Z_3}. \tag{4}$$

The constants of (3) and (4) are real and positive since they are cofactors of terms in the leading diagonal of the discriminant of a resistance network. The determinant being symmetrical, there is the following relation among them:

$$(ad_1-a_1d+bc_1-b_1c)^2=4(bd_1-b_1d)(ac_1-a_1c). \tag{5}$$

The function to be studied is, then, a rational function of two variables, having positive real coefficients determined by the resistances alone. Furthermore, if one reactance is kept constant while the other is varied, the function is bilinear. The particular property of the bilinear function, which has been studied in great detail, of interest here, is that by it circles are transformed into circles.[3]

When, as in this case, the variable in a bilinear function is a pure imaginary, the function may be rewritten in a form which gives directly the analytical data needed. For suppose

$$w=\frac{u+vz}{u_1+v_1z} \tag{6}$$

where z is a pure imaginary and the coefficients are complex. This is

$$w=\frac{v}{v_1}+\frac{u-u_1v/v_1}{u_1+v_1z}. \tag{7}$$

Multiplying the second term by a factor identically unity,

$$w=\frac{v}{v_1}+\frac{u-u_1v/v_1}{u_1+v_1z}\times\frac{v_1'(u_1+v_1z)+v_1(u_1'+v_1'z')}{u_1v_1'+u_1'v_1} \tag{8}$$

where primes indicate conjugates, or

$$w=\frac{uv_1'+u_1'v}{u_1v_1'+u_1'v_1}+\frac{uv_1-u_1v}{u_1v_1'+u_1'v_1}\times\left(\frac{u_1'+v_1'z'}{u_1+v_1z}\right). \tag{9}$$

[3] G. A. Campbell discusses, in the paper cited, the theorem that if a single element of any network be made to traverse any circle whatsoever, the driving-point impedance of the network will also describe a circle.

Now, as z is varied, the first term is constant. In the second term the first factor is constant and the second factor varies only in angle, since the numerator is the conjugate of the denominator. The first term, therefore, is the center, and the absolute value of the first factor of the second term is the radius, of the circle in which w moves as z takes all imaginary values.

One Variable Reactance Giving Circular Locus

The significance of the equations may be made apparent by a study of Fig. 2, which shows the impedance S when one of the reactances, say Z_3, is made zero. We have, then,

$$S=\frac{A+A_{22}Z_2}{A_{11}+A_{11\cdot 22}Z_2}=\frac{a+bZ_2}{a_1+b_1Z_2} \tag{10}$$

and the trivial case $ab_1-a_1b=0$ is excluded. This is of the type of (6). When Z_2 varies over all pure imaginary values, S traces out a

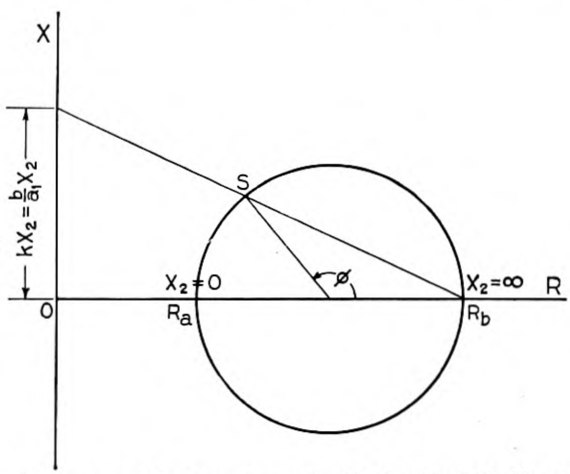


Fig. 2—Locus of the Impedance S with One Variable Reactance

circle, which (9) shows has its center on the resistance axis. Its intercepts on the resistance axis are

$$S=\frac{a}{a_1}=R_a,\text{ say, when } Z_2=0 \tag{11}$$

and

$$S=\frac{b}{b_1}=R_b,\text{ when } Z_2=\infty. \tag{12}$$

But in a symmetrical determinant

$$A_{11}A_{22}-A_{12}^2=AA_{11\cdot 22}; \tag{13}$$

therefore

$$ab_1<a_1b \tag{14}$$

or

$$\frac{a}{a_1}<\frac{b}{b_1} \tag{15}$$

whence

$$R_a<R_b. \tag{16}$$

To find the value of S when Z_2 has some value, say $Z_2=iX_2$, it 's only necessary to mark the circular locus with a scale in terms of Z_2. This may be done directly by using (9) to determine the angle, ϕ, which the radius of the circle makes when $Z_2=iX_2$. It is simpler to use the fact that a line passing through R_b and the point S has an intercept on the reactance axis of

$$X_0=\kappa X_2 \tag{17}$$

where $\kappa=b/a_1$.

The factor κ is determined by the resistances; therefore the scale, as well as the locus, is completely fixed by the resistances. Since κ is always positive, as X_2 is increased the circle is traversed in a clockwise sense; for positive values of X_2 the upper semi-circle is covered; for negative values, the lower. That is, when Z_2 is an inductance the impedance of the network varies on the upper semi-circle from R_a to R_b as the frequency is increased from zero to infinity. When the magnitude of Z_2 is changed the same semi-circle is described but each point (except the initial and final ones) is reached at a different frequency. When Z_2 is a capacity the lower semi-circle, from R_b to R_a, is traced out.

We know that, in general, the value of a pure reactance [4] increases algebraically with frequency, and that its resonant and anti-resonant frequencies alternate, beginning with one or the other at zero frequency. When Z_2 is a general reactance, therefore, as the frequency increases the entire circle is described in a clockwise sense between each consecutive pair of resonant (or anti-resonant) frequencies. For example, if Z_2 is made up of n branches in parallel, one being an inductance, one a capacity and the others inductance in series with capacity, as the frequency increases from zero to infinity the circle is traced out completely $n-1$ times commencing with R_a.

[4] See: A Reactance Theorem, R. M. Foster, *Bell System Technical Journal*, April, 1924, pages 259–267; also: Theory and Design of Uniform and Composite Electric Wave-Filters, O. J. Zobel, *Bell System Technical Journal*, January, 1923, pages 1–47, especially pages 35–37.

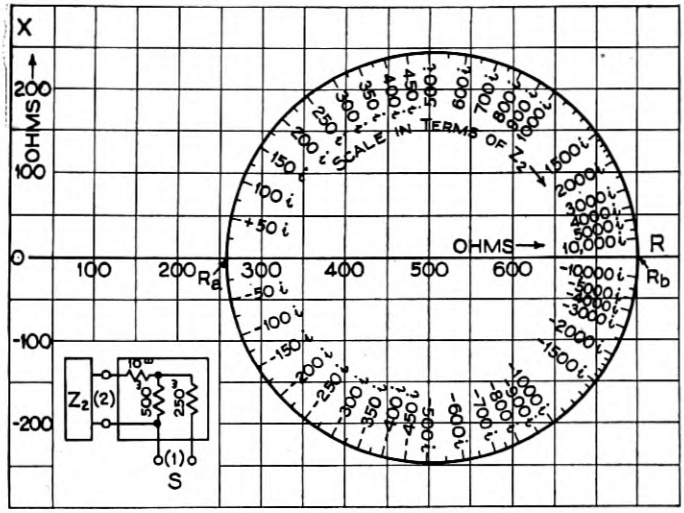


Fig. 3—Impedance of Resistance Network Containing One Variable Reactance

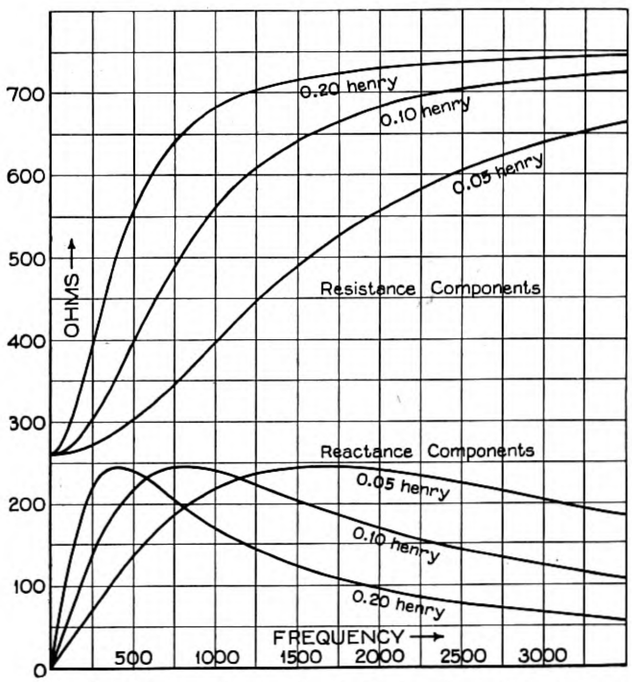


Fig. 3a—Components of Impedance in Fig. 3 when Z_2 is an Inductance Having the Values 0.05, 0.10, and 0.20 Henry

In Fig. 3 is shown the impedance locus for the particular network given on the diagram. The circle is marked in terms of Z_2. From it, certain properties of S may be read at once: the resistance component, R, varies between 260 and 750 ohms, and the reactance

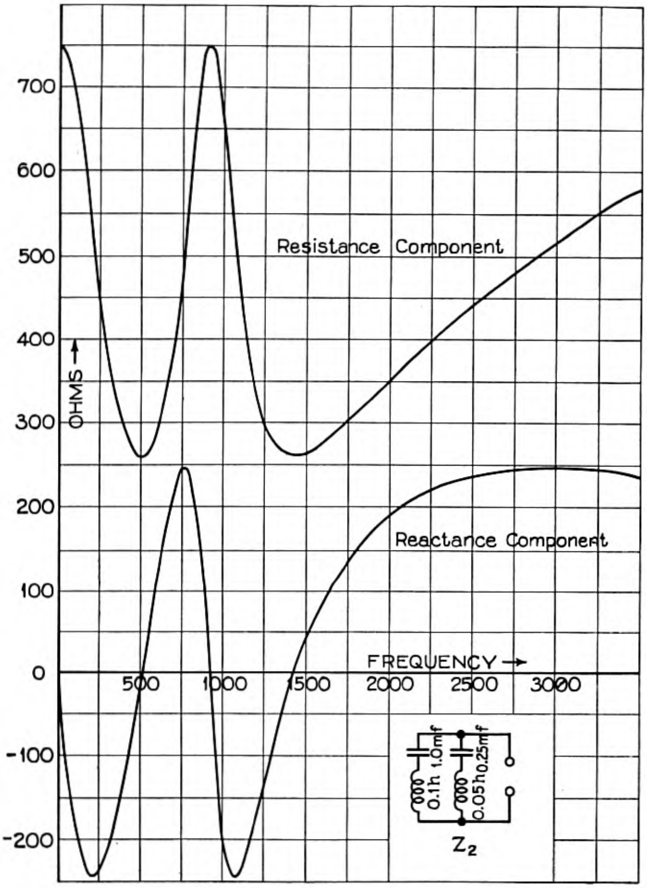


Fig. 3b—Components of Impedance in Fig. 3 when Z_2 is Doubly-Resonant

component, X, is not greater than 245 ohms nor less than -245 ohms, attaining these values when Z_2 is $+510i$ and $-510i$, respectively.

When the variation of the reactance Z_2 with frequency is known the variation of R and X with frequency may be found by using the scale on the circle. For a particular reactance network, the scale may be marked directly in terms of frequency, or if it is desired to compare the behavior of R and X when different reactance networks are substituted, the impedance locus may be marked with the frequency scale for each reactance network in some distinctive manner.

However, to show in the usual way some of the types of R and X curves represented by the locus of Fig. 3, as well as to avoid needless complication of what is intended as an illustrative rather than a working drawing, Figs. 3a and 3b have been prepared by direct projection from Fig. 3. In Fig. 3a are shown the R and X curves plotted against frequency when Z_2 is an inductance. In Fig. 3b are shown similar curves when Z_2 is a doubly-resonant reactance. The R component has a minimum at each resonant frequency and a maximum at each anti-resonant frequency, while the X component becomes zero at resonant and anti-resonant frequencies alike. The number of examples from this one resistance network might be multiplied endlessly; it is believed, however, that these are sufficient to show the great amount of information to be obtained in very compact form from one simple figure in the complex plane, and the especial superiority of the complex plane in displaying the characteristic common to all the curves of Figs. 3a and 3b: namely, that R and X at any frequency, with any reactance network, are such that the impedance lies on one circle.

Two Variable Reactances Giving Eccentric Annular Domain

Returning to the more general impedance of (4) it is seen that in each case short-circuiting and open-circuiting the terminals (2) and (3) one at a time, and varying the reactance across the other terminals, yields a locus for S which is a circle of the type just discussed. These circles are determined as follows:

Circle	Extremities of Diameter	Scale Factor κ
$Z_2=0$	R_a and R_c	c/a_1
$Z_2=\infty$	R_b and R_d	d/b_1
$Z_3=0$	R_a and R_b	b/a_1
$Z_3=\infty$	R_c and R_d	d/c_1

where $R_c=c/c_1$ and $R_d=d/d_1$. An examination similar to that in (13)–(15) shows that

$$R_a \leq R_b \leq R_d, \tag{18}$$

$$R_a \leq R_c \leq R_d. \tag{19}$$

It may furthermore be assumed without loss of generality, since it is merely a matter of labelling the reactances Z_2 and Z_3, that $R_b \leq R_c$.

Hence, the four critical points of the impedance are always in the following order:

$$R_a \leq R_b \leq R_c \leq R_d. \tag{20}$$

These circles are shown in Fig. 4. By means of the appropriate scale factors κ each may be marked in terms of the reactance which is left in the circuit.

Now suppose Z_3 is kept constant at some value which is a pure imaginary, and Z_2 is varied over the range $-i\infty \leq Z_2 \leq +i\infty$. We may rewrite (4) in the normal form (6):

$$S = \frac{a + cZ_3 + (b + dZ_3)Z_2}{a_1 + c_1Z_3 + (b_1 + d_1Z_3)Z_2}. \tag{21}$$

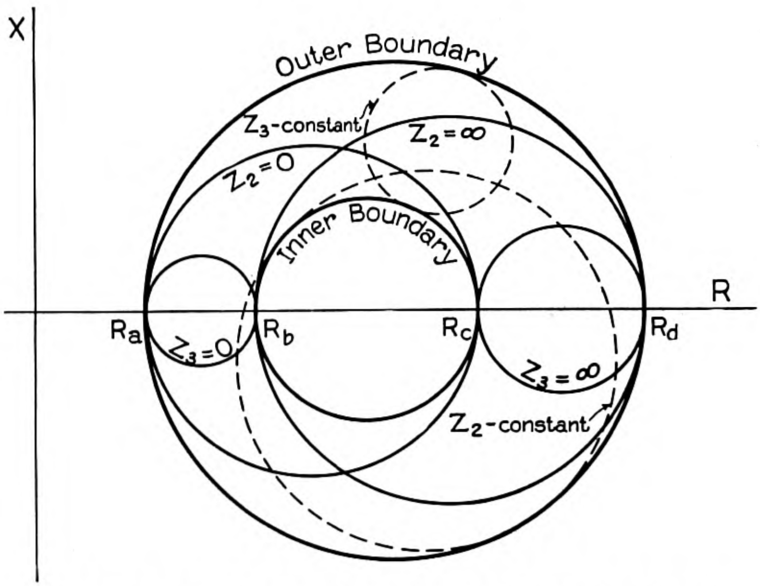


Fig. 4—The Region to which S is Restricted and the Critical Circles

The locus of S is one of a family of circles, each circle corresponding to a value of Z_3 and completely traced out by complete variation of Z_2. The properties of each circle may be found by substitution in (9).

Similarly, if Z_2 is held constant while Z_3 varies, the locus of S is one of another family of circles.

By the use of (9), keeping (5) in mind, it may be shown that the circles of each of these families are tangent to two circles determined by the resistance network alone. Both families are tangent internally to a circle centered on the resistance axis, extending from R_a to R_d. Both are tangent to a circle centered on the resistance axis, extending from R_b to R_c, in such a way that the Z_3-constant circles are tangent *externally* and the Z_2-constant circles are tangent *enclosing* the circle from R_b to R_c. These relationships are illustrated in Fig. 4.

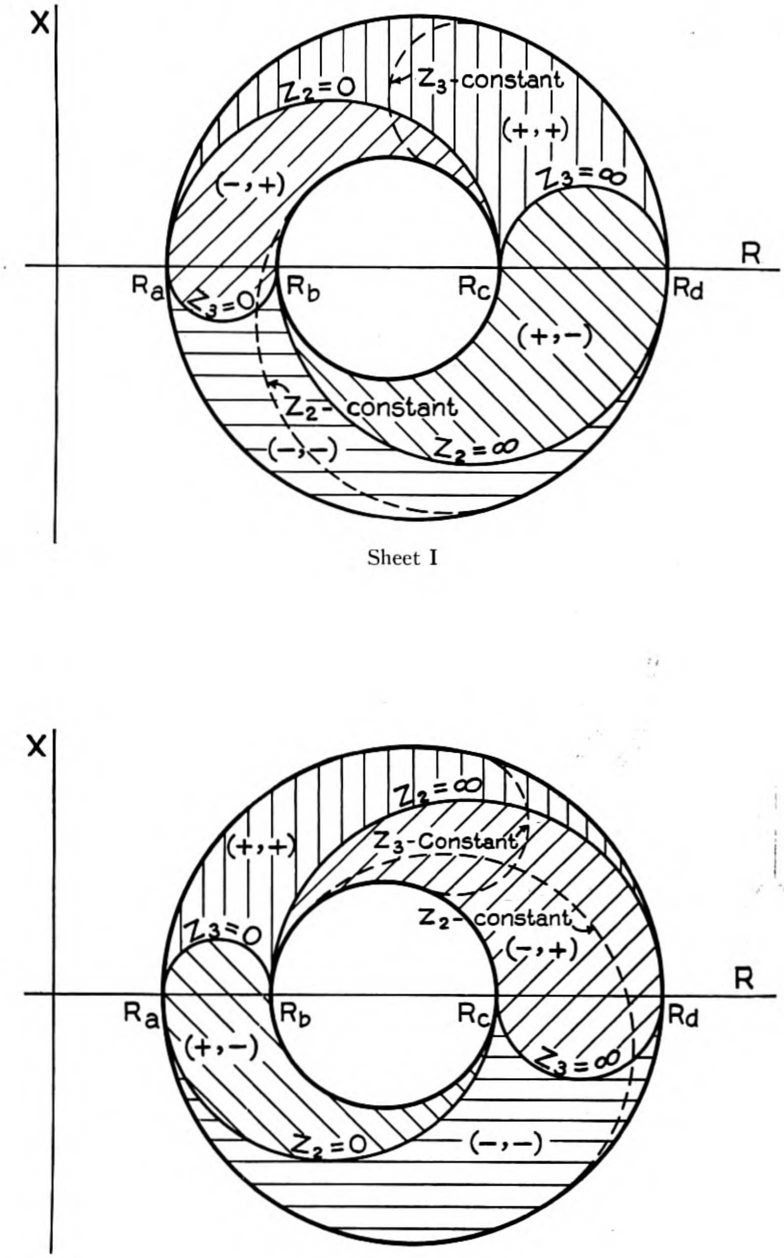
The circles R_a to R_d and R_b to R_c are, therefore, outer and inner boundaries, respectively, of the region mapped out by the two families of circles generated when first one and then the other reactance is treated as a parameter while the remaining reactance is treated as the variable. No matter what reactances may be attached to terminals (2) and (3), the resistance component R, measured at terminals (1), is not greater than the resistance when terminals (2) and (3) are open and not less than the resistance when terminals (2) and (3) are short-circuited, and the reactance component X, measured at terminals (1), is not greater in absolute value than half the difference of the resistances measured when terminals (2) and (3) are open and short-circuited. That is,

$$R_a \leq R \leq R_d, \tag{22}$$

$$|X| \leq \tfrac{1}{2}(R_d - R_a). \tag{23}$$

The two families of circles (Z_2-constant and Z_3-constant) intersect and may be used as a coordinate system from which the components of S may be read for any pair of values Z_2, Z_3. To avoid intersections giving extraneous values of S resort is made to a doubly-sheeted surface, analogous to a Riemann surface, for which the two boundary circles are junction lines. That is, the impedance plane is conceived of as two superposed sheets, transition from one to the other being made at the boundary circles. Thus, in Fig. 5, where the two sheets are separated, each Z_2-constant circle is shown running from the outer to the inner boundary in Sheet I (using the clockwise sense), and from the inner to the outer boundary in Sheet II, while the Z_3-constant circles run from the inner to the outer boundary in Sheet I and are completed in Sheet II.[5]

[5] It may be mentioned that the inner and outer boundaries are impedance curves traced out when $Z_2Z_3 = \dfrac{A_{12}A_{13}}{A_{12\cdot33}A_{13\cdot22}}$ and $\dfrac{Z_2}{Z_3} = \dfrac{A_{13}A_{12\cdot33}}{A_{12}A_{13\cdot22}}$, respectively.



Sheet I

Sheet II

Fig. 5—The Doubly-Sheeted Surface

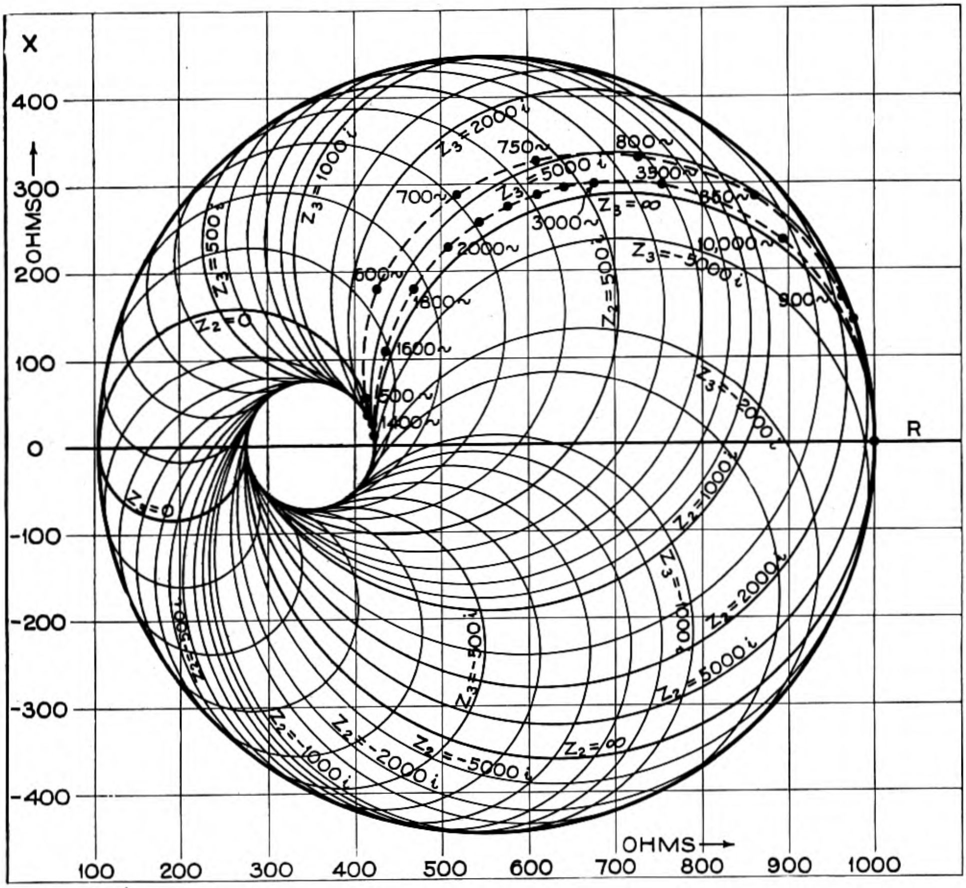


Fig. 6—Sheet I

Sheets I and II of Fig. 6, taken together, show the impedance domain of the network at the bottom of the opposite page, made up of three fixed resistances and two variable reactances. The dashed curve, appearing in four distinct parts, two on each sheet, shows the impedance S when Z_2 is the doubly-resonant circuit of Fig. 3b, and Z_3 is an inductance of 1.0 henry. Points on this curve are labelled in terms of frequency

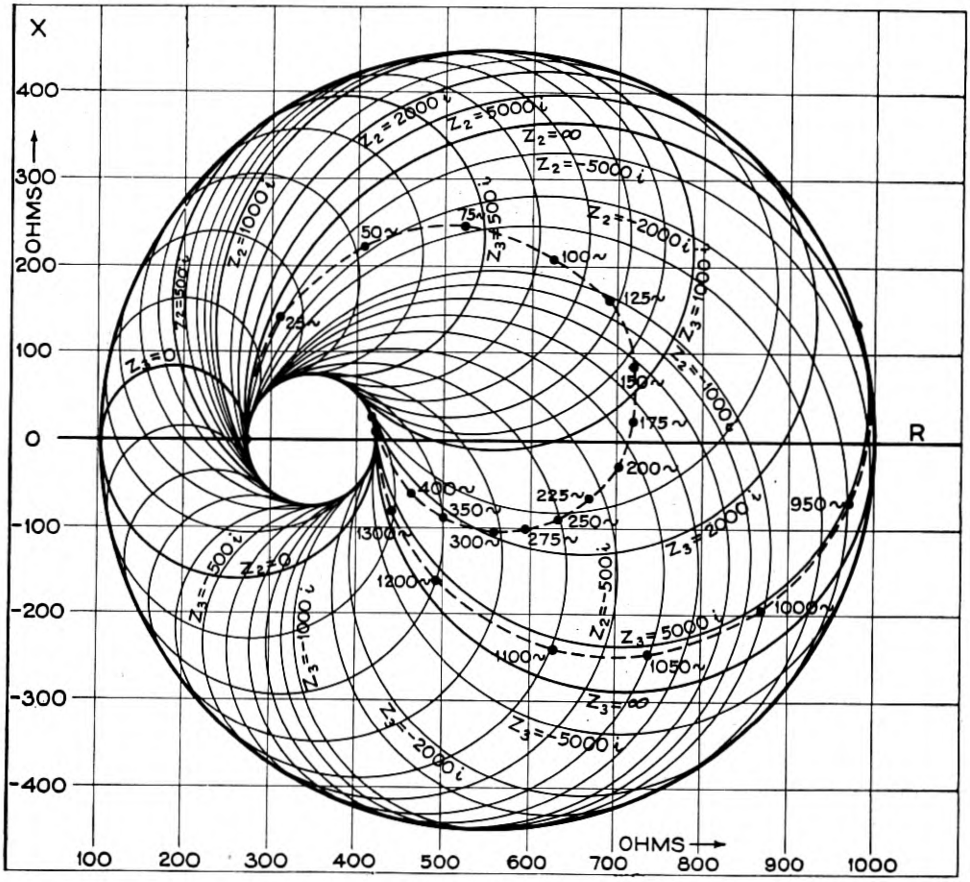
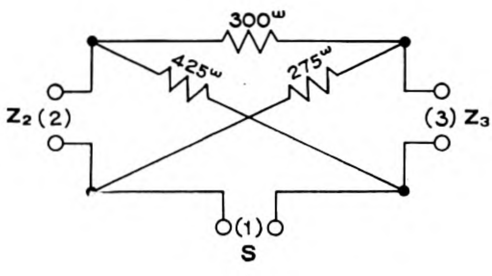


Fig. 6—Sheet II



The numbering of the sheets is, of course, arbitrary. If the upper half of the $Z_2=0$ circle is put on Sheet I, the arcs of the other critical circles are determined as follows:[6]

Circle	On Sheet I	On Sheet II
$Z_2=0$	Upper half	Lower half
$Z_2=\infty$	Lower half	Upper half
$Z_3=0$	Lower half	Upper half
$Z_3=\infty$	Upper half	Lower half

Each sheet, then, is divided into four sub-regions, indicated on Fig. 5 by the signs of the reactances for which S is within them. When Z_2 and Z_3 are composed of single elements the sub-regions in which S falls at any frequency are as follows:

Z_2	Z_3	Sub-Region	At Frequency Zero	Infinity
Inductance	Inductance	$(+,+)$	$S=R_a$	$S=R_d$
Inductance	Capacity	$(+,-)$	R_c	R_b
Capacity	Inductance	$(-,+)$	R_b	R_c
Capacity	Capacity	$(-,-)$	R_d	R_a

The course of S over the complete frequency range may be shown by a curve through the appropriate intersections of the Z_2-constant and Z_3-constant circles, as in the following example.

The impedance region for a particular bridge network is illustrated in the two sheets of Fig. 6. The arcs of Z_2-constant and Z_3-constant circles in each sheet form a curvilinear grid superposed on the R,X grid of the complex plane. For example, if $Z_2=200i$ and $Z_3=900i$, the value of S is read from Sheet I as $327+i291$, and S has this value irrespective of the structure of Z_2 and Z_3.

An impedance curve (dashed) is shown in Fig. 6 representing the variation of S with frequency when Z_2 is the doubly-resonant reactance

[6] When the sheets are numbered in this way, the point S falls on Sheet I or Sheet II according to the following table, in which k_1 and k_2 are the critical values for the product and quotient of Z_2 and Z_3, respectively, given in Footnote 5:

(Z_2, Z_3)	$(+,+)$	$(+,-)$	$(-,+)$	$(-,-)$
On Sheet I, if	$Z_2/Z_3<k_2$	$Z_2Z_3>k_1$	$Z_2Z_3<k_1$	$Z_2/Z_3>k_2$
On Sheet II, if	$Z_2/Z_3>k_2$	$Z_2Z_3<k_1$	$Z_2Z_3>k_1$	$Z_2/Z_3<k_2$

For the network of Fig. 6, $k_1=116{,}875$ and $k_2=0.972111$.

used in Fig. 3b and Z_3 is an inductance of 1.0 henry. This impedance curve has four parts, two in each sheet. It starts on the resistance axis at the intersection of the $Z_2=\infty$ and $Z_3=0$ circles. As the frequency increases from zero the first part of the curve is traced out in Sheet II. At 25 cycles the impedance is approximately $310+i140$. The reactance component has a maximum of about 250 ohms at about 70 cycles, the resistance component has a maximum of about 720 ohms at about 160 cycles, the reactance component has a minimum of about -110 ohms at about 300 cycles, and finally at about 480 cycles the curve reaches the inner boundary, whereupon it changes to Sheet I. It remains in Sheet I up to a frequency of about 910 cycles, the resistance component having a minimum and the reactance component a maximum, which may be read from the diagram. The impedance between 910 cycles and approximately 1,390 cycles lies on Sheet II, and from 1,390 cycles to infinite frequency on Sheet I. The resistance component has a total of three maxima and three minima, and the reactance component three maxima and two minima, following the cyclical order: R-minimum, X-maximum, R-maximum, X-minimum.

An interesting exercise is to observe the effect on the impedance curve of changing the value of the inductance Z_3. The curve intersects the Z_2-constant circles at the same frequencies in each case, but the points of intersection are moved in a clockwise or counterclockwise sense as Z_3 is increased or decreased. With each such change parts of the impedance curve disappear from one sheet and reappear on the other. For instance, with a decrease of the inductance Z_3 the first loop of the impedance curve on Sheet II shrinks, and with sufficient decrease in inductance may become too small to plot, although it does not disappear entirely.

It is evident that if Z_2 and Z_3 are formed of reactance networks of greater complication the impedance curve may be very involved. But no matter how tortuous its path, it is restricted to the impedance region, that is, to the ring-shaped region between the non-intersecting boundary circles determined by the resistance network alone.

My thanks are due to Dr. George A. Campbell for his stimulating, continued interest, and to Mr. R. M. Foster for suggestions on every phase of this work.

The Vibratory Characteristics and Impedance of Telephone Receivers at Low Power Inputs

By A. S. CURTIS

THE ordinary telephone receiver is one of the most sensitive known detectors of weak alternating currents over a considerable part of the audible frequency range. Its high sensitivity, combined with its simplicity and convenience, have led to its general adoption as the detecting element in the AC impedance bridge and other measuring apparatus employing the nul method. There are also a number of cases outside of the laboratory where a knowledge of the behavior of the receiver operating near its minimum audible power input is of importance. In apparatus developed during the World War, such as that for detecting and locating submarines, in radio reception, and in the reception of various other sorts of signals, the receiver is frequently operated near the threshold of audibility. While it is in general possible to employ a vacuum tube amplifier to render weak signals more easily audible, considerations of cost or increased complication often make it impracticable to do so. In any case, if it is desired to reduce to the limit the minimum audible signal, it is necessary to know the constants of the receiver working on these low power inputs, in order to design intelligently its circuits and other associated apparatus.

Current literature dealing with the sensitivity of telephone receivers indicates that the relation between the impedance and vibratory characteristics of the receiver at currents near minimum audibility to those as ordinarily determined in the laboratory, is not generally known. It would, therefore, seem of interest to publish the results of an experimental determination of receiver characteristics at very low currents. Such an investigation was carried on in 1918 and 1919, using the Western Electric No. 509 radio receiver (the present standard Western Electric Receiver for radio use). The work, however, was done, not merely with the idea of determining the characteristics of this particular instrument, but for the purpose of ascertaining the behavior of receivers in general, near minimum audibility.

Inasmuch as the damped impedance of the receiver—that is the impedance with the diaphragm held motionless—is very close to the impedance obtained with the instrument on the ear, it is commonly used as the basis of circuit calculations. A knowledge of its value for weak currents is therefore of importance. Measurements were first made of the damped impedance of six instruments at a frequency of 1,000 cycles for a wide range of input current, and later the work was extended to the measurement of the vibratory characteristics. A bridge network was used for measuring the current supplied to the impedance bridge and from the circuit constants the current through the receiver under test could be calculated. The re-

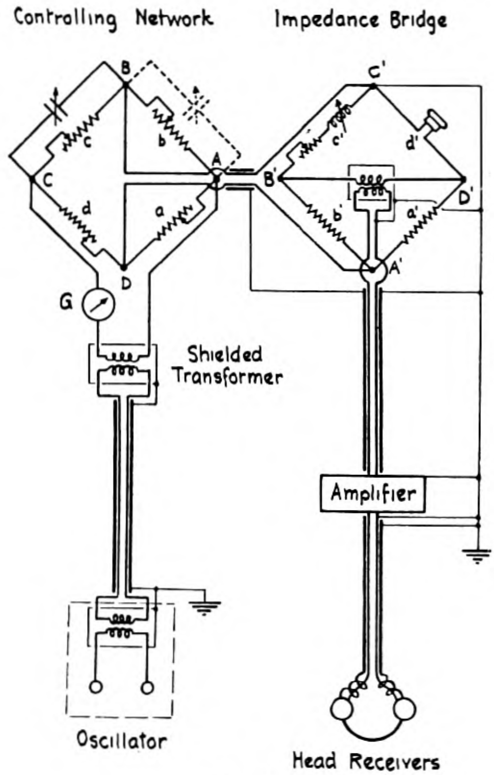


Fig. 1

sistances in the various arms of the controlling bridge network were chosen so as to furnish an essentially constant current through the receiver under test, although its impedance might vary through a rather wide range. With the extremely small values of currents involved, it was necessary to amplify the power to the bridge balancing receivers approximately 100 TU. For this amount of amplification, it was obviously necessary to take extreme precautions in grounding and shielding the apparatus, in order to reduce to inaudibility the effect of stray fields from the source of current supply. This was success-fully done and the impedance bridge measured impedance accurately with currents as low as 10^{-9} amperes, through the receiver under test. The correctness of the point of balance of the bridge was established by measurements of standard impedances over the range of currents employed in the receiver tests. A schematic diagram of the circuit is shown in Fig. 1.

For measurements of damped impedance, the receiver was placed in a small sound-proof box, with its diaphragm damped by a micrometer depth gauge, which was carefully adjusted so as just to impinge upon the diaphragm. It was necessary to insulate the receiver from mechanical agitation, since minute voltages generated in it were sufficiently amplified to cause an excessive noise in the head receivers.

Fig. 2 shows the damped effective resistance and reactance of the six instruments, taken at 1,000 cycles, plotted on semi-logarithmic

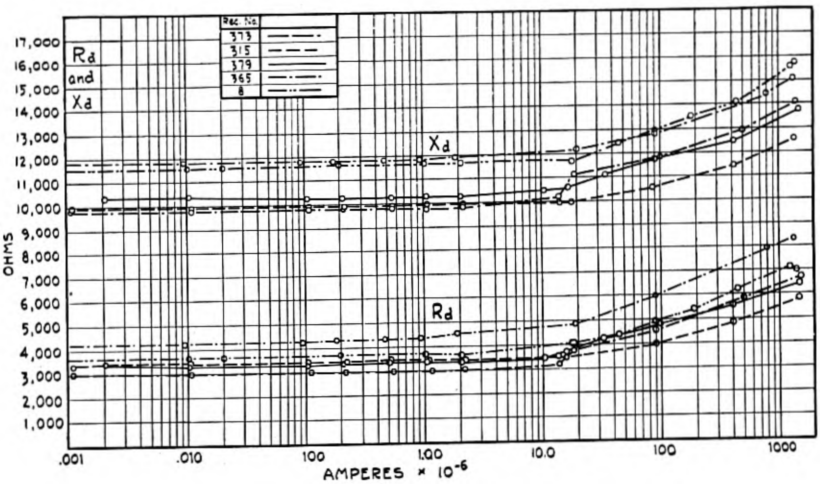


Fig. 2

paper. It will be seen that below approximately 10^{-6} amperes, the impedance is constant. However, above this value both the effective resistance and the reactance show a consistent increase with the current. The minimum current employed (10^{-9} amperes), is between two and three times the minimum audible current for this type of instrument, but from the data taken there is no reason to suppose that the impedance would vary for smaller currents. This receiver has a winding of 11,000 turns, and it can, therefore, be assumed that this type of structure will have constant impedance below a magnetomotive force of .01 ampere turns. For laboratory measurements on this instrument a current of 2×10^{-5} amperes is ordinarily used, and it will be noted that the impedance at extremely low currents is not greatly different.

It is generally known that, in the case of either a steady or an alternating field, the permeability and the shape of the hysteresis

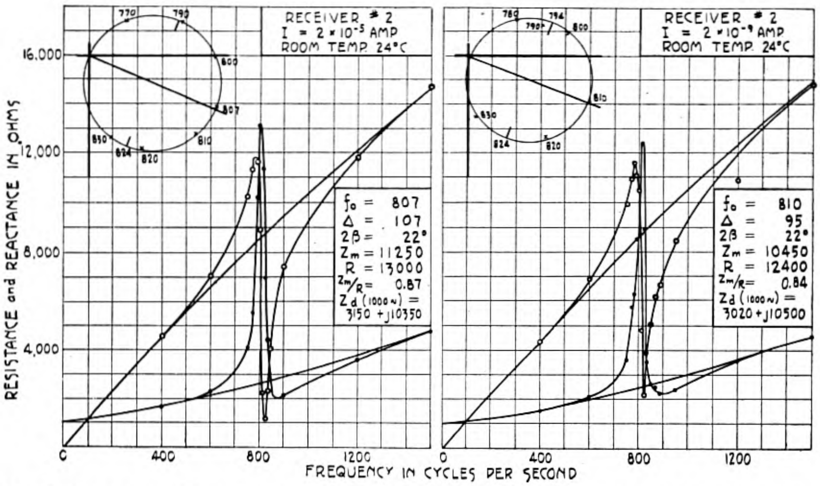


Fig. 3—f_o = natural frequency; Δ = logarithmic decrement per second; 2β = depression angle of principal diameter; Z_m = maximum motional impedance; R = free resistance at resonance; Z_o = damped impedance.

loop for ordinary magnetic materials reach limiting values as the magnetomotive force is reduced, that is, further reductions of the magnetomotive force have no effect on these magnetic characteristics. The results cited above show that this condition obtains for a weak alternating field when it is superimposed on a relatively strong steady field.

In the measurements of free impedance for determining the vibratory characteristics the small sound-proof box could not be used on account of the proximity of its walls. Accordingly, the receiver and the impedance bridge were placed in a large sound-proof booth with padded walls where the effect of reflection of sound waves was very small. With the diaphragm of the instrument free to vibrate, its efficiency as a sound detector was materially increased and the noise in the head receivers due to the slightest movements of the observer became so serious that it was not feasible to take data with currents of less than 2×10^{-9} amperes.

Fig. 3 shows impedance characteristics, with their associated circles, of the same receiver with currents of 2×10^{-9} and 2×10^{-5} amperes. It will be seen that the differences between these curves are insignificant when one considers the low precision of motional impedance data in the absence of extreme precautions with regard to constancy of temperature, etc. Moreover, other impedance analyses at intermediate values of current agree with the above within the precision of the measurements.

To summarize the results, it may be said that the characteristics of receivers remain substantially unaltered as the current is reduced to the point of minimum audibility. In taking impedance measurements, it is well to use a current which is low enough to be on the flat part of the curve. This can usually be done without the use of amplifiers between the impedance bridge and balancing receivers. The fact that the vibratory characteristics of the receiver remain unaltered as the power input is reduced to the threshold of audibility throws an interesting light on the behavior of the diaphragm material under very small motions. Calculations of the minimum audible amplitude near resonance, based on the fact that the constants of the material remain unchanged, show it to be of the order of 10^{-9} centimeters. This motion is less than the mean molecular diameter of the diaphragm material.

Some Contemporary Advances in Physics—VIII
The Atom-Model, First Part[1]

By KARL K. DARROW

A. Introductory Remarks About Atom-Models

MORE than any other word of the language, the word *atom* is implicated with the history of human speculations concerning the nature of things. It is introduced when people cease to content themselves with observing, and begin to philosophize. There are many of the fundamental and essential writings of the literature of physics in which it does not appear, or appears without warrant. These are the descriptions of things observed, the accounts of experiments, the records of measurements, on which the edifice of theoretical physics is founded. There are many articles of what is commonly called the "theoretical" sort in which it does not occur. Such are the papers on the motions of planets, on the vibrations of elastic solids, on the currents in electrical networks, on the courses of light-rays through optical systems—papers which are essentially descriptions, although they give the impression of being something greater and deeper because they relate to idealized cases, and are phrased in the laconic language of mathematics. When the word *atom* appears justifiably in a discourse, it means that the author has departed from the safe routine of describing observed and observable events, however selectively, however skilfully, however intelligently. It signifies that he has gone beyond the limits of observation, and has entered upon the audacious adventure of constructing by the side of the real universe an ideal one, which shall act as the real one does, and be intelligible through and through.

Atoms are the building stones of this art-world or image-world, which is intended to represent the actual world, imperfectly indeed for the time being, perhaps completely at some distant day. Some few experiments, it is true, prove (as well as anything can prove anything else) the existence of very minute particles of matter having the minute charges, the minute masses, the minute magnetic moments

[1] This part, the first of two composing the article, is devoted chiefly to the facts of observation which the favorite atom-model of the physicists of today—the atom-model known by the names of Rutherford and of Bohr—is designed to interpret. A brief description of this atom-model is included; but the detailed account of the peculiar features, of the strange and important limitations which are imposed upon it to adjust it to all the phenomena mentioned, is reserved for the second part. Owing to the great quantity of information which it is desirable to present, the article needs all the benefit it can derive from a careful and obvious organization, and I have sacrificed fluency to a quite formal arrangement under headings and sub-headings.

which it is found expedient to ascribe to the atoms. These experiments are enormously important, for they invest the atom with a reality which nothing else could give it. To some they have given the hope that all the properties of the atom may one day be demonstrated unquestionably by direct evidence. There is little reason to expect that we shall see that day. The atom is no longer entirely a product of the scientific imagination; but neither is it entirely an object of experience. Most of its properties are invented, not discovered. Whether this invented and imagined entity is "real" is a difficult question. Perhaps it is best to evade such a question by asking the questioner what he means by "reality". As a matter of fact, it is not possible to discuss atomic theories thoroughly without raising and settling such formidable questions as, what is a theory? and, what is an explanation? and even, what is reality? perhaps eventually, what is truth? I do not aspire to answer these questions. But there are some common misconceptions about atoms which it is prudent to clear away at the beginning.

In the first place, one does not utter an atomic theory by saying that a substance is made up of small pieces, each exactly like a large piece of the substance in every respect except size. We should achieve nothing by saying that iron is made of black lustrous conductive magnetic atoms, or that glass is built of colorless transparent brittle insulating atoms, or that an apple consists of white soft sweet juicy atoms. The atoms must be endowed with fewer and simpler properties than the substance they are meant to compose, else they are futile. One must select some of the properties of the substance to be attributed to its atoms, and set the others aside to be explained by those.

Again, it is not obvious which properties should be selected for the atom; these depend largely on the fantasy of the atom-builder. However, certain qualities such as viscosity and plasticity, conductance for heat and conductance for electricity, opacity and transparency and lustre, warmth and flavor and fragrance, are not usually assigned to atoms. In general, the more a quality of a substance varies with its state, the less it is suited to be made an atomic quality. Ferromagnetism is a quality which one would assign almost instinctively to the iron atom; but it is possible to deprive iron altogether of this quality by a simple heat treatment, and hence it is not generally supposed to be a feature of the atom. But the rule is not an absolute one. The visible radiations from gaseous iron are supposed to be characteristic above all other things of the atom itself, yet they cease when the iron is condensed. It is supposed that in the condensed phases the atoms are so close together that they distort one another—a permissible idea if used with discretion, yet an atomic theory could easily become a meaningless form of words if this device were employed without limit. Of all the properties of matter, mass alone appears to be entirely exempt from change. For this reason all atom-models involve mass as an essential property of the atom; and this is the only respect in which they all agree.

Few and simple, therefore, must be the properties of the atom; yet we must not rush to the other extreme, and contrive atoms simplified into uselessness. The chemists know of eighty-eight different elements, sufficiently unlike to be distinguished; and we all know how great is the contrast between carbon and gold, hydrogen and lead, fluorine and helium. It is scarcely likely that such differences as these can be explained by atoms which are simply hard pellets differing only in size and shape and weight, like those of Lucretius and Newton, or by atoms which are abstract centres of force, like those of Boscovich. We are forced to invent atoms more complicated that these; and from this it is not far to say that we must imagine a structure for the atom; and from this scarcely farther to say that we must imagine an atom built of parts.

At this point we meet with a clamor from a number of excellent people, many of them otherwise quite innocent of Hellenic culture, who have it firmly fixed in their minds that *atom* is the Greek word for *indivisible*; whence they conclude that when the physicist speaks of subdividing his atoms, he is contradicting his own terms, he is violating the rules of his own game, and has forfeited his right to be heard.[2] The premise may be right, but the conclusion is absurd. If some of the properties of gold are explained by assuming it made of atoms with fewer properties, and later the explanation is improved and extended by assuming these atoms made of still smaller particles with still fewer properties, the second step is not less legitimate than the first. It may be contended, with some reason, that the name *atom* should be transferred at once to the smaller particles. At best this would be one of the changes which are desirable in principle but cause more trouble than they are worth. The contention is, however, weakened by the fact that some at least of the smaller particles of which we imagine gold atoms to be made are not imagined to be peculiar to gold, but are conceived as particles of a fundamental substance common to all elements. That the "atoms" of the many

[2] I should have put this passage even more strongly, but that Schuster tells that Kelvin himself inveighed on one occasion against the idea of subdividing atoms. He was answered by a young man who said, "There you see the disadvantages of knowing Greek." This seems as good an answer as any.

elements should be systems of "atoms" of one or a few fundamental materials is a thoroughly pleasing idea, although at present an unrealized ideal. It is unknown how far our descendants will find it expedient to dissect the atom; but it is certain that they will not be stopped by etymology.

Another fact about atom-models is that they are not always displaced by their successors; several may and do persist side by side, each adapted to a certain set of facts and observations. Every atom is designed in view of a very small fraction of the available knowledge about properties of matter; and this applies to the latest model as well as the earliest. The chemists of the nineteenth century were most impressed by the immutable weight of matter and by the laws of chemical combination; hence their atoms were merely weighted particles equipped with hooks to catch the hooks of other atoms. To the physicists of fifty years ago the physical properties of gases seemed the easiest phenomena to interpret, and they imagined atoms as rigid elastic spheres with radii of some 10^{-8} centimetre; by the masses and motions of such atoms they explained the pressure, elasticity, viscosity, diffusion and specific heats of gases. The physicists of the next generation attended chiefly to the emission, the refraction, the dispersion of vibratory radiations by luminous gases, and conceived the atom as a framework holding vibrators, like a belfry with a carillon of bells. This third model is inferior to the second in explaining the properties of gases, inferior to the first in explaining the laws of chemical combination; each of the three is superior in its own field to the atom-model to which this article is chiefly devoted, and which in its turn is primarily adapted to a field of its own. Still other atom-models have been devised, endowed with other properties, to account for other phenomena; and it is altogether probable that many more will be presented before the eventual one is attained, if it ever is. For instance, we may some day behold an atom-model devised to explain the conduction of electricity in solids, very competent in its field and quite unlike these others. In the eventual atom-model the essential qualities of all of these, and of many others, must be happily combined; it does not matter about the inessential ones.[3]

[3] Now and then an article appears in a physical or chemical journal, in which an oddly unconventional atom-model is proposed to interpret some such property of matter as the thermoelectric effects, or supra-conductivity, or valence, or some other with which the Rutherford-Bohr atom-model has not as yet been matched. It is easy for a physicist to ignore such articles, on the ground that any model departing from that of Rutherford and Bohr must be wrong. This is certainly a mistaken policy. Any partially competent atom-model deserves to be examined with care; its essential features must reappear in the eventual model. But, of course, the essential feature is not always the conspicuous one.

In awaiting that eventual atom-model, it is best to regard the atoms of the present day as mutable and transitory. Like railway time-tables, atom-models should be inscribed "subject to change without notice." Nothing is irrevocable in physics, except the record of past events; and we who have seen the undulatory theory of light assailed and shaken may well hesitate to put unquestioning faith in any atom-model. Even if there is no danger of change, it is a virtue to keep data and theories sharply separated in one's mind. In no field is this more difficult and important than in the field of this article, where the very language used to describe the data is saturated with the spirit of a particular conception of the atom, and it is customary to expound the theory before the facts. For these reasons I shall go to the opposite extreme, and treat the contemporary atomic theory with an exaggerated reserve which in many places will seem excessive to the reader and in some to the writer himself.

The favorite atom-model of the physicists of today is a structure of electrons, congregated about a positively-charged nucleus. The data which this atom is designed primarily to interpret were discovered before 1913, or else since 1913 by methods developed before that time. These discoveries are due largely to Rutherford, whose name the model often bears. The sections of this article which are labelled *B*, *C* and *D* are devoted to these data, and to the inferences from them. In 1913 a great change in the situation was wrought by a brilliant idea of Niels Bohr. Bohr did not discover new data; he taught a new way of interpreting old ones, he showed men how to read spectra. Through this interpretation of spectra, and through data which were discovered by men inspired with his idea, a previously-unknown property of matter was disclosed. This is expressed by saying that each atom possesses many distinct Stationary States. The largest section of this First Part of the article, the section *E*, is devoted to the knowledge of these Stationary States. Had these been discovered earlier, an atom-model might have been devised to explain them and them alone. Rutherford's atom-model was already in the field, and it was modified so that it might interpret the new knowledge. To these modifications, of which some are of a remarkable simplicity and beauty, the Second Part of this article will be devoted.

B. The Electron [4]

The *electron* is the atom of negative electricity. An individual electron can be captured upon a droplet of oil or mercury, or a minute

[4] This section is drastically curtailed, for the chief facts about the electron should by this time be common knowledge. Millikan's book "The Electron" (now in its second edition) may be consulted.

solid particle, and the amount of its charge determined. This amount is $4.774.10^{-10}$ in electrostatic units, according to Millikan. It is designated by the symbol e. When a magnetic field is applied to a stream of electrons all moving with the same speed, the electrons are deflected all to the same degree, which shows that they all have the same mass. This mass is practically equal to 9.10^{-28} in grammes, unless the electron is moving at a very uncommonly high speed, in which case it is appreciably greater.

These facts of experience are about all that is definitely known or needs to be known about the electron, in order to appreciate its role in modern atomic theory. There is no good way of determining its size, although the length of its mean free path in certain gases indicates, perhaps definitely proves, that it is much smaller than an atom. If the electron is a spherule of negative electricity uniformly dense, then its radius cannot be less than 2.10^{-13} cm, for if it were, the electromagnetic mass of the spherule would exceed the observed mass of the electron.[5] This size is much smaller than the one which it is expedient to attribute to the atom, happily for us, since otherwise it would be difficult to conceive of atoms containing electrons.

Since electrons can be coaxed or forced out of substances of every kind—elements and compounds, metals and non-metals, liquids and solids and gases—the atoms are supposed to contain one or more electrons apiece. This argument was formerly fortified by the fact that the light emitted from glowing gases is in many respects such as oscillating electrons would emit. This second argument is for the present under a cloud.[6]

[5] This is a short way of saying that, if the electron were a particle of smaller radius than 2.10^{-13} cm., more energy would have to be supplied to it to increase its speed than is actually required. For, in order to set an electrified particle into motion, energy must be supplied to build up the magnetic field which surrounds a moving electric charge; this energy U is additional to the kinetic energy $\frac{1}{2}mv^2$ required to set the mass m associated with the charge into motion with speed v, and it may be regarded as the kinetic energy associated with an extra "electromagnetic" mass $2U/v^2$ which the particle possesses by virtue of its charge. This quantity $2U/v^2$ can be calculated, for a given size and shape of the electron; if we make the electron too small, $2U/v^2$ comes out larger than its observed mass, which is a *reductio ad absurdum*. This illustrates the rather surprising fact that we are not permitted to imagine the electron as an infinitely small particle, a mere geometrical point loaded with an infinitely concentrated charge and mass. Speculations about its size and shape and the distribution of charge within it are not necessarily trivial; some may even be verifiable. We also meet with this dilemma: how does the electron, a piece of negative electricity of which each part should repel every other, keep from exploding?

[6] Perhaps I ought to mention that F. Ehrenhaft of Vienna has been ardently contending for about fifteen years that there is no such thing as an electron. He maintains that he can demonstrate negative charges much smaller in amount than

C. Positively-Charged Particles Accepted as Atoms[7]

Positively-charged particles are found in abundance in gases in which an electrical discharge is or has lately been maintained, and they may be produced under well-controlled circumstances by pouring a stream of electrons with properly-adjusted speeds into a gas, and in other ways. Only the ratio of the charge to the mass can be determined for these particles, not the charge individually nor the mass individually. But particles of apparently the same substance show distinct values of this ratio, which stand to one another as the numbers 1, 2, 3, . . . and the intermediate values do not occur. This supports the quite natural idea that these particles are atoms which have lost one or two or three or more of their electrons. If we make this supposition, we thereby assume values for the charges, and can calculate the masses of the particles from these and the observed values of the charge-mass ratio. The masses lie between 10^{-24} and 10^{-21} (in grammes) for particles occurring in the vapors of the various chemical elements, and they lie in the same order as the combining-weights of the chemical elements. This is powerful testimony that the particles indeed deserve the name of "atoms".

There is one sort of positive particle for which the charge can be measured directly. This is the alpha-particle, which cannot be produced at will but is supplied by Nature from radio-active substances. Counting the number of these particles emitted from a bit of radioactive substance in a given time, and measuring the total electrical charge lost by the substance in the same time, and dividing the latter figure by the former, Rutherford and Regener obtained the charge of the alpha-particle, which is twice the electron-charge (with reversed sign) within the limits of experimental error. This suggests that the alpha-particle is an atom of something or other, which has lost two electrons. As an evacuated tube into which alpha-particles are admitted is presently found to contain helium, the "something or other" is supposed to be helium. The mass of the alpha-particle can be determined directly from its charge and charge-mass ratio. It amounts to $6.60 \cdot 10^{-24}$, and this agrees with the mass inferred in the foregoing way for the positive particles found in helium.

$4.774 \cdot 10^{-10}$. Anyone interested in his case may find it presented in the April, 1925, number of the *Philosophical Magazine*. The question is for experimental physicists to discuss; but it is not likely that the edifice of modern physics is liable to be ruined by a flaw at its very foundation, such as this would be.

[7] The material of this section may be found much more extensively presented in my fourth article, in which I have also written about isotopes, a subject omitted here for the sake of brevity.

The alpha-particle is supposed, like the electron, to be much smaller than an atom; partly because it can go through a thin sheet of metal, chiefly because of evidence to be expounded in the next paragraph.

Collisions between alpha-particles and other particles of similar mass are occasionally observed; the mass of the struck particle can be deduced from the directions in which it and the alpha-particle fly off after the impact, assuming only that conservation of momentum and conservation of kinetic energy prevail during the impact. In this way it is possible to determine the masses of tiny particles (presumably atoms) of hydrogen, helium, oxygen and nitrogen (perhaps eventually of other elements) in terms of the mass of the alpha-particle, which is determined from its charge-mass ratio and its charge, which are determined directly. If all the properties of the elements could be explained by atoms possessing no features except charge and mass, all the foundations of science might be laid down already.

The alpha-particle is one of the most valuable and powerful instruments in the physicist's equipment. It is a sort of hyper-microscope, penetrating and revealing the arrangements of systems so minute that microscopic objects are universes compared with them. Rutherford's development of the technique of using the alpha-particle is to be ranked among his greatest works.

Positively-charged particles with masses as low as that of the electron have never been observed; the least massive of the known positively-charged particles has 1,840 times the mass of the electron.

D. The Nuclear Atom-Model

Since we have met with positively-charged particles which are accepted as atoms deprived of one or more of their electrons, and since these incomplete atoms are much greater in mass than the electrons, it is natural to suppose that the completed atom consists of a positively-charged particle or *nucleus* in which almost its entire mass is concentrated, and one or more electrons which compensate the charge of the positive particle but add little to the mass of the atom. If we further suppose that the dimensions of the electrons and of the positively-charged particle are small in comparison with the distance between them, we invent the *nuclear atom-model.*[8]

The direct evidence for the nuclear atom-model consists of a very

[8] Commonly known as the Rutherford atom-model, after the physicist who invented it and discovered most of the evidence for it; occasionally as Nagaoka's, after another physicist who suggested it; occasionally as the Saturnian model, as some have supposed that the electrons lie in flat rings around the nucleus like the rings of Saturn around that planet.

small but a beautiful and convincing series of experiments, of which the first and the most were performed by Sir Ernest Rutherford and his pupils.[9] These experiments are designed to show that the orbit of a minute charged particle (usually an alpha-particle), flying through a thin film of metal, is in certain cases very like the hyperbolic orbit of a comet around the sun. Such an orbit is the path of a particle moving near to an immobile particle, for instance a light particle moving close to a much more massive one, which attracts it or repels it by a force varying inversely as the square of their distance apart. If these experiments show what they are designed to show, then they indicate that the atom includes a particle much more massive than an electron, bearing an electric charge, and sufficiently isolated from the other charges in the atom (such as the electrons) so that its field of force in a measurable space around it is not disturbed by theirs. We cannot, however, trace the entire path of an individual flying charged particle as it swings around through an atom, and are forced to make up for this deficiency by a statistical study of the visible portions of the paths of a great multitude of charge particles.

Let us consider exactly what these experiments show; for whatever they do prove is the most securely proved of all the beliefs about atoms. In the first place, they show that there is a nucleus; and a vacant space surrounding it, in which an inverse-square force centred upon the nucleus prevails; and they indicate the dimensions of this vacant space. This commences within 10^{-12} cm. of the nucleus, which is another way of saying that the diameter of the nucleus is less than 10^{-12} cm.; and it extends beyond a distance given (to take instances) as 14.10^{-12} cm. for platinum and 10^{-9} cm. for argon, which is another way of saying that nearly all of the negative charge of the atom lies still farther out from the nucleus. If the negative charge is indeed subdivided into electrons, then the atom is formed like a hollow cloud of electrons, with a massive positively-charged nucleus at the centre of the interior hollow.

The diameter of this cloud of electrons is not furnished by the experiments on alpha-particle deflections; but considering that the distance between adjacent atoms locked into a crystal lattice is generally a small multiple of 10^{-8} cm., it cannot be much greater than 10^{-8} cm. unless we are prepared to admit interpenetration or violent distortion of atoms; nor does it seem likely that the diameter is very much smaller than this amount. I have already mentioned that some of the properties of gases are adequately explained by assuming

[9] For the mathematical theory of these experiments, the second article of this series may be consulted.

that the atoms are elastic rigid spheres with a diameter of about 10^{-8} cm. Unlike as an elastic rigid sphere and a cloud of electrons seem, this agreement between so differently made estimates is probably no mere coincidence. It will be noticed that all of the figures about sizes at which we have arrived in such various ways (diameters for the electron and the nucleus, for the vacant space inside the electron-cloud, for the entire atom) are quite compatible with one another. If the value derived for the diameter of the interior hollow had been ten times the spacing of atoms in a crystal, or a tenth the diameter of a spherule of electricity with the same electromagnetic mass as an electron, we should indeed be in trouble.

In the second place, these studies of the deflections of alpha-particles yield numerical values for the nuclear charge: $(77.4 \pm 1)e$ for platinum, $(46.3 \pm 0.7)e$ for silver, $(29.3 \pm 0.7)e$ for copper, $19e$ for argon, $6.5e$ for "air" (a sort of statistical average of the values for oxygen and nitrogen).[10] To these must be added the value $+2e$ for the nuclear charge of helium; for we have already seen the evidence that the alpha-particle is what is left of a helium atom when two electrons are removed, and these last-cited experiments show that it is itself a nucleus, hence a helium nucleus. This nuclear charge must be balanced by negative charges within the atom; if this balancing negative charge is subdivided into electrons, then the numerical factors of e occurring in these numerical values are equal respectively to the number of electrons belonging to each atom. We thus have fairly accurate estimates of the number of constituent electrons within each of four or five atoms.

These estimates agree, within their experimental uncertainties, with the famous and splendid idea of van den Broek and Moseley: that the number of electrons in each atom, and the nuclear charge measured as a multiple of the electron-charge, "is the same as the number of the place occupied by the element in the periodic table". This idea is also supported by rough measurements of alpha-particle deflections by a few other atoms, and by the extent to which different atoms scatter X-rays; but the most important of the additional evidence will find its appropriate place in the second section of this article.

These conclusions are almost all that can be deduced from the data. The arrangement of electrons within the electron-cloud is almost

[10] References for these data are given in the fourth article of this series. The data obtained by E. S. Bieler (Proc. Roy. Soc., 105A, pp. 434–450, 1924) show incidentally, if I do not misread his article, that the nuclear charges of Mg and Al have the desired values $12e$ and $13e$, respectively, within a few per cent. Rutherford's studies of encounters between alpha-particles and hydrogen atoms prove a nuclear charge of e for the latter.

entirely concealed. It is not altogether inaccessible; for the deflections suffered by alpha-particles and electrons flying through atoms are influenced by the electrons of the atom, not by the nucleus exclusively; and from the degree in which the observed deflections differ from what the nucleus alone would compel, it is possible to draw some conclusions about the way in which the electrons are arranged. The mathematical difficulties, as the reader will readily admit, are tremendous; the problem of determining the path of a flying electron through a cloud of electrons, probably themselves in motion, is enough to make the best of mathematicians despair; yet some progress in this direction has already been achieved, as I narrated in the second article of this series. Again, the scattering of X-rays by atoms should depend on the manner in which their electrons are arranged; and some measurements and some deductions have already been made, although the researches have been in abeyance for some years, probably because the newest discoveries about X-ray scattering make it extremely doubtful what the mechanism of the effect really is.

The study of deflections of alpha-particles by atoms has thus brought precious guidance to the atom-builder, and imposed severe limitations upon him, yet only partial ones. He is constrained to erect his atom according to certain fundamental rules, and yet has an extremely free hand in arranging the details. He is practically compelled to build the atom of an element which occupies the Nth place in the periodic system, out of N electrons and a much more massive nucleus with a positive charge Ne. The data which I have cited do not absolutely enforce these numerical values; but there is no other model which they permit which could possibly rival this one in respect of convincing simplicity. He may not make the electrons go more than a few times 10^{-8} cm. from the nucleus; he is constrained to leave a small vacant space around the nucleus, and within this space he may not tamper with the inverse-square law of force (a restriction which has eliminated several favored atom-models of the decade before 1910).[11] Having conformed to these restrictions he

[11] Except that he may and must alter the inverse-square law of force to just the extent that further and more delicate experiments of this type require. Thus Bieler (*l.c. supra*) concludes, from a study of deflections of alpha-particles passing close to the nuclei of aluminium atoms, that within about 10^{-12} cm. of the aluminium nucleus the inverse-square repulsion which it exerts upon an alpha-particle is supplemented by an attractive force—perhaps an inverse-fourth-power attraction, just balancing the repulsion at a distance of $3.44 \cdot 10^{-13}$ cm. from the centre of the nucleus. Rutherford earlier found anomalies in the encounters between hydrogen nuclei and alpha-particles, which suggested to Darwin that the latter might be considered as a disc-shaped hard particle, or an oblate spheroid of semi-axes 4.10^{-13} and 8.10^{-13} cm.; this would repel hydrogen nuclei according to the inverse-square law so long as it did not actually strike them.

may do very nearly as he pleases with the electrons and the region they occupy. No data can be invoked to support him nor to confute.

Having expounded the merits of the nuclear atom, I will proceed to undo my work in part by pointing out its great and grave defect. No less a defect than this, that it is impossible. It cannot exist. Even if it were brought into existence miraculously at an instant, it could not survive, for it carries the seeds of its own dissolution within itself. For if at that initial instant all of the electrons were at rest relatively to the nucleus, they would immediately start towards it, fall into it, and expire. Of course, this consequence is so obvious that the notion of stationary electrons would not even occur to anyone having a bowing acquaintance with mechanics. Such a person would immediately assume that the electrons were in motion around the nucleus as the planets are around the sun; he would convert the nuclear atom-model into what I might call a *sun-and-planets atom-model,* the nucleus playing the role of the sun, the electrons those of the planets. Such an idea is alluring in the extreme; it implies that Nature acts similarly in great things and in small, copying the solar system within the atom; and this is most acceptable, partly because of its philosophical beauty and partly because it enables us to use the intellectual methods and habits acquired in the study of astronomy, relieving us of the labor of acquiring new ones. Unfortunately it is as untenable as the idea that the electrons stand still. For owing to the radiation of energy which continually goes on from accelerated electrified particles, an electron cannot describe a circle or an ellipse about a nucleus, as a planet may about the sun; it can only describe a narrowing spiral, ending in a collision between electron and nucleus. The nuclear atom is not stable nor enduring; and the dilemma is complete.

The only recourse is to make some entirely new and unprecedented assumption; for instance, that the electrons, in spite of everything, can stand still in certain positions without falling into the nucleus; or that they, in spite of everything, can revolve interminably in certain closed orbits without spiralling into the nucleus. Such a modification of the nuclear atom is, of course, essentially a denial of it. An atom composed of masses and electrostatic charges, plus certain restrictive rules or arbitrary assertions, is no longer simply an atom composed of masses and electrostatic charges. Instead of giving to our ultimate particles a few properties selected from among the ones which matter *en masse* displays to our senses or our instruments, we have to invent some new ones for them. This seems regrettable, but only because our expectations were too high.

Another circumstance leads us to another dilemma. Suppose that we could circumvent that difficulty about the revolving electron, which radiates part of its energy at each revolution and slides down a spiral path into the nucleus; suppose that we could find justification for saying that no radiation occurs, that the electron like a planet may revolve forever in an ellipse. If two atoms collided, as in a gas they must very frequently do, would not the electrons all be disarranged, disorganized, flung over from one orbit into another? This we should certainly expect; yet if it happens, no two atoms in a gas can be exactly alike, nor can any atom retain its character for more than a fraction of a second. If this is so, then the various sharply-defined properties of a gas must, each and every one of them, be statistical properties—not themselves properties of individual atoms, but the results of other properties of individual atoms, held in different amounts by different atoms and all averaged together. In some cases this is unobjectionable; the pressure and the temperature of a gas are sharply definite properties, resulting from the mass and the motion of the atoms, and the latter of these properties is not necessarily the same for any two atoms at the same moment nor for any atom at different moments. But one would be reluctant to treat the spectrum of a gas as such a property; according to all the traditions of physics this is one of the properties of the individual atoms. But the spectrum is very constant, sharp, immutably defined; we must therefore assume either that it depends only on the number of electrons in the atom and not upon their motion nor position, an idea which would be difficult to carry through; or that the electrons are ineluctably constrained to certain orbits or certain positions, so that the atom retains its personality and its character.

We have now made the acquaintance of two ideas which will be exceedingly prominent in the second division of this article. The nuclear atom-model is of itself unstable; therefore stability must be enforced upon it by outright assumption, it must be made stable by fiat. But this stability may not be extended to all conceivable arrangements or configurations of the model; it must be reserved for one or a few, that the atom may possess a fixed character and a personality.

We now arrive at the phenomena by means of which these vaguely-expressed ideas are to be sharpened and hardened into definite doctrines.

E. The Stationary States

E 1. The Direct Evidence for the Stationary States

Imagine a tube filled with gaseous helium, and containing a hot filament from which electrons emerge. By means of an accelerating potential applied between the filament and a fine-meshed gauze close in front of it, the electrons are speeded up, and pass through the gas with an energy which is accurately controlled by the accelerating-potential. A third electrode is maintained at a potential only slightly higher than that of the filament. To reach this electrode, the electrons must sacrifice nearly all of the energy which they acquired in coming up to the gauze. If they lose little or no energy in their progress through the gas, they can win their way to the third electrode, like water rising again to the level of its source. If, however, they lose a notable amount of energy to the atoms with which they collide, they cannot reach the third electrode, as water which has turned a mill-wheel cannot climb again to the level whence it fell.

By measuring the current into the third electrode in the helium-filled tube, it is found that if the electrons have an amount of energy lower than 19.75 equivalent volts, they lose scarcely any of it in their progress through the gas; but if the energy of an electron is just equal to 19.75 equivalent volts, it may and frequently does lose its energy altogether; and if the energy of an electron surpasses 19.75, it may and frequently does surrender just 19.75 equivalent volts to the gas, retaining the residuum itself. Imagining that the electron collides with atoms of helium on its way across the gas, we conclude that the helium atom can receive exactly 19.75 of these units of energy, no lesser quantity and (within certain limits) no greater. From similar experiments it appears that the mercury atom can receive 4.66 equivalent volts of energy, no smaller amount and (within certain limits) no larger. It appears that the sodium atom can receive 2.1 equivalent volts, no less and (within certain limits) no more—and the list can be extended to some thirty elements.

Another way of saying the same thing is this: the helium atom may exist (at least transiently) in its normal state, *or also* in a second state in which its energy is greater by 19.75 equivalent volts than in its normal state,—but not, so far as we can find evidence, in any state with any intermediate value of energy. Let us call this second state an "excited state." The mercury atom then has, in addition to its normal state of undefined energy, an excited state of energy greater by 4.66 equivalent volts. The sodium atom has, in addition to its normal state, an excited state of energy greater by 2.1 equivalent volts—and so with a number of others. I give these and a few other values in the following table:

TABLE I

	He	Ne	Na	Cs	Mg	Hg.
Energy-value of the						
Normal state.........	0	0	0	0	0	0
First excited state.....	19.75	16.65	2.1	1.45	2.7	4.66
Other excited states....	20.55	18.45			4.4	4.86 5.43 6.7
Ionized atom.........	24.5	21.5	5.12	3.9	7.6	10.4

It will be noticed that values are given for several excited states in the same column; these rest upon evidence of the same sort as does the first excited state, so that in general the atom must be considered to possess not one only, but several possible states in addition to its normal state.

It will be noticed also that values are given for the "ionized atom." These are the amounts of energy just sufficient (when applied by means of an impinging electron) to detach an electron from the atom. When electrons with so much energy or more are poured into the gas in question, positively-charged particles, such as I previously mentioned and characterized as the residues of atoms deprived of an electron apiece, appear in it. It is not absurd to call this an "excited state." If it takes just 24.5 equivalent volts of energy to detach an electron from a helium atom, then the system formed of an ionized helium atom and a free electron has a potential energy of 24.5 equivalent volts. Any experiment, therefore, in which the energy required to detach an electron from an atom is measured—any experiment for determining the *ionizing-potential*, as this energy when expressed in equivalent volts is called—is essentially an experiment for locating one of the excited states of the atom.

In this sense the energy-values of the last line in Table I are to be taken. I introduce them here for two reasons. In the first place, the fact that this energy-value is greater than any of the others in the same column suggests this interpretation for the excited states: that they correspond each to a certain *partial* lifting-out of an electron, to a certain stage of *incomplete separation*, while the energy-value of the ionized atom corresponds to the *total* lifting-out or to the *complete separation*. This idea is fortified by the fact that a helium atom may be ionized by two consecutive blows from electrons each with

20 equivalent volts of energy, if the blows fall closely enough together—as if the energy spent in raising the atom to its first excited state were paid into account, and could be used toward detaching the electron when the deficiency is supplied. This fact is exceedingly important for the theory, and I mention it here as a passing anticipation. In the second place it is desirable—for a reason which will presently appear—to measure the energy-values of the normal and of the excited states not from the energy of the normal state, as I have done in Table I, but from the energy of the ionized atom as zero-value. This is done in Table II.

TABLE II

	He	Ne	Na	Cs	Mg	Hg
Energy-value of the						
Ionized atom......	0	0	0	0	0	0
Non-ionized atom					−3.2	− 3.7
Excited states....	− 3.95	− 3.0				− 4.97
						− 5.54
First excited state.	− 4.75	− 4.85	−3.0	−2.45	−4.9	− 5.74
Normal state.....	−24.5	−21.5	−5.1	−3.9	−7.6	−10.4

With this convention, all the energy-values for the non-ionized atom become negative—a source of confusion, but not of nearly so much confusion as the previous convention would eventually entail. It is well to remember tenaciously that, in at least nine cases out of ten in the literature, the energy-values of the normal state and the excited states are referred to the energy of the ionized atom as zero, and that they all should always bear the minus sign, though generally it is left off.

For the excited states and for the normal state, I will employ the common general name of *Stationary States;* occasionally, for the sake of variety, the alternative name *levels*. Another common word is *term*, the origin of which will appear in the next section.[12]

As the reader will be forced to make himself familiar with schematic representations of the Stationary States, he may as well begin at once with a simple one. Fig. 1 is a diagram showing the stationary states listed for helium in the foregoing tables. The levels are represented by horizontal lines, separated by distances proportional to the

[12] Anyone who reads the physical literature of today soon becomes familiar with the phrase "the electron is in the . . . orbit" used instead of "the atom is in the . . . state." This phrase expresses theory rather than facts of observation, and does not always express theory adequately; I have avoided it in this article.

differences between their energy-values (usually, however, these distances are distorted for convenience). The energy-values, expressed in equivalent volts, are affixed to the lines; on the left, they are measured from the normal state of the neutral atom as zero of

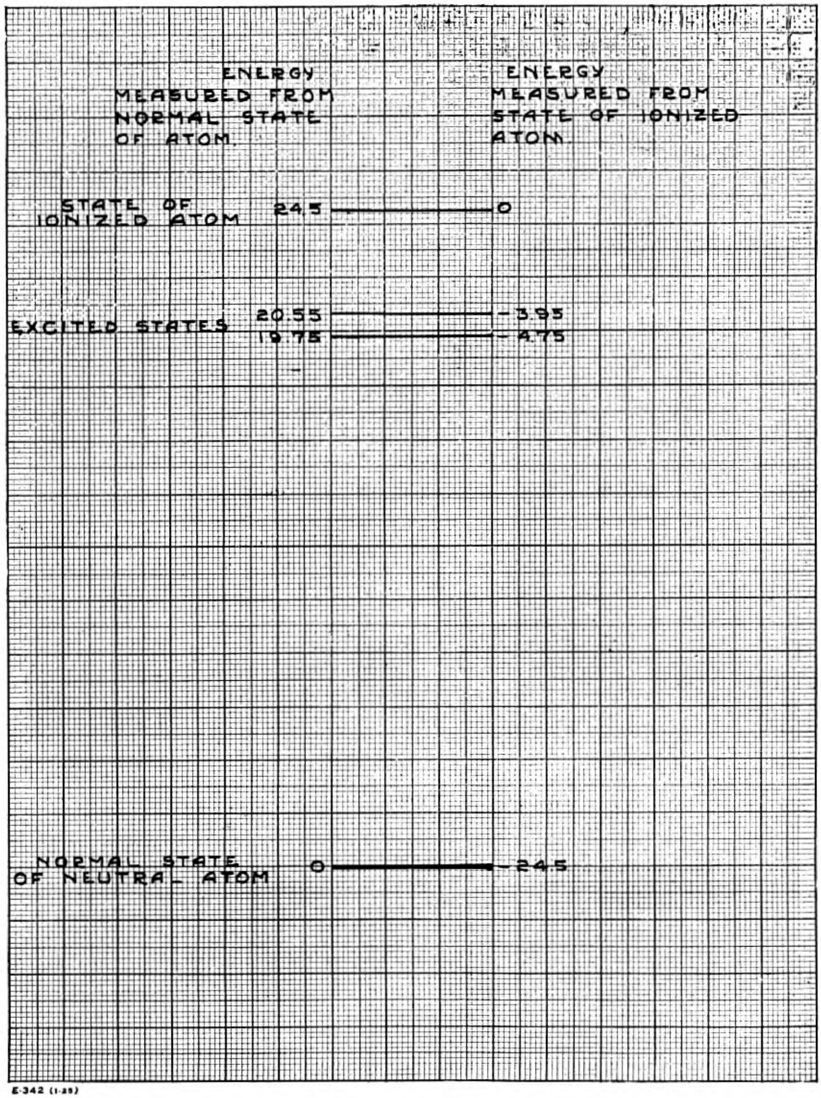


Fig. 1—Diagram of the stationary states of helium, determined by the method of electron-impacts

energy; on the right, they are measured from the state of the ionized atom (which is the more common practice)[13].

E 2. Bohr's Interpretation of Spectra

In 1912, the evidence to which the foregoing section is devoted was still entirely undiscovered, the Stationary States were unknown. That evidence was sought and found because Niels Bohr had divined the Stationary States in developing a new and brilliant interpretation of spectra. Until then, all physicists had wished to interpret the frequencies forming the spectrum of an atom as the natural resonance-frequencies of an elastic system. Bohr supplanted this idea with an idea of his own, one of the most novel, fecund and potent in all the long evolution of physics. Several of the ideas incorporated in the contemporary atom-model are due to Bohr; among them all this is the primary and fundamental one, and certainly the most secure.

Consider the spectrum of hydrogen. In the visible region, this spectrum consists of a "line-series"—that is to say, a procession of lines converging upon a limit, falling at intervals ever narrower and narrower, these intervals so smoothly diminishing that they bear witness to a common character and a mutual origin of all the lines.

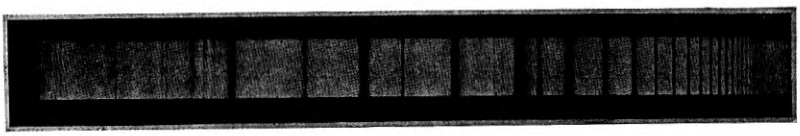


Fig. 2—Balmer series of lines in the hydrogen spectrum. (R. H. Curtiss, from Foote & Mohler, "Origin of Spectra")

This line-series is shown in Fig. 2. Not only to the eye is it of a wonderful regularity; the frequencies of its consecutive lines are bound together in a simple numerical law. They are equal successively to

$$\nu_{lim}-R/3^2,\ \nu_{lim}-R/4^2,\ \nu_{lim}-R/5^2,\ \nu_{lim}-R/6^2,\ \text{etc.} \qquad (1)$$

[13] This method of locating stationary states by observing transfers of energy from electrons at atoms is called the *method of inelastic impacts*; for the impacts of electrons against atoms are elastic (by definition) so long as there is no transfer of energy into the internal economy of the atom, and are inelastic when such transfers occur. When an atom returns into its normal state from an excited state, it usually emits radiation; hence a method for detecting the first commencement of radiation is usually (perhaps not always) equivalent to a method for detecting the first commencement of inelastic impacts. As it is generally easier to set up apparatus for detecting radiation than to seek evidence for elastic impacts, direct observations upon these last are not so abundant as they should be. Nobody really knows how many stationary states might be discovered by the method of inelastic impacts, although Franck and Einsporn detected over a dozen for mercury (of which those given in Table II are some). In fact they detected more than could conveniently be ascribed to mercury atoms, so that it was necessary to attribute some of them to molecules.

in which

$$\nu_{lim} = \text{frequency of the limit of the series} = R/4$$

R standing for a certain constant. There is another series of lines in the ultraviolet part of the same spectrum, whereof the frequencies are equal consecutively to

$$\nu = \nu_{lim} - R/4,\ \nu_{lim} - R/9,\ \nu_{lim} - R/16, \text{ etc.} \tag{2}$$

in which

$$\nu_{lim} = R,$$

R having the same value as before. The utter simplicity of the terms to be subtracted from ν_{lim} in each of these cases, not to speak of the related form of the expressions for ν_{lim}, suggests like simple laws in other fields of physics that in this formulation of the facts something highly important has been partially unveiled. There are certain other series in the spectrum of hydrogen, and inspecting them all one is led to the rule that *every frequency emitted by the hydrogen atom can be calculated by inserting different pairs of integers in the places of* m *and* n *in the formula*

$$\nu = R\left(\frac{1}{m^2} - \frac{1}{n^2}\right). \tag{3}$$

The case of the ionized-helium [14] atom is quite as simple. Every frequency emitted by this atom can be calculated by assigning different pairs of integer values to the constants m and n in the formula

$$\nu = 4R\left(\frac{1}{m^2} - \frac{1}{n^2}\right). \tag{4}$$

Line-series have been found in the spectra of many other elements. Some of them are as strikingly outstanding as the line-series in the spectrum of hydrogen, and converge upon limits scarcely less easy to locate; for instance, the "principal" series of the spectrum of sodium (Fig. 3). Most are by no means so obvious; often they are involved in the midst of a luxuriant jungle of unrelated or otherwise-related lines. Most spectra conceal their structures from the unpractised eye, as a tone-poem of Strauss its themes or an opera of the Ring its *Leitmotiv* from the inexperienced ear. Long training and a skilled judgment are required in the deciphering of spectra, except in the few untypically simple cases; and usually the arrangement of lines into series which the spectroscopist presents must be

[14] The reader may take this, for the time being, simply as the name of a particular element.

accepted by the theorist without question and without suggestion, for he is not competent to analyze the data for himself.

Having grouped a certain number of lines into a series, having guessed as well as possible the convergence-frequency ν_{lim} of this series, the spectroscopist has still the task of finding the numerical

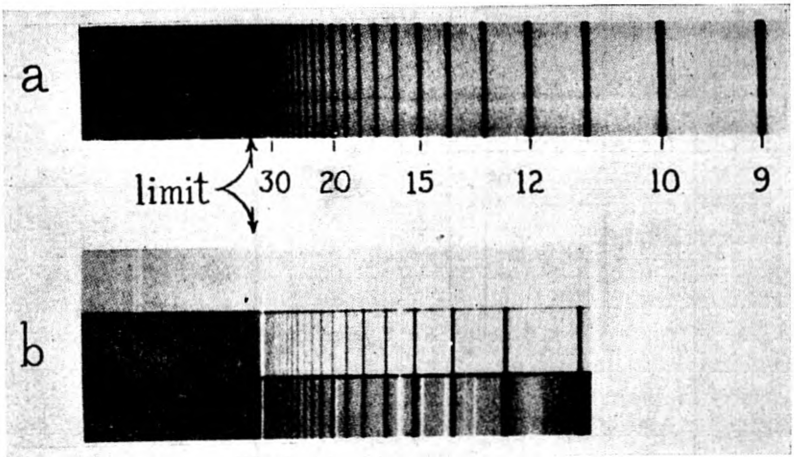


Fig. 3—Principal series of sodium (two photographs). (G. R. Harrison, *Physical Review*)

law to which the consecutive frequencies conform. As a matter of course, all the frequencies can be expressed by a formula generalized from (1) and (2):

$$\nu_i = \nu_{lim} - f(i) \tag{5}$$

in which i is the order-number distinguishing each line, and $f(i)$ is a different quantity for each of the lines, which approaches zero as we pass along the series to the limit. This means nothing by itself; the question is, does the function $f(i)$ have a simplicity comparable with the simplicity of the subtrahenda in (1) and (2) which suggested that they are the symbols of something deeply important? In general, the function $f(i)$ is not so simple as the function which occurs in the series of the spectra of hydrogen and ionized helium. In many cases, however, it is almost as simple, in others a little more complicated, in others a little more complicated yet, and so forth; so that the eventual result is this, that the formula (3) appears to be the proper way of describing the lines of series spectra, even in cases where the series is so irregular and the form of the function $f(i)$ so intricate that if it were the only series in existence, no one would attach any particular importance to it.[15]

To the physicists of a generation ago, who regarded the spectrum frequencies as natural vibration-frequencies of the atom, and tried hard to invent a mechanical model of which the vibration-frequencies should conform to the formula (3) or the more general formula (5), the character of these formulae was an insurmountable obstacle. Elsewhere [16] I have given a brief account of the vain attempts to contrive such a model. Bohr abandoned this procedure altogether; and taking equation (3), he multiplied both sides of it by Planck's constant h $(=6.56 \cdot 10^{-27})$.

$$h\nu = hR\left(\frac{1}{m^2} - \frac{1}{n^2}\right). \tag{6}$$

The significance of this act depends on the meaning of h. Planck had found it expedient, in developing an adequate theory of radiation, to assume that solid hot bodies are populated with multitudes of

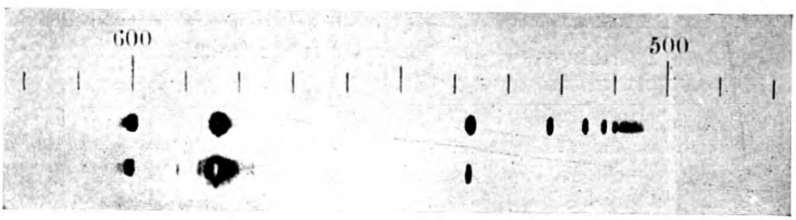


Fig. 4—Principal series of helium (singlet system). (T. Lyman, *Astrophysical Journal*)

oscillating electrons of all the various frequencies, possessing a very curious and inexplicable property; this being, that an oscillator vibrating with frequency ν can emit radiant energy of that same frequency ν only in units or *quanta* of amount $h\nu$. Einstein had found it expedient, in describing the photoelectric effect and other phenomena, to assume that radiant energy of the frequency ν goes about in units or quanta of the amount $h\nu$, emitted integrally, absorbed integrally, travelling integrally. Suppose then that we assume that the quantity $h\nu$, standing on the left-hand side of the equation (6), represents the amount of radiant energy emitted by the hydrogen

[15] As a matter of fact, the series-limit is not generally so obvious to the eye that it can be located at once; it is determined after and by means of a careful choice of the most suitable form for the function $f(i)$. This is one of the difficulties of the spectroscopist's task.

[16] In the seventh article of this series (footnote 9).

atom in the process of pouring out radiation of the frequency ν. The right-hand side is the difference between two terms. One term is the energy of the hydrogen atom before it emits the radiation of frequency ν; the other is the energy of the atom after the emission is concluded. *The radiation of frequency ν is emitted by reason of a transition between two stationary states of the hydrogen atom; the energies of these states are equal to the terms whereof the frequency ν is the difference, each term multiplied by h.* The terms of the spectrum formulae are the energy-values of the stationary states of the atom, when translated into the same units by multiplying them by h. When translated into proper units, *the terms are energies, and the energies are terms*. This is Bohr's great and memorable idea.

Once this idea is accepted, the known stationary states of the atom increase enormously in number. The paltry one, two, or half-dozen, which are all that have been detected by observing the energy-losses of rebounding electrons, are multiplied into hundreds and thousands. The accuracy with which each energy-value is known is augmented tenfold or a hundredfold, sometimes far more; for spectroscopic measurements are among the most accurate in physics, although the necessity of extrapolating the observed frequencies to arrive at the series-limit neutralizes some of their precision.

One point must be kept clearly and always in mind, at the peril of infinite confusion. *The energy-values which the spectrum terms supply are* not the energy-values of the stationary states measured from the normal state, as might seem natural; they are *the energy-values measured from the state of the ionized atom.* These being negative, it is the negative term-value which is significant. Equation (6) must therefore be rewritten in this fashion:

$$h\nu = Rh\left(-\frac{1}{n^2}\right) - Rh\left(-\frac{1}{m^2}\right). \tag{7}$$

The energies of the successive stationary states of the hydrogen atom are $-Rh$, $-Rh/4$, $-Rh/9$, $-Rh/16$, and so forth, relatively to the energy of the ionized atom as zero. They are not Rh, $Rh/4$, $Rh/9$, and so forth, relatively to the normal state of the atom as zero. Anyone who entertains this last idea is doomed to trouble.

The stationary states of the hydrogen atom are shown in Fig. 5, which is constructed like Fig. 1, with the energy-values of the various levels measured downwards from the state of the ionized atom, and affixed on the right. The distances from the various levels to the zero-line are proportional to these energy-values (this feature will henceforth be found too inconvenient to maintain).

The energy-value of a stationary state, when obtained by analyzing a spectrum, is generally given not in equivalent volts, but in a unit called the "wave-number." This unit is $1/hc$ times as great as an erg, and $300hc/e$ (about .0001237) times as great as an equivalent volt. When the energy-values of two stationary states are expressed in

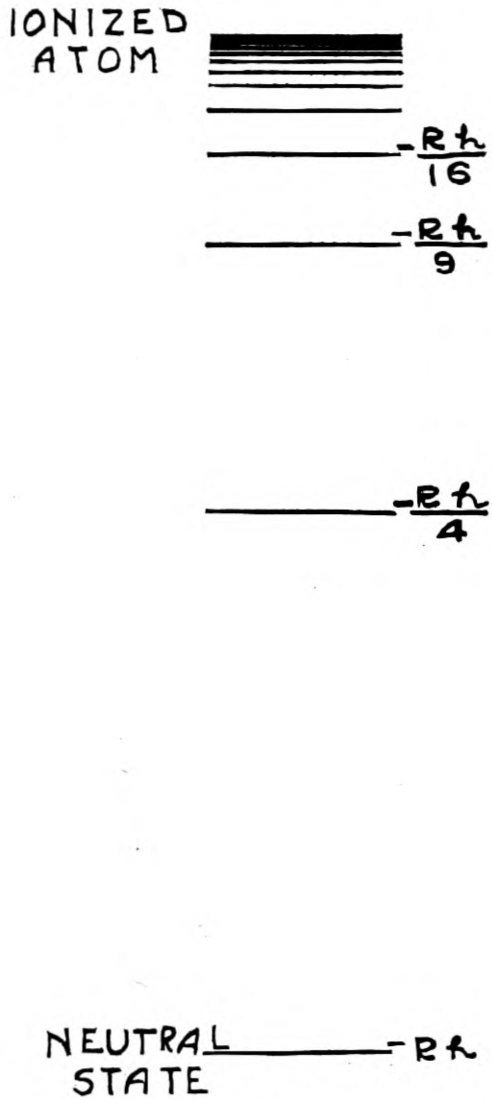


Fig. 5—Diagram of the stationary states of hydrogen, deduced from its spectrum

this unit, the difference between them is equal to $1/c$ times the frequency of the line which corresponds to the transition from one to the other.

A spectrum-line corresponding to a transition between two stationary states is symbolized, on a diagram of stationary states, by an arrow connecting the dashes (or whatever marks are used) which symbolize the two levels. This is illustrated in Fig. 6.

I pause at this point to remark that each of what I have been calling the "stationary states" is in fact usually a group of two or more stationary states, often but not always exceedingly close together; just as many stars in the sky are in fact groups of stars too close together for any but an excellent telescope to discriminate. This will be discussed at length in a later section; at present it is expedient to regard each of these groups as one stationary state.

The experimental test of Bohr's method for identifying stationary states consists in comparing the stationary states inferred from the spectrum, according to Bohr's procedure, with the stationary states derived directly by the study of electron-impacts. The agreement is perfect wherever the experiments by the latter method can be carried out. By a curious fatality, this is impracticable for hydrogen and ionized helium, as neither sort of atom occurs in gas quiescent enough for experiments on energy-transfers from electrons to atoms.

For about fifteen other elements, the comparison has been made for two or more of the Stationary States. Every energy-value given in Table II was obtained by the method of electron impacts, and confirmed by analyzing the spectrum of the element.

E 3. The Classification of Stationary States by Utilizing "Rules of Selection"

I have said that every line in a spectrum, at least of those arranged in series, may be represented by an arrow connecting two stationary states. If arrows are drawn from every one of the stationary states to every other, will every arrow correspond to a line actually observed in the spectrum? Every line has an arrow; does every arrow have a line? By no means; the answer is definitely and strongly negative. If the wave lengths deduced from all the possible arrows are sought in the spectrum, most of them are found unoccupied by lines. The great majority of the apparently possible transitions either do not occur at all, or if they do occur, the energy which is liberated is disposed of in some other way than by radiation. There is reason for believing that the atom may embrace this last alternative if it col-

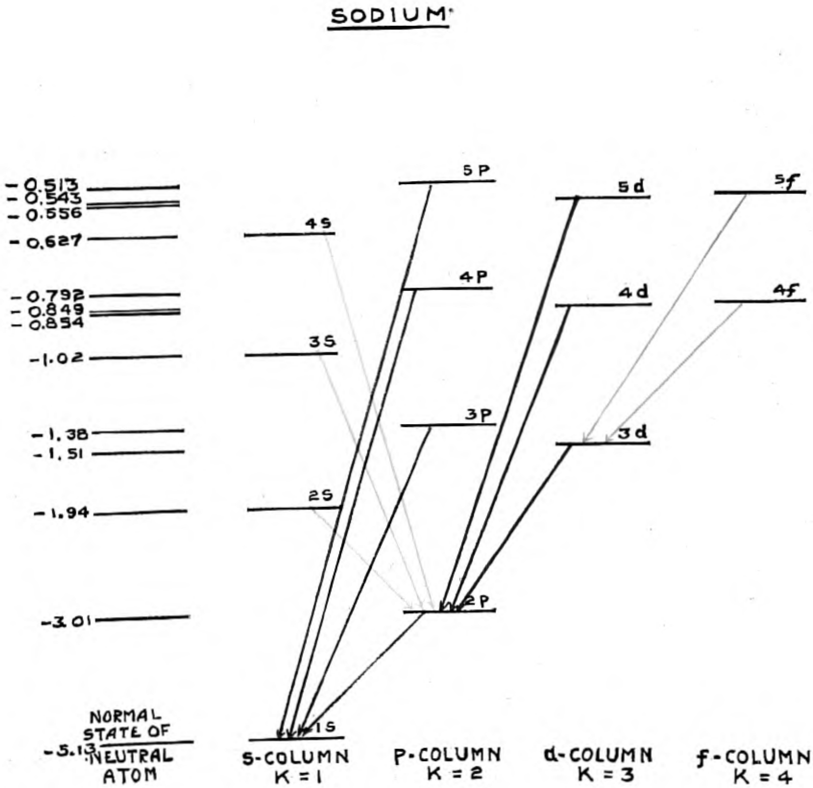


Fig. 6—Diagram of the stationary states of sodium, sorted out into columns by applying the selection-principle. Arrows represent various lines (blue for principal, yellow for sharp, red for diffuse and green for Bergmann series)

lides with another atom or with an electron. Otherwise, it seems that if the atom cannot radiate the energy liberated in a transition, the transition itself cannot happen at all. If, therefore, the line corresponding to an arrow is missing, the transition corresponding to the arrow must be inhibited by some agency as yet unknown. Many transitions must be inhibited, for many lines are missing.

These missing lines are precious to the student of spectra and to the architect of atom-models. Whatever explanation is devised for the stationary states must include a reason for the occurrence of some transitions and the non-occurrence of some others. This is good rather than bad fortune, since if such a reason is demanded, it may be found in one and not in another of two competing theories which otherwise would stand on an equal footing; the missing lines may even suggest a theory. At all events they suggest a system of classification; and, while the hardened theorizer may regard a system of classification as merely the forerunner of a theory, a theory is itself often nothing more than a classification stated in the language of an artificial analogy. It is, in fact, possible to arrange the stationary states, not in a single column as in Figs. 1 or 5, but in several as in Fig. 6; this arrangement being so contrived, that any transition can be identified in a moment as belonging among those which occur, or among those which are missing, whichever its case may be.

The mere fact that such an arrangement can be contrived shows that the missing lines are not distributed at random, but subject to some sort of a rule. Such rules are known as *principles of selection.* The missing lines are commonly called *verboten* lines by the German physicists, possibly because that was the most conspicuous word in the official German language before the war. It is not a happily-chosen word, neither are the English equivalents "forbidden" and "prohibited"; since while we know that the lines are missing, we do not definitely know what circumstance is responsible; and, whatever that circumstance may be, it is highly unconventional for a physicist to say that it "forbids" the lines. The same objection applies with extra force to the phrase "forbidden by the selection-principle". It is much better to accept the fact that certain lines are missing as a fact of experience, and the selection-principles as rules of experience whereby the facts are codified.

E 4. The Families of Stationary States (for Other Atoms than Hydrogen)

There is a far-reaching contrast between the spectra of all atoms but hydrogen and ionized helium, on the one hand, and the spectra of these two atoms on the other. The selection-principles at first accentuate this contrast, and later to a certain extent aid to explain it away. I commence with the atoms other than hydrogen, and take sodium as the specific instance.

A few of the stationary states of the sodium atom are exhibited in a single column on the left of Fig. 6. The energy-value of each level, measured from the energy of the ionized atom as zero, is affixed at the left; but the practice of drawing the levels at distances proportional to their energy-values has had to be discarded for the sake of lucidity. In this case, the distances are proportioned to the differences between the logarithms of the energy-values. Drawing arrows from each of the levels to every other, and ascertaining which of them correspond to actual and which to missing lines, we find that the missing lines are such that the stationary states can be sorted out into several families, to be arranged in parallel columns as on the right of Fig. 6. There are at least seven of these, but it is of no advantage to us to consider more than the first four. The feature of this arrangement is, that *transitions between stationary states in adjacent columns correspond to actual lines; but the lines corresponding to all other transitions are missing.*

This is a principle of selection. It may be phrased in an equivalent but pregnant way, in this manner. Let me attach to the several columns the numerals 1, 2, 3, 4 . . . , as they are indicated at the bases; and let me use k as the general symbol for each and all of these numerals. Then this particular selection-principle may be phrased thus:

> *The only transitions which correspond to actual spectrum lines are those in which* k *changes by unity*; $\Delta k = \pm 1$.

The numeral k bears the ponderous name of *azimuthal quantum-number*. This is a name derived from theory and not from experience, as will be made clear in due time. The principle of selection which has just been stated is the selection-principle for the azimuthal quantum-number.

Exceptions to this rule occur; the *verboten* lines, like other *verboten* things, occasionally evade the prohibition. This happens particularly when the atoms are subjected to intense electric fields, or to violent spasmodic electrical discharges in which strong transient fields are produced; in these circumstances great numbers of the missing lines leap suddenly into sight. In Fig. 11 some of these lines appear elicited by a strong electric field. Some lines corresponding to changes of k by two units or by none, which by the foregoing rule should be absent, do actually occur even when there is no obvious reason whatever for thinking that the atoms are subject to unusual stresses.[17] The exceptions, however, are not numerous enough to jeopardize the rule.

Two other features of the columns should be pointed out; first, that the successive levels in each column are not scattered at random, but form a converging series approaching the top of the column as limit (their energy-values form a sequence converging to zero); and second, that there is nothing arbitrary about the order of the columns, since the column at the extreme left admits of transitions to only one other column and therefore is unmistakable, and all the others follow after it in an immutable order.

E 5. A Digression About Notation

The symbol for a transition between two stationary states, and for the spectrum line which corresponds to that transition, consists of the symbols for the two states separated by an arrow, or a dash, or a semicolon, or any convenient mark. The final state is commonly written first. Thus the line due to the transition from a state B to a state A is designated thus: $(A)-(B)$. Chess-players will be reminded of the "Continental" system of describing moves at chess, in which symbols for the squares from which and to which the piece is moved are written down one before the other.

The notation for spectrum lines thus flows easily and naturally from the notation for stationary states. This notation is not in principle very difficult, but it has become confused and confusing, largely because of the alterations which have been wrought upon it to make it express not the facts, but divers theoretical interpretations of the facts. Alterations in names and notations generally produce an evil effect in physics even when justified in the highest degree, for the old systems and the new persist side by side and cause interminable trouble; all the more is this so when the alterations are based on uncertain grounds and impermanent. The notation for stationary states has already suffered much in this manner, and probably the worst is yet to come.

The classification of levels which I have just described enables and requires us to give a twofold symbol to each level; the symbol must designate the column in which the level stands, and its order-number or serial number in that column. The columns are generally desig-

[17] Foote, Meggers and Mohler observed a line corresponding to a change of two units in k (the line $(1,s)-(3,d)$, in the notation to be explained in section E5) under circumstances in which it seemed impossible to believe in an abnormally large electric field.

nated by the letters s, p, d, f (or their capital, or Gothic, or Greek equivalents).[18] A spectroscopist using these symbols generally writes the serial number of the level before the letter, with a comma between, thus: $(1,s)$ and $(2,p)$ and $(3,d)$. Or the columns may be designated by their values of the numeral k, which is then commonly written as a subscript to the serial number. These symbols have at least the advantage of being comparatively fixed. It is far otherwise with the serial numbers. One might expect that the level having the greatest energy-value in a particular column would be called Number 1, and the successive ones Number 2, Number 3, and so forth towards the convergence-limit. Unfortunately (though for not a bad reason) the habit is to designate the first levels of the successive columns by the order-numbers 1, 2, 3 and 4, successively; so that their respective symbols are $(1,s)$; $(2,p)$; $(3,d)$ and $(4,f)$. These are the symbols I have affixed in Fig. 6; but they are not the only ones, as the order-numbers have jumped up and down several times to satisfy the exigencies of new atom-models. It would be unprofitable to confuse the reader with further details, at least at this point. The important things to remember are three: that the symbol for each stationary state must contain one index for its column and another for its place in its column—that the former index is usually one of the specified letters—that the latter index is a number, usually beginning with 1, 2, 3, 4 for the first level in the s, p, d, f columns, respectively, and ascending along the column in unit steps.

E 6. Names and Features of the Most Noted Line-Series

Every line in every series, according to Bohr's fundamental idea, corresponds to a transition or "combination" between two stationary states of the atom—to a transition from an initial state to a final state. The atom possesses more energy in the initial state than in the final state (we are speaking of emission-spectra only). Hence the energy-value of the initial state, reckoned as it usually is from the energy of the ionized atom as zero, is algebraically higher and arithmetically lower than the energy-value of the final state.

The various lines of any one line-series have this in common: they correspond to transitions from various initial states which however all lie in one and the same column, into one final state which is the same for all and lies in an adjacent column. Each line-series thus

[18] The symbol b is sometimes used instead of f. For the columns following to the right of the f-column there are various notations, particularly f', f'', f''' and g, h, i. See also footnote 21.

belongs to one particular final state, and to one particular column of initial states.

The line-series consisting of transitions into the state (1,*s*), or *terminating upon* (1,*s*) as the phrase sometimes is, bears the name of *principal series.* Its consecutive lines are: (1,*s*) – (2,*p*); (1,*s*) – (3,*p*); (1,*s*) – (4,*p*) and so forth. They are signified by the blue arrows of

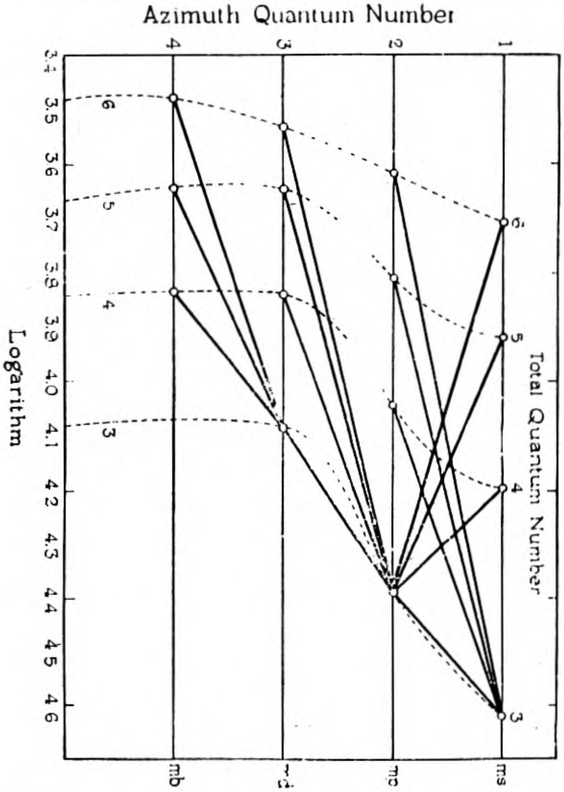


Fig. 7—Another way of mapping the stationary states of sodium

Fig. 6. The general symbol for this series is (1,*s*) – (*m*,*p*); which will be quite intelligible. The (1,*s*) level is the normal state of the atom; consequently, the various lines of the principal series correspond to transitions, by which the atom regains its normal state after a temporary exile from it. It is probably for this reason that the series is prominent enough to have received the name *principal* from the spectroscopists.

Two series terminate upon the (2,*p*) level. One of these consists of transitions from various levels of the s-column. This is the *sharp* (or second) *subordinate* series, and its symbol is $(2,p)-(m,s)$. The other series consists of transitions from various levels of the d-column; it is the *diffuse* (or first) *subordinate* series, and its symbol is $(2,p)-(m,d)$. Yellow and red arrows signify these series, respectively, in Fig. 6. Of the two line-series terminating upon the $(3,d)$ level, only one has been endowed with a name; this is the series $(3,d)-(m,f)$, known alternatively as the *Bergmann* or the *fundamental* series (the second name is a bad one) and symbolized by green arrows in Fig. 7.

These series seem to be the only ones which impressed themselves strongly enough upon the minds of spectroscopic experts to receive names [19] from them. However, many other series have been identified, and emphasized, especially since Bohr's manner of thinking took root among the students of spectra; for instance, series terminating upon $(2,s)$ and $(3,s)$, which are conspicuous in the spectrum of helium, and such line-series as $(3,d)-(m,p)$, and $(4,f)-(m,d)$.

Several rules about line-series, which are very prominent in accounts spectra, become self-evident when the rules governing the stationary states are mastered (of course, this is only because the latter rules are based upon the former). For instance, there is a rule that the sharp and the diffuse series have the same limiting-frequency; and there is a rule that the difference between this limiting-frequency and the limiting-frequency of the principal series is equal to the frequency of the first line of the principal series. The reader may derive these by inspecting Fig. 6.

Such rules do not apply to the spectra of hydrogen and of ionized-helium, which are profoundly different from the spectra of sodium and other elements; and it is perilous to attach such names as *principal* or *subordinate* to the line-series of those first elements. The stationary states of those elements are known by their energy-values, and the series by the names of their discoverers or interpreters.

E 7. *Further Analysis of the Stationary States of Hydrogen and Ionized Helium; Fine Structure*

In our earlier analysis of the spectrum of hydrogen and the spectrum of ionized helium, we inferred from each of these spectra a family of stationary states, the energy-values of which follow one upon the other in a very regular procession governed by a simple numerical law. This makes it practically impossible to divide up these stationary states into classes; all of the levels for each of the atoms must

[19] The reader will recognize, in the initials of these names, the letters s, p, d, b, and f used to designate the several columns of levels.

inevitably be arranged in a single column, as it was done in Fig. 5. But in this arrangement the selection-principle of the foregoing paragraph is apparently contravened. For, when the levels of the sodium atom were arranged into columns, the transitions between levels belonging to one and the same column were among the inhibited transitions, the lines corresponding to these were among the missing lines. But the transitions between the levels in the single column which contains all of them for the hydrogen atom, correspond to the actual lines which constitute the entire hydrogen spectrum.

This discord is only apparent. It vanishes when we recall the fact, already once mentioned as a forewarning and then neglected for ease of exposition, that the stationary states of the hydrogen atoms are compound—that what has been called a "stationary state" in the preceding pages is really an ensemble of adjacent stationary states. Every line of the Balmer series, the series $R(1/m^2-1/2^2)$, is actually a close doublet; the frequency-differences between the components of all the doublets are approximately the same. Interpreted in the new fashion, this means that what we have called the stationary state of energy $-Rh/4$ is actually a pair of "component" stationary states very close together—so close together, that if the energy of one were exactly $-Rh/4$, the energy of the other would depart from that value by less than one part in forty thousand. Further in analyzing the spectrum of hydrogen we cannot go, probably because the minute details (if there are any) of the structure of its lines overtax the resolving-power of our spectroscopes. The spectrum of ionized helium, however, is spread out in a more generous scale; and some of its lines were analyzed by Paschen. Among these were the lines of frequency $4R(1/3^2-1/4^2)$; $4R(1/3^2-1/5^2)$; and $4R(1/3^2-1/6^2)$. They were resolved respectively, into six, five, and three components; and the line $4R(1/4^2-1/5^2)$ resolved into four.

Interpreted in the new manner, these data mean that what we have called the stationary states of energy-values $-4Rh/9$, $-4Rh/16$, $-4Rh/25$, and $-Rh/36$, are really ensembles of "component" stationary states lying very closely together. It would scarcely be possible to infer from these data, independently and without extraneous guidance, just how many "components" belong to each of the four ensembles. Fortunately or unfortunately, Paschen's measurements were preceded and inspired by a specific prediction of the number of components in each ensemble—a prediction that what we have called the nth stationary state should be a group of n "component" stationary states. This prediction is graphically set forth in the second column of Fig. 8, in which the level of energy-value

$-4Rh$ is drawn as a single dash, the next level as two dashes, the next as three, and so forth. Paschen's data were therefore compared with this prediction.

The data and the prediction were found compatible. If arrows are drawn from every "component" stationary state to every other "component" stationary state, it is found that each of the lines which was observed corresponds to one of the arrows (but it is necessary to assume that, in some places, two or more adjacent lines are fused apparently into one by reason of the insufficient resolving-power of the spectroscope). Some of the arrows, however, correspond to missing lines. Evidently some sort of inhibiting agency is at work; some sort of a selection-principle is adumbrated. Furthermore, some and perhaps all of the missing lines appear when the electric field strength acting upon the radiating atoms is increased, and this, it will be remembered, is the behavior of the missing lines in the sodium spectrum. Whether the selection-principle could ever have been inferred from these data alone seems doubtful. Naturally one proceeds to try out the same principle as served for the previous case. Can the component stationary states of the ionized-helium atom be sorted out into parallel columns, in such a manner that transitions between levels in adjacent columns correspond to actual, all the other transitions to missing, lines?

This is attempted in the manner shown in Fig. 8. The result is fairly satisfactory. The lines due to transitions between levels in adjacent columns should by this principle be visible, and they are. The lines corresponding to transitions between levels in the same column, or more than one column apart, should be missing; and some of them are, but also some of them undeniably can be seen. To account for these unwelcome guests, it is necessary to assume that some of the radiating atoms are subject to a strong electric field which might, but would not be likely to, exist in the discharge. This is an uncomfortable solution; but there are other numerical agreements between the prediction and the data, which it is not expedient to describe at this point, but which are good enough to excuse that deficiency to some extent. *En somme*, the evidence presents no insuperable objection to our arranging the component stationary states of the ionized-helium atom in parallel columns, and declaring that the only transitions which occur (except in strong electric fields) are those between members of adjacent columns; and this is just what we did with the sodium atom, and can in general do with every other kind of atom whereof the spectrum has been interpreted. This being granted, we can assert that the spectra and the stationary states of the ionized-helium atom (and presumably those of the hydrogen atom) are not so radically different from those of the sodium atom as they seemed to be; some of the apparent differences being traceable to the fact that corresponding levels in the *f*, the *d*, the *p*

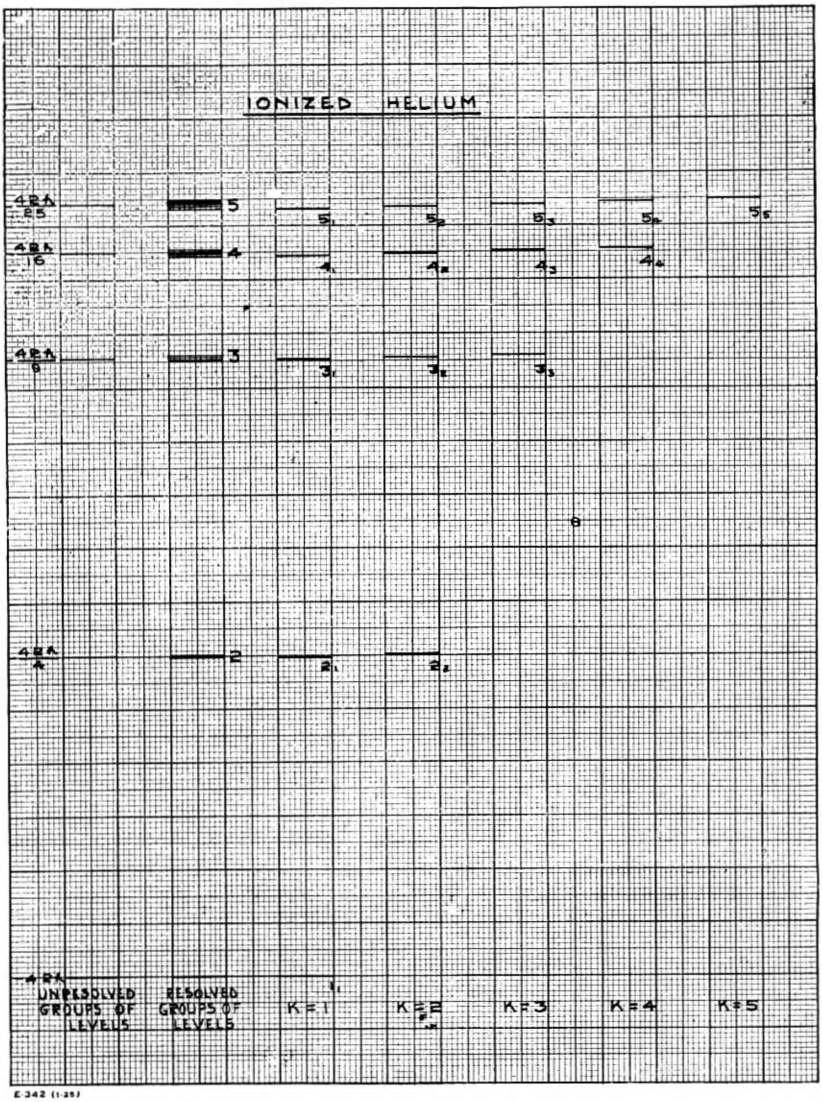


Fig. 8—Diagram of the stationary states of ionized helium, resolved to account for the fine structure of the spectrum lines

and the *s*-columns, which in the sodium atom are widely separated, are in the former atoms so closely crowded together that lines, which in the sodium spectrum are far apart, are in the former spectra packed into all-but-irresoluble groups. This is probable, but not certain. Further data about other lines in the ionized-helium spectrum would be gratefully received.[20]

The notation for the various "component" stationary states of the ionized-helium atom is shown in Fig. 8. The successive columns are denoted by the numerals 1, 2, 3, 4 . . . for which the general symbol is k, as previously. This numeral is written as a subscript to the serial number of the level in its column, which commences with 1 in the first column, 2 in the second, 3 in the third, and so forth. By inspecting the figure, the reader will see a reason for using these different values of the serial-number for the first levels of the different columns. The serial-number is designated by n and called the *total-quantum-number*. The numeral k is called the *azimuthal-quantum number*, as before. These heavily long names are imposed by the theory and not by the data.

E 8. Further Analysis of the Stationary States of Other Elements than Hydrogen and Ionized Helium; Multiplets

Having performed a two-stage analysis of the spectra of ionized helium and of hydrogen, we return to the spectra of the other elements for a second attack.

Let us consider the reasons for making these analyses in two stages. When the mid-Victorian physicist trained his spectroscope upon a tube full of glowing hydrogen, he saw the spectacle of Fig. 2—the converging procession of distinct bright lines, of which the frequencies form that delightfully smooth numerical progression which we have already met. Later physicists with better instruments discovered that each of these "lines" was in fact a pair of lines. Now in strict truth, this discovery showed that the "lines" of the Balmer series were no lines at all; for a doublet is not a line. But the physicists continued to refer to the "lines" of the Balmer series, chiefly no doubt because to anyone equipped with an ordinary spectroscope the doublets do appear as single lines. By itself this is little reason; but the usage is not altogether faulty. Few people would hesitate to admit that each of these doublets is not a couple of casual neighbors, not two

[20] It would be particularly interesting to settle beyond question whether the missing lines demand the selection-principle already explained in section E4, rather than the one to be explained in section E8. This is one of the reasons for wanting to produce and examine the spectrum of doubly-ionized lithium, in which the evidence would probably be much clearer.

unrelated lines fortuitously close together, but a pair of lines sharing some deeply fundamental quality in common. This is indicated chiefly by the facts that the distance (measured in frequency) between the components of a doublet is the same for all the doublets, and very small compared with the distance between consecutive doublets. For this reason the doublets are treated as entities, and they require a name; which is what physicists have preserved for them, in continuing to call them "lines." "Doublet" would be better than "line", and "group" would be better yet; but we cannot ever be sure that even the apparently-single lines are not very close groups, and yet it would be silly to call every line a group. Sirius appears as a double star in a few of the most powerful telescopes, but nobody would insist on calling it a double star when pointing it out in the night sky.

All this is not so trivial as it sounds. It is easy enough to speak of doublets when looking at lines which appear single except when viewed in the most powerful spectroscope, and then are resolved into components much closer together than the nearest similar line is to either. Such lines occur not in the spectra of hydrogen and ionized helium only, but in the spectra of sodium and other elements generally. But the spectroscopist is constantly applying such names as "doublet" and "triplet" and "quadruplet", and the inclusive name "multiplet" to groups of lines which lie far apart in the spectrum, with scores of others intervening. Here his function is not to split apparent lines into narrow groups, but to unite widely-scattered lines into wide groups. This he does not because of propinquity of the lines, but because of resemblances or analogies or fixed intensity-relations between them, or because he finds it possible to construct a series of such groups with identical frequency-differences between corresponding lines within them, or because of analogies with other elements with more perspicuous spectra, or theoretical predictions, or intuitions or clairvoyance. Groups such as these are not generally termed lines, except in very abstract discussions; it is difficult to call a group a line, when it is clearly resolved by any instrument worthy the name of spectroscope. But they are like the lines of the Balmer series, treated as entities because their lines are believed to share some deeply fundamental quality in common.

What I have said about lines and groups of lines is transferable in substance to stationary states and groups of stationary states. What we had originally called the levels of hydrogen and ionized helium, with their energy-values $-Rh/n^2$ and $-4Rh/n^2$ ($n=1, 2, 3 \ldots$), were resolved into groups of levels in order to interpret the fine structure of the lines. But owing to the propinquity and to certain numerical relations of the levels in a group, and to certain qualities of the transitions between them, it was felt that the levels of each group share some deeply fundamental quality in common. For this reason we used a system of classification in which each level is represented by two symbols, one for its group and one for its place in its group; and we numbered the levels in succession, not 1 and 2 and 3 and 4 and 5 and so forth, but 1_1 and 2_1 and 2_2 and 3_1 and 3_2 and 3_3 and so forth. Interpreting the groups of lines in the spectra of sodium and other atoms, we infer groups of levels. The levels in one of these groups are often far apart. They may be eighteen or more in number, other levels may lie between; but by reason of the resemblances between the lines whence they were inferred, by reason of certain numerical relations between the levels themselves, they are believed to have some deeply fundamental quality in common. If this is vague, so also at times is the interpretation.

The statements in the foregoing sections about the stationary states of sodium are now to be understood as relating to groups of stationary states. It is the *groups of stationary states which are arranged in parallel columns, designated by numerals* k, *such that no transition takes place unless in it* k *changes by one unit.* It is the group of stationary states which is marked by a pair of numerals, one to designate its column and the other its place in its column; or by a letter to designate its column and a numeral to designate its place in its column. It is the group of stationary states which is denoted by (3_2) or $(1,s)$ or $(5,d)$.

To denote a particular stationary state we must add, to the symbols for its group, a third symbol for its place in its group. This symbol is generally a numeral, hung on as a subscript to the letter designating the column (thus: $(2,p_1)$ and $(2,p_2)$) or as an additional subscript to the two numerals (thus: 3_{21} and 3_{22}).[21] The most common general symbol for this numeral is j. Geometrically, the stationary states may be represented by lines or dots arranged, not in one row of several parallel columns as in Fig. 7, but in several rows of parallel columns. Readers with three-dimensional imaginations in good working order may develop this idea *ad libitum*. The systems for assigning the values of j are shifted around every few months to correspond to new atom-models, and are scarcely worth memorizing.

[21] The notation suggested by Saunders and Russell, evidently in concord with a number of other experts, is built in this way: Designate the column to which a group belongs by the letters suggested in section E5, capitalized (i.e., S, P, D, F, G, H for $k=1, 2, 3, 4, 5, 6$); write the serial-number of the group before the letter, and append the value of j as a subscript to the letter. If it is desired to state what sort of a system (cf. section E10) a level belongs to, one may add an index to the left of the letter and above it.

The best of them, however, are adjusted so as to express a new and additional selection-principle, which is coequal with the other selection-principle we met a few pages above.

This principle is derived in the same way as the first one. The groups of levels are established by inference from the groups of lines; then arrows are drawn from every level to every other, the corresponding spectrum-lines are sought, and most of them are not found. Some of these missing lines are those which would contravene the first selection-principle, as they correspond to transitions in which the numeral k changes by more than one unit, or not at all. Putting these aside, there are still a number of missing lines, to which the first selection-principle has offered no objection. Now it is found possible to choose the numeral j in such a manner that the only transitions which correspond to actual spectrum lines are those in which j changes by one unit or not at all ($\Delta j = O, \pm 1$). Furthermore it is possible to adjust the values of j in such a manner that the lines corresponding to transitions, in which j is initially zero and remains unchanged, are missing.

This is the *selection-principle for the inner quantum number*; for the numeral j, when adjusted in this manner, is known as the inner quantum number. This again is a name imposed by theory and not by the data of experience.

As the two selection-principles are effective concurrently, the pair of them may be fused into this one:

Of the three numerals n, k *and* j, *which specify a stationary state completely, two* (k *and* j) *may be so chosen that the only transitions which correspond to actual lines are those in which*: *first,* $\Delta k = \pm 1$; *second,* $\Delta j = O, \pm 1$; *third,* j *is not zero both before and after the transition.*

This complicated rule is evidently the sign of some very important principle, the full nature of which thus far escapes us. It will probably seem difficult to grasp and fix in mind; but difficulty of this sort is likely to abound in the physics of the near future. Not so many years ago the physicist's path lay among differential equations; the defter he was in integrating hard specimens of these, the better he was fitted for his profession. I should not care to say that this is no longer true; but he will probably have to cultivate a sense for problems such as this.

It remains to give some idea about the number of stationary states in the various groups. For sodium, as laid out in Fig. 6, the groups in the s-column are merely single levels (this sounds like a contradiction in terms, but may be borne for the sake of the generality); the groups in the other columns are pairs of levels, or "doublet terms."

This is the common character of the alkali elements Li, Na, K, Rb and Cs, which occupy the first column of the periodic table; probably also for the noble metals which share this column, but the data are few. For elements of the second column of the periodic table there are two complete systems of stationary states, each having its own s-column, its own p-column, its own d-column, and all the rest. In one system, all the groups in every column reduce to single levels; it is a *singlet system*; in the other, all the groups in the s-column are single levels, all the groups in the other column are triads of levels or "triplet terms;" it is a "triplet system." The complexity mounts up stage by stage as we cross the periodic table of the elements from left to right, and soon becomes terrific.

E 9. Effect of Magnetic Field on the Stationary States

When a magnetic field is applied to a radiating gas, most of the lines of its spectrum are replaced by triplets (Fig. 9), or by even richer groups of lines (Fig. 10). By a somewhat loose usage the lines are said to be *resolved* into three or more components. This is the "Zeeman effect." There is a multitude of empirical rules about these components, their spacings, the way in which their number and their spacings vary from one line to another, and other features. According to the new fashion, however, we focus our attention not on the component lines, but on the stationary states which are inferred from them.

The effect of a magnetic field may be described by saying that it replaces each stationary state (with a few exceptions) by two or more new ones. Each of these new states requires four symbols to designate it; the symbols n, k and j for the original stationary state, and a new symbol m to denote its place in the resulting group. As heretofore, when every stationary state is connected with every other by an arrow and the corresponding lines are sought, it is found that some of the lines are missing. Still another selection principle is therefore to be sought, and the values of the new numeral m are to be so adjusted—if possible—that the selection-principle can be read easily from them. When so adjusted m is called the *magnetic quantum-number*.

In certain cases the empirical rules for the components whereby the magnetic field replaces the individual lines are simple; and the derived rules for the new stationary states which arise out of the original ones when the magnetic field is applied are correspondingly simple. These are the cases of "normal" Zeeman effect (the adjective "normal" may be an entirely misleading choice). Let ΔU_m

represent the energy-difference between the new stationary state denoted by the index m, and the original stationary state. The rules are comprised in the formula,

$$\Delta U_m = m\omega Hh \tag{8}$$

and in the selection-principle. In the formula H stands for the magnetic field; ω is a factor equal within experimental error to $e/4\pi\mu c$ (μ = mass of the electron) and commonly identified with it. m has two or more values spaced one unit apart (for instance, 1 and 0, or $\frac{1}{2}$ and $-\frac{1}{2}$, or 1 and 0 and -1).

The selection principle is as follows: *The only transitions which correspond to actual lines are those in which* m *changes by unity or not*

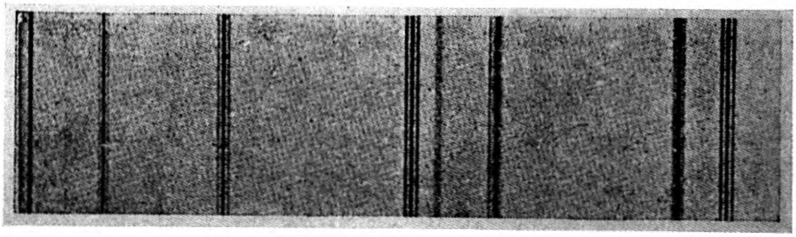


Fig. 9—Effect of magnetic field on spectrum lines. (P. Zeeman, *Journal of the Franklin Institute*)

at all: $\Delta m = 0, \pm 1$. This is the *selection-principle for the magnetic quantum number*.

If we allow m to assume only two values, this principle becomes nugatory. If on the other hand, we adopt the principle, m can assume any number of values whatever, provided only they are spaced at unit

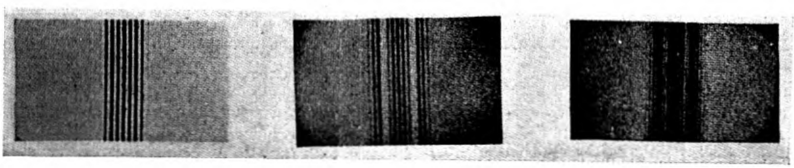


Fig. 10—More complicated effects of magnetic fields on spectrum lines. (P. Zeeman, *l.c.*)

intervals; it makes no difference with the observed lines whether there are two or two hundred new stationary states for every original one. This is convenient for theorizing. In dealing with the Zeeman effect in general, and not merely with these special "normal" cases, it is necessary to assume that ω is not restricted to the particular value just given, but depends on the stationary state in question; and that m depends on the value of j for the stationary state in question.

Very strong magnetic fields treat a group of stationary states as if they were one single state—as if they were first all fused together into one, and this one then resolved according to equation (8). This is the *Paschen-Back effect.* It evidently means a great deal.

The light emitted from a gas exposed to a magnetic field is polarized. Some of the new lines are circularly polarized about the direction of the magnetic field as axis; others are plane-polarized, with the electric vector parallel to the direction of the magnetic field. The lines corresponding to transitions in which m changes by one unit are all polarized in the former way; the lines corresponding to transitions in which m does not change are all polarized in the latter way.[22]

E 10. *Interrelations of Multiplets and Zeeman Effect*

I insert this section chiefly for the benefit of such readers as may be preparing for a thoroughgoing study of atomic theory. Others may do well to pass it over, as the statements it contains can scarcely be apprehended with any vividness, except by the aid of pencil and paper and hours of reiteration. For those who omit this section I will merely say, that the material described in it goes far to show that the numerical values which we have been assigning to k and j are not quite arbitrary, but are determined by something fundamental; although the ones heretofore assigned are not necessarily the most expressive.

I begin with a description of the various known systems of stationary states, condensed into Table III. To make this table clear I will explain the fourth line; this line contains the statement that a "quartet system" of stationary states consists of an s-column of single levels, a p-column of groups of three levels each, and a d-column, an f-column, and additional columns of groups of four levels each.

TABLE III

Name of System	s	p	d	f	f'	f''
Singlet	1	1	1	1	1	1
Doublet	1	2	2	2	2	2
Triplet	1	3	3	3	3	3
Quartet	1	3	4	4	4	4
Quintet	1	3	5	5	5	5
Sextet	1	3	5	6	6	6
Septet	1	3	5	7	7	7
Octet	1	3	5	7	8	8

[22] The effect of a magnetic field on resonance-radiation, discovered by Wood and Ellett, will be described in the Second Part.

Elements of the first column of the periodic table possess a doublet system of stationary states; elements of the third column, a doublet system and in addition a quartet system. It is inferred that elements of the fifth column possess these and a sextet system in addition; elements of the seventh, these three and an octet system in addition. Elements of the second column of the periodic table possess a singlet system and in addition a triplet system. It is inferred that elements of the fourth column possess these two and a quintet system in addition; elements of the sixth column, these three and a septet system; elements of the eighth, these four and a nonet system. These inferences have been partially verified. For titanium, in the fourth column of the periodic table, the triplet and quintet systems have been discovered; for vanadium (fifth column) the quartet and sextet; for chromium (sixth) the quintet and septet; for manganese (seventh) the quartet, sextet, and octet; for iron (eighth) the triplet, quintet, and septet. Apparently it is by no means certain that the unmentioned systems are really missing, as the difficulties of analyzing these complex spectra are terrific.

There are certain rules governing the number of levels in a group, and the effect of a magnetic field upon these levels. These rules were discovered chiefly by Landé; I give them in his notation. I recall, to begin, that we have designated each group of levels by a numeral k, which is 1 for all the groups in the s-column, 2 for all groups in the p-column, 3 for all groups in the d-column, and so forth. We have further distinguished the different levels in a group by assigning them different values of another numeral j; the manner in which these values of j are chosen was described in section E8. Landé introduces a numeral K which is smaller than k by $\frac{1}{2}$; K thus is $\frac{1}{2}$ for all groups in the s-column, $1\frac{1}{2}$ for all groups in the p-column, and so forth. He also introduces a numeral J which is greater than j by $\frac{1}{2}$; and a numeral R which is $\frac{1}{2}$ for every level belonging to a singlet system, 2/2 for every level belonging to a doublet system, 3/2 for every level belonging to a triplet system, and so forth.

These are Landé's rules:

(1) The total number of levels in a group characterized by the numeral K, belonging to a system characterized by the numeral R, is twice the smaller of the two numerals R and K (that is, it is $2R$ if $R<K$; $2K$ if $R>K$; $2R=2K$ if $R=K$).

(2) In the formula (8) for the Zeeman effect, the factor ω is equal to $e/4\pi\mu c$ multiplied by a factor g, which depends on the numerals R, K, and J for the level in question in the following manner:

$$g=3/2+(R^2-K^2)/2(J^2-\tfrac{1}{4}) \tag{9}$$

(3) In the same formula, the magnetic quantum-number m depends on the numeral J for the level in question; it assumes $2J$ values altogether, commencing at the maximum value $(J-\frac{1}{2})$ and going downwards across zero to $(-J+\frac{1}{2})$.

These rules form a beautiful little problem for the designer of atom-models. They have often been tested and verified (it is not easy to find out just how far), and at present are widely used in the deciphering of spectra. It appears, however, that some spectra—particularly those of the inert gases—are too complicated even for these rules, and possess a structure even more elaborate. Considering how difficult it is to grasp the structures already described, one may be excused for feeling some dismay at the prospect.

E 11. Effect of Electric Field on the Stationary States

When an electric field is applied to a radiating gas, the lines of its spectrum are replaced by groups of lines, often rich and complicated.

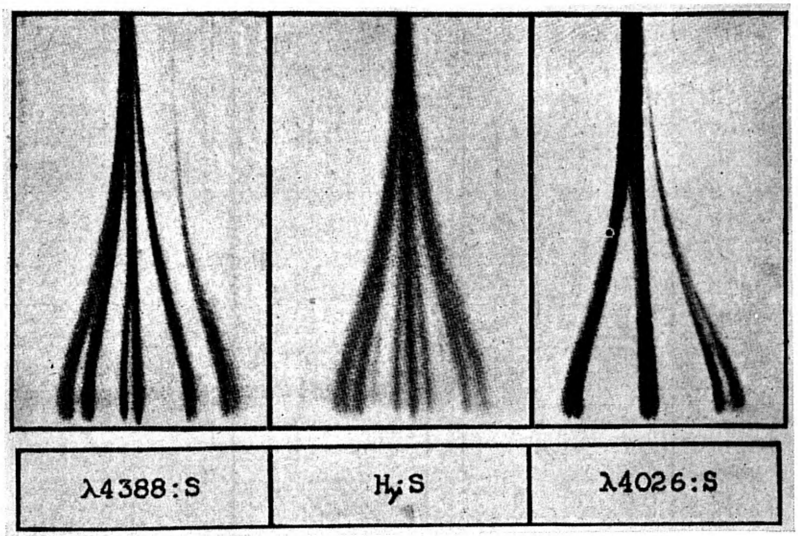


Fig. 11—Resolution of spectrum-lines into groups, displacement of lines, and emergence of missing lines, produced by a strong electric field (increasing from the top downwards to nearly the bottom of the picture). (J. S. Foster, *Physical Review*)

(Fig. 11.) From these we infer, as heretofore, that the stationary states are replaced by groups of stationary states. The atom-model proposed for hydrogen and ionized helium has been extraordinarily successful in describing the effect of electric field upon their spectra, and therefore I shall violate the rule I have heretofore followed, and postpone the description of the phenomena until the theory is stated. Atoms of other kinds are affected in at least two ways; the stationary states are displaced, and the "missing lines" are evoked, as I have said already.

E 12. Intensity-Ratios

The relative intensities of the various lines of a doublet, or triplet, or multiplet are often equal within the (fairly large) uncertainties of measurement to simple ratios, such as 1 : 2, 2 : 3, 3 : 4. This happens too often to be easily put down as a mere coincidence, and indicates that the occurrence of transitions is governed by simple laws. Our selection-principles are themselves indications of the same type, since they may be taken as signifying that the intensity-ratio of certain lines to certain others is zero. This problem may be more difficult than the ones I have stressed hitherto, since each line involves two stationary states and is not a quality of one only. This applies to other properties of lines, such as their sharpness or diffuseness.

E 13. Excitation of Individual Frequencies

So long as an atom is conceived as a belfry full of bells of various pitches, it would probably be argued that a shock to the atom would set all the bells to jangling, and a gas bombarded by electrons would emit all of its natural frequencies if any. The interpretation of spectra to which these pages are devoted leads to a very different idea. A spectrum-line of frequency ν is emitted when the atom passes from a stationary state B to a stationary state A. The energy-value of state B by itself does not determine ν; this is controlled by the difference between the energy-values of B and A, which is $h\nu$. But the energy-value of B has everything to do with whether or not the frequency ν is emitted under given conditions; for it will not be emitted at all unless the atom is first put into state B. If the gas is bombarded with electrons of energy insufficient to raise an atom from its normal state to state B, then the line in question, and all of the other lines which result from transitions from B to other levels of lower energy-value, will fail to appear. If the energy of the electrons is raised past the critical value (the difference between the energy-value of B and the energy-value of the normal state) all of these lines suddenly appear.

This is illustrated by Fig. 12, relating to magnesium. An electron striking a magensium atom and having an energy equal to 3.2 equivalent volts is able to put the atom into a particular excited state; the atom emits radiation of wavelength 4571 in returning to its normal state. To get the atom to emit another sort of radiation, the electron must possess 6.5 equivalent volts to put it into another excited state. Any excited state can be reached if the electron has 10 equivalent volts to pass over to the atom.

In a gas sustaining an electrical discharge, the atoms are subject to stimuli of such variegated force and type that the distinctions between different lines are not so clearly marked; but it can be seen

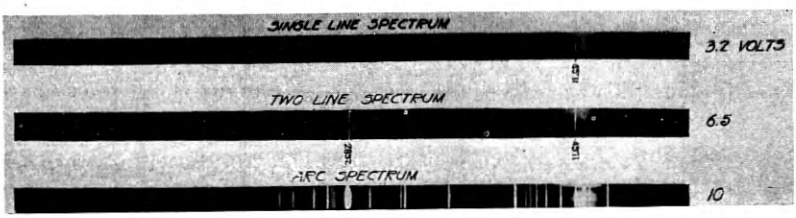


Fig. 12—Successive excitation of lines requiring electron-impacts of successively greater violence to bring atoms into the necessary initial states. (Foote, Meggers, and Mohler, *Philosophical Magazine*)

that mild discharges favor lines for which the initial level is adjacent or close to the normal level, while other lines require a more violent stimulus. Furthermore, when a gas is steadily heated to higher and higher temperatures, various lines of its spectrum appear in more or less the order of the stationary states which are the initial states of the transitions responsible for these lines. Accordingly a "temperature classification" of spectrum lines has been developed at Mount Wilson Observatory and elsewhere, and is valuable in deciphering intricate spectra.

E 14. Absorption-Spectra

An atom which will emit a frequency ν when it is originally in a state B and passes over into a state A, will absorb light of the same frequency if it is initially in the state A. This has the important consequence that the lines which a gas absorbs, when lying at rest and unexcited, are those which it emits in passing from any and every other state *into the normal state*. The lines emitted when an atom passes from one of its stationary states into another which latter is not the normal state, are not absorbed by the gas lying quiescent and undisturbed. For this reason helium and neon and argon are quite transparent to all visible light, although they have many emission-lines in this region of the spectrum; for each of these lines corresponds to a transition into some other than the normal state and the lines which correspond to transitions into the normal state lie far off in the ultraviolet. But if such a gas is made the theatre of a self-sustaining electrical discharge, the other lines likewise are absorbed; for the discharge puts the atoms of the gas temporarily but frequently into various abnormal states. This incidentally is one of the bits of evidence that an atom may sojourn for a finitely long time in another stationary state than the normal one. If the gas is heated, the same effect occurs; for the violent collisions between atoms in a hot gas occasionally bring atoms into excited states.

By observing the absorption-spectrum of a quiescent gas one learns which lines in the emission-spectrum correspond to transitions into the normal state—a valuable piece of information in the cases of elements of which the spectra are complicated and obscure.

E 15. Spectra of Ionized Atoms

In a violent electrical discharge, such as a spark, the gas emits many lines which cannot be fitted into the system of series of the usual spectrum of the gas. These may also be produced by bombarding the gas with electrons possessing more than enough energy to ionize its atoms. They are believed to emanate from ionized atoms, or from atoms deprived of one electron. The spectrum of ionized-helium has been very important in these pages. In very violent sparks many more lines emerge, and these are associated with atoms deprived of two, three, or even more electrons.

The spectrum of the ionized atom of an element resembles, in its system of series, and in more minute details, the spectrum of the neutral atom of the element preceding it in the periodic system. The spectrum of an atom deprived of n electrons resembles the spectrum of the neutral atom preceding it by n places in the periodic system. This confirms the belief that the spectrum and the other properties of an element are determined chiefly by the number of electrons which its atom contains.

E 16. X-ray Spectra

The difference between the X-ray spectra to which we now come, and the "optical" spectra which we have been discussing seemed profound and vital in the era of very defective knowledge, but it has faded steadily away with the deepening of understanding. Twelve or fifteen years ago the contrast was multiform and very sharp; for the optical spectra were produced chiefly by maintaining an electrical discharge in a gas, the X-ray spectra invariably by bombarding a solid body with exceedingly fast-moving electrons or with other X-rays; the optical frequencies could be diffracted and refracted, the X-rays not at all or almost imperceptibly little; the optical frequencies were all inferior to 3.10^{15}, the X-ray frequencies all clearly more than a thousand times as great. Since then, rays of almost all the intermediate frequencies and with intermediate properties have been generated in a variety of ways, and the distinction is no longer trenchant, except between the extremes. To make it so, one must seek a theoretical reason—and perhaps there is none to be found.

There is, however, apparently good ground for introducing a theoretical distinction. I have pointed out heretofore that the energy which an atom loses, when it radiates one of the lines of its "optical" spectrum, is less than the ionizing-energy. Or, turning this statement around and amplifying it a little: the energy which an atom absorbs, when it absorbs one of the rays of its optical spectrum, is less than what is required to detach the loosest electron from it. Therefore it is possible to assume, at least as a trial hypothesis, that the energy is spent in lifting the loosest electron partway out—a hypothesis fortified by the fact that, when the atom has just absorbed some energy in this manner, the electron can be detached by supplying the atom with enough extra energy to bring the total amount up to the ionizing-energy. But if we take one of the typical X-ray frequencies, and multiply it by h to ascertain how much energy the atom gains in the process of absorbing that frequency, we find that the quantity $h\nu$ exceeds the ionizing-energy tremendously. This circumstance makes it quite out of the question to imagine that the X-rays are due to changes in the position or the motion of the loosest electron alone. We may therefore define the X-ray frequencies as those which cannot be explained as due to transitions of the loosest electron, from one motion or position to another, unaccompanied by other changes. By this definition, every frequency ν for which the quantum-energy $h\nu$ is greater than the ionizing-energy, goes into the X-ray spectrum. For the remaining frequencies the question is more dubious, perhaps never quite to be settled unless and until complete theoretical classification of all the lines is attained. In this section, however, I shall speak only of frequencies hundreds or thousands of times greater than the ionizing-frequency.

Gazing upon typical X-ray emission spectra one sees that they consist of groups of lines with wide intervals between. Going from higher frequencies towards lower, the groups are known successively as the K group, the L group, the M group and the N group. The word *series* is more commonly used than *group*; but this is a misfortune, for it suggests a dangerously misleading analogy with the series in the optical spectra which we have studied with so much care.[23] The process of measuring these lines and classifying them

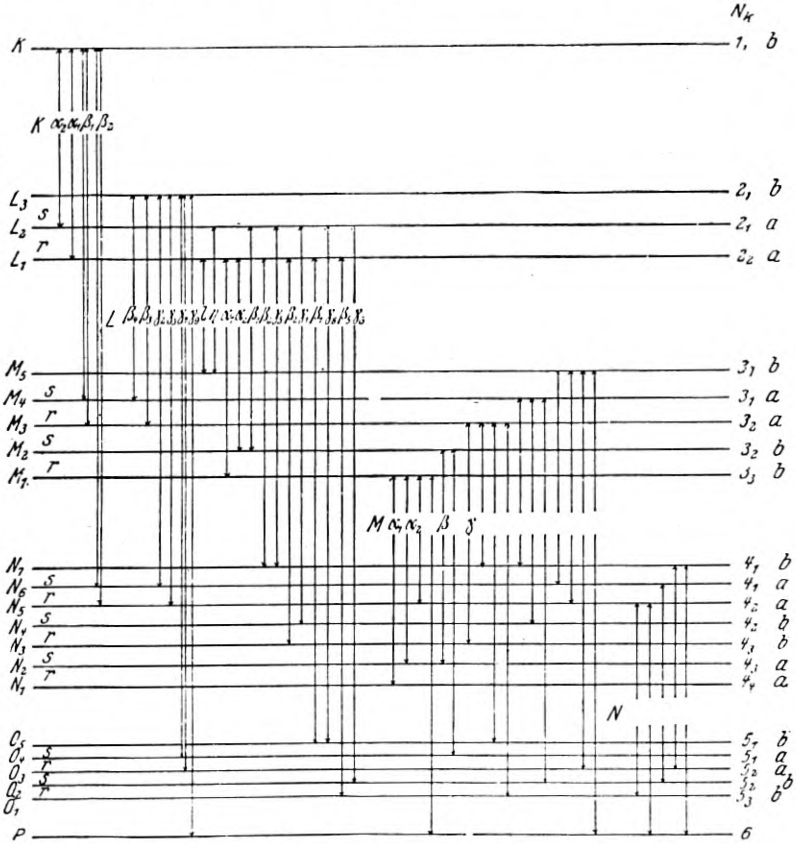


Fig. 13—Diagram of stationary states designed to account for the X-ray spectrum of uranium. (From Siegbahn, after Coster)

was carried out after the dissemination of Bohr's great idea that each line-frequency should be multiplied by h and the product interpreted as the difference between the energy-values of two stationary states of the atom. The complete analysis of an X-ray spectrum

[23] In fact the usage is inverted. A series, in the optical spectrum, is a set of lines having the same *final* state in common; but the "*k-series*" is a group of lines having the same *initial* state in common, the L-series a set of 3 groups corresponding to 3 initial states.

thus culminates in a diagram of stationary states, as for the optical spectra.

Such a diagram is shown in Fig. 13, which is for an element far up in the periodic system, therefore with a rich system of X-ray lines and stationary states. In comparing it with one of the diagrams made for optical spectra, it must be remembered that its scale is enormously more compressed—the distance from top to bottom corresponds to about one hundred thousand equivalent volts. Each line in the X-ray spectrum corresponds to an arrow between two of the levels, but not every arrow corresponds to a line. Again there is a selection-principle, and this selection-principle is partly expressed by attaching a double index to each of the levels. When the indices are assigned as in Fig. 13, transitions between levels for which the second numeral differs by one unit include the only ones which actually occur. But this is not the complete selection-principle; it is necessary to add that in any actually occurring transition, the first numeral must change by one or more units; and further, that transitions may occur only between levels to which different letters are attached. The first numeral is designated by n, the second by k; they are called the total and the azimuthal quantum-number.

The levels are also frequently known by letters with subscript numerals, as the diagram shows. The letters by now are pretty definitely fixed, but the subscripts are still being shuttled around. The notation for the X-ray lines is in a terrible state.

A curious and evidently important feature of these levels is, that when an atom is put into any one of them—say into the K level, or the L_1 level, or the L_2 level—it extrudes an electron. Or, in other words, each of these stationary states is a state in which the atom lacks one of its electrons—like the "ionized-atom" state from which we previously measured the energy-values in dealing with the optical spectra. All of them, at least the highest ones, are in fact "ionized-atom states." Since, however, they are all different, it is natural to suppose that a different electron is missing, or that an electron is missing from a different place, in each of the different cases. Apparently an atom cannot enter into a stationary state with so high an energy, and remain neutral.

We must pause to consider from what standard state the energy-values of these stationary states are measured. In the previous case of the optical spectra, the energy-values of the stationary states were measured, so to speak, *downwards* from the state of the ionized atom to the normal state of the neutral atom; the energy of the ionized atom was set equal to zero, that of the neutral atom in its normal state then had a certain negative value, all the other energy-values were negative and scattered between these two. In this case of the X-ray spectra, the energy-values of the stationary states are measured *upwards* from the normal state of the neutral atom, to which the energy-value zero is assigned, while all the other energies are positive. In Fig. 13 this zero-line must be imagined just under the level marked P.

The exact position of this zero-line for the high energy stationary states is not very accurately known; although the distance between any two levels is determined with all the usually very great exactitude of X-ray wavelength-measurements, the distance from any level to the zero-line is uncertain within a few tens of volts. This uncertainty is not great enough to be important when dealing with the high-frequency X-rays.

This point being attended to, we are now in position to consider the striking difference between X-ray emission-spectra and X-ray absorption spectra—striking indeed when one looks at typical photographs, apparently altogether a different matter from the contrast between optical emission-spectra and optical absorption-spectra, yet in principle very much the same thing. In dealing with optical spectra, I remarked that while an atom *may* absorb any frequency which it can emit—while the complete absorption-spectrum of a gas is identical with its complete emission-spectrum, yet the absorption-spectra one ordinarily sees contain only a small selection of the emission-lines. This occurs because when a gas is being examined for its absorption-spectrum in the laboratory, by sending light through it, it is generally in an untroubled and quiescent condition, each of its atoms being in the normal state; therefore it absorbs only such frequencies as provoke transitions from the normal state to the various excited states, and not such frequencies as would induce transitions from one excited state to another, for few or none of the atoms are in any one of the excited states to start with. Such also is the case with the X-ray spectra. Quiescent atoms absorb only such X-ray frequencies as produce transitions from the normal state into one of the stationary states designated by K, or L_1, or L_2, and so forth—they do not absorb such frequencies as would produce the transitions from L_1 or L_2 to K, for instance, for the atoms are not initially in the states L_1 or L_2. This is quite the same behavior as is observed in the response of atoms to radiations in their optical spectra. It is much more pronounced, however; for, while it is possible to make a gas absorb frequencies which produce transitions from one excited state to another, by maintaining the gas in a state of intense electrical excitation, this has never been done with metals or gases exposed to X-ray frequencies.

Atoms therefore do not absorb such X-ray frequencies as are represented by the downward-pointing arrows in Fig. 13. They do absorb such frequencies as would be represented by arrows drawn from the very bottom of the diagram—a little below the level marked P—up to the various levels; and (it may seem, unexpectedly) they also absorb frequencies somewhat higher than these. This however does not mean that the atom may be put into an excited state of higher energy than the K state, for instance; it means simply, as direct evidence proves, that the extruded electron receives the extra energy and goes away with it. Owing to this fact, the X-ray absorption-spectrum consists not of sharp absorption-lines at the several frequencies corresponding to a transfer of the atom into the K-state, the L_1-state, and so forth, but of continuous bands commencing with sharp edges at these frequencies, and trailing out gradually towards higher frequencies.

Another curious feature of the X-ray spectra is that transitions from the various excited states of high energy-values, such as the K-state and the L-states, directly into the normal state, apparently do not occur.

E 17. Band-spectra

Band-spectra are the spectra of molecules,—that is to say, of clusters of two or more atoms, such as appear in certain gases. This is proved by the fact that they are displayed by gases which are known in other ways (gramme-molecular volume, specific heat) to consist of molecules; by the fact that the band-spectrum of such a gas disappears when the gas is heated to the point where its molecules are dissociated into atoms; and by the general successfulness of the quantitative theory based on the assumption that they are due to molecules. Occasionally band-spectra are displayed by gases which are not otherwise known to contain molecules, such as helium and potassium; it is supposed that they are due to molecules too few to be detected by the other accepted methods. Usually they are easy to distinguish at first glance from the optical spectra of atoms, although there are exceptions, such as the band-spectrum of the hydrogen molecule. Like the spectra we have discussed, they consist of lines; the term "band-spectrum" describes the manner in which these lines are grouped. Again like the spectra we have discussed, they are analyzed according to Bohr's fundamental principle, by interpreting the lines as the results of transitions between stationary states.

F. Magnetic Moments of Atoms

Of the enormous and chaotic variety of facts about the magnetic properties of materials, only a few of the least conspicuous have been serviceable to atom-builders; the notorious ones have helped very little or not at all. The famous and characteristic magnetic properties of iron, nickel, cobalt, depend on the arrangement of the atoms and on the temperature of the metal, and cannot safely be attributed to the atoms themselves. Diamagnetism, an inconspicuous and rarely-mentioned quality of certain elements, is in some instances quite independent of temperature, and may well be a property of the atoms. Paramagnetism, an almost equally inconspicuous quality of certain other elements, depends on temperature, but in such a way that it may sometimes be explained by assuming that each atom has a characteristic magnetic moment, the same for all the atoms of a substance. The value of this magnetic moment of the atom may be calculated from measurements on the paramagnetism of the substance; the process of calculation involves certain assumptions, at least one of which is at the present open to question.

Direct measurements upon the magnetic moments of certain atoms are now being made by Gerlach; and they are among the most important achievements of these years. In a small electric oven, a metal such as silver is vaporized; a beam of the outflowing atoms, passing through a small orifice in the wall of the oven and through others beyond this one, eventually travels across a strong magnetic field with a strong field-gradient and falls upon a plate. Suppose that each atom is a bar-magnet, oriented with its length parallel to the magnetic field. If the field were uniform, the bar-magnet would not be deflected, it would travel across the field in a straight line; for although its north pole would be drawn sidewise by a force, its south pole would be pushed by an exactly equal force in the exactly opposite direction. That the atom may be drawn aside, the field must be perceptibly different at two points as close together as the two poles of the magnet. When one considers how small an object the atom is, it is clear that the field must change very rapidly from one point of space to another, its gradient must be enormous. Gerlach succeeded in contriving so great a magnetic field with so great a gradient that the beam of flying atoms was perceptibly drawn aside. The most-deflected atoms are those of which the magnetic axes are most nearly parallel to the magnetic field. From their deflections, the field, and the field-gradient, the magnetic moment of the atom can be computed very simply. The values thus obtained are of the order of 10^{-10} in *CGS* units.

I shall comment in the second part of this article upon other inferences from these experiments, which are as valuable as the experiments upon the transfer of energy from electrons to atoms. At this point it is sufficient to realize that these experiments prove that atoms, or at least the atoms of some elements, possess magnetic moment. If magnetic moment is due to electric current flowing in closed orbits, as Ampere and Weber guessed a century ago, the atom must be supposed to contain such currents; if the atom consists of a nucleus and electrons, some at least among the electrons must be supposed to circulate. And if the electrons are assumed to circulate in a particular manner the magnetic moment of the atom so designed can be computed, and thereupon tested by experiment.

This completes the list of the phenomena, the properties of matter, which are used in designing the contemporary atom-model. Nobody will require to be convinced that it is not a list of all properties of matter, nor of all phenomena. These are not among the obvious and familiar qualities of matter; and no one meets any of them in everyday life, nor perceives any of them with his unaided senses. They are phenomena of the laboratory, discovered after a long and painstaking development of laboratory technique. Lucretius did not know them, and they were inaccessible even to Newton and to Dalton. They are a very limited selection from among the phenomena of nature, but not for that the less important. The atom-model which is devised to explain them is at best a partial atom-model; thus far it serves for no other phenomena than these, but these it does interpret with an elegance and a competence quite without precedent among atom-models. I have said that some of these phenomena are explained by conceiving an atom made of a positively-charged nucleus and a family of electrons around it; but this conception is not tenable if unmodified. It can be modified so as to interpret the rest of these phenomena; but this means little by itself. The important fact is this, that the modifications which are demanded appear in some cases to be endowed with a beauty and a simplicity, which indicate that they are the expressions of an underlying principle of Nature. To these the following article will be devoted.

Transatlantic Radio Telephone Transmission [1]

By LLOYD ESPENSCHIED, C. N. ANDERSON
and AUSTIN BAILEY

Synopsis: This paper gives analyses of observations of long-wave transmission across the Atlantic over a period of about two years. The principal conclusions which the data seem to justify are as follows:

1. Solar radiation is shown to be the controlling factor in determining the diurnal and seasonal variations in signal field. Transmission from east to west and west to east exhibit similar characteristics.

2. Transmission in the region bordering on the division between the illuminated and the darkened hemispheres is characterized by increased attenuation. This manifests itself in the sunset and sunrise dips, the decrease in the persistence of high night-time values in summer and the decrease in daylight values during the winter.

3. Definite correlation has been found between abnormal radio transmission and disturbances in the earth's magnetic field. The effect is to decrease greatly the night-time field strength and to increase slightly the daylight values.

4. The limit of the high-night-time value of signal field strength for transatlantic distance is essentially that given by the Inverse Distance Law. The normal daylight field strengths obtained in these tests can be aproximated by a formula of the same form as those earlier proposed but with somewhat different constants.

5. The major source of long wave static, as received in both England and the United States, is indicated to be of tropical origin.

6. In general, the static noise is lower at the higher frequencies. At night the decrease with increase in frequency is exponential. In day-time the decrease with increase in frequency is linear in the range of 15 to 40 kilocycles. The difference between day and night static is, therefore, apparently due largely to daylight attenuation.

7. The effect of the static noise in interfering with signal transmission, as shown by the diurnal variations in the signal-to-noise ratio, is found to be generally similar on both sides of the Atlantic.

8. Experiments in both the United States and England with directional receiving antennas of the wave antenna type show an average improvement in the signal-to-static ratio of about 5 as compared with loop reception.

IT will be recalled that something over two years ago, experiments in one-way radio telephone transmission were conducted from the United States to England.[2] In respect to the clarity and uniformity of the reception obtained in Europe, the results represented a distinct advance in the art over the transatlantic tests of 1915. However, they were carried out during the winter, which is most favorable to radio transmission, and it was realized that an extensive favorable to radio transmission, and it was realized that extensive study of the transmission obtainable during less favorable times would be required before the development of a transatlantic radio telephone service could be undertaken upon a sound engineering basis.

[1] Presented before the Institute of Radio Engineers, May 6, 1925.

[2] "Transatlantic Radio Telephony," Arnold and Espenschied, *Journal of A.I.E.E.*, August, 1923. See also, "Power Amplifiers in Transatlantic Telephony," Oswald and Schelleng, presented before the Institute of Radio Engineers, May 7, 1924.

Consequently, an extended program of measurements was initiated to disclose the transmission conditions obtaining throughout the twenty-four hours of the day and the various seasons of the year. The methods used in conducting these measurements and the results obtained during the first few months of them have already been described in the paper previously mentioned. The results there reported upon were limited to one-way transmission from the United States to England upon the telephone channel. Since then the measurements have been extended to include transmission on several frequencies in each direction from radio telegraph stations in addition to the 57 kilocycles employed by the telephone channel.

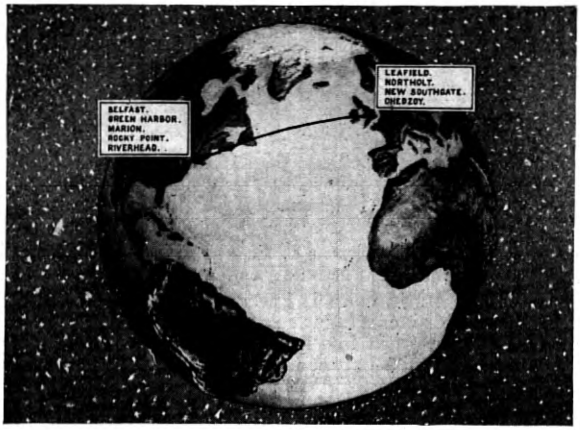


Fig. 1

The present paper is, therefore, in the nature of a report upon the results thus far obtained in work currently under way. It seems desirable to make public these results because of the large amount of valuable data which they have already yielded, and because of the timely interest which attaches to information bearing upon the fundamentals of radio transmission. The carrying on of this extensive measurement program has been made possible through the cooperation of engineers of the following organizations: in the United States—The American Telephone and Telegraph Company and the Bell Telephone Laboratories, Inc., with the Radio Corporation of America and its Associated Companies; in England—The International Western Electric Company, Inc., and the British Post Office.

Measurement Program

The scene of these transatlantic experiments is shown in Fig. 1 The British terminal stations will be seen to lie in the vicinity of London and the American stations in the northeastern part of the United States. The United States transmitting stations are the radio telephone transmitter at Rocky Point, and the normal radio

Fig. 2—Exterior of Riverhead Radio Receiving Station

telegraph transmitters at Rocky Point and Marion, Mass. The measurements of these stations were made at New Southgate and at Chedzoy, England. The British transmitting stations utilized in measuring the east to west transmission were the British Post Office telegraph stations at Leafield and at Northolt. The receiving measurements in the United States were initiated at Green Harbor, Mass., and continued at Belfast, Maine and Riverhead, L. I.

The Riverhead receiving station, shown in Figs. 2 and 3, is typical of the receiving stations involved in the measurement program. The interior view of Fig. 3 shows the group of receiving measurement apparatus at the right and the loop at the left. The three bays of apparatus shown are as follows: That at the left is the receiving set proper which is, in reality, two receiving sets in one, arranged so that one may be set for measurements on one frequency band and the other set upon another band. The set is provided with variable filters which accounts for the considerable number of condenser dials. The second bay from the left contains voice-frequency output apparatus, cathode ray oscillograph and frequency meter. The third bay

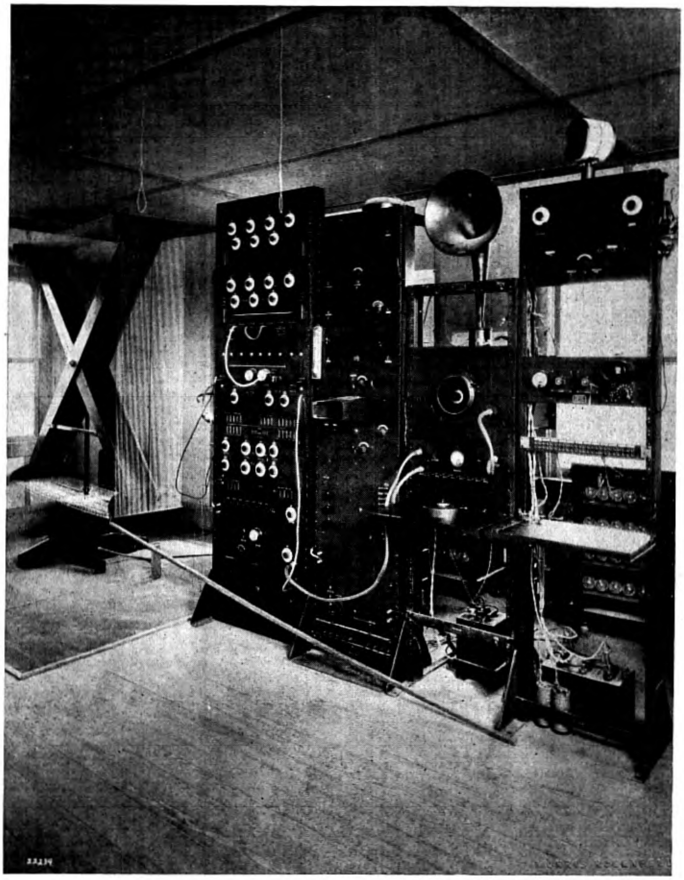


Fig. 3—Interior Riverhead Station

carries the source of local signal and means for attenuating it, and the fourth bay contains means for monitoring the transmission from the nearby Rocky Point radio telephone transmitter.

The measurements are of two quantities: (1), the strength of received field, and (2) the strength of received noise caused by static. The particular frequencies upon which the measurements were taken (given in the chart of Fig. 4) lie in a range between 15 and 60 kc. The arrows indicate the single frequency transmissions which were employed for signal field strength measurements, those at the left indicating the frequencies received in the United States from England, and those at the right, the frequencies received in England from the United States. The black squares in the chart denote the bands in which the noise measurements were taken. In general the measure-

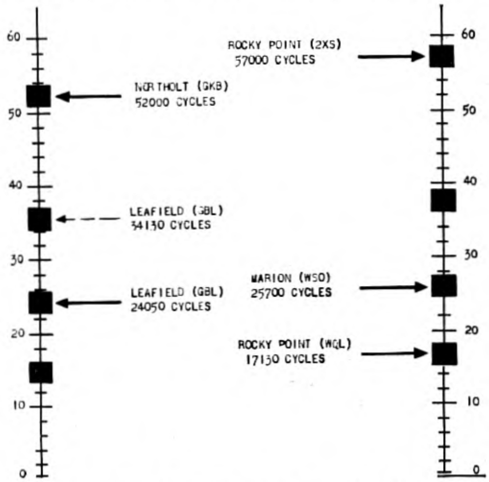


Fig. 4—Frequency distribution of measurements. Black squares denote band in which noise measurement was taken

ments of both field strength and noise have been carried out on both sides of the Atlantic at hourly intervals for one day of each week. The data presented herewith are assembled from some 40,000 individual measurements taken during the past two years in the frequency range noted above. The transmitting antenna current has been obtained for each individual field strength measurement and all values corrected to a definite reference antenna current for each station measured. The data have been subject to careful analysis in order to disclose what physical factors, such as sunlight and the earth's magnetic field, affect radio transmission.

Measurement Methods

Although it will not be necessary to describe in any detail the type of apparatus employed in making these measurements, as this information has already been published,[3] a brief review of the methods involved will facilitate an understanding of the data.

[3] Radio Transmission Measurements, Bown, Englund, and Friis. Proceedings I.R.E., April, 1923.

In general the method employed in measuring the signal field strength is a comparison one. A reference radio-frequency voltage of known value is introduced in the loop antenna and adjusted to give the same receiver output as that from the distant signal. This is determined either by aural or visual means. Under such conditions equal voltages are introduced in the antenna from local and distant sources, and by calculating the effective height of the loop the field strength of the received signal is determined.

In the noise measurements, static noise is admitted through a definite frequency band approximately 2,700 cycles wide. A local radio-frequency signal of known and adjustable voltage is then in-

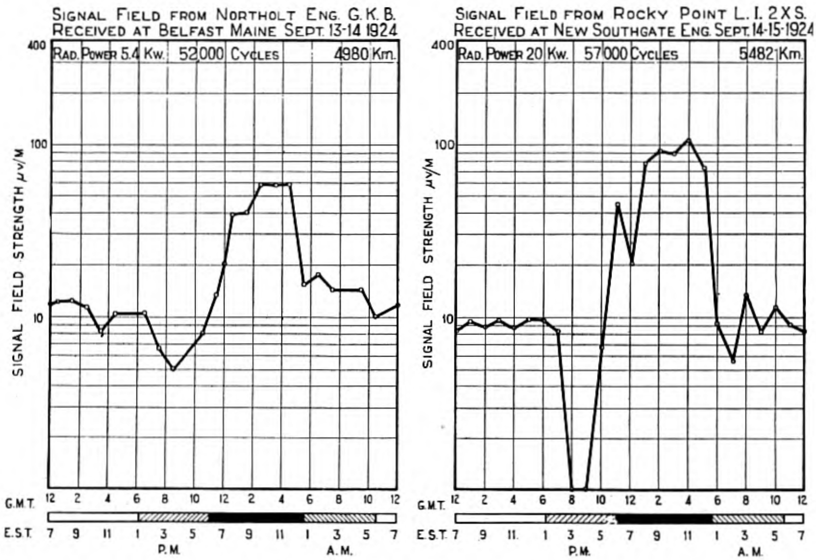


Fig. 5—Diurnal variation in signal field

troduced. The radio-frequency source of this signal is subjected to a continual frequency fluctuation so that the detected note has a warbling sound. This is done in order that the effect of static upon speech can be more closely simulated than by using a steady tone. The intensity of the signal is then adjusted to such a value that further decrease results in a rapid extinction. The comparison signal is then expressed in terms of an equivalent radio field strength. Thus the static noise is measured in terms of a definite reference signal with which it interferes and is expressed in microvolts per meter.

Signal Field Strength

The curves of Fig. 5 are given as examples of the field strength measurements covering a single day's run. The curves have been constructed by connecting with straight lines the datum points of measurements taken at hourly intervals. It will be evident that

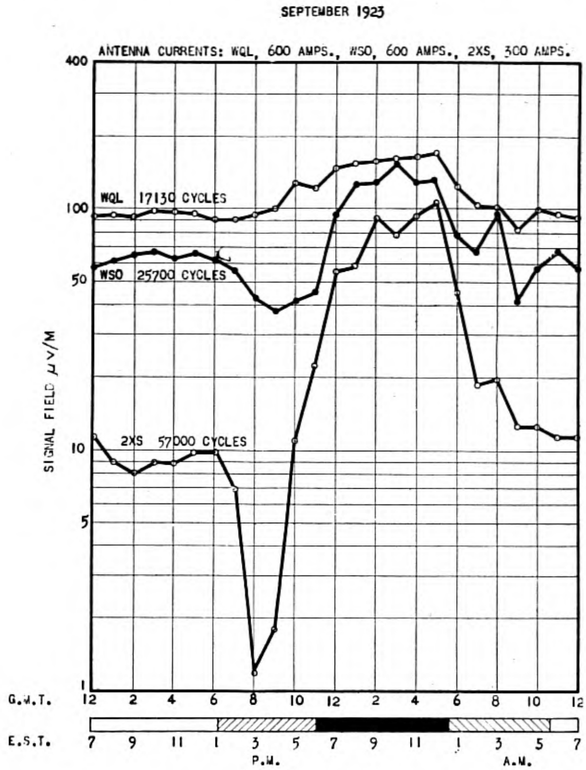


Fig. 6—Monthly average of diurnal variation in signal field transmission from American stations on various frequencies received at New Southgate, England, September, 1923

they portray the major fluctuations occurring throughout the day, but that they are not sufficiently continuous to disclose, in detail, the intermediate fluctuations to which the transmission is subject.

Diurnal Variation. The left-hand curve is for transmission from England to America on 52 kilocycles, and the right-hand one for transmission from America to England on 57 kilocycles. These curves illustrate the fact, which further data substantiate, that both transmissions are subject to substantially the same diurnal variation. The condition of the transatlantic transmission path with respect to daylight and darkness is indicated by the bands beneath the curves. The black portion indicates the time during which the transatlantic path is entirely in darkness, the shaded portions the time during which it is only partially in darkness, and the unshaded portions the time during which daylight pervades the entire path.

The diurnal variation may be traced through as follows:

1. Relatively constant field strength prevails during the daylight period.

2. A decided drop in transmission accompanies the occurrence of sunset in the transmission path between the two terminals.

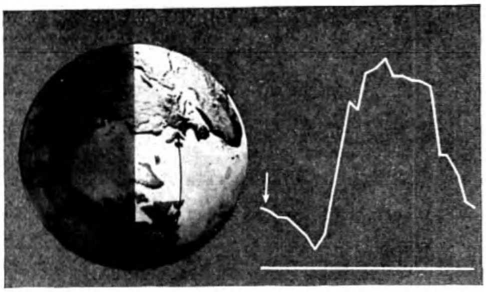
3. The advent of night-time conditions causes a rapid rise in field strength to high values which are maintained until daylight approaches.

4. The encroachment of daylight upon the eastern terminal causes a rapid drop in signal strength. This drop sometimes extends into a morning dip similar to, but smaller than, the evening dip. After this, relatively steady daylight field strengths again obtain.

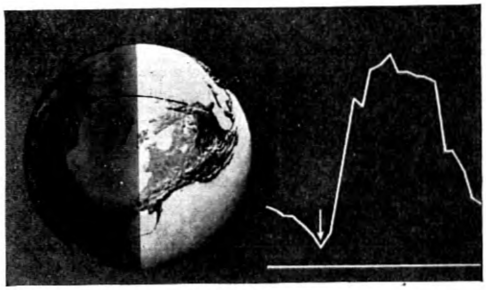
Three or four curves similar to Fig. 5 are obtained each month. By taking the average of such curves for the month of September, 1923, the lower curve on Fig. 6 is obtained. The upper curves are for similar averages of measurements made on the lower frequencies. These curves show clearly that the range of the diurnal fluctuation is less for the lower frequencies. This is because of the lesser daylight absorption.

The mechanism by which the transatlantic transmission path is subjected to these daily and seasonal controls on the part of the sun, would be more evident were we enabled to observe the earth from a fixed point in space. We should then be able to see the North Atlantic area plunged alternately into daylight and darkness as the earth rotates upon its axis, and to visualize the seasonal variation of this exposure to sunlight as the earth revolves about the sun. Photographs of a model of the earth showing these conditions have been made, and are shown in Fig. 7. The first condition is that for January, in which the entire path is in daylight. The curve of diurnal variation is shown in the picture and that part which corresponds to the daylight condition is indicated by the arrow, In the next position the earth has rotated so that the London terminal is in darkness while the United States terminal is still in daylight. This corresponds to the evening dip, the period of poorest transmission. With the further rotation of the earth into full night-time conditions for the entire path, the received signal rises to the high night-time values. These high values continue until the path approaches the daylight hemisphere as indicated in the fourth position. As the path enters into sunlight, the signal strength drops with a small dip occurring when sunrise intervenes between the two terminals.

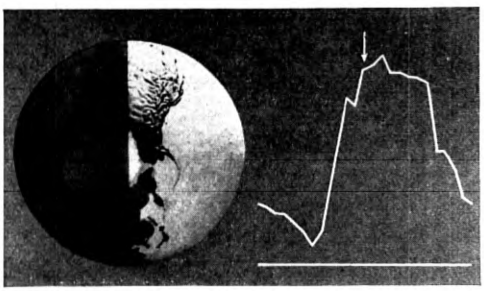
Seasonal Variation. By assembling the monthly average curves for all months of the year, the effect of the seasonal variation on the



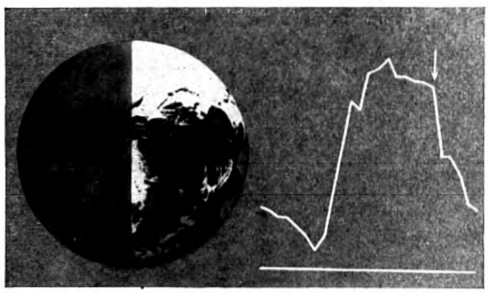
a



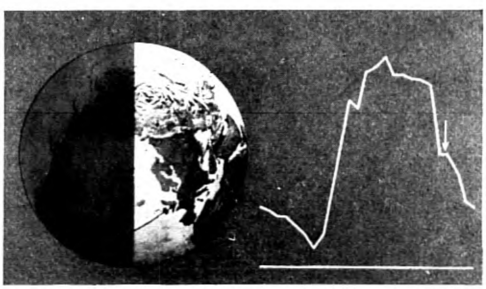
b



c



d



e

Fig. 7—Signal Field January—Variation with exposure of transmission path to sunlight

diurnal characteristic becomes evident. This is shown in Fig. 8, the data for which actually cover two years.

The outstanding points to be observed in this figure are:

1. The continuance of the high night-time values throughout the year.

2. The persistence of the high night-time values for a longer period in the winter than in the summer months.

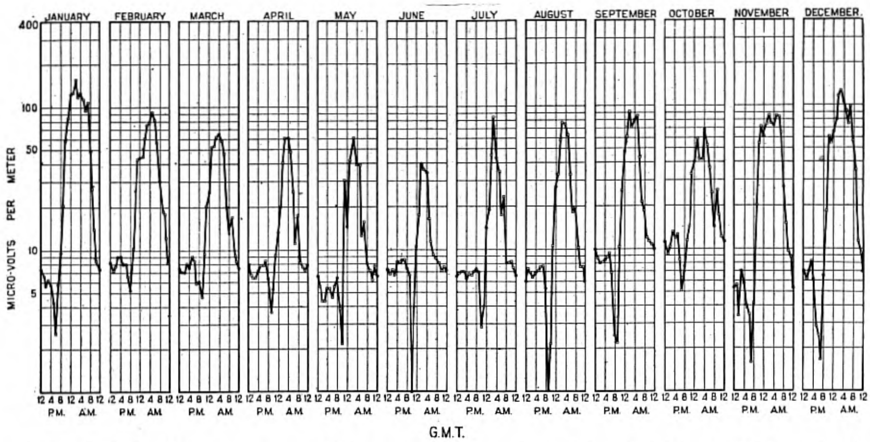


Fig. 8—Monthly averages of diurnal variation in signal field, Rocky Point, L. I. (2 X S) to New Southgate, England, 57,000 cycles—Ant. Current, 300 Amps—5480 Km. 1923–1924

3. The daylight values show a comparatively small range of variation.

4. The extreme range of variation shown between the minimum of the sunset dip and the maximum of the high night-time values is of the order of 1 to 100 in field strength. This is equivalent to 1 to 10,000 in power ratio.

It will be recalled that the cause of the seasonal changes upon the earth's surface resides in the fact that the earth's axis is inclined and not perpendicular to the plane of its orbit about the sun. As the earth revolves about the sun, the sunlit hemisphere gradually extends farther and farther northward in the spring months and by the summer solstice reaches well beyond the north pole, as indicated in Fig. 9. As the earth continues to revolve about the sun, the sunlit hemisphere recedes southward until at the winter solstice it falls considerably short of the north pole and extends correspondingly beyond the south pole. Since the transatlantic path lies fairly high in the northern latitude, it is not surprising that the transmission conditions disclose a decided seasonal influence. The effect of this seasonal influence in shifting the diurnal transmission characteristic is better shown in Fig. 10. This figure consists of the same monthly average diurnal curves as are assembled in Fig. 8, arranged one above the other instead of side by side.

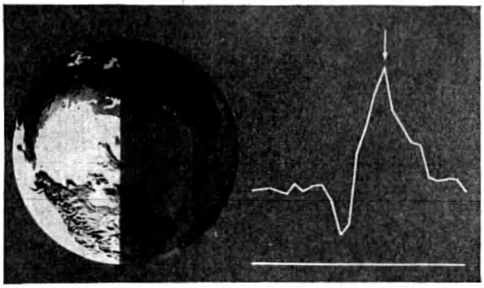


Fig. 9—Signal Field June—Night conditions showing proximity of transmission path to sunlit hemisphere

In particular, there should be noted:

1. The time at which the sunset dip occurs changes with the change in time of sunset.

2. Similarly, the time at which the morning drop in field strength occurs changes with the time of sunrise.

3. The period of high night-time values, bounded between the time of sunset in the United States and the time of sunrise in England, is much longer in the winter than in the summer months.

It is also to be observed that, as a rule, full night-time values of signal field strength are not attained until some time after sunset at the western terminal and that they begin to decrease before sunrise at the eastern terminal. In other words, the daylight effects appear to extend into the period in which the transmission path along the earth's surface is unexposed to direct rays of the sun. The effect of this is that with the advance of the season from winter to summer the time at which the high night-time value is fully attained occurs later and later whereas the time at which it begins to fall off occurs earlier and earlier, until the latter part of April when these two times coincide. At this time, then, the transmission path no sooner comes into the full night-time conditions than it again emerges. As the season further advances into summer, the day conditions begin to set in while the night-time field strength is still rising. The proximity to the daylight hemisphere, which the transatlantic path reaches at night during this season of the year is illustrated in Fig. 9.

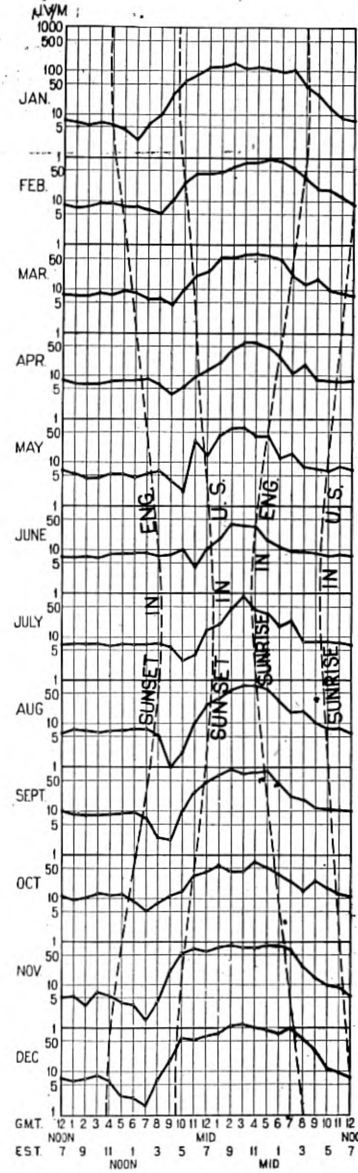


Fig. 10—Monthly averages of diurnal variation of signal field, Rocky Point, L. I. (2 X S) to New Southgate, England; 20.8 K.W. radiated power, 57,000 cycles, 1923–1924

As the sunlit hemisphere recedes southward after the summer solstice a time is reached, about the middle of August, at which the full night-time values are again realized. Beyond this time they are sustained for increasing periods of time. It is of interest to note that at these two times of the year, the last of April and the middle of August, direct sunlight exists over the darkened hemisphere some 500 kilometers above the great circle path.

For all of the conditions noted above, namely, sunset, sunrise, and summer approach of the transmission path to the northern boundary of the night hemisphere, the path lies in a region wherein the radiation

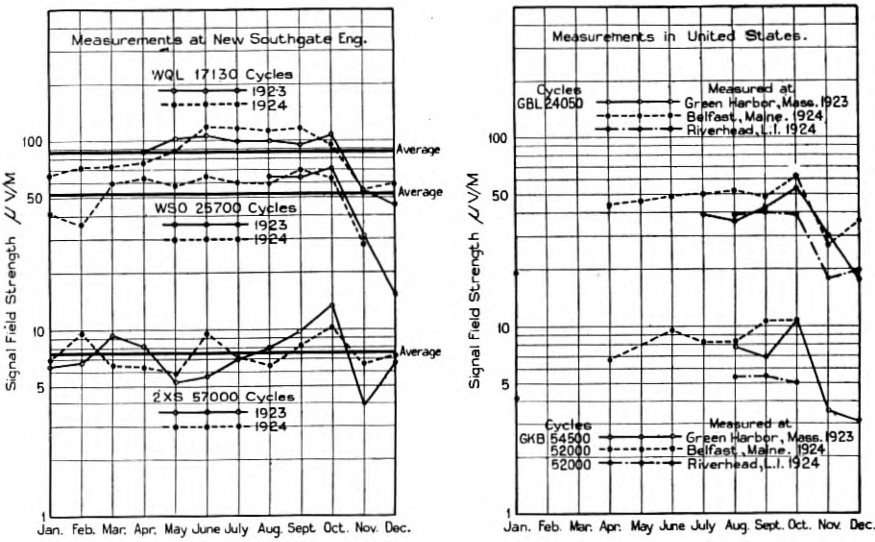


Fig. 11—Monthly averages of daylight field strength

from the sun grazes the earth's surface at the edge of the sun-lit hemisphere. The transmission path also approaches this region during daylight in the winter months, as will be seen by reference to the first position of Fig. 7 for the month of January. The results of measurements for the months of November, December and January for all of the frequencies measured show definite reductions in the daylight field strengths. This reduction is evident in Fig. 8 for the 57-kilocycle transmission, but shows up more strikingly in the curves of Fig. 11. The effect of each of these conditions, in which the transmission path approaches the region in which the solar emanation is tangential to the earth's surface, will be observed to be that of an *increase* in the transmission loss. The fact that in one instance this occurs in daylight would seem to suggest for its explanation the presence of some factor in addition to sunlight, such as electron emission.

Field Strength Formulae. The two major phases of the diurnal variation of signal field strength which lend themselves to possible predetermination are the daylight values and the established night-time values. As to the night-time values our data show, within the limits of experimental error, that the maximum values do not exceed that defined by the inverse distance law. This fact seems to support the viewpoint [4] that the high night-time values are merely the result of a reduction of the absorption experienced during the day. Fig. 11 presents the monthly averages of the *daylight* field strengths for the various frequencies on which measurements were taken. The chart at the left is for reception in England and that at the right for reception in the United States.

The difficulty in predicting by transmission formulae, values to be expected at any one time will be evident and the best that can be expected is to approximate the average. The formulae of Sommerfield, Austin-Cohen and Fuller take the form

$$E\mu v/M = \frac{377HI}{\lambda D}\, e^{-\frac{\alpha D}{\lambda^x}}$$

where the coefficient $\frac{377HI}{\lambda D}$ represents the simple Hertzian radiation field and the exponential $e^{-\frac{\alpha D}{\lambda^x}}$ the attenuation factor. From theoretical considerations, Sommerfeld (1909) gave $\alpha = .0019$ and $x = 1/3$. In the Austin-Cohen formula α is given as .0015 and $x = 1/2$. Fuller gives $\alpha = .0045$ and $x = 1.4$. The Austin-Cohen formula was tested out experimentally chiefly with data obtained from the Brant Rock station (1911) and from the Arlington station by the U.S.S. *Salem* in February and March, 1913. Fuller derived his .0045 value of α from 25 selected observations from tests between San Francisco and Honolulu in 1914.

An attempt has been made to determine the constants of a formula of the above form which would approximate averages of some 5,000 observed values of field strength over this particular New York to London path and over the frequency range of 17 kc. to 60 kc. For each transmitting station a series of comparatively local measurements were taken to determine the power radiated. By combining these local measurements with the values obtained on the other side

[4] See also "Radio Extension of Telephone System to Ships at Sea," Nichols and Espenschied, Proc. I. R. E., June, 1923, pages 226–227.

of the Atlantic we found that approximately $\alpha=.005$ and $x=1.25$. The transmission formula then becomes

$$E\mu v/M=\frac{377HI}{\lambda D}\,e^{-\frac{.005D}{\lambda^{1.25}}}$$

or in terms of power radiated

$$E=\sqrt{P}\,\frac{298\times10^3}{D}\,e^{-\frac{.005D}{\lambda^{1.25}}}$$

where

E = Field strength in microvolts per meter
P = Radiated power in kw.
D = Distance in km.
λ = Wave length in km.

The table shown on next page summarizes the data relative to daylight transmission.

Correlation Between Radio Transmission and Earth's Magnetic Field

In analyzing the measurements we were impressed by the occasional occurrence of marked deviations from the apparent normal diurnal characteristic. A series of measurements which includes an example of this condition is represented in the upper curves of Fig. 12. The curves of the first four days exhibit the normal diurnal characteristic as did the curves of the preceding measurements. The next test of February 25–26 exhibits a marked contrast with that of two days previous. Such abnormality continues in greater or less degree until partial recovery in the test of April 29–30.

Comparison of these data with that of the earth's magnetic field for corresponding days shows a rather consistent correlation. This will be evident from inspection of the magnetic data plotted below in the same figure. Both the horizontal and vertical components of the earth's field are shown. The first decided abnormality occurs February 25–26. The three succeeding periods show a tendency to recover followed by a second abnormality on March 25–26 and again one on April 22–23. It is of interest to note that within limitations of the intervals at which measurements were taken, these periods correspond roughly to the 27-day period of the sun. Coincidences similar to those described above have been found for other periods. Except for this coincidence of abnormal variations in earth's magnetic field and radio transmission, exact correlation of the fluctuations has not been found possible.

TRANSATLANTIC RADIO TELEPHONE MEASUREMENTS

Transmitting Terminal	Receiving Terminal	Freq.	Distance Km.	Power* Radiated Kw.	Daylight Field Strengths Observed			Daylight Field Strengths Calculated		
					1923	1924	Av.	Austin-Cohen	Fuller	This Paper
2 X S	New Southgate, Eng.	57,000	5,482	20.6	7.5	7.65	7.6	6.9	21.2	7.8
WSO	New Southgate, Eng.	25,700	5,282	8.95	(Aug.-Dec.) 48.7	(Jan.-Nov.) 54.6	52.7	16.6	78.5	50.2
WQL	New Southgate, Eng.	17,130	5,482	12.	(Apr.-Dec.) 86	87.3	86.8	27.7	116.	86.
GBL	Green Harbor, Mass.	24,050	5,149	4.06	(July-Jan.) 34.2			13.2	59.	39.
	Belfast, Maine	24,050	4,885	4.06		(Apr.-Dec.) 51(?)		15.6	54.7	41.8
	Riverhead, L. I.	24,050	5,363	4.06		(Aug.-Dec.) 31.5		11.4	55.2	34.5
GBL	Green Harbor, Mass.	34,130	5,149	4.85	(July-Jan.) 16.1			9.5	41.2	22.6
GKB	Green Harbor, Mass.	54,500	5,241	7.9	(Aug.-Dec.) 6.1			5.6	18.6	7.1
	Belfast, Maine	52,000	4,980	5.4		(Apr.-Oct.) 9.1		6.15	20.	9.05
	Riverhead, L. I.	52,000	5,457	5.4		(Aug.-Oct.) 5.3		4.2	15.	5.9

* Computed from local observations using formula of this paper.

NOTE: Measurements of transmission from Rocky Point (2 X S) on 57,000 cycles measured at Mexico City, July, 1924, give an average daylight field strength of 39.4 mv/M. Calculated value 42.5 mv/M.

The magnetic data have been supplied through the courtesy of the United States Geodetic Survey. Similar data taken in England were obtained from the Kew observatory and show similar results.

The contrast in the diurnal variations of radio transmission before and after the time a magnetic storm is known to have started, is

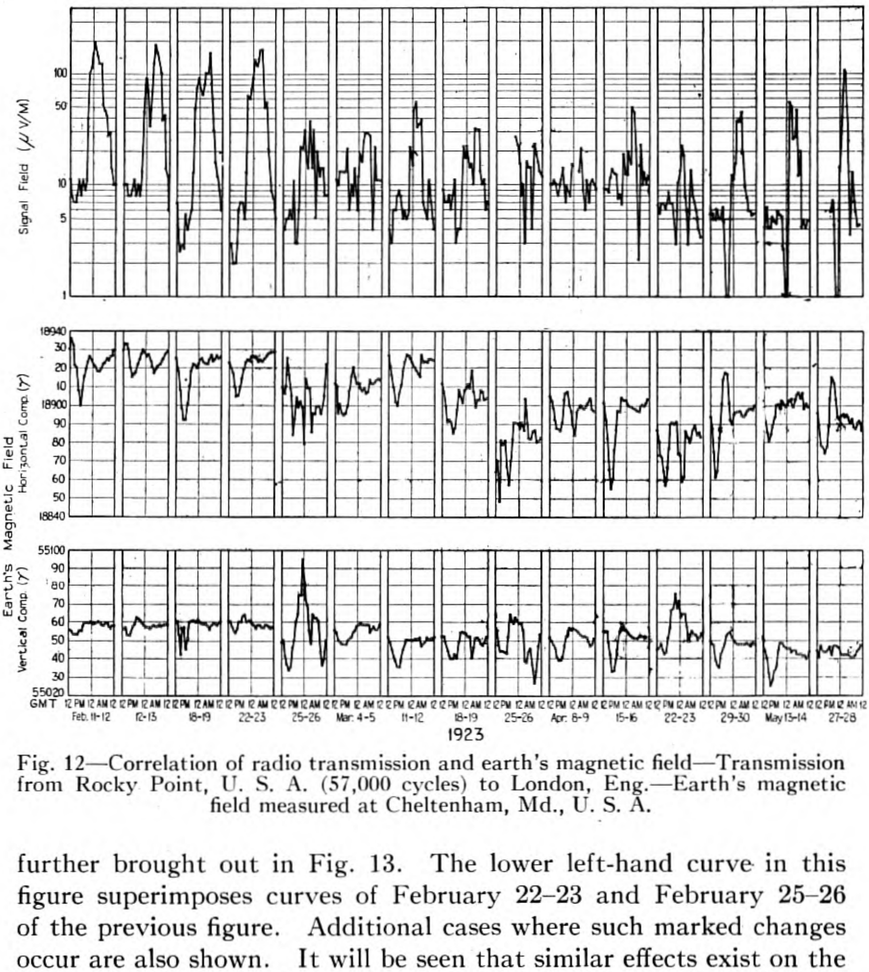


Fig. 12—Correlation of radio transmission and earth's magnetic field—Transmission from Rocky Point, U. S. A. (57,000 cycles) to London, Eng.—Earth's magnetic field measured at Cheltenham, Md., U. S. A.

further brought out in Fig. 13. The lower left-hand curve in this figure superimposes curves of February 22–23 and February 25–26 of the previous figure. Additional cases where such marked changes occur are also shown. It will be seen that similar effects exist on the lower frequency of 17 kc. All of these examples are for days of other than maximum magnetic disturbance. In general the effect is to reduce greatly the night-time values and slightly increase the daylight values. The higher peaks in the daylight field strength of Fig. 11 are due to the high daylight values which prevailed at the time of these disturbances.

Noise Strength

Next to field strength the most important factor in determining the communication possibilities of a radio channel is that of the interfering noise. The extent to which noise is subject to diurnal and seasonal variations is therefore of first order of importance.

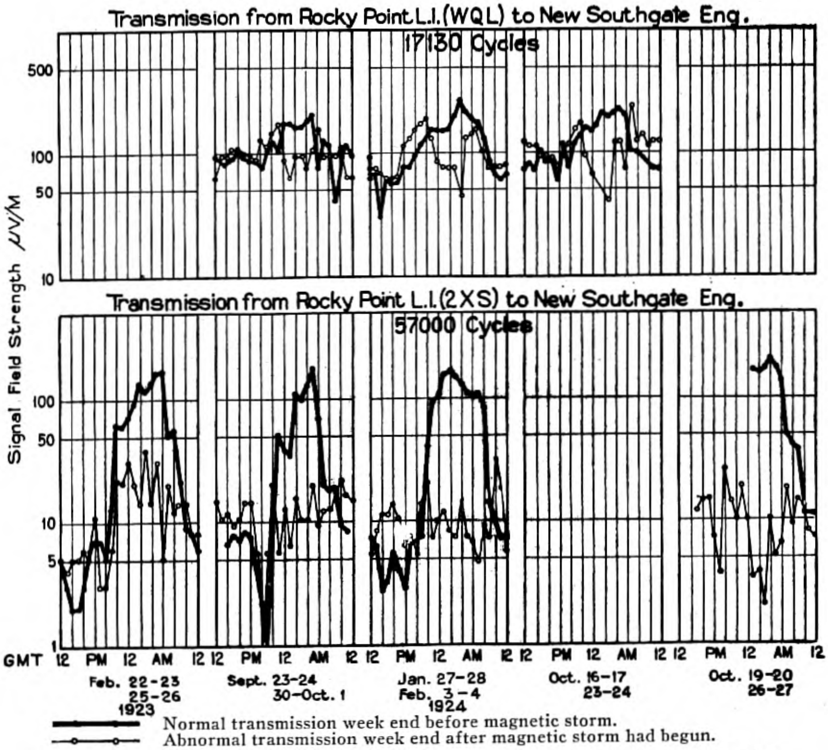


Fig. 13—Correlation between radio transmission and variations in earth's magnetic field

Diurnal Variation. An example of the diurnal characteristic of the noise for both ends of the transatlantic path is given in Fig. 14. One curve is shown for each of the several frequencies measured. The outstanding points to be observed are:

1. The rise of the static noise about the time of sunset at the receiving station, the high values prevailing at night, and the rather sharp decrease accompanying sunrise. The curve for 15 kc. shows the existence of high values also in the afternoon. During the summer months high afternoon values are usual for all frequencies in this

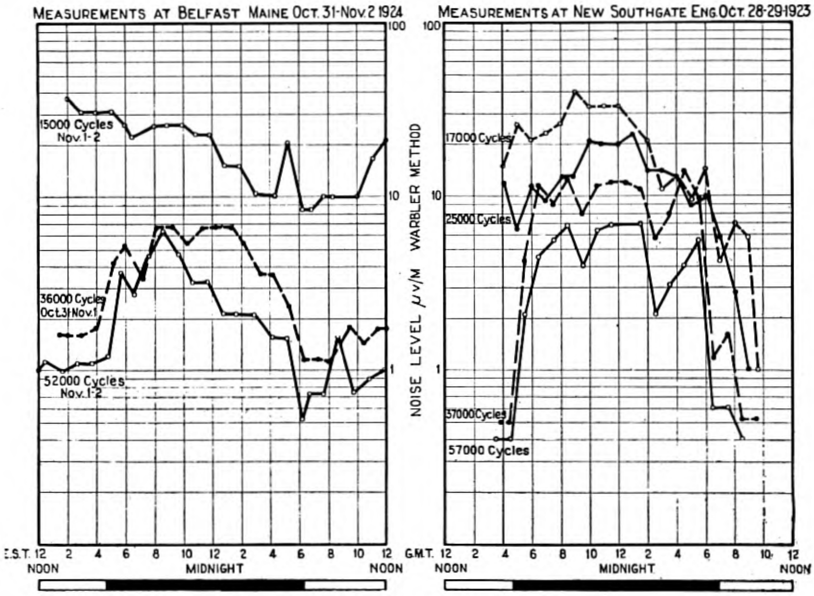


Fig. 14—Diurnal variation in noise

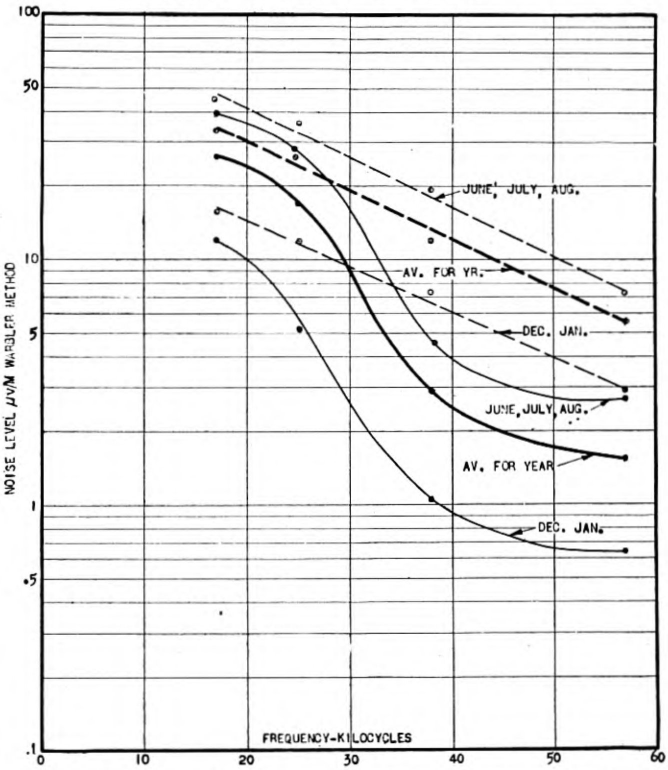


Fig. 15—Frequency distribution of noise, New Southgate, England—
Night time————Day time ————1923–1924

range. They extend later into the fall for the lower frequencies, and hence are in evidence on the date on which these measurements were taken, October-November.

2. In general the noise is greater the lower the frequency.

Noise as a Function of Frequency and of Receiving Location. The distribution of static noise in the frequency range under consideration is depicted in Fig. 15 for the case of reception at New Southgate, England. The set of full-line curves is for daylight reception and the set of dash-line curves for night-time reception. The values obtaining during the transition period between day and night have been excluded. For both conditions three curves are shown, one the average of the summer months, another the average of winter months and the third, the heavy line, the average for the entire year. The curves represent averages for all of the measurements taken during both 1923 and 1924. In considering curves of this type it should be remembered that they represent an average of a wide range of conditions and at any one time the distribution of static may differ widely from that indicated by the curves. Also it should be realized that the extreme difference between winter and summer static is much greater than the difference between the averages.

A similar study of frequency distribution was made at two locations in the United States, Belfast and Riverhead. The results obtained at these two locations together with those for New Southgate, England, are presented in Fig. 16 for a period during which data were obtained for all three places. The similarity of the three sets of curves shows that there is an underlying cause common to both sides of the Atlantic which may account for the difference between the daytime and night-time static on the longer waves. It will be evident from the curves that for frequencies around 20 kc. there is not very much difference between the day and night static noise but that at the higher frequencies in the range studied, the daylight values become considerably less than the night-time values. Actually the divergence between the night-time and the daytime noise curves up to about 40 kc. is an exponential one. This suggests that the lowering of the daylight values may be largely due to the higher absorption which occurs in the transmission medium during the day. There is a further interesting point to be noted concerning both figures, namely, that the night-time values decrease exponentially with increase in frequency. Since these night-time values are but little affected by absorption in the transmitting medium, the distribution of the static energy as received, also roughly represents the distribution of the static power generated.

The curves of Fig. 16 show also the substantial difference in the noise level which exists at the three receiving points. As has been experienced in practice, the New Southgate curve indicates that England is less subject to interference than northeastern United

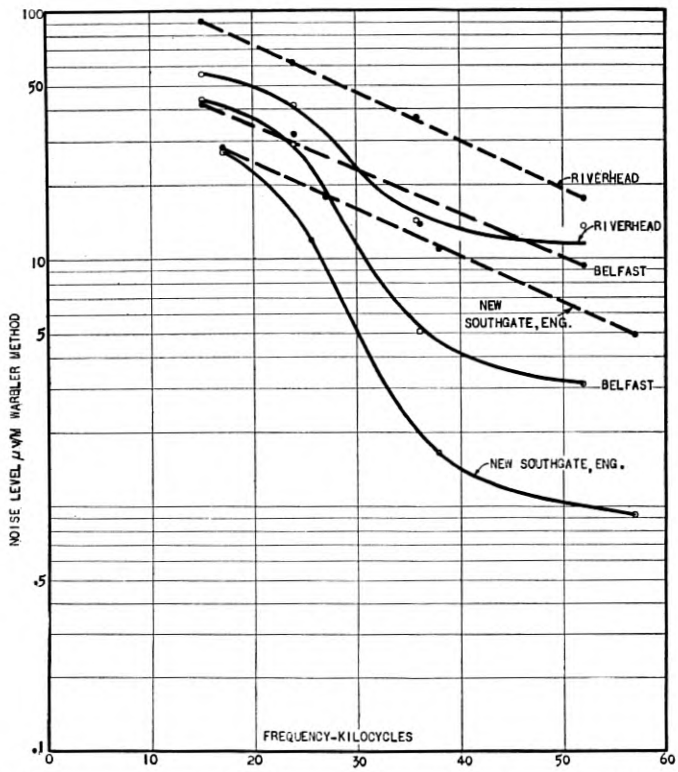


Fig. 16—Frequency distribution of noise, New Southgate, Eng., Belfast, Maine Riverhead, L. I.—Night time————Day time————Aug.–Dec., 1924

States. In the United States the superiority of Belfast over Riverhead is also consistent with the better receiving results which in general have been experienced in Maine. There should be noted also the fact that the curves for these three locations lie one above the other in the inverse order of the latitudes. This is in keeping with other evidence which points towards the tropical belt as being a general center of static disturbance on the longer wave lengths. Further evidence on this point is presented below in connection with the seasonal variations of noise.

Seasonal Variation. Curves showing the diurnal variation in noise level for each month of the year together with the variation in time of sunset and of sunrise, are shown in Fig. 17. Each curve is the average of all the measurements taken during that particular month in 1923 and 1924. The diurnal variations are generally similar for the different months in respect to the high night-time values which are limited to the period between the times of sunset and sun-

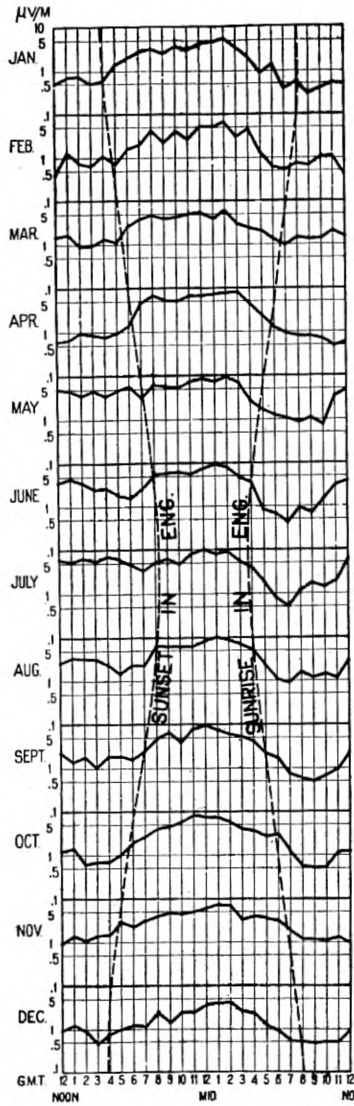


Fig. 17—Monthly averages of diurnal variation of noise, New Southgate, England—57,000 cycles—1923–1924

rise in England. There is a certain deviation, however, which it is well to point out. During the summer months the rise in night-time static starts several hours before and reaches high values at about sunset in England, whereas in the winter-time, the night-time static begins to rise at about sunset and reaches high values several hours later. A similar effect is observed for the sunrise condition wherein

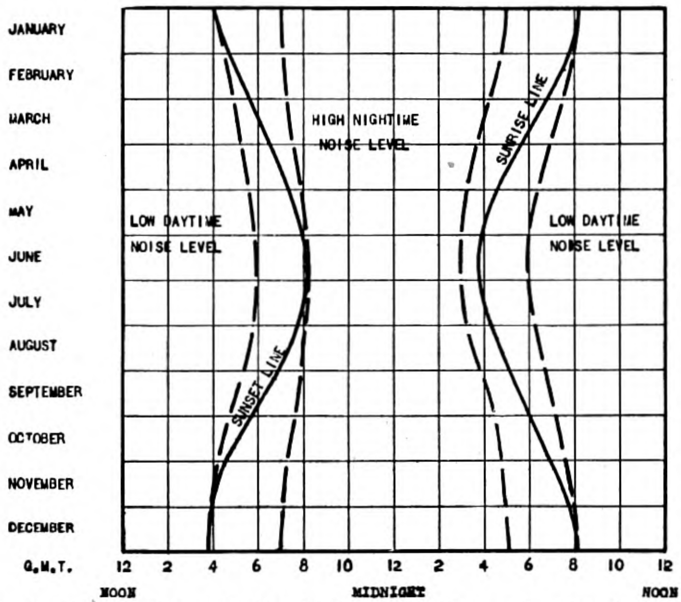
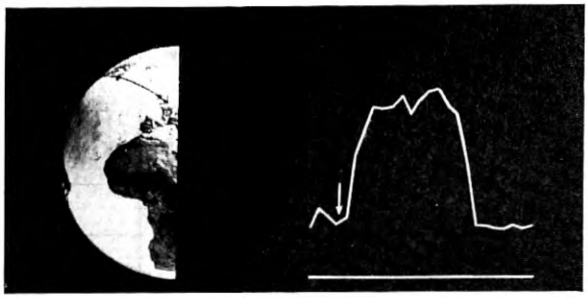


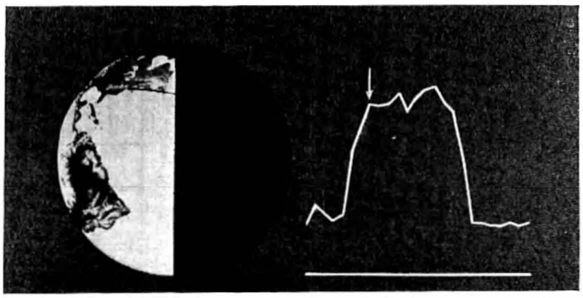
Fig. 18—Seasonal variation in distribution of daytime and night time noise with respect to sunset and sunrise, New Southgate, England—1923–1924

the reduction of static sets in during the summer months about the time of sunrise, reaches low daylight values several hours later, and in the winter the reduction commences several hours before sunrise and reaches low daylight values at sunrise. In other words, the rise to high night-time values occurs earlier with respect to sunset in the summer than in the winter, and conversely the fall from high night-time static to the lower daylight values occurs later with respect to sunrise, in the summer than in the winter.

This is more definitely brought out in Fig. 18 which combines the data for all frequencies measured. The dash-lines associated with the sunset curves, delineate the beginning and the attainment of the night-time increases and those associated with the sunrise curve delineate the beginning and the attainment of the low daylight values. This discloses the fact that sunset and sunrise at the receiving



a



b

c

Fig. 19—Noise at New Southgate, England, in January—Variation with exposure of equatorial belt to sunlight

point does not completely control the rise and fall of the high night-time static. It has been found that the discrepancy can be accounted for if sunrise and sunset are taken with respect to a static transmission path as distinguished from the receiving point alone, and if the assumption is made that the effect of sunlight upon the static transmission path is similar to that on usual radio transmission.

Major Regional Source of Static Noise

A broader conception as to the causes underlying the diurnal and seasonal variation is obtained by considering the time of sunset and sunrise over a considerable area of the earth's surface. Fig. 19 shows a series of day and night conditions for three representative parts of the diurnal noise characteristic at England for January. It will be seen that the rise to high night values does not begin until practically the time of sunset in England with over half of Africa still in daylight. By the time the high night-time values are reached, as indicated in the second phase, darkness has pervaded all of the equatorial belt to the south of England. Incidentally at this time sunset occurs between the United States and England, resulting in very poor signal transmission. The third phase of this series shows the noise having just reached the low daytime value and, although the sun is just rising in England, the African equatorial belt is in sunlight, subjecting the static transmission path to high daylight attenuation.

The sunset conditions which existed for the afternoon and evening of the day upon which the diurnal measurements of Fig. 14 were taken are shown in Fig. 20. The hourly positions of the sunset line are shown in relation to the evening rise of static in London. The coincidence between the arrival of sunset in London and the *start* of the high night-time noise on the higher frequencies is evident. By the time the high night-time values are reached, about 7 o'clock G.M.T., the equatorial belt to the south of London is in darkness.

Fig. 21 shows the sunrise conditions in relation to the decrease in static from the high night-time values to the lower daylight values. The decline starts about 5 or 6 o'clock an hour or two before sunrise, and is not completed until several hours later, at which time daylight has extended over practically the entire tropical belt to the south of England which corresponds in general to equatorial Africa.

Another fact presented in the previous figures which appears to be significant in shedding light upon the source of static, is that noise on the lower frequencies rises earlier in the afternoon and persists later into the morning than does the noise on the higher frequencies. This could be accounted for on the basis that the limits of the area from which the received longer wave static originates, extend farther along the equatorial zone than they do for the higher frequencies.

The inclination of the shadow line on the earth's surface, which is indicated in the previous figure for October 28, shifts to a maximum at the winter solstice, recedes to a vertical position at the equinox and then inclines in the opposite direction. These several positions are illustrated in Fig. 22. The set of three full lines to the right shows the position which the sunset shadow line assumes upon the earth's surface for each of three seasons—winter solstice, equinox, and summer solstice. Likewise, the dash-line curves show the position assumed by the sunrise line for the corresponding seasons. The

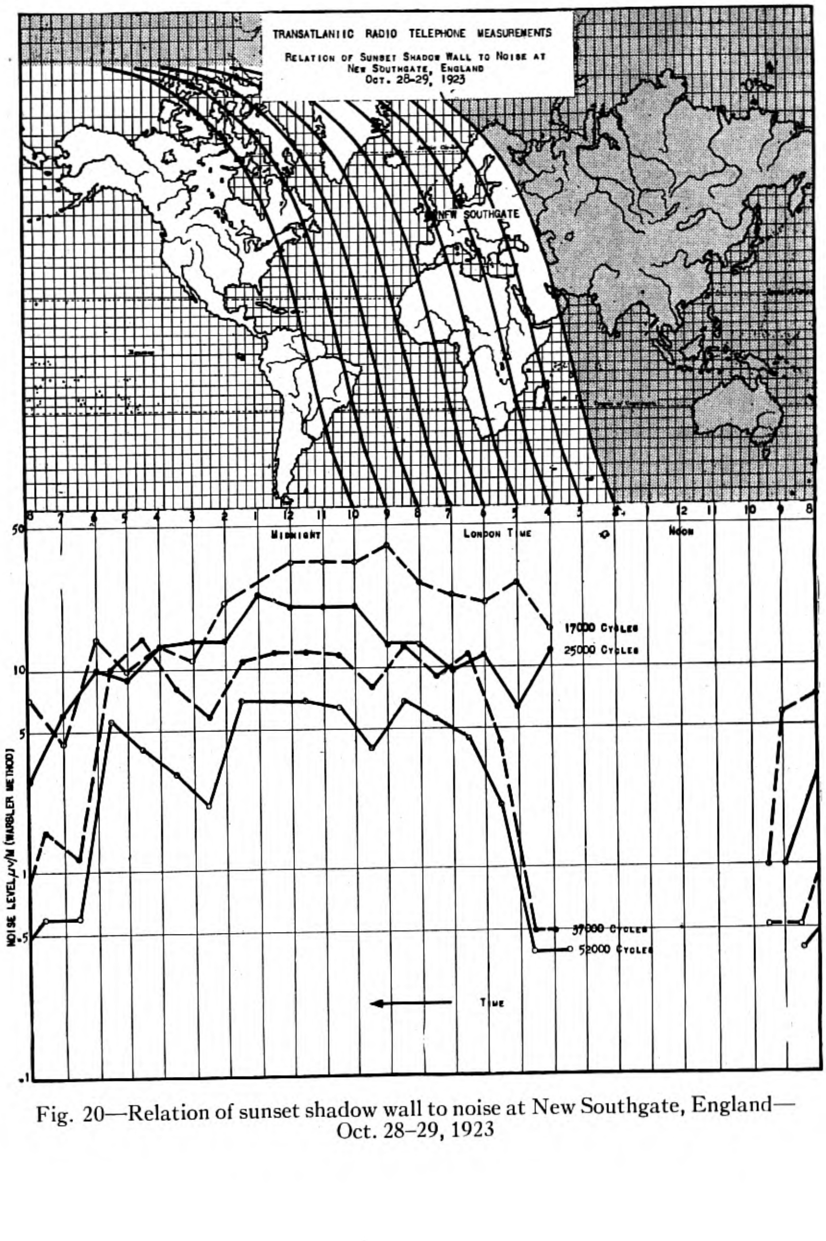


Fig. 20—Relation of sunset shadow wall to noise at New Southgate, England—Oct. 28–29, 1923

particular time of day for which each of the sunset curves is taken, is that at which the static in London *begins* to increase to large night values. In winter, this occurs about sunset, at the equinox about one hour earlier, and in summer about two hours earlier, as illustrated in Fig. 18. Correspondingly, the time for which each of the sunrise curves is taken, is that at which the high night-time values have reached the lower daylight values. From Fig. 18 it will be evident

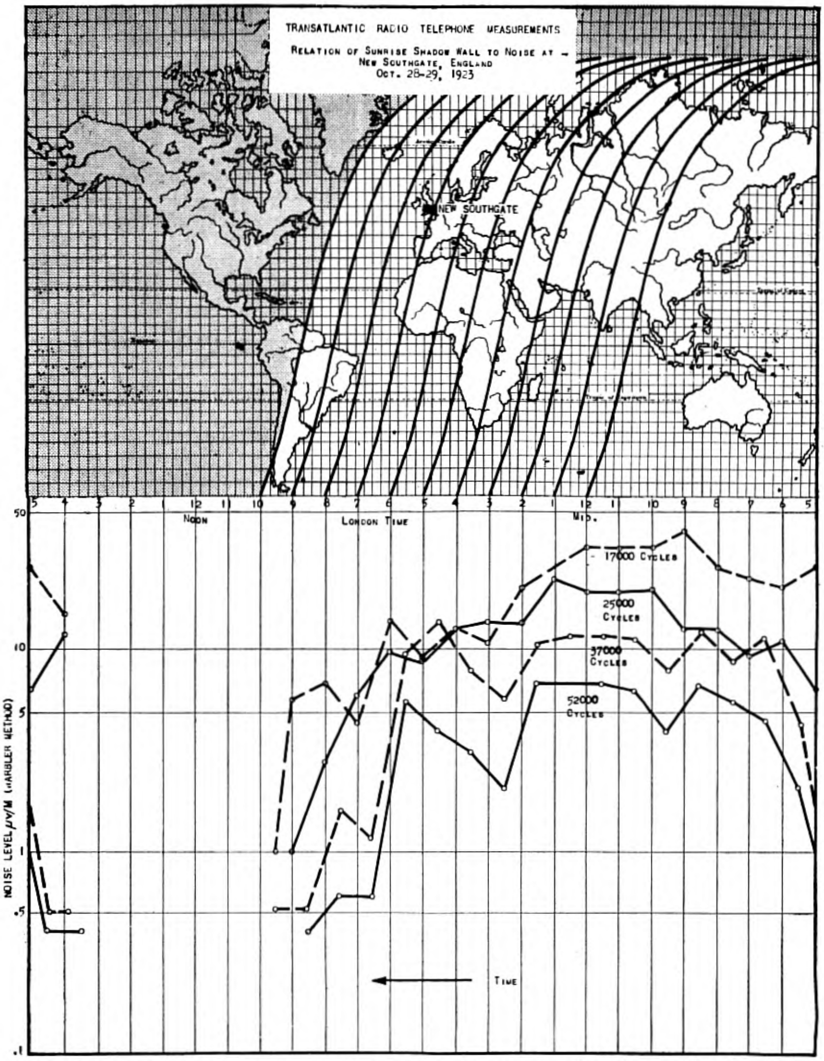


Fig. 21—Relation of sunrise shadow wall to noise at New Southgate, England—Oct. 28–29, 1923

that this occurs during the winter at about sunrise, at the equinox about an hour later, and during the summer some two hours later.

It will be observed that the two sets of curves, one for sunset and the other for sunrise, intersect at approximately the same latitude, the sunset curves southeast and the sunrise curves southwest of

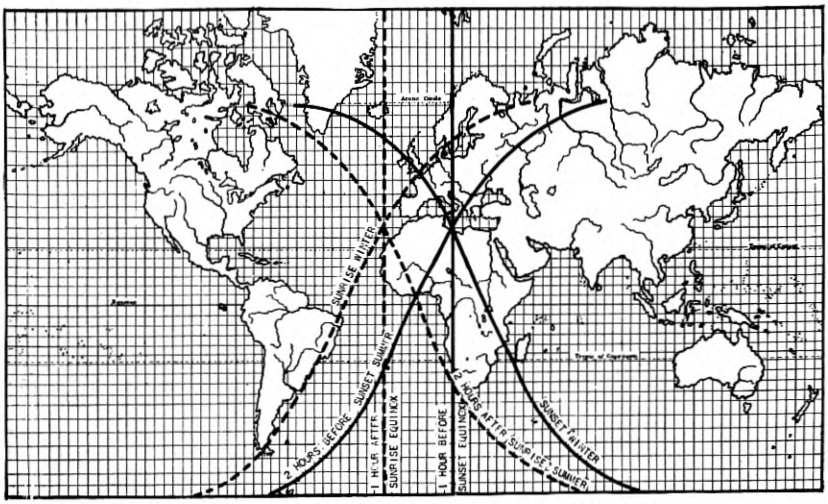


Fig. 22—Position of sunset lines at sunset dip and sunrise lines at sunrise dip in noise level in England for various seasons

England. If it is assumed that the effect of the shadow wall upon the transmission of static is similar to that upon signal transmission across the Atlantic, namely, the high night-time values commence when the shadow wall is approximately half-way between the terminals, the crossing of the lines upon the chart may be taken as having significance in roughly determining the limits of the tropical area from which the major static originates. The crossing of the sunset lines indicates that the eastern limit of the area which contributes most of the static to England is equatorial East Africa. The crossing of the sunrise lines indicates that the corresponding western limit is somewhere in the South Atlantic, between Africa and South America. In other words, from these data the indications are that there is a more or less distinct center of gravity of static, which extend along the tropical belt, and that most of the long-wave static which affects reception in England comes from the equatorial region to the south of England, namely, equatorial Africa. This is exclusive of the high afternoon static prevailing during the summer months.

The data obtained in the United States indicate that generally similar conditions exist there as to the relation between sunset and sunrise path and the major rise and fall of static. This relationship is shown in Fig. 23, which shows in the upper half the course of the

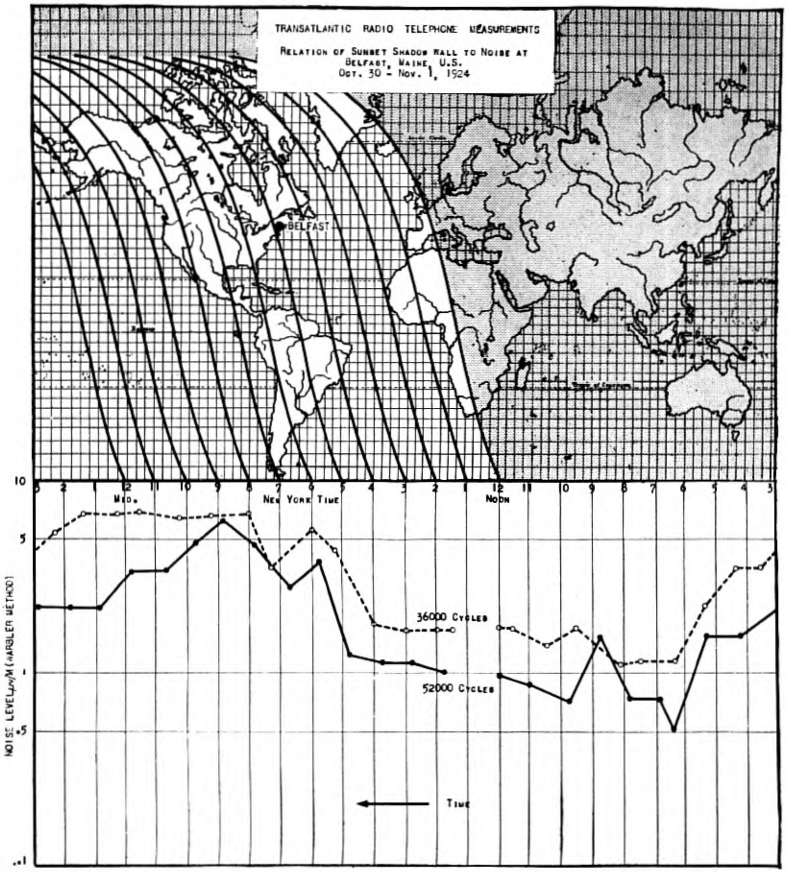


Fig. 23—Relation of sunset shadow wall to noise at Belfast, Maine, U. S.—Oct. 30—Nov. 1, 1924

night-time belt as it proceeds from Europe to America and the corresponding rise in the static noise. The noise level curves are the same as those shown in Fig. 14 for reception at Belfast, Maine. The rise commences about one hour before and continues for one hour or so after sundown. This is for the fall season of the year. A similar chart for the sunrise conditions is given in Fig. 24. Although high night-time values started to fall off some five hours before sundown in Belfast, the more rapid drop was within some two hours in advance. While these curves are for but a single day, they are fairly representative of the average of a greater amount of data. The change in the inclination of the sunset-sunrise curves with the

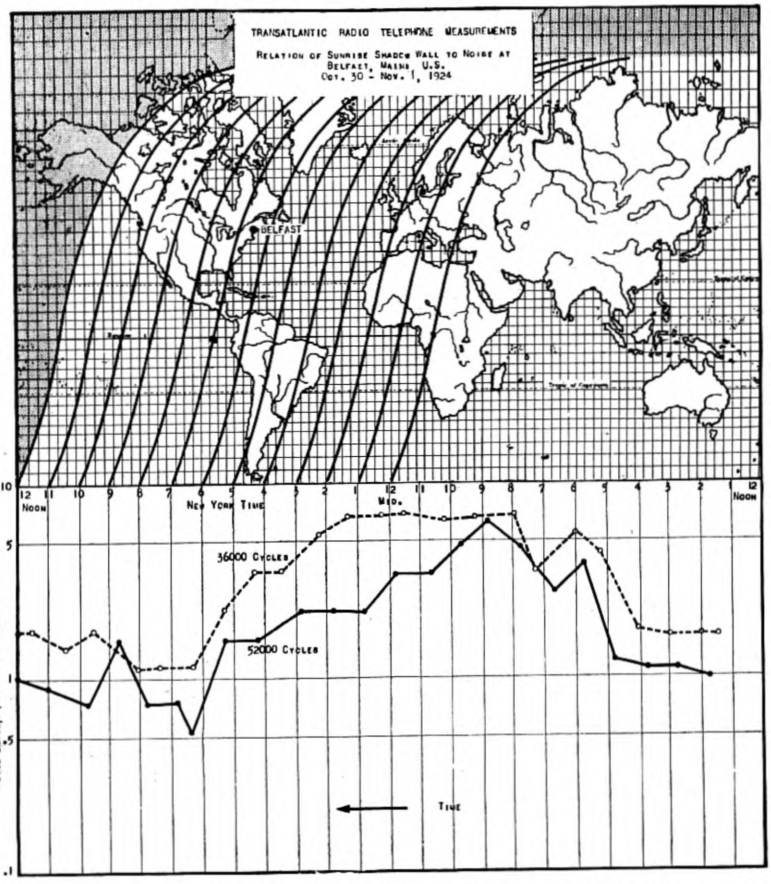


Fig. 24—Relation of sunrise shadow wall to noise at Belfast, Maine, U. S.—Oct. 30—Nov. 1, 1924

season of the year effects changes for American reception somewhat similar to those shown for reception in England, except that for the summer months the coincidence is less definite. It may be that this is because of the somewhat lower latitude of the United States terminal and of the reception of a greater propoıtion of the static from the North American continent.

In general, therefore, the American results accord with those obtained in England in indicating quite definitely that a large proportion of the static received on the longer waves is of tropical origin.

Signal to Noise Ratio

It is, of course, the ratio of the signal to noise strength which determines the communication merit of a radio transmission channel.

Variation with Frequency. A comparison for representative summer and winter months is given in Fig. 25 of the signal-to-noise ratio

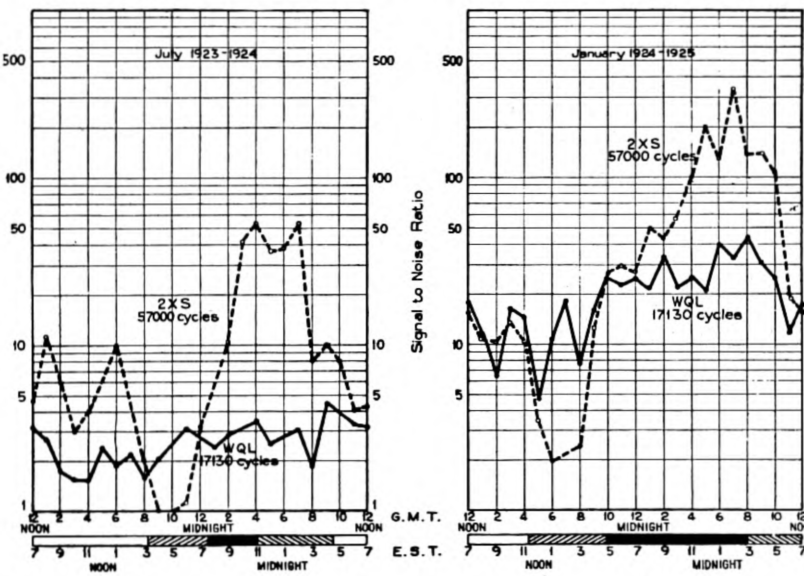


Fig. 25—Variation of signal to noise ratio with frequency. Corrected to same antenna input power (68.5 KW) in Rocky Point antenna—Reception at New Southgate, England

for the two extreme frequencies measured. Both of these transmissions were effected from the same station, Rocky Point, and similar antennae were employed. Comparison is made of the overall transmission by correcting the values of the two curves to the same antenna power input, the power of both channels being scaled down to 68 kilowatts, the power used in the telephone channel during the early parts of the experiment. This chart shows clearly the greater stability in signal to noise ratio obtainable on the lower frequency channel. While for certain periods of the day the higher frequency gives a much better ratio, it is subject to a much more severe sunset

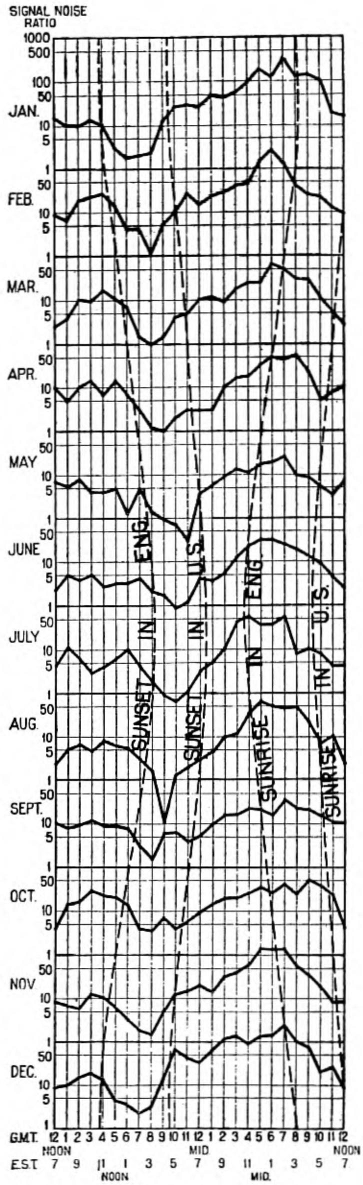


Fig. 26—Monthly averages of diurnal variation of signal to noise ratio; Rocky Point, L. I. (2 X S) received at New Southgate, England; 20.8 KW radiated Power —57,000 cycles—5480 Km—1923-24

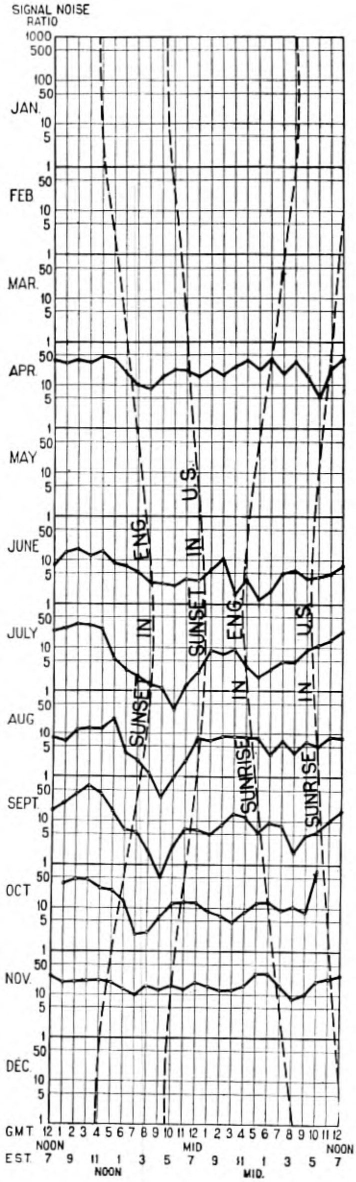


Fig. 27—Monthly averages of diurnal variation of signal to noise ratio, Northolt, Eng. (GKB) received at Belfast, Maine—20.8 KW radiated power—4980 Km—52,000 cycles—1924

decline than is the lower frequency. During the summer time, afternoon reception in England is better on the higher frequency channel. This is because of the considerably greater static experienced at this time on the lower frequency. The higher signal-to-noise ratio prevailing during the winter month of January as compared with the summer month of July is evident. This is due primarily to higher summer static.

Seasonal Variation in England and United States. For the 57-kilocycle channel there is shown in Fig. 26, for each month of the year, signal-to-noise ratios of two years' data. These show a distinct dip corresponding to the sunset dip of the signal field strength. The night-time values are generally high in accordance with the high night-time signal strength but the maximum values are shifted toward the time of sunrise. This is due to the fact that the noise rises earlier in the afternoon and declines earlier in the morning than do the corresponding variations in signal strength.

Fig. 27 presents the signal-to-noise ratios for such data as have thus far been obtained upon transmission from England to the United States on a frequency of 52 kilocycles. The low values obtained about sunset are, of course, due to the evening dip in field strength. In general, the night-time ratios do not reach high values as do those for England because the early morning signal field strength begins to fall off while the noise level is still high. Comparisons of the signal-to-noise ratios obtained at New Southgate and at Belfast show that the Belfast values are somewhat higher for that part of the day, corresponding to forenoon in the United States and afternoon in England. This is because the forenoon static in the United States is lower than the afternoon static in England.

Directive Receiving Antennae

The picture which has been given of the transmission of static northward from the tropical belt suggests that the signal-to-noise ratio might be materially improved by the use of directional receiving systems. This is, of course, what has actually been found to be the case in commercial transatlantic radio telegraphy wherein the Radio Corporation has made such effective use of the wave antenna devised by Beverage. The expectations are confirmed by measurements which have been made in the present experiments using such wave antennae.

A year and a half ago the British Post Office established a wave antenna with which to receive from the Rocky Point radio telephone transmitter. More recently a program of consistent observations in directional reception of east-to-west transmission was also undertaken in which were employed, wave antennae built by the Radio Corporation of America for radio telegraph operation upon lower frequencies.

An indication of the improvement which the wave antenna gives in signal-to-noise ratio is had by reference to Fig. 28. The set of

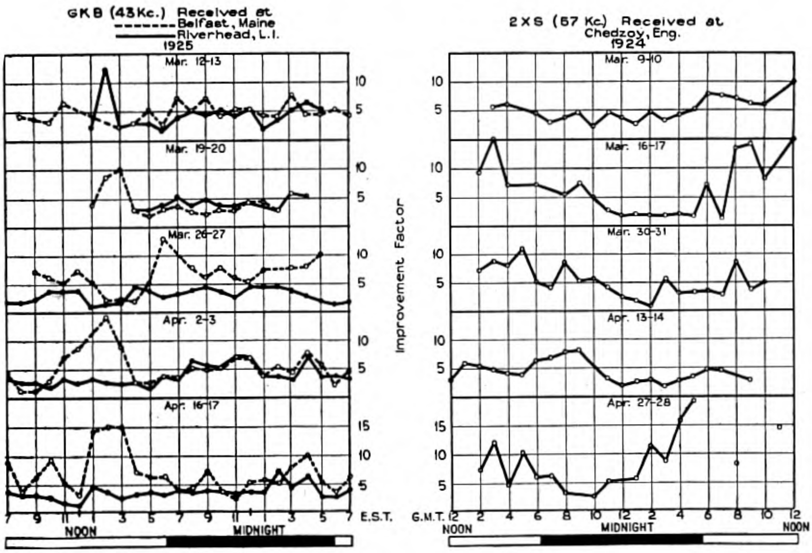


Fig. 28—Improvement in signal noise ratio of wave antenna over loop reception

curves to the right is for reception at Chedzoy, England, and those at the left for reception at Belfast and Riverhead in the United States. The improvement is measured in terms of the signal-to-noise ratio obtained on the wave antenna, divided by the signal-to-noise ratio measured on the loop. For the particular days and frequency indicated, the improvement in England will be seen to vary over a considerable range, averaging about 5. Data for reception in England is for 1924 while that for the United States is for the corresponding period of 1925. The United States results will be seen to be generally similar to those obtained in England. Although these experiments are still in an early stage, the results do give a measure of the order of improvement which can be expected.

Test of Words Understood. Perhaps the most convincing measure of the efficiency of directional receiving systems for transatlantic transmission is the improvement effected in the reception of intelligible words. Fig. 29 shows the improvement which the wave antenna in England has made in the ability to receive certain test words spoken from Rocky Point. For this purpose there was transmitted from Rocky Point a list of disconnected words. A record

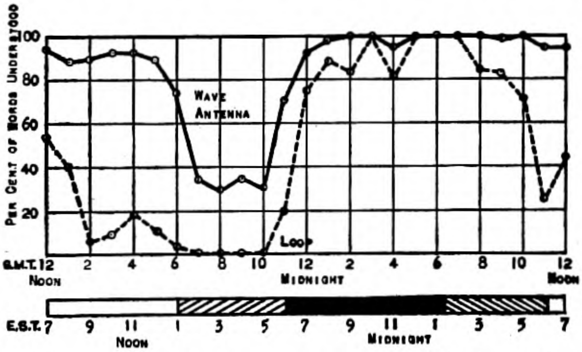


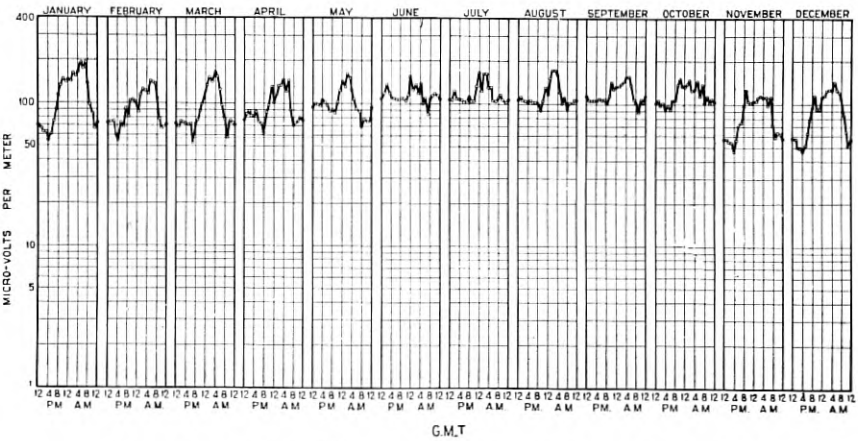
Fig. 29—Comparison of reception on wave antenna and loop. Per cent of words understood—Reception of Rocky Point (2 X S) at Chedzoy, England, March, 1924

was made at Chedzoy of the percentage of the words understood for reception on the loop and on the wave antenna. This constitutes a convenient method of rough telephone testing. It will be appreciated, however, that it would be possible to understand a greater proportion of a conversation than is represented by these results. The curves show that it was possible to receive, for example, 80% of the words for but 9 of the 24 hours on the loop, whereas with the wave antenna reception continued for 18 hours.

APPENDIX

Transatlantic Radio Telephone Measurements
1923, 1924, 1925

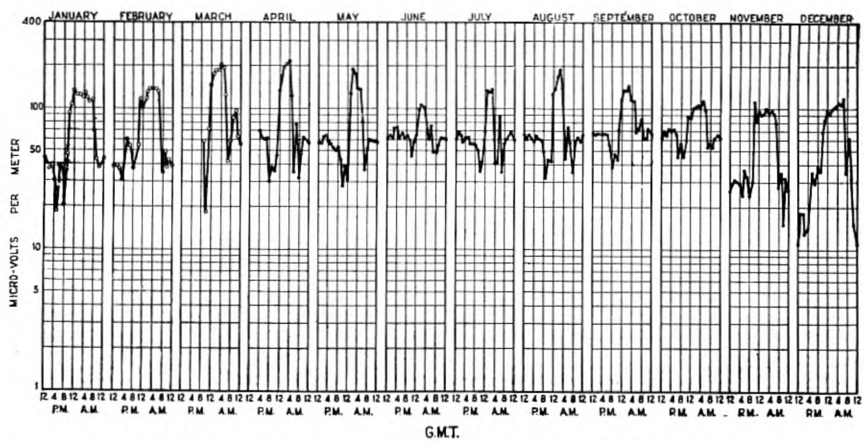
Month by Month Record of Noise and Field Strength



Monthly Averages of Diurnal Variation of Signal Field Strength
Rocky Point, L. I., U. S. A. (WQL) Measured at New Southgate, England
Corrected to 600 Amperes Antenna Current

5,480 Km. 17,130 Cycles

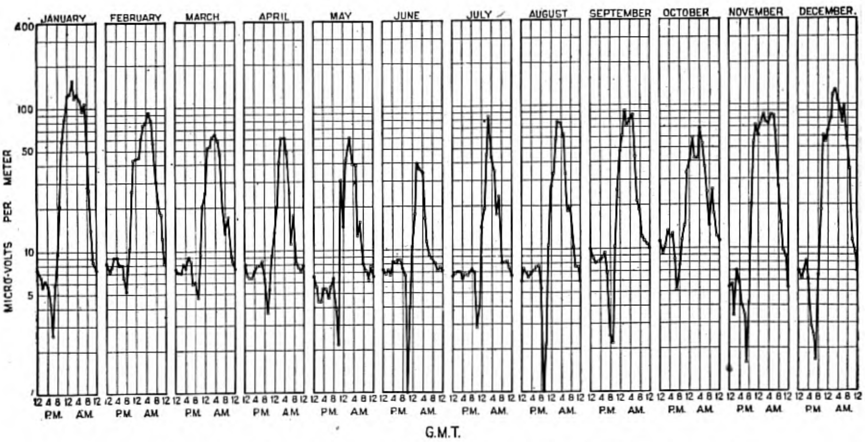
April, 1923—Feb., 1925



Monthly Averages of Diurnal Variation of Signal Field Strength
Marion, Mass., U. S. A. (WSO) Measured at New Southgate, England
Corrected to 600 Amperes Antenna Current

5,280 Km. 25,700 Cycles

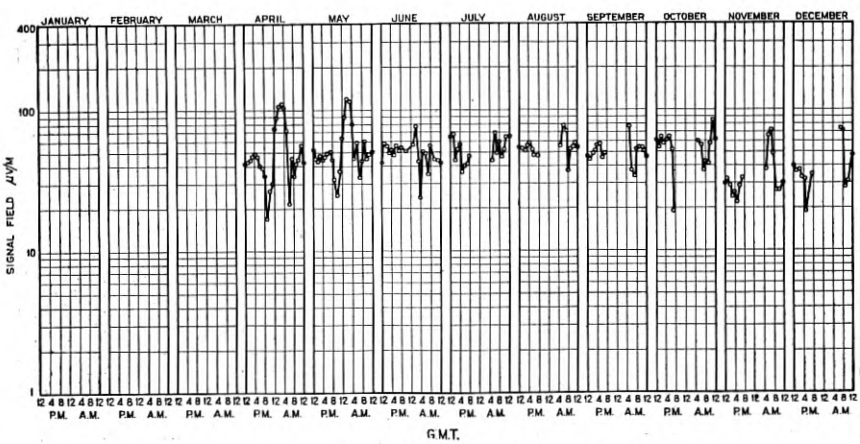
Aug., 1923—Feb., 1925



Monthly Averages of Diurnal Variation of Signal Field Strength
Rocky Point, L. I., U. S. A. (2XS) Measured at New Southgate, England
Corrected to 300 Amperes Antenna Current

5,480 Km. 57,000 Cycles

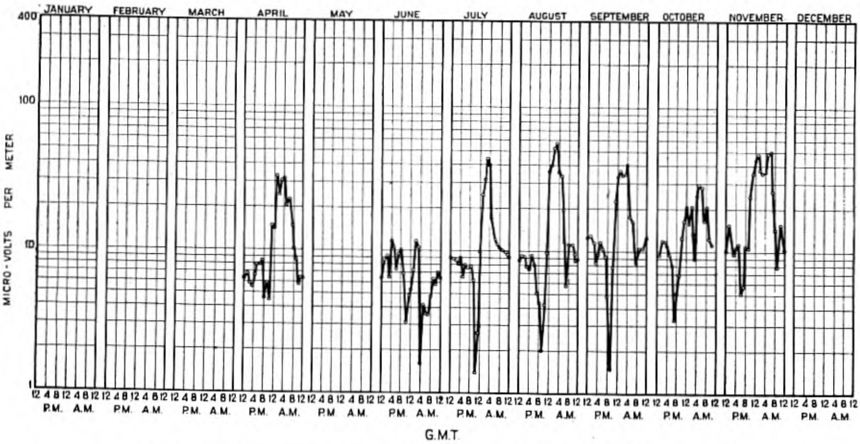
Jan., 1923—Dec., 1924



Monthly Averages Diurnal Variation of Signal Field Strength
Leafield, England (GBL) Measured at Belfast, Maine
Corrected to 300 Amperes Antenna Current

4,980 Km. 24,050 Cycles

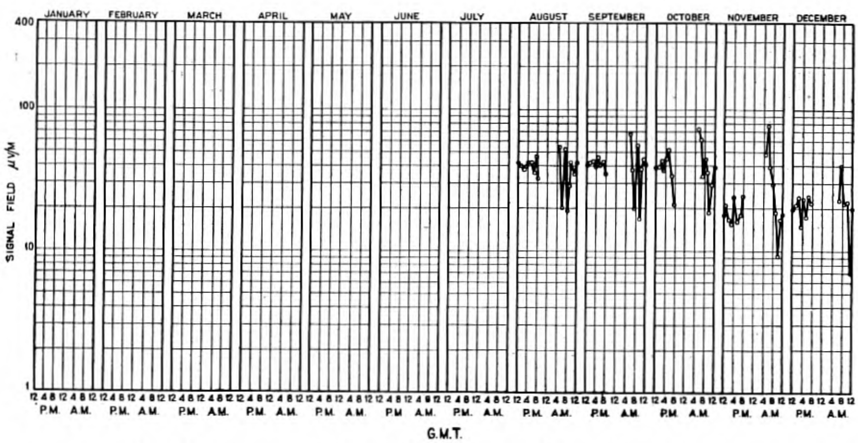
1924



Monthly Averages Diurnal Variation of Signal Field Strength
Northolt, England (GKB) Measured at Belfast, Maine
Corrected to 100 Amperes Antenna Current

4,885 Km. 52,000 Cycles

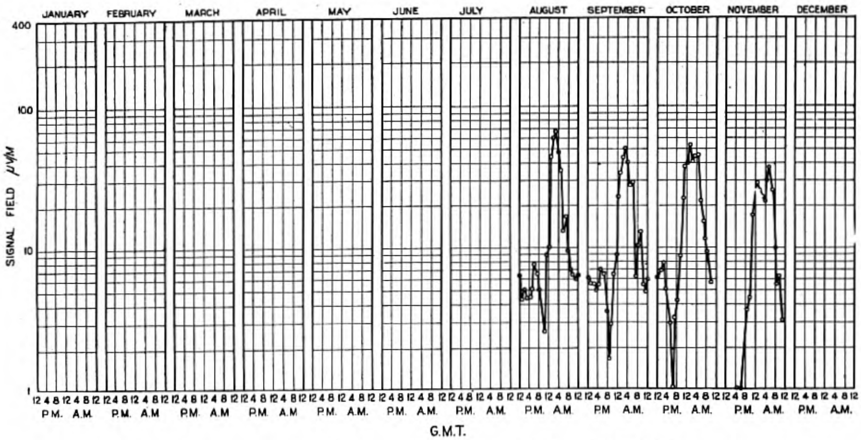
1924



Monthly Averages Diurnal Variation of Signal Field Strength
Leafield, England (GBL) Measured at Riverhead, L. I.
Corrected to 300 Amperes Antenna Current

5,360 Km. 24,050 Cycles

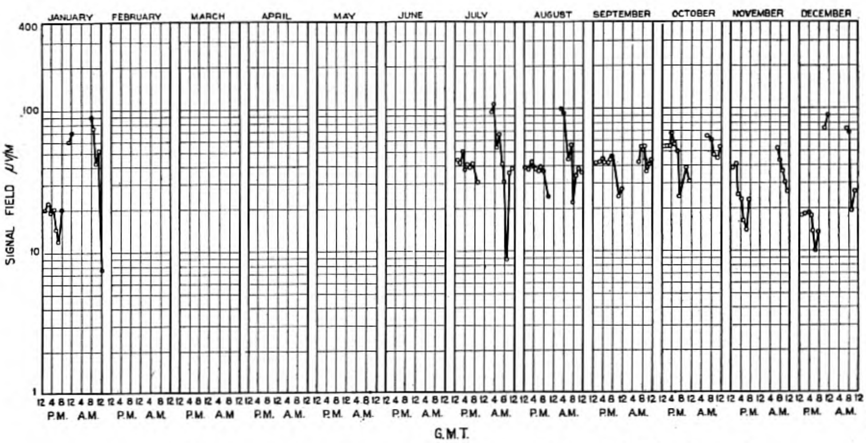
1924



Monthly Averages Diurnal Variation of Signal Field Strength
Northolt, England (GKB) Measured at Riverhead, L. I.

5,460 Km. **52,000 Cycles**

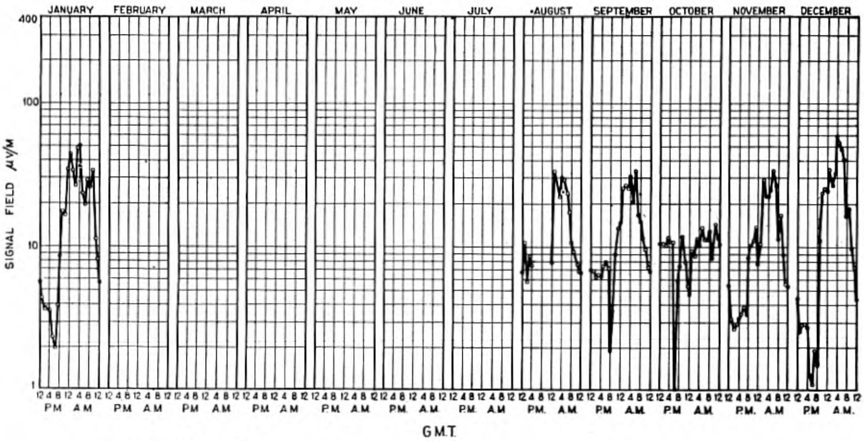
1924



Monthly Average of Diurnal Variation of Signal Field Strength
Leafield, England (GBL) Measured at Green Harbor, Mass.
Corrected to 300 Amperes Antenna Current

5,150 Km. **24,050 Cycles**

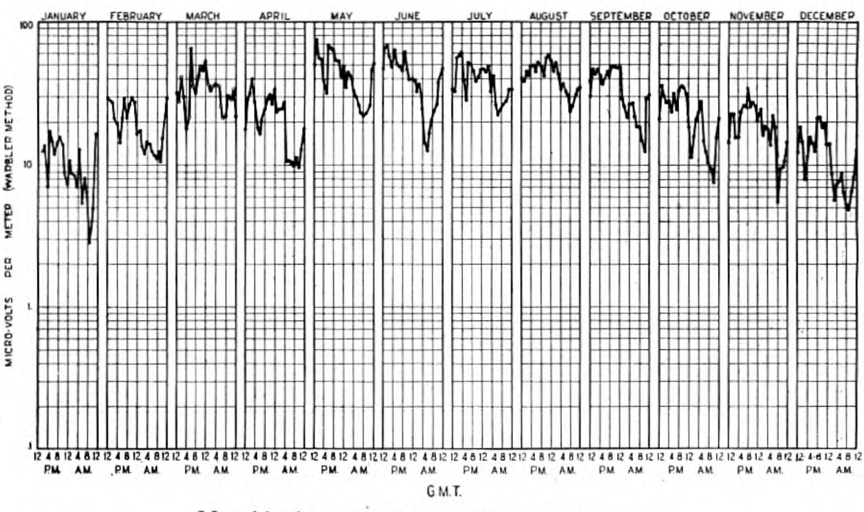
July, 1923—Jan., 1924



Monthly Average of Diurnal Variation of Signal Field Strength
Northolt, England (GKB) Measured at Green Harbor, Mass.
Corrected to 100 Amperes Antenna Current

5,240 Km. 54,500 Cycles

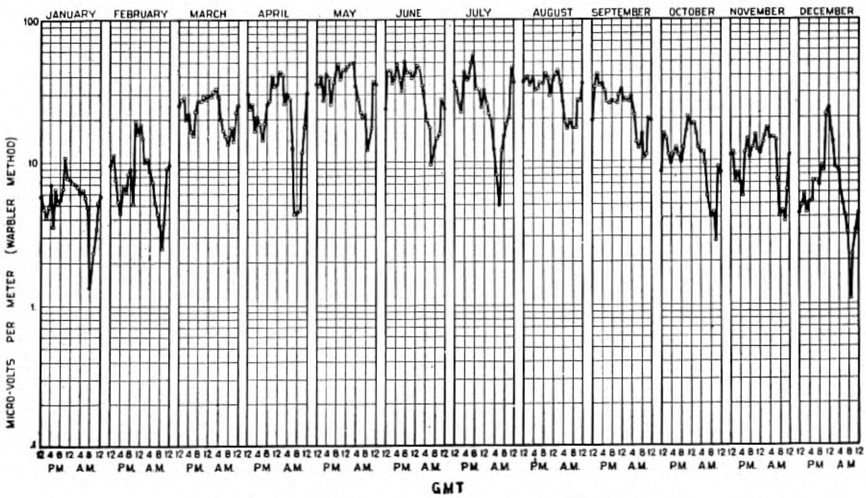
Aug., 1923—Jan., 1924



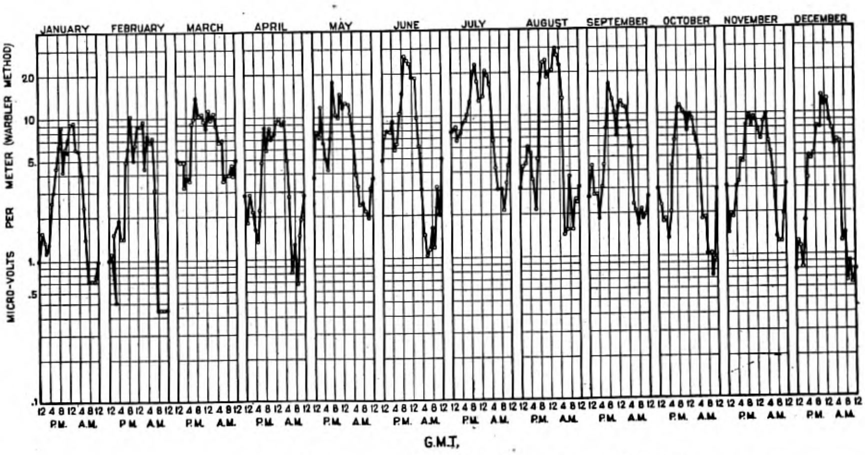
Monthly Averages of Diurnal Variation of Noise

New Southgate, England 17,000 Cycles

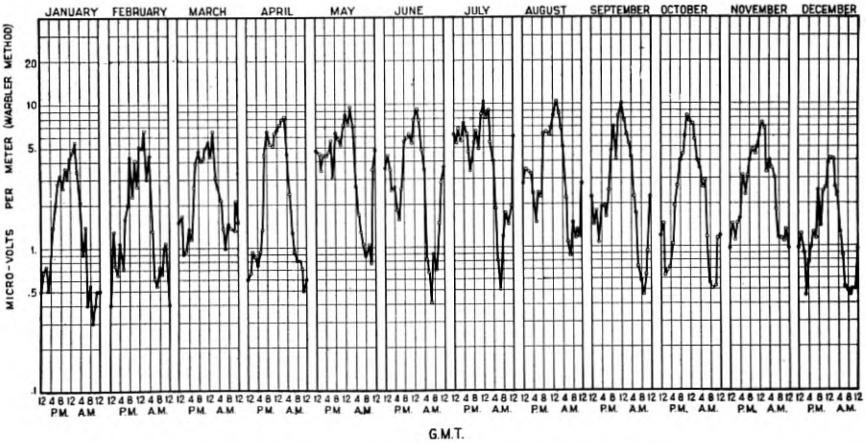
April, 1923—Feb., 1925



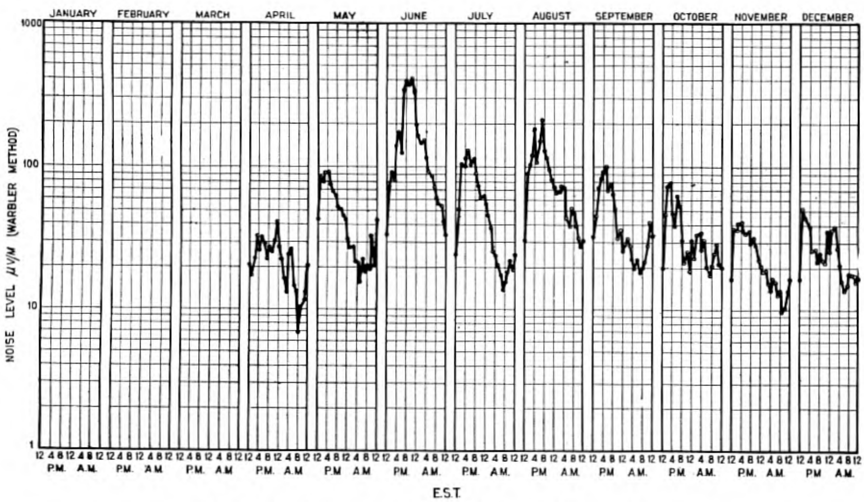
Monthly Averages of Diurnal Variation of Noise
New Southgate, England
25,000 Cycles
Aug., 1923—Feb., 1925



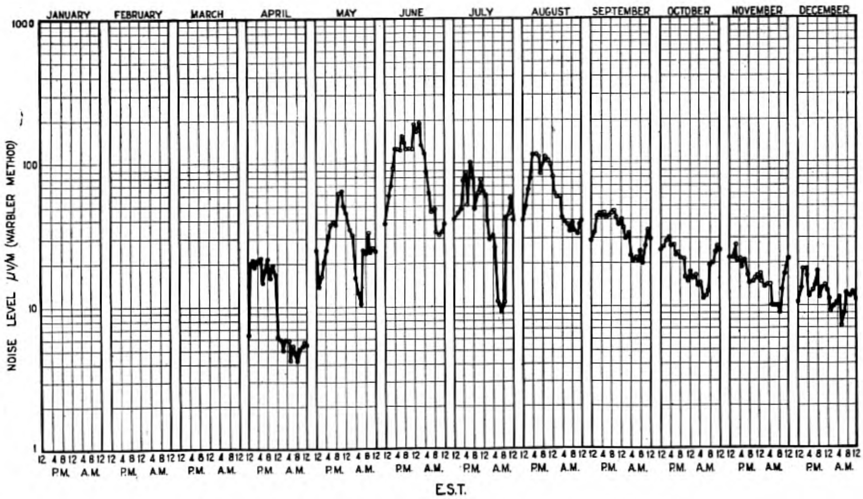
Monthly Averages of Diurnal Variation of Noise
New Southgate, England
37,000 Cycles
Oct., 1923—Feb., 1925



Monthly Averages of Diurnal Variation of Noise
New Southgate, England 57,000 Cycles
1923—1924

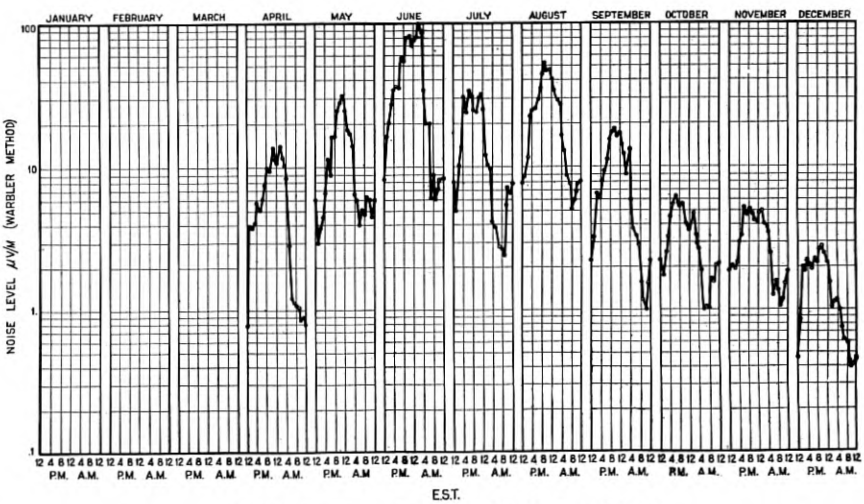


Monthly Average of Diurnal Variation of Noise
Belfast, Maine 15,000 Cycles
1924



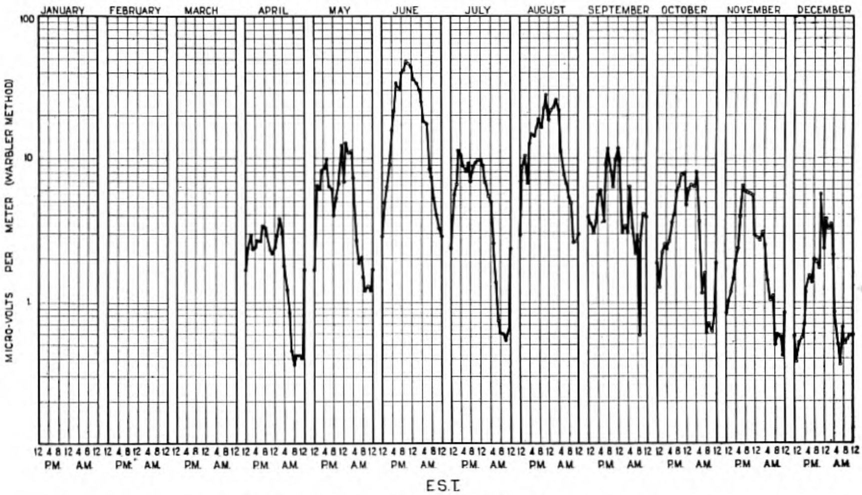
Monthly Average of Diurnal Variation of Noise

Belfast, Maine 1924 24,000 Cycles



Monthly Average of Diurnal Variation of Noise

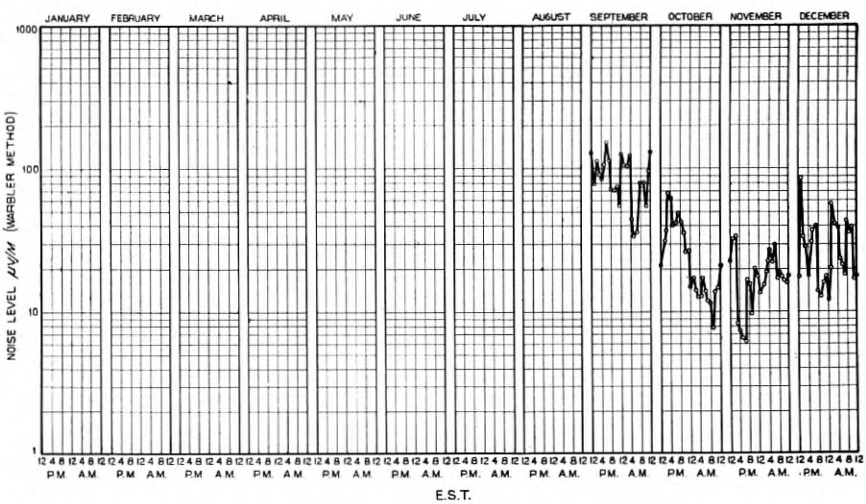
Belfast, Maine 1924 36,000 Cycles



Monthly Average of Diurnal Variation of Noise

Belfast, Maine 52,000 Cycles

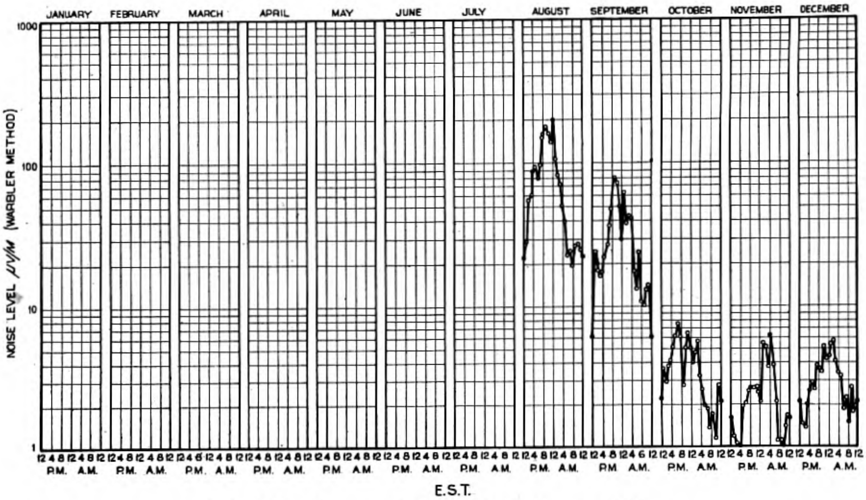
1924



Monthly Average of Diurnal Variation of Noise

Riverhead, L. I. 15,000 Cycles

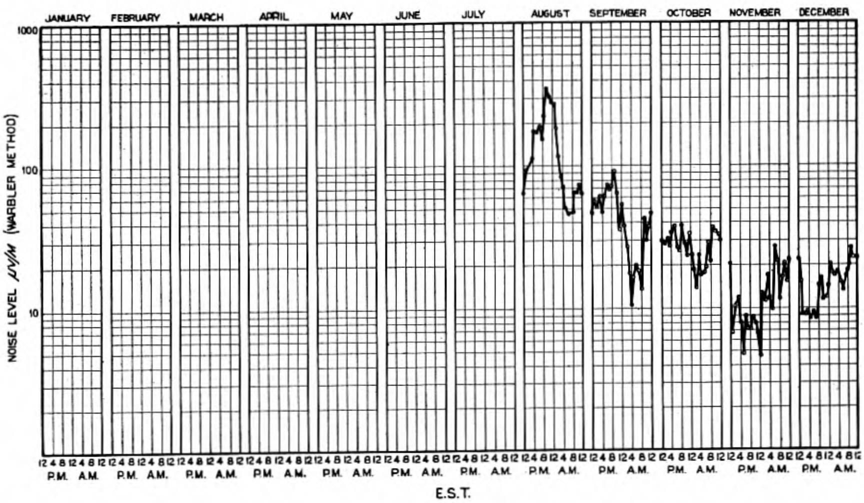
1924



Monthly Average of Diurnal Variation of Noise

Riverhead, L. I. 36,000 Cycles

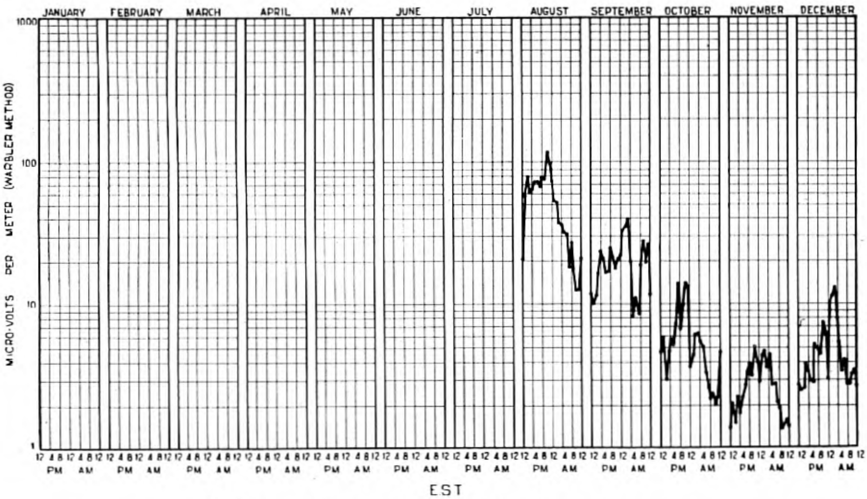
1924



Monthly Average of Diurnal Variation of Noise

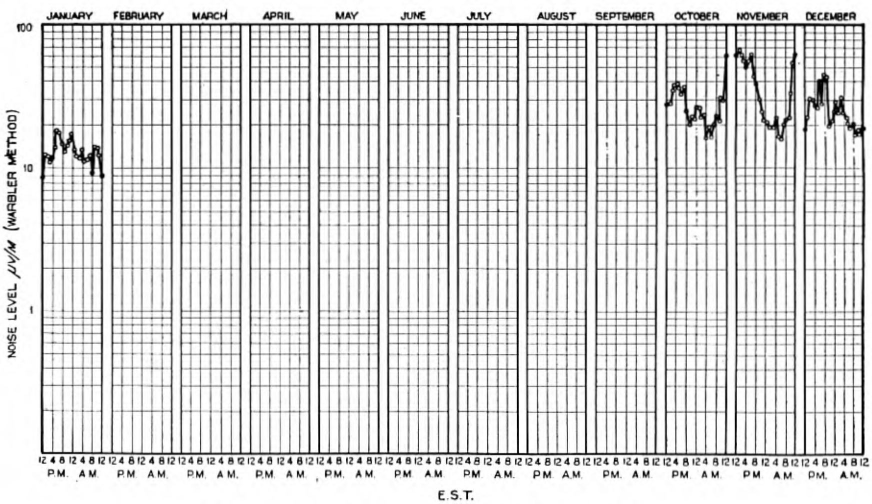
Riverhead, L. I. 24,000 Cycles

1924



Monthly Average of Diurnal Variation of Noise

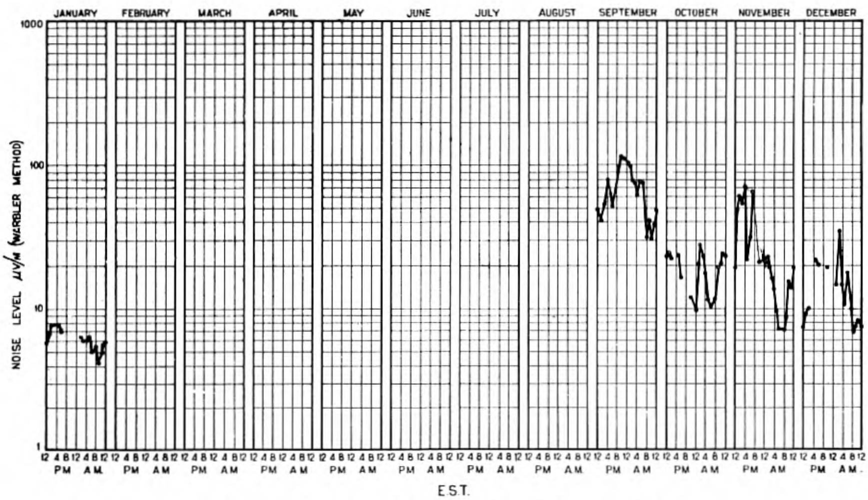
Riverhead, L. I. 52,000 Cycles



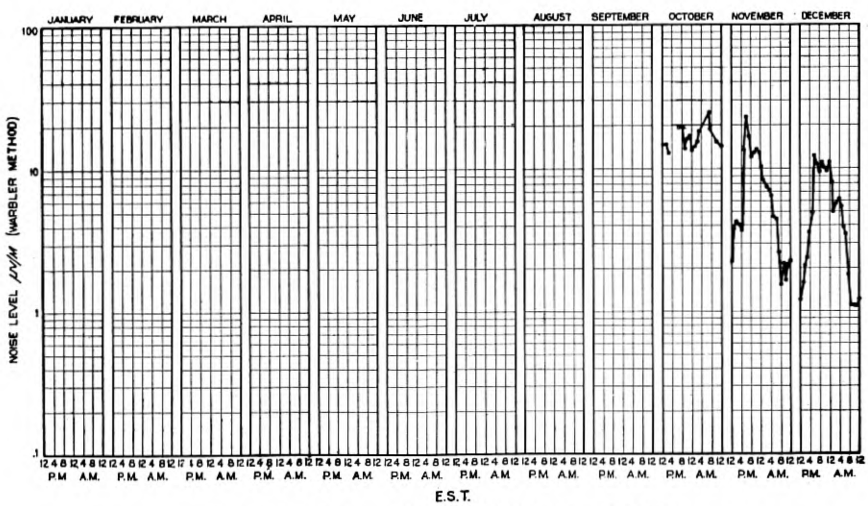
Monthly Average of Diurnal Variation of Noise

Green Harbor, Mass. 15,000 Cycles

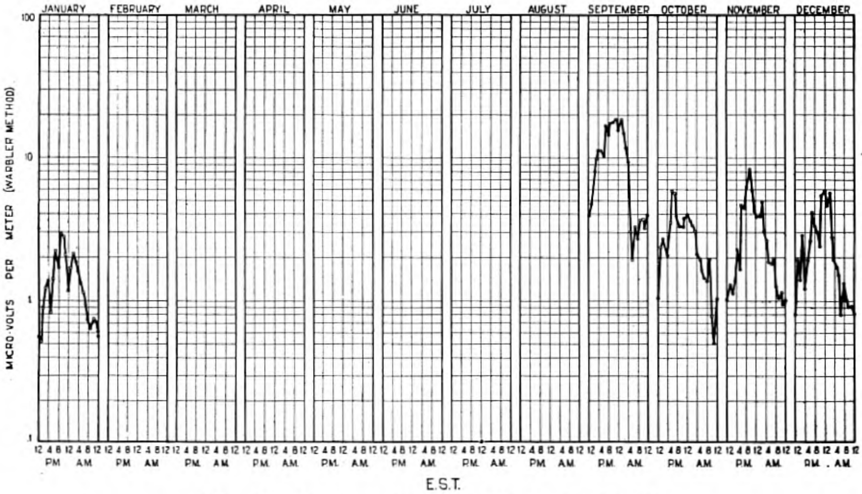
Oct., 1923—Jan., 1924



Monthly Average of Diurnal Variation of Noise
Green Harbor, Mass. 24,000 Cycles
Sept., 1923—Jan., 1924



Monthly Average of Diurnal Variation of Noise
Green Harbor, Mass. 34,000 Cycles
1923



Monthly Average of Diurnal Variation of Noise

Green Harbor, Mass. **55,000 Cycles**

Sept., 1923—Jan., 1924

Abstracts of Bell System Technical Papers Not Appearing in this Journal

Radioactivity.[1] A. F. KOVARIK and L. W. MCKEEHAN. This review of progress in radioactivity forms one of a series of monographs prepared by committees of the National Research Council. It outlines the experimental and theoretical advances in the subject since 1916, the date of the last compendium. The section headings are: I. Introduction, II. Radioactive Transformations, III. Alpha-Rays, IV. Beta-Rays, V. Gamma-Rays, VI. Nuclear Structure and Radioactive Processes, VII. Radioactivity in Geology and Cosmology, VIII. The Effects of Radioactive Radiations upon Matter. The references to periodical literature are particularly detailed.

Echo Suppressors for Long Distance Telephone Circuits.[2] A. B. CLARK and R. C. MATHES. This paper gives a brief description of a device which has been developed by the Bell System for suppressing "echo" effects which may be encountered under certain conditions in telephone circuits which are electrically very long. The device has been given the name "echo suppressor" and consists of relays in combination with vacuum tubes which are operated by the voice currents so as to block the echoes without disturbing the main transmission.

A number of echo suppressors have been operated on commercial telephone circuits for a considerable period, so that their practicability has been demonstrated.

The Telephone Transmission Unit.[3] DR. F. B. JEWETT. The adoption by the Bell System of the TU as a telephone transmission unit aroused considerable active discussion in foreign circles, namely, by Colonel Purves, Engineering Chief of the British Post Office Department, and Dr. Breisig of the German Telephone Administration. In this short paper, Dr. Jewett explains certain words and expressions which, when accurately defined, he believes will eliminate misinterpretations such as seem to have led to the controversies over the Bell System TU.

Dr. Jewett also points out that the numerical size for a transmission unit is controlled by two factors, first, the magnitude should be such that computation is convenient, and second, the magnitude should be such as to permit telephone engineers and operating people to most

[1] Bulletin National Research Council, Vol. 10, part 2, March, 1925, 203 pages.

[2] Journal A. I. E. E., Vol. 44, page 618, 1925.

[3] London Electrician, Vol. 94, page 562, 1925

readily comprehend the ratios corresponding to any given number of units. Since it is desirable that every unit be based on a decimal system of notation, unless there is some very important reason why it should not, the TU based on the decimal system was chosen. Satisfactory experience during the past year and a half is pointed to as showing the wisdom of having chosen the TU.

A Suspension for Supporting Delicate Instruments.[4] A. L. JOHNSRUD, Bell Telephone Laboratories, Incorporated, New York. A description, with diagram, is given of a modified Julius suspension designed especially to eliminate disturbances due to vertical vibrations from the building structure. The frame holding the instrument is supported by a system of tape-wound coil springs, which, because of the tightly wound friction tape, damp out mechanical vibrations. The frame with its balancing weights, is heavy (about 120 pounds), and so proportioned in mass that a twisting or tilting impulse, necessary at times in adjusting the instrument, disturbs its moving system only in a secondary degree. This is a second feature of this suspension. Surprisingly effective kinetic insulation is achieved. Quadrant electrometers and a moving magnet galvanometer have remained undisturbed even when heavy trucks were passing on the street seven floors below. This type of suspension, developed some years ago through the efforts of Mr. H. C. Harrison and Mr. J. P. Maxfield, has been adapted for use throughout the Bell Telephone Laboratories in a variety of ways.

Power Amplifiers in Transatlantic Radio Telephony.[5] A. A. OSWALD and J. C. SCHELLENG. The paper describes the development of a 150-kilowatt (output) radio frequency amplifier installation built for transatlantic telephone tests. The characteristics of the single-sideband eliminated-carrier method of transmission are discussed with particular reference to its bearing upon the design of the power apparatus. A classification of amplifiers is proposed in which there are three types distinguished from each other by the particular portion of the tube characteristic used. The water-cooled tubes employed in these tests are briefly described, special consideration being given to their use in a large installation. The system is then shown in outline by means of a block diagram, the elements of which are subsequently discussed in greater detail. The theory, electrical design, and mechanical construction of the last two stages of the amplifier are outlined, including the output and antenna circuits. Means employed to prevent spurious oscillations are described. The method

[4] Journal Opt. Soc. of Am., Vol. X, No. 5, pp. 609–611, May, 1925.

[5] Proc. of I. R. E., Vol. 13, page 313, June, 1925.

used in increasing the transmission band width to a value much greater than that of the antenna is explained. The power requirements of a single sideband installation are outlined and a description of the six-phase rectifier, used as a source of high potential direct current is given, together with a brief theoretical treatment of its operation. Circuit diagrams, photographs, and a number of characteristic curves are discussed.

Production of Single Sideband for Transatlantic Radio Telephony.[6] R. A. HEISING. This paper describes in detail the equipment and circuit used in the production of the single sideband for transatlantic radio telephony in the experiments at Rocky Point. The set consists of two oscillators, two sets of modulators, two filters, and a three-stage amplifier. The oscillators and modulators operate at power levels similar to those in high-frequency communication on land wires. The three-stage amplifier amplifies the sideband produced by these modulators to about a 500-watt level for delivery to the water-cooled tube amplifiers.

The first oscillator operates at about 33,700 cycles. The modulator is balanced to eliminate the carrier; and the first filter selects the lower sideband. In these transatlantic experiments the second oscillator operated at 89,200 cycles, but might operate anywhere between 74,000 and 102,000 cycles. The second modulator, which is also balanced, is supplied with a carrier by the second oscillator and with modulating currents by the first modulator and first filter. The second filter is built to transmit between 41,000 and 71,000 cycles, so that by varying the second oscillator, the resulting sideband, which is the lower sideband produced in the second modulating process, may be placed anywhere between these two figures. Transmission curves for the filters are given as well as some amplitude-frequency performance curves of the set.

A Null-Reading Astatic Magnetometer of Novel Design.[7] RICHARD M. BOZORTH. This instrument is designed for measuring the magnetic properties of very small amounts of material in the form of fine wires, thin tapes, or as thin deposits (electrolytic, evaporated, sputtered) supported on non-magnetic forms. The specimen, 4 cm. long, is mounted parallel to the line joining the two needles, so that its poles produce the maximum torque on the suspended needle system, the position of which is read by mirror and scale. The effect of the magnetizing coil on the needles is annulled once for all by the suitably placing of an auxiliary coil, and the magnetic effect of the

[6] Proc. of I. R. E., Vol. 13, page 291, 1925.
[7] J. O. S. A. and R. S. I. 10, 591-8 (May, 1925).

sample itself is balanced by passing a measured current through a third coil. The applied field and the induced magnetization are then proportional to the electric currents passed through the magnetizing coil and the balancing coil, respectively. A hysteresis loop is shown, obtained from an iron wire weighing 3 mg.

An Explanation of Peculiar Reflections Observed on X-Ray Powder Photographs.[8] RICHARD M. BOZORTH. There has been previously reported (J. O. S. A. and R. S. I. 6,989-97; 1922) the existence of "anomalous" reflections of X-rays, observed when analyzing substances by the method of Debye-Scherrer and Hull. These reflections are now explained in accordance with the well-known laws governing X-ray reflections. It is shown that the molybdenum X-ray spectrum as ordinarily used, although it is filtered by zirconium screens, contains in addition to the characteristic $K\alpha$ radiation a considerable amount of general radiation. Although usually not effective, this general radiation becomes important when the sample being analyzed is composed of crystal grains of certain sizes. The effect under discussion is caused by reflection of this general radiation from the principal atom planes of these crystals. Several experiments, and a geometrical analysis of the positions and orientations of the diffraction effects, confirm this conclusion.

[8] J. O. S. A. and R. S. I. 9, 123-7 (August, 1924).

Contributors to this Issue

Frank Gill, European Chief Engineer of the International Western Electric Company. Mr. Gill has had long experience as a telephone engineer, first, with the United Telephone Company in London, then with the National Telephone Company and later as a consulting engineer. At the outbreak of the war, he was called upon to undertake important work in the Ministry of Munition for which he was later awarded the Order of the British Empire. As European Chief Engineer of the International Western Electric Company, he is taking a leading part in the discussion and study of conditions necessary for the establishment of an adequate long distance telephone service through Europe.

Oliver E. Buckley, B.Sc., Grinnell College, 1909; Ph.D., Cornell University, 1914; Engineering Department, Western Electric Company, 1914–1917; U. S. Army Signal Corps, 1917–1918; Engineering Department, Western Electric Company (Bell Telephone Laboratories), 1918—. During the war Major Buckley had charge of the research section of the Division of Research and Inspection of the Signal Corps, A. E. F. His early work in the Laboratories was concerned principally with the production and measurement of high vacua and with the development of vacuum tubes. More recently he has been connected with the development and applications of magnetic materials and particularly with the development of the permalloy-loaded telegraph cable.

Harvey Fletcher, B.S., Brigham Young, 1907; Ph.D., Chicago, 1911; instructor of physics, Brigham Young, 1907–08; Chicago, 1909–10; Professor, Brigham Young, 1911–16; Engineering Department, Western Electric Company, 1916–24; Bell Telephone Laboratories, Inc., 1925—. During recent years, Dr. Fletcher has conducted extensive investigations in the fields of speech and audition.

Charles W. Carter, Jr., A.B., Harvard, 1920; B.Sc., Oxford, 1923; American Telephone and Telegraph Company, Department of Development and Research, 1923—.

Arthur S. Curtis, Ph.B., 1913; E.E., 1919; Sheffield Scientific School; Instructor in Electrical Engineering, Yale University, 1913–17; Engineering Department, Western Electric Company, 1917–24; Bell Telephone Laboratories, Inc., 1925—. Mr. Curtis' work has been connected with the development of telephone instruments.

Karl K. Darrow, S.B., University of Chicago, 1911; University of Paris, 1911–12; University of Berlin, 1912; Ph.D., in physics and mathematics, University of Chicago, 1917; Engineering Department Western Electric Company, 1917–24; Bell Telephone Laboratories, Inc., 1925—. Mr. Darrow has been engaged largely in preparing studies and analyses of published research in various fields of physics.

Lloyd Espenschied, Pratt Institute, 1909; United Wireless Telegraph Company as radio operator, summers, 1907–08; Telefunken Wireless Telegraph Company of America, assistant engineer, 1909–10; American Telephone and Telegraph Company, Engineering Department and Department of Development and Research, 1910—. Took part in long distance radio telephone experiments from Washington to Hawaii and Paris, 1915; since then his work has been connected with the development of radio and carrier systems.

Austin Bailey, A.B., University of Kansas, 1915; Ph.D., Cornell University, 1920; assistant and instructor in physics, Cornell, 1915–18; Signal Corps, U. S. A., 1918–19; fellow in physics, Cornell, 1919–20; Corning Glass Works, 1920–21; asst. prof. of physics, University of Kansas, 1921–22; Dept. of Development and Research, 1922—. Dr. Bailey's work while with the American Telephone and Telegraph Company has been largely along the line of methods for making radio transmission measurements.

C. N. Anderson, Ph.B., M.S., University of Wisconsin, 1919; Technical Asst. U. S. Naval forces in France, 1917–19; instructor Engineering Physics, University of Wisconsin, 1919–20; General Electric Co., 1920–21; Fellow to Norway, American-Scandinavian Foundation, 1921–22; American Telephone and Telegraph Co., Dept. of Development and Research, 1922—. Mr. Anderson's work has been chiefly on radio transmission.