Analysis of the Ionosphere *

By KARL K. DARROW

The ionosphere is a region in the very high atmosphere from which radio signals are reflected, a fact which is adequately explained by assuming that region to be populated with free electrons. In exploring the ionosphere, signals of a wide range of frequencies are successively sent upward, and the time elapsing before the return of the echo is measured. The delay of the echo mulitplied by $\frac{1}{2}c$ is called the virtual height of the ceiling for the signal. The data appear in the form of curves relating virtual height of ceiling to frequency of signal. These curves are peculiar in shape and vary remarkably with time of day, time of year and epoch of the solar cycle. By theory they can be translated into curves relating electron-density to true height above ground. The theory is approximative, but the results are accurate enough to be of value. The magnetic field of the earth affects the data remarkably, making it possible to test the theory and to evaluate the field-strength at great heights. The free electrons are supposed to be liberated from the air-molecules by ionizing agents, of which the chief but not the only one is ultra-violet light from the sun.

THE very title of this article embodies the assumption that in the upper reaches of the atmosphere there is a host of ions. By "upper reaches" here is meant, a region of the atmosphere so high that no man ever entered it, nor even a balloon with instruments. The ions therefore have never been observed by normal electrical means. They are postulated as the explanation of two things mainly: the echoing of radio signals from the sky, and that small portion of the earth's magnetic field which fluctuates with time.

The idea that these things require explanation, and the idea of the sort of postulate that is required to explain them, can both be followed back for many years. What was lacking in the early days was the notion of mobile electrified particles, that is to say, of "ions," in the air. That notion did not even exist, when in the eighties Balfour Stewart desired to imagine a conducting layer in the upper air for explaining magnetic fluctuations. It was only just being formed, when in 1902 Kennelly and Heaviside independently desired to imagine a conducting layer in the upper air for explaining why wireless signals can travel around the world. To speak of a conducting layer in the

* This paper, in abbreviated form, appears in the current issue of *Electrical Engineering*.

eighties, when air was regarded as an almost perfect insulator and nothing was known that could make it conductive, was certainly audacious. To speak of it in 1902 was still ingenious but no longer daring, for by then it could reasonably be expected that the researches on ions lately begun by Thomson and so many others would justify the notion.

Never was an expectation better founded. Within a few years those researches had made it sure that the upper atmosphere must be conductive, because of containing the raw material required for making ions and one at least among the agents capable of making them: to wit, atoms and molecules, and ultra-violet light from the sun capable of ionizing them. The problem then became: what distribution and what kinds of ions must be postulated for the upper atmosphere, to explain (for instance) the reflection of radio signals?

This problem could not even be attacked, without great forward strides in both the art of experimentation and the mathematical theory. These strides were rapidly made in the middle and late twenties. Had theory alone gone ahead, it would have been little more than a pretty exercise in mathematics. Had the art of experimentation progressed by itself, the experimenters would at least have found some interesting correlations of the data with such variables as time of day and epoch of the solar cycle and presence of magnetic storms; but the lack of theory would have been sorely felt. But theory and experiment advanced together, and the interplay between the two has seldom been so well exemplified.

The advance in the art of experiment lay not so much in the invention of new apparatus (though this has not been wanting) as in turning away from the practical problem of sending signals to great distances, and instead designing the experiments for the purpose, first of proving the ionosphere and then of "sounding" it. Three methods were invented for this purpose, all based upon the fact that wireless waves when sent into the sky come bouncing back from it. Two of these will be scarcely more than mentioned in the pages to follow, since an already great and ever-increasing proportion of the data is obtained by the third. In this third a sharply-delimited signal or pulse or wave-group is sent up, and a short time (a few milliseconds) later it is detected coming back, like an echo from a cliff: the delay of the echo is measured. This is done for many signals, and the delay is plotted against the mean frequency of the wireless waves composing the signals; and curves so plotted constitute the ultimate data. Usually the signal is sent vertically upward, the echo comes vertically downward; and there is the quaint situation, that wireless telegraphy is chiefly famed for bridging great distances over the earth, but its foundations are best studied with sender and receiver side by side.

Electromagnetic signals thus find a mirror or a ceiling overhead; and the theory interprets this mirror as consisting of the ions, and especially the free electrons, diffused in the upper air. This perhaps seems singular, in view of the tenuity of the air and the lightness of the individual electrons. It might have seemed better, at least in the days of the Greeks, to propose that the dome of the sky is a hard metallic mirror—of well-polished silver, for instance. Well, in effect that is what *is* proposed. A mirror of silver reflects not by virtue of its hardness, but because of electrons diffused like a gas through the pores of the metal. A metal is a container for an electron-gas, and in the upper air there is an electron-gas without a container; and both of them reflect.

The theory is strictly classical, in the sense of the word prevailing in physics. No relativity, no quantum theory, no suggested revision of the concepts of space and time, afflict the student thereof. It is the working-out of the basic principle of Maxwell and Lorentz, that the passage of electromagnetic waves through a medium is controlled by the electric current which the waves themselves evoke in the medium. Under the influence of the electric field in the waves, the ions swing in sympathetic vibration, and form a part of that current. They thus react upon the waves, alter the speed thereof, and bring about the reflection. The motion of the ions is simple-harmonic, so that the mathematics of the theory is simple and familiar—so long, at least, as no account is taken of any forces acting on the ions other than that due to the field of the waves themselves. Here is the explanation of the echoing of radio signals, and hence follows the procedure for translating the data of echoes into statements about the distribution of the ions in the atmosphere. It is not difficult to describe or explain, and will be carried through in this article.

From this point the theory ramifies in two directions. Two things modify the sympathetic vibrations of the ions: the collisions between the ions and the neutral molecules of the air, and the earth's magnetic field. By their influence on the vibrations, they modify the speed of the waves, and therefore the conditions of the echoing. The theory extended in either direction continues to be easy in one sense, for neither the physical concepts nor the mathematical operations are unfamiliar; but becomes very hard in another, for the algebraic expressions are often of fearful complexity, impossible to remember and hard even to keep straight when written out. When it is extended in both ways at once the expressions become so intricate, that nearly every investigator when taking account of either influence simply ignores the other. On the whole, the mathematical developments have far outrun the data. Yet there are important connections between experiment and theory, including for instance the proof that the ions which principally reflect the signals are free electrons.

After the theory come what I will call, for contrast, the speculations. The analysis of the ionosphere being made and accepted, a host of questions arise. Must we assume additional agents of ionization, other than the ultra-violet light of the sun? The answer to this question being certainly "yes," one must inquire how to distinguish that part of the ionization which is due to sunlight from the rest, and what are the causes of the rest. Why the distinctive distribution-in-height of the ions, amounting to what is called "the stratification of the ionosphere"? What assumptions must we make about the composition of the atmosphere in its dependence on height? or (as the question is more commonly put) what information can we derive about the composition of the atmosphere? How far can we go in interpreting the fluctuations of terrestrial magnetism, and (as later will be apparent) in mapping out the earth's magnetic field? The possible questions even rise to the realm of astronomy, and the suggested answers form a part of the theory of the sun as a potent source of radiations of all kinds, luminous and electrical and material. The implications of the ionosphere seem to be almost limitless, but a severe limit will nevertheless be set by space and time upon this article.

Methods of Experiment, and a Simplified Picture of the Ionosphere Adduced for Illustrating Them

The ionosphere is a canopy of ions overarching the earth, and in Fig. 1 it is represented by a model, very simplified indeed and yet instructive. Here it is shown as consisting of two "layers" marked E and F, with an ion-density which is uniform in each, and greater in F than in E. It is time to become familiar with the symbol N used for number of ions per unit volume: this picture shows N having the constant values N_E and N_F ($> N_E$) in E-layer and F-layer respectively, and the value zero between.

The lines which are broken at the layer-edges are paths of wireless signals or waves sent out from the source at S—sent out obliquely, for transmission over long distances. There is a path reflected from E, a path reflected from F and a path which penetrates both. These correspond to relatively low, medium, and high frequencies respectively: as examples I will give the values 1, 10 and 100 mc. (megacycles, i.e. millions of cycles per second) corresponding to wave-lengths of 300, 30 and 3 metres. Here already the reader meets the fact that the height at which such a signal is reflected, or the question whether it shall be reflected at all, depends on the frequency of the waves and the density of the ions. For every frequency there is what I shall call a "mirror-density": signals are reflected as soon as they reach the lowest level in the ionosphere where that mirror-density is attained. The higher the frequency, the higher the mirror-density. The formula

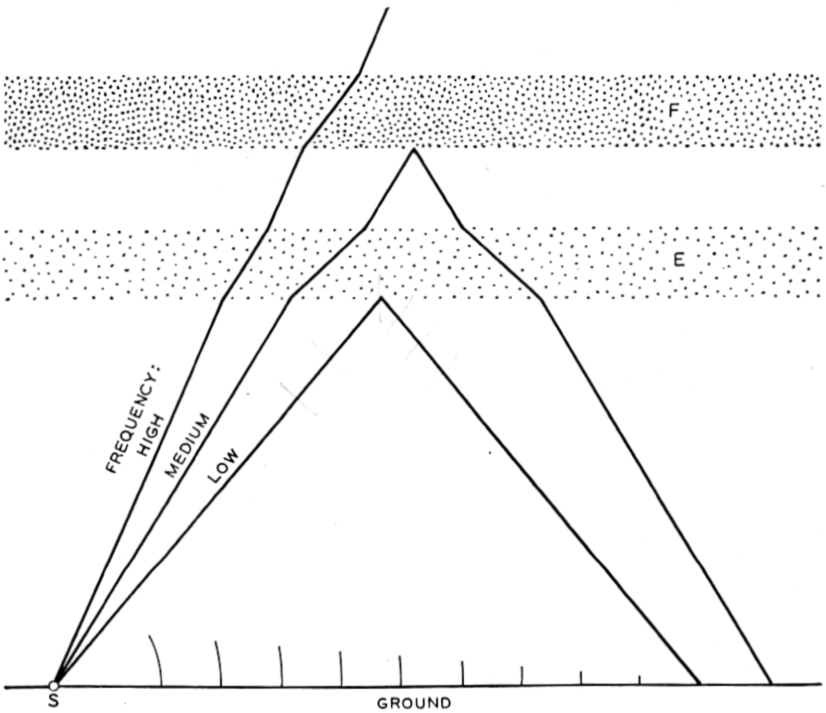

Fig. 1

will soon be derived and shown; but for the moment, let us inquire how the altitude of either layer can be measured, in terms of the simple model of Fig. 1.

One way of measuring the altitude is very obvious. Suppose the observer to go a known distance away from the aerial, and measure there the angle which the reflected wave or "sky wave" makes with the horizontal as it comes down to him. If this can be done, then clearly he can get the altitude by the simplest trigonometry—the altitude of the *E*-layer or the *F*-layer, according to the frequency which he uses. It *can* be done, and it was done by Appleton and Barnett in 1925. What they measured was the angle between the directions of propagation of the sky wave and the "ground wave," of which the wave-fronts are shown in Fig. 1 creeping along the ground with a rapid attenuation. This ground wave, by the way, is the only one by which radio transmission could be effected but for the ionosphere; and it is seldom detectable beyond a few hundred miles.

Another scheme is much more complicated, and owing to its supersession I may be excused for giving only the merest outline of it. It is a clever way of putting to useful service the very great inconvenience known as "fading." This term refers to a train of signals which dies out and revives and keeps on fluctuating over and over again, in a most irregular fashion. This sort of thing occurs in the region where the sky wave and the ground wave both arrive and overlap one another, and it has been traced to what in optics is called the "interference" of the two. If conditions were absolutely stable, then in taking a walk in the region of overlapping one would pass through several maxima and minima of intensity. Since conditions are never absolutely stable, the observer need not take the walk; while he stands at any fixed point, the maxima and the minima float past him while the ionosphere wavers in the sky and this is "fading." But imagine the conditions relatively stable for a time, and the observer standing still; and suppose that the engineer at the sending aerial changes the wave-length by a small and known amount—then, several maxima and minima will float past the observer, and by counting them he can (though this is not at all obvious!) get a datum which enables him to figure out the altitude of the reflecting layer in the sky. This method also was invented by Appleton, and can be found explained in the literature under the name "wave-length-change method."

The third of the methods has crowded out the others, and henceforth will figure alone in these pages. It is the "echo-method," still sometimes called by the clumsy name of "group-retardation method." Anticipated by Swann, it was realized by Breit and Tuve at the Carnegie Institution of Washington.

What is sent up to the sky is here a short sharp signal; if it could be heard, it would be called a click. What comes back is the echo of the signal. Passing over the receiving device, it produces a short sharp kick on an oscillograph-record. Some of these are shown in Fig. 2. The kicks marked E_1 and F_1 are due to signals echoed from the layers E and F respectively. Those marked $F_2 \cdots F_5$ are due to "multiple echoes"; the signal has traveled two to five times the entire journey from ground to ionosphere to ground again, the surface of the earth being itself a good reflector. Those marked *G* are due to the signal spreading along the ground itself. If the sender and the receiver are practically side by side, as usually is the case, the kicks *G* occur at the instants of departure of the signals. The record is moving laterally with the speed intimated by the wavy line beneath, and accordingly the distance along it from a *G*-kick to the following echo-kick is a measure of the "delay of the echo."

The delay of the echo is an indication of the altitude of the mirror where it was reflected—the layer *E* or *F*, as the case may be. Signals of relatively low frequency being reflected from *E* while those of medium frequency are echoed at *F*, one adjusts the frequency according

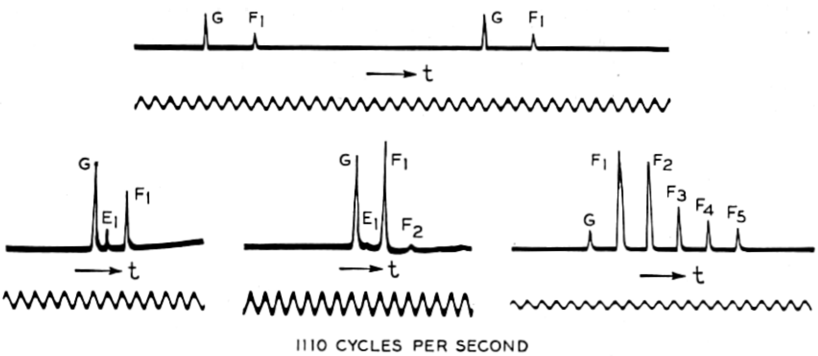

Fig. 2—Echoes. *G*, original signal; E_1, F_1, echoes returning after a single reflection from *E* and *F* respectively; $F_2 \cdots F_5$, echoes which have suffered two to five reflections at *F*-layer. (Appleton and Builder.)

to the layer which one wishes to locate. If there should be not two but several layers of the ionosphere, each having a greater *N*-value than the one beneath it, one would locate them all with appropriate frequencies. If there is a continuously-rising distribution of *N* with height in the ionosphere, one may plumb it by varying the frequency continuously. Now we are at the principle of the echo-method; but before it is used, there are many details to clear up.

First as to the "signals," a term which (it must have been noticed) replaced the term "waves" in the foregoing paragraphs. These signals are wave-trains indeed, but not the long continuous uniform trains tacitly assumed in the description of the other two methods. Those methods are adapted to trains of indefinite length, but not the echo-method, which for an obvious reason requires wave-trains of limited length—and the more limited, the better.

Now, a limited wave-train is equivalent to an infinity of infinitely long wave-trains, with an infinite variety of frequencies. The amplitudes and the frequencies of these constituent waves are chosen so that the waves reinforce one another over the length of the signal, counteract one another for all time before and after the signal. To choose them thus is always mathematically feasible, whatever shape of signal be prescribed. Whether these constituent waves should be regarded as "physically real" is a question that was discussed long before the days of quantum theory and other modern puzzles. Anyhow, by taking them as such, one arrives at verifiable statements about the signals, and this is all that matters.

Let us now conceive the signal as a chopped-off segment of a sine-wave-train of frequency f_0; and let us compare its travel with the travel of a limitless wave-train of frequency f put equal to f_0. One feels that the signal ought to follow the same path through the ionosphere as would the limitless train, and ought to move along that path with the same speed as would the wave-crests of the limitless train. This is true if, but only if, the speed u of the wave-crests in unlimited trains is independent of f. But when wireless waves are traveling through the ionosphere, u varies very much with f, according to a law which will later be worked out; and this makes a remarkable difference.

The difference as to path is not serious. The infinite wave-trains which form the signal are most intense at frequencies very close to f_0, and this is sufficient to make the signal follow nearly (though not without some deviation and distortion) the path which the infinite train of frequency f_0 would follow by itself. We may therefore regard the broken lines of Fig. 1 as the paths of signals or of waves, indifferently.

The difference as to speed *is* serious. It is not reduced by the preponderance of component wave-trains very close to f_0, and it does not tend to vanish as this preponderance is increased by lengthening the signal. It remains serious in the artificially-simplified case of just two component wave-trains of small frequency-difference Δf, where the signals become the "beats" well known in acoustics and in radio. In this case the beat-speed or signal-speed approaches a limit as Δf approaches zero. This limit, the "group-speed" denoted by v, is always used for the signal-speed, though for actual signals it is but an approximation, and the signals themselves become distorted as they travel. Contrasted with v is the "wave-speed" or the "phase-speed" of the wave-crests of the unlimited wave-trains, already denoted by u.

Since it is entirely the signal-speed which determines the delay of the echo, the wave-speed may seem a pointless side-issue. This quantity u is, however, essential in the theory from which are derived first the conditions of echoing and then the dependence of v upon f_0. Postponing the topics of signal-speed and echo-delay in order to return to them later with better preparation, we now take up the theory.

Theory of Wave-Speed, Total Reflection and Group-Speed in the Ionosphere

It will now be proved from Maxwell's theory, combined with the concept of mobile ions, that total reflection of wireless waves must occur in the ionosphere at the level where the ion-density attains a certain value depending on the wave-frequency.

The famous equations of Maxwell melt together into a wave-equation. The waves which it describes consist of an oscillating electric field which I will denote by $E_0 \sin nt$, and an oscillating magnetic field which we are permitted to ignore. When of high enough frequency these are the waves of light, as Maxwell knew; when of the frequency-range with which we are now concerned they are the waves of radio, as Maxwell was never to know because of his premature death. In the wave-equation there figures of course the wave-speed u. Here then is a paraphrase of the great idea of Maxwell: *the square of the wave-speed varies inversely as the current-density provoked by unit amplitude*[1] *of the oscillating field.*

Now we see at once that in the ionosphere the wave-speed must be affected by the presence of the free electrons, since they are set into oscillation by the waves and therefore make a contribution to the current-density.

At this point those who were educated in the electronic era (an ever-increasing fraction of the population) are in some danger of falling into a serious error. One may in fact assume that the electrons form the whole of the current, and deduce that in vacuo the oscillating field provokes no current at all, and the wave-speed must therefore be infinite—an absurd conclusion! Maxwell was wiser. He understood, and made it a part of his theory, that wherever there is an electric field which is changing in time the rate-of-change of that field is equivalent to a current. This he called the "displacement-current," and for the case of vacuum he said that the displacement-current-density is precisely equal to the rate-of-change of the field, multiplied by $1/4\pi$.

[1] I introduce the words "unit amplitude" to shield the reader from drawing the false inference that wave-speed depends upon wave-amplitude.

In vacuo, therefore, there is a current-density $(1/4\pi)$ times the rate-of-change of $E_0 \sin nt$, and it has an amplitude of $(nE_0/4\pi)$ and is 90° ahead of the field in phase. To this current-density corresponds the speed of light in vacuo, the well-known constant c. The speed of light in the non-ionized lower regions of the atmosphere differs so little from c that we need never bother with the difference, which henceforth will be ignored.

When the waves pass out of ordinary air into the ionosphere, there is still the displacement-current but now in addition there is the current borne by moving electrons. Here is a second pitfall. It may seem obvious that the electron-current must add on to the displacement-current, creating a total current-density greater than that in vacuo and therefore lowering the wave-speed. Not so at all! The point is, that when the electrons are truly free, the field sets them into oscillation in such a curious way that when they become adjusted, they are oscillating with their velocities 90° behind the field in phase. Their contribution to the current, being proportional to their velocity, is also 90° behind the field, and hence in perfect opposition of phase to the displacement-current.

Therefore the electron-current density—call it I_e—is to be *subtracted* from the displacement current-density! Accordingly I write,

$$\frac{u^2}{c^2} = \frac{n(1/4\pi)E_0}{n(1/4\pi)E_0 - I_e}. \tag{1}$$

The reader may suppose that the factor $\cos nt$, common to both currents, has been divided out.[2] The quantity I_e is clearly proportional to E_0 and also to our familiar N the density of electrons, and in fact the reader can undoubtedly work out with ease that it is equal to NE_0e^2/mn. Here e and m stand for the charge and mass of the ion, as is customary. Therefore we find:

$$\frac{u^2}{c^2} = \frac{1}{1 - 4\pi Ne^2/mn^2} = \frac{1}{1 - Ne^2/\pi mf^2}. \tag{2}$$

The wave-speed is greater in the ionosphere than it is in vacuo or ordinary air. I now recall from the most elementary optics the principle that when two media adjoin in which light has different wave-speeds, and light passes through their common boundary into the medium where its speed is greater, it is refracted away from the normal to the boundary. Accepting for the moment the over-simplified model of the

[2] Actually Maxwell's theorem does refer to the amplitudes of the currents—but if the currents are not exactly 0° or 180° apart in phase, the amplitude of one must be taken as a complex quantity.

ionosphere in Fig. 1, and considering the lower frequencies, we have the non-ionized lower atmosphere and the E-layer for these media. The paths of the waves are drawn accordingly. Now I further recall that total reflection occurs for all values of the angle of incidence i greater than that given by the equation:

$$\sin i = c^2/u^2. \tag{3}$$

Thus we see that for any frequency whatever, total reflection must occur when the waves impinge with sufficient obliqueness upon the ionosphere; but (so long as c/u does not sink to zero) total reflection will not occur if the waves rise vertically, or in a direction sufficiently near to the vertical.

The waves thus penetrate or are reflected back from the ionosphere, according as their angle of incidence thereon is less or greater than a certain critical value. Here is the explanation of what is called "skip-distance": the sky-wave is perceived beyond a certain distance from the source, but not within that certain distance.[3]

But all this seems to have nothing to do with the usual conditions of experiment, in which, as I intimated, the signals are sent up vertically! It is indeed a fact that in optics, no case is known in which total reflection occurs at vertical incidence. Yet equations (2) and (3) predict that if ever c^2/u^2 should vanish, total reflection would extend even to vertical incidence. Now there is nothing mathematically impossible or physically unplausible about the condition for the vanishment of c^2/u^2, which is simply that f should be equal to f_c given thus:

$$f_c^2 = Ne^2/\pi m \tag{4}$$

or alternatively that N should be equal to N_c given thus:

$$N_c = \pi m f^2/e^2. \tag{5}$$

Here we have the basic formula of the analysis of the ionosphere; for it is assumed that vertically-rising waves or signals of any frequency f climb until they reach the lowest level at which N is equal to N_c, and there they find their mirror or their ceiling, and are converted into echoes which return. Equation (5) is the formula for the "mirror-density" for signals of frequency c, to which I above referred.

It sounds all right to say that c^2/u^2 is zero when $N = N_c$, and negative when $N > N_c$; but it is disconcerting to notice that this

[3] Notice incidentally that owing to the curvature of the earth and its overhanging ionosphere, the angle of incidence can never rise to 90°; it follows that waves of frequency beyond a certain value (ordinarily around 30 mc.) never suffer total reflection.

amounts to saying that the phase-speed is infinite when $N = N_c$, imaginary when $N > N_c$. However, the concept of phase-speed is of such a quality of abstractness, that even these statements imply nothing absurd in the physical situation. The signal-speed itself remains safely finite and real.

The signal-speed is strictly indefinite, since the signal distorts itself as it proceeds. However, the practice is to identify it with the group-speed v, which, as I intimated (page 462), is the speed of the beats formed by two superposed wave-trains differing infinitesimally in wave-length, each such beat being a very special type of signal. The formula is,

$$v = u - \lambda(du/d\lambda) = u \Big/ \left(1 - \frac{n}{u}\frac{du}{dn}\right). \tag{6}$$

It is difficult to visualize or derive without a diagram,[4] but the derivation may be summarized as follows. Imagine two superposed wave-trains of phase-speeds u and $u + du$, wave-lengths λ and $\lambda + d\lambda$; consider two consecutive wave-crests A, A' of one and two consecutive wave-crests B, B' of the other; transpose temporarily to a frame of reference in which the former wave-train is stationary. At a certain place and time A and A' will coincide, and the maximum of one of the beats will be right there. Let the time $d\lambda/du$ elapse; when it has elapsed, the crests B and B' will be coinciding and the maximum of the beat will have moved on by one entire wave-length. The beat therefore travels with speed $\lambda du/d\lambda$ in the temporary and with speed $u - \lambda(du/d\lambda)$ in the original frame of reference (the minus sign is evident when the reasoning is gone through in detail).

Combining (6) with (2) one finds:

$$v = c^2/u; \tag{7}$$

the greater the phase-speed, the slower the signal! Relativists will be pleased to observe that according to this formula, the signal never attains any speed greater than c; students of quantum mechanics may be misled by its superficial resemblance to a formula relating phase-speed to group-speed for de Broglie waves, with which it has nothing to do. Students of the ionosphere should remember its approximative character. Almost all that needs to be known for the purposes of this article is, that as a signal climbs into the ionosphere it goes more and more slowly, the nearer N approaches to that value N_c where the signal finds its ceiling.

[4] Cf. this journal, 9, 173 (1930), or my *Introduction to Contemporary Physics*, 2nd edition, p. 147.

CHARACTERISTIC CURVES OF THE IONOSPHERE: THE (h', f) CURVES

Now that we have the concept of a signal ascending until it reaches the ceiling where $N = N_c$, we will consider first the to-be-expected relation between true height of ceiling and frequency of signal, then the relation between delay of echo and frequency of signal. By following this order we pass from the unobservable to the observed, which is the reverse of the customary way, but nevertheless has its advantages. Let z stand for height over ground when used as independent variable, with N depending on it; h for the height of the mirror or ceiling for signals, when expressed as a function of the signal-frequency f.

Take first the oversimplified model of the ionosphere appearing in Fig. 1: N having the values N_E over one range of heights and N_F ($> N_E$) over another range at a higher elevation, and the value zero elsewhere. It is evident that the (h, f) curve for such an ionosphere would consist of two horizontal lines or "branches," extending respectively from abscissa 0 to abscissa $f_E = \sqrt{N_E e^2/\pi m}$ and from abscissa f_E to abscissa $f_F = \sqrt{N_F e^2/\pi m}$ respectively. The latter would lie higher than the former; there would be a jump or gap between the branches. The names "E-branch" and "F-branch" for these last are obvious, and so is the usage "penetration-frequency of the E (or F) layer" for f_E or f_F; "critical frequency" is also used.

Next we approach closer to the truth by supposing that N rises continuously with increase of z across the E-layer and also across the F-layer, N_E and N_F now representing the highest N-values found in the respective layers. The two branches of the (h, f) curve would then be no longer horizontal, but slanting or probably curving upwards toward the right.

In the foregoing paragraph it was tacitly assumed that N still vanishes between the layers; but now let us approach still closer to the truth by postulating the sort of dependence of N on z shown in Fig. 3A. Here N drops with further increase of height after the "crown" of the E-layer is reached, but it does not fall to zero. It might, however, just as well fall to zero so far as reflections are concerned, for the signals which could be reflected from these regions never reach them. Regarding the curve of Fig. 3A as a sequence of hills and valleys, we see that the valleys contribute no echoes. The E-branch of the (h, f) curve refers to the left-hand side of the first hill; the F-branch refers to the left-hand side of the second hill, and not even to all of that, but only to the portion which rises above the first hill. Thus Fig. 3B, with its upturning branches and its gap, represents the (h, f) curve for the ionosphere of Fig. 3A, without in the least depending on the dashed parts of the $N(z)$ curve of Fig. 3*A* or indicating anything whatever about those parts except that they do not rise above the ordinate N_E.

Now if the signal and the echo traveled fro and to with the speed c, the delay T of the echo multiplied by $\frac{1}{2}c$ would be the height of the ceiling. This, however, is not the case, since the signal-speed depends on N. We must therefore denote the product $\frac{1}{2}cT$ by another symbol h', and make an inquiry into the probable dependence of h' on f, taking into account our vague knowledge as to the dependence of signal-speed on N.

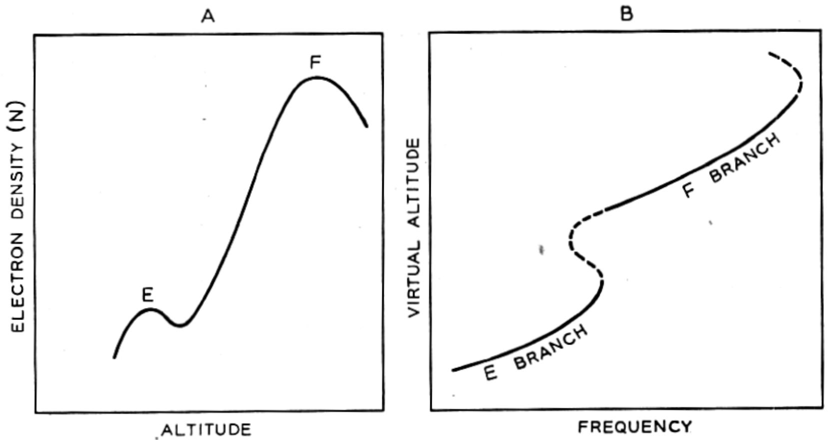

Fig. 3—*A*. The "curve of inference": conjectural dependence of number N of electrons per unit volume on true altitude h. *B*. The "curve of data": dependence of virtual altitude h' of ceiling (one-half the delay of the echo, multiplied by c) on frequency f.

It is easily seen that h' must be greater (or at least no less) than h, and that the excess of h' over h must be larger, the farther the signal travels through regions where N is almost but not quite equal to N_c. The (h', f) curve must therefore lie above the (h, f) curve, and farthest above it in the immediate neighborhood of the gap on both sides. There will still be an E-branch and an F-branch, but the upturns toward the right-hand ends of these branches will be exaggerated, and an upturn running to the left will be introduced into the left-hand end of the F-branch. It is conceivable that these upturns may become so large, that the (h', f) curve will appear to show a peak where the (h, f) curve would show a gap.

With the remark that h' is known as "virtual altitude," "virtual height," "equivalent height," or "effective height," I turn now to examples of the characteristic (h', f) curves of the ionosphere.

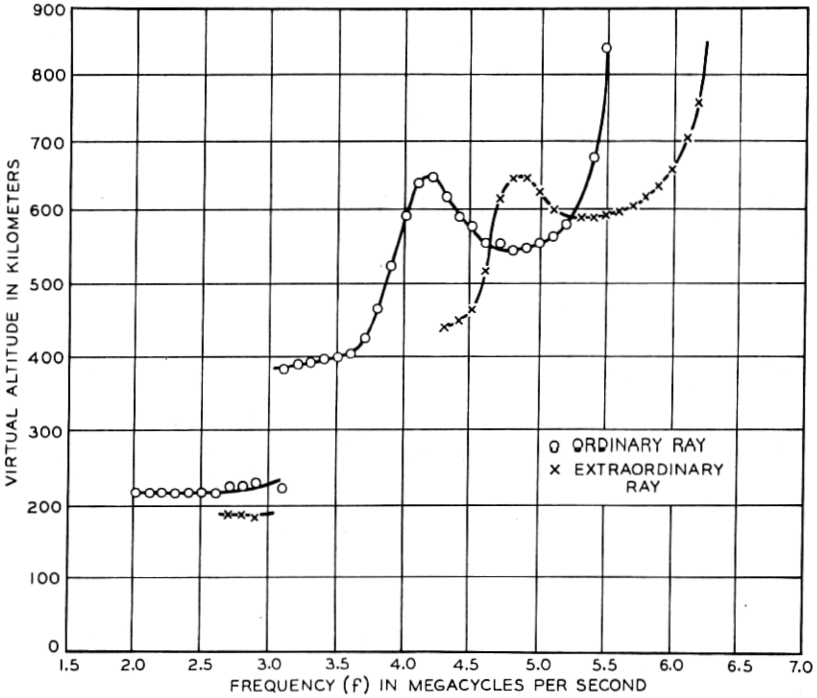

Fig. 4—Example of characteristic or (h', f) curves, showing gap between E-branch and F-branch, and crinkle in F-branch indicating presence of F_1-layer. Duplication of curve due to earth's magnetic field (page 479). (Appleton.)

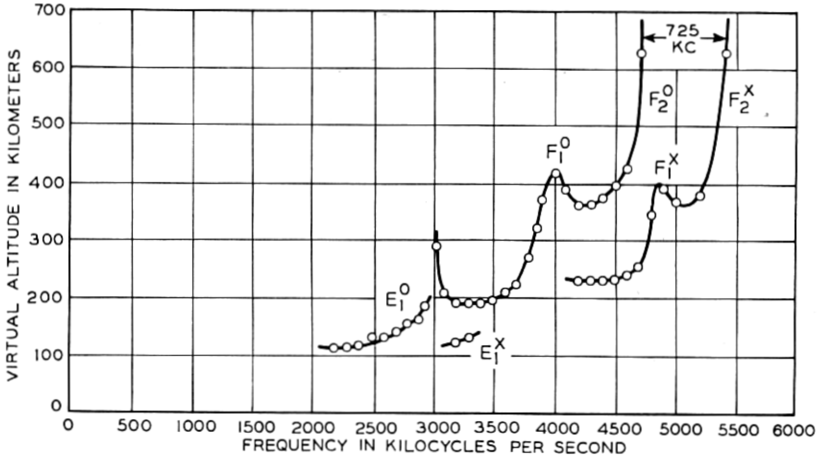

Fig. 5—Another example of (h', f) curve, showing gap and crinkle. (Schafer and Goodall.)

Figures 4 and 5 show two examples of these curves, from data obtained while the sun was high in the sky. Actually there are two curves in each of the figures; the appearance is that of a single curve, repeated with a sidewise shift. I mention that this repetition is due to the earth's magnetic field, but ask the reader to ignore for the present the right-hand curve and fix his attention on the left-hand one. Here he will see the *E*-branch, the gap, and the *F*-branch. The

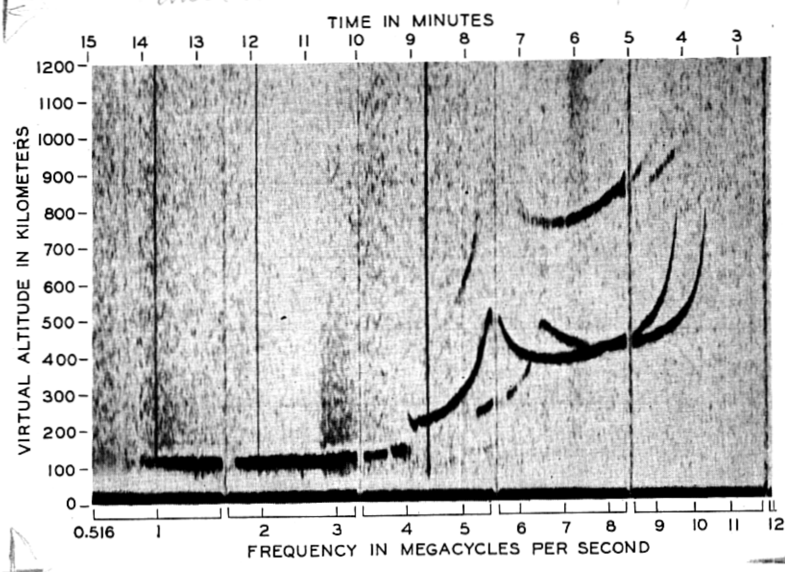

Fig. 6—Characteristic (h', f) curves obtained with the multi-frequency apparatus; sun high in sky, F_1 crinkle apparent. (Carnegie Institution of Washington.)

upturns to right and left of the gap are striking on Fig. 5, insignificant in Fig. 4. The *F*-branch is deformed by an enormous hump or crinkle. This is supposed to correspond to a second gap, the upturns on right and left being so pronounced as to give a perfect semblance of a peak; indeed one sees in Fig. 5 how readily the gap between *E* and *F* might have been drawn as a peak. Curves of this sort are therefore taken as evidence for three layers in the ionosphere, denoted by *E* and F_1 and F_2. Sometimes there are signs of a fourth, lying between *E* and F_2, and denoted by *M* or E_2.

So great is the interest in curves like these, and so much do they vary from time to time and from place to place, that lately there have been more than a score of stations over the world engaged in making them. At some of these the tracing of the curves is speeded up and made incessant by a remarkable machine developed at the

Bureau of Standards and the Carnegie Institution. Automatically sending out the signals ten times in a second, and changing the frequency by (on the average) 1600 cycles between each signal and the next while the photographic film is moved a tiny bit from left to right, this "multi-frequency apparatus" traces the (h', f) curve over the

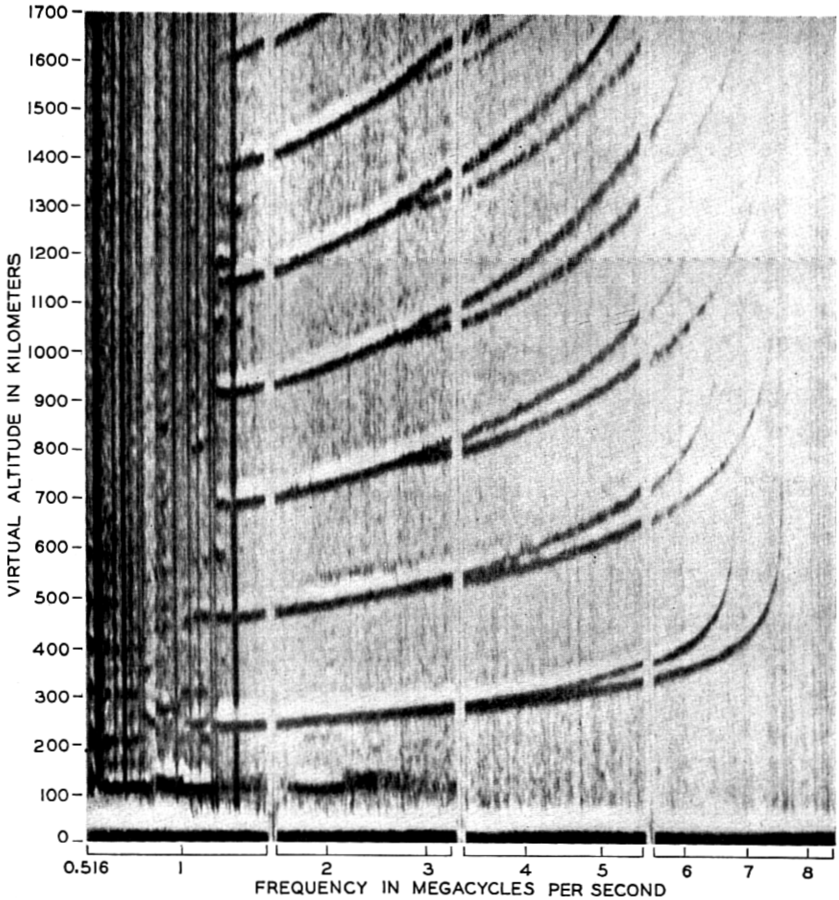

Fig. 7—Characteristic (h', f) curves obtained with the multi-frequency apparatus; sun low in sky, F_1 crinkle missing. (Carnegie Institution of Washington)

frequency-range between 0.516 mc. and 16 mc. in fifteen minutes, and then goes right back and does it over and over again. Figures 6, 7 and 8 show individual curves thus automatically taken, and Fig. 15 a sequence of them spanning several hours of the day.

In Fig. 6 are curves with crinkles in the *F*-branch, similar to those of Figs. 4 and 5. In Fig. 7, however, the crinkle is missing, and the

F-branch sweeps smoothly and slowly upward from its commencement. (The forking signifies that here are two similar curves lying side by side as in the previous figures, but overlapping so much that over a large part of their course they are not distinct.) Many observations have concurred in showing that the crinkle is present only when the

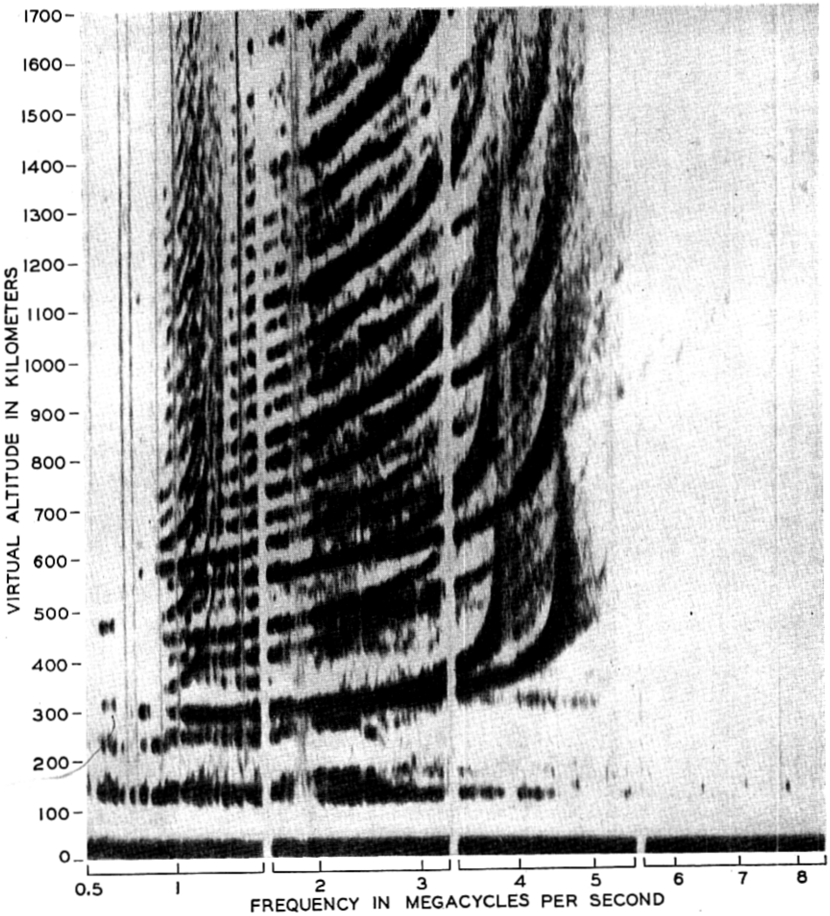

Fig. 8—Multiple echoes. (Carnegie Institution of Washington.)

sun is high in the sky (within some 40° of the zenith)—therefore absent by night and at the beginning and end of day, and indeed absent all day in winter where the latitude is high. This is our first example of the dependence of the ionosphere on sunlight, a very important feature.

In all of these photographs the curves are repeated several times along the vertical direction. This signifies echoes which have traveled four, six, eight or more times between ground and ionosphere, being reflected by both. Figure 8 shows a wonderful multitude of such echoes.

The curves of Fig. 9 are sketches generalized from many data. Contrasting those on the left with those on the right, we see the

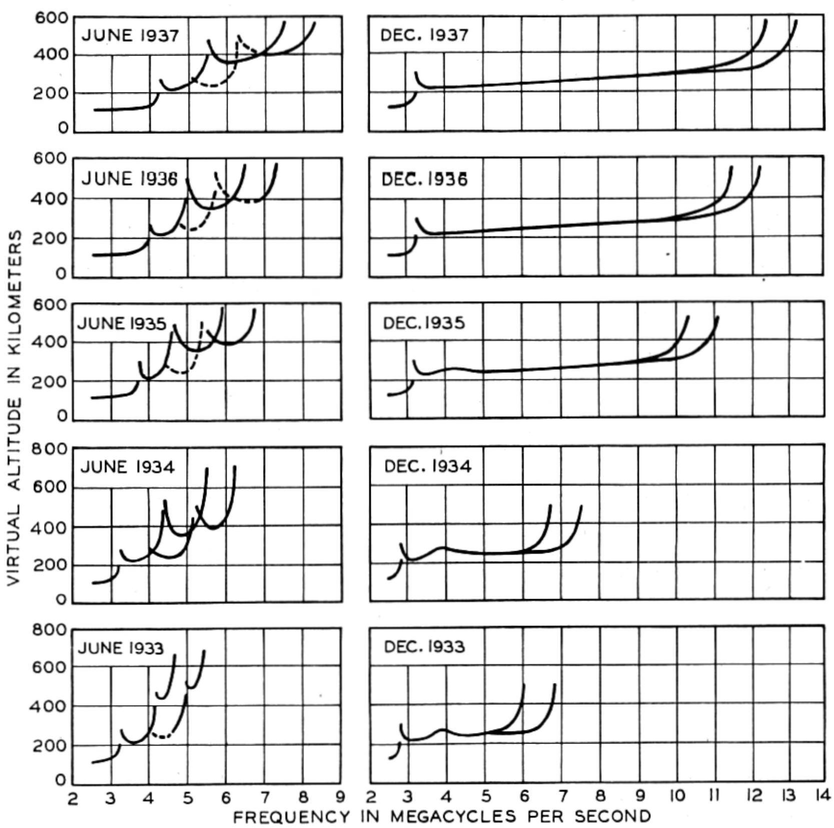

Fig. 9—Dependence of (h', f) curve on season and on the sunspot cycle (Smith, Gilliland and Kirby: National Bureau of Standards).

crinkle prominent in the ones, missing or feeble in the others. This is the difference between summer and winter. All of the data were taken near noon, but though the District of Columbia is not exactly in polar latitudes, the sun in December does not rise far enough in Washington's sky to bring out that feature of the ionosphere of which the crinkle is the sign. Running the eye along either column, one

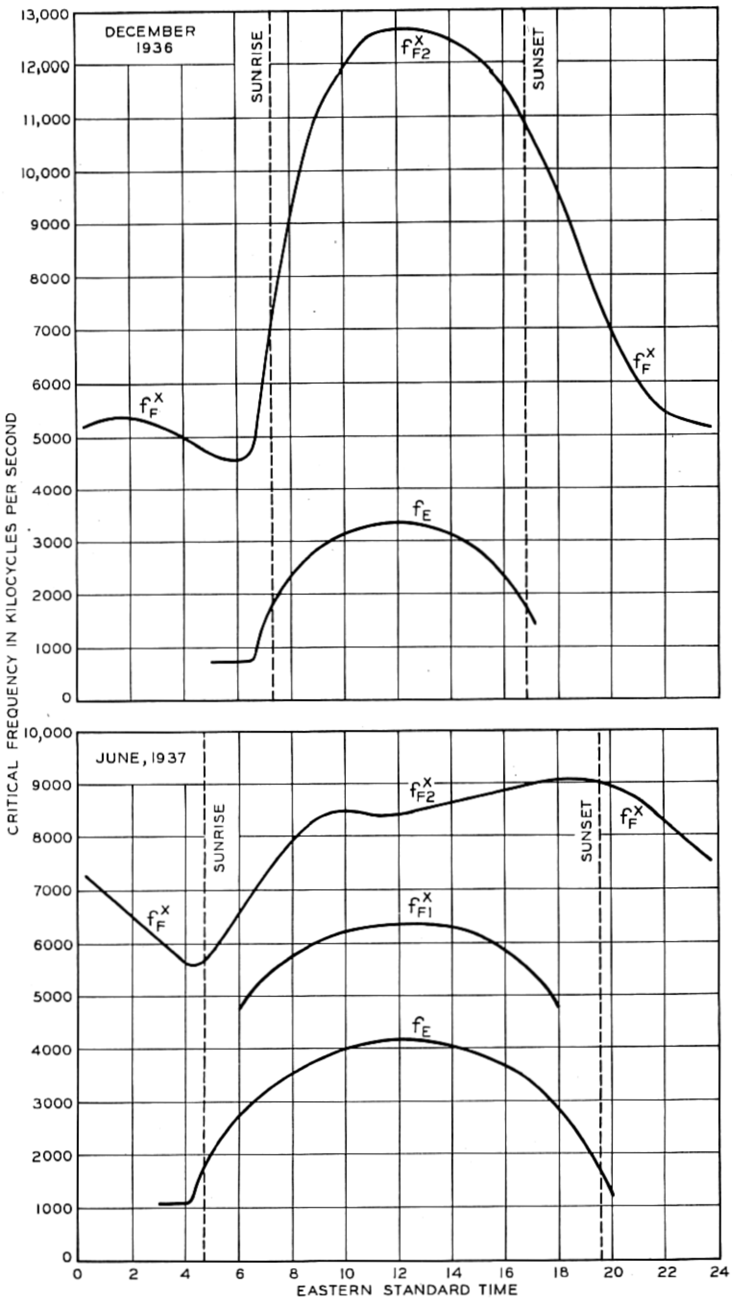

Fig. 10—Dependence of critical frequencies on season and hour of day. (Smith, Gilliland and Kirby.)

sees the curve lengthening out to the right as year follows year. During this series of years the sun spots were growing more frequent: the sun was in the ascending part of that eleven-year cycle of its fever, of which the sun spots are one of the manifestations, while the form of these curves is another. Do not, however, misread this statement as meaning that the curve lengthens out, when and only when there are sun spots on the solar disc! there is no such correlation.

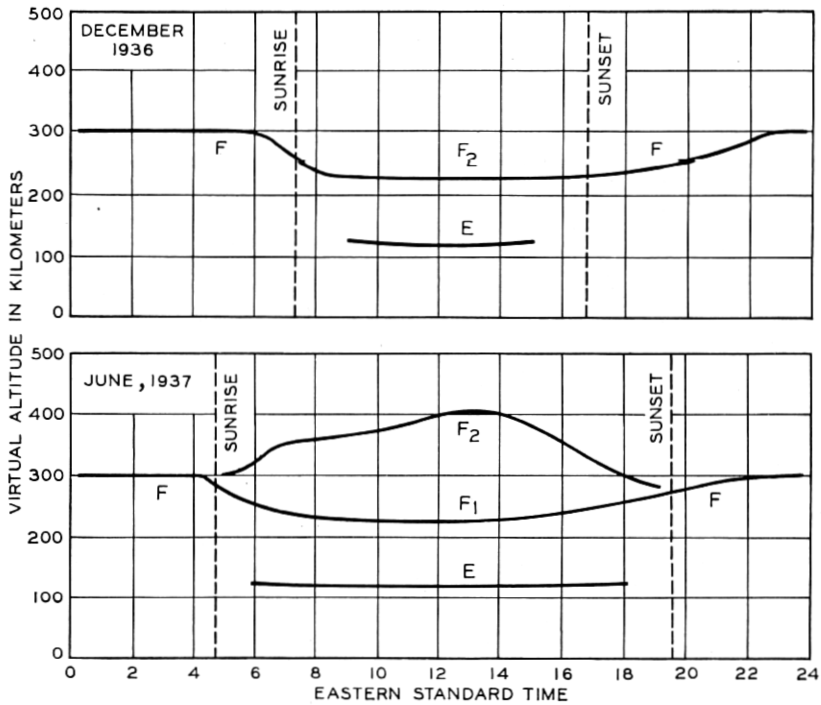

Fig. 11—Dependence of virtual altitude of the crowns of the principal layers on season and hour of day. (Smith, Gilliland and Kirby.)

Another way to study dependence on the sun is to pick out salient features of the (h', f) curves, and see how they vary. Such features are the abscissae and the ordinates of the right-hand ends of the branches and of the top of the F_1 crinkle. The abscissae are the penetration-frequencies for the layers E, F_1 and F_2; the ordinates are the virtual heights (not the true heights) of what I have called the crowns of these layers. All are shown, in their dependence on season and hour of day, in Figs. 10 and 11.

It may have struck the reader that for the last five paragraphs there has been no allusion to the theory, the correlation of the (h', f)

curves with the sun having been presented as if for its own sake. Much work in the field does stop at this point, and is not without value in spite of its stopping there. The theorist indeed may care for an (h', f) curve only as material for deducing the (N, z) curve—the distribution-in-height of the ions—to which he aspires. It is, however, fortunate that this is not the only value of the (h', f) curves, for as we now shall see, the derivation of the (N, z) curves from them is full of difficulties.

Perhaps the greatest of these difficulties springs from the dependence of the signal-speed on N, which is to blame for the difference between virtual height h' and true height h. Even if it is fully justifiable to identify signal-speed with group-speed v, the difficulty is not banished. It resides in the fact that $(h' - h)$ depends not on things already known but on the very thing one is striving to find out, to wit, the distribution of ion-density in the ionosphere. The value of h for any particular h' depends indeed not on the value of N at that height alone, but on the values of N at all inferior heights. The problem is somewhat like having to solve for x an equation in which x appears badly entangled on both sides of the equals sign. The mathematical technique is difficult and approximative.

Another major difficulty resides in the fact that when N varies by an appreciable fraction over a distance equal to a wave-length of the waves, the consequences of the theory become a good deal more complex than those embodied in the simple equations (4) and (5). For instance, partial reflection may occur at a level where N is rising rapidly, though as yet far below the mirror-density N_c. One cannot therefore say that whenever an echo is observed on a frequency f, there *must* somewhere exist an electron-density related to f by (5). The literature is full of allusions to mystifying echoes, some of which are ascribed to partial reflection. Figure 2 shows an E-echo and an F-echo received from the same signal; and many a (h', f) curve shows the E-branch running along for quite a distance underneath the F-branch, instead of stopping at just the abscissa where the F-branch begins.[5] Yet on the other hand, the E-layer may be denser at some places than it is at others of equal altitude, and parts of a signal may be reflected from the places of high density while other parts slip between these and go on to the F-layer. The assumption that N depends on z only, which hitherto has been taken for granted in this paper as it is in most theory, is in fact very assailable; and people are

[5] This is so well-known a phenomenon that lengthy papers have been written about it under the name of "abnormal E-ionization," though it seems too common to deserve the adjective "abnormal."

beginning to study the distribution of N in the horizontal plane, e.g. by using obliquely-sent as well as vertically-emitted signals.

Now we turn briefly to a difficulty affecting not the relation between h' and h, but the relation between N_c and f presented as equation (5). This equation was based on the tacit assumption that the electric force on a single electron is the same as though there were no other electrons at all in the ionosphere. The assumption has been doubted, and quite a polemic has ranged about it. The question is in fact a special case of one of the most pestiferous questions of all mathematical physics, occurring for instance in the theory of magnetized bodies and of bodies polarized electrically: when a great many similar atoms side by side are exposed together to an external field, how is the force suffered by any one of them modified by the presence of its equally-affected neighbors? One strongly-held position is, that there *is* such a modification which manifests itself in a factor 3/2, to be multiplied into the right-hand member of equation (5). A test experiment has been devised, and the early results have favored this theory. The presence or absence of this factor alters in equal proportion all the ordinates, but does not modify in the least the trend of the $N(z)$ curve; but the student specially interested in numerical values of N must discover, from each paper wherein such are given, which formula was used in computing them.

After uttering all these warnings about the theory underlying the (N, z) curves, I will risk a few statements about the curves themselves.

The shape of the (N, z) curve, when the sun is low in the sky and there is no crinkle in the $(h'f)$ curve, is roughly that of Fig. 3*A*. If the sun is within some 40° of the zenith and the crinkle is present in the $(h'f)$ curve, the theory indicates not that the F-peak of Fig. 3*A* has split into two, but rather that a bulge has appeared on the left-hand side of the F-peak. The letters F_1 and F_2 are then applied to the bulge and the peak, respectively. If the shape of the (h', f) curve indicates yet another layer between E and F_1, it appears as a small hump in the valley between the peaks of Fig. 3*A*.

As for the N-values, those of most interest are those corresponding to the crests of the peaks; or to define them better by staying closer to the data, they are the ones corresponding to the points on $(h'f)$ curves which adjoin the gaps or lie at the tops of the crinkles. These may be called the values corresponding to the "crowns" of the several layers.

At Huancayo in the Peruvian Andes, at a typical summer noon, N has the values $1.8 \cdot 10^5 - 3.3 \cdot 10^5 - 1 \cdot 10^6$ at the crowns of the layers E, F_1, F_2: so says Berkner. At Slough near London, at noon on a certain day of early spring (1933) Appleton found $1.2 \cdot 10^5$ at the crown of E and $3.8 \cdot 10^5$ at the crown of F, the F-layer being at that time and place not differentiated into F_1 and F_2. These figures are not far apart, if E be compared with E and F_1 with F; but with a little search I could have found plenty of values differing much more greatly, as is attested by Figs. 9 and 10. From the former of these we have already deduced that critical frequencies vary as the sunspot cycle proceeds: I now add that from minimum to maximum of the cycle just ending, N at the crown of the E-layer increased by three-fifths while N at the crown of F_2 went up no less than fourfold! From Fig. 10 we infer, by squaring the values of critical frequencies, how great is the change of these N-values with hour of day. Sudden unaccountable changes also occur; one evening over Cambridge (Massachusetts) the N-value for the E-layer was more than tenfold the values given above, being ascertained by Mimno as $2.8 \cdot 10^6$!

I therefore summarize, as precisely as seems justifiable: the N-values at the crowns of the layers vary with hour of day and time of year and year of the sunspot-cycle very markedly, not to speak of sudden unexplained fluctuations; and 10^5 to 10^6 electrons per cc. is a good figure to keep in mind for the order of magnitude thereof.

To terminate this section I show Fig. 12, in which the delay of the echo for a certain frequency (2 mc.) is plotted against time during the hours preceding and following dawn. Interpreting with the aid of Fig. 3*A*: during the night the E-peak was too low to echo back the signals of this frequency, which accordingly climbed farther and found their mirror in F; but at 6:35 A.M. very sharply, the E-peak increased in height to just the extent needed to intercept them. Or since confusion may arise from using the word "height" in two senses, I express what went on in an exacter way: during the night the electron-density at the crown of the E-layer was inferior to the mirror-density for $f = 2$ mc., but with the oncoming of day it rose, and at 6:35 A.M. very sharply it attained and overpassed that mirror-density.

I recall that Fig. 11 exhibits how the virtual altitudes of the layer-crowns vary with hour of day and season of the year in the sky over Washington. It is evident that E is a fixture of the ionosphere with a virtual height surprisingly steady at the close neighborhood of 120 km, while F rises and falls in the course of a winter day, rises and falls and divides and merges again during a day of summer.

Now we must take brief notice of a difficult subject: the forkings and the doublings of the (P', f) curves, and the theory which finds their source in the earth's magnetic field.

A magnetic field, the earth's or any other, should have no effect whatsoever on radio waves so long as these are traveling in air composed entirely of neutral molecules. When, however, the waves are setting electrons into motion, the moving electrons are affected by the field, which has a twisting action on their paths. We have seen already that the moving electrons react, so to speak, upon the waves, raising the wave-speed thereof. By altering the motions of the electrons, the magnetic field will influence at second hand the waves themselves. But will the result be perceptible? In view of the fact

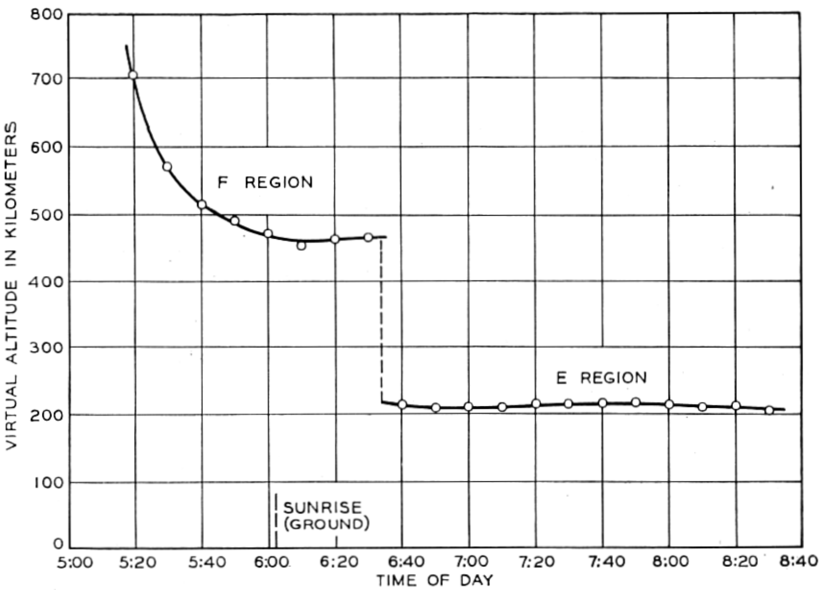

Fig. 12—Ceiling for signal of fixed frequency shifting near sunrise from F to E as the electron-density of the E-layer increases with increase of light. (Appleton.)

that the earth's magnetic field is very feeble by comparison with the fields between the poles of our electromagnets great or small, or even with those around the horseshoe magnets which are playthings, one might well think the influence not worth the trouble of computing. But those who first undertook to compute it—Nichols and Schelleng in America, Appleton independently in England, in the winter of 1924–25—found it a serious influence, and very well worth the trouble.

The problem is one of those which are not very hard to state, but can be very tedious to solve except in special cases which may or may not be of practical importance. For this problem it happens that two of the special cases can be solved with relative ease, and one at least is realizable in practice. I cannot venture to give the theory of even this one, but at least I will attempt to describe what happens.

The special case occurs when the waves are traveling at right angles to the field. Since they travel vertically,[6] it is necessary to find a place on the earth where the field is horizontal. Such places are found in the equatorial regions only, and these are not precisely crowded with universities or engineering experiment stations. However, the Carnegie Institution of Washington was inspired, several years ago, to set up a station in just such a place: Huancayo, in the Andes of Peru. Here they established long straight horizontal antennae, one running north-and-south, another east-and-west, and yet another northeast-and-southwest. In the waves which mount from these to the ionosphere and then come bouncing back, the electric field $E_0 \sin nt$ —henceforth to be called "the electric vector"—is faithful to the direction of the antenna. They are called "plane-polarized waves."

When the north-south antenna is used, the electrons are impelled to and fro in the north-south direction which is that of the magnetic meridian. Now as is well known, an electron moving parallel to a magnetic field behaves just as it would if there were no such field at all. These waves ought therefore to behave according to the theory which we set up while we were still disregarding the magnetic field. They are the so-called "ordinary waves" or "*o*-waves."

When the east-west antenna is used, the electrons of the ionosphere are impelled to and fro in the east-west direction, which is transverse to the earth's magnetic field. This is just the condition for the maximum amount of meddling by the field in the motion of the electrons. The meddling consists in bending the electron-paths into curiously twisted arcs. *The action of the magnetic field is tantamount to strengthening the electron-current-density I_e parallel to the electric vector.* It will be recalled (from page 464) that it is I_e which for small N-values cancels a part of the displacement-current and so speeds up the waves, and for a certain critical N-value cancels the whole of the displacement-current and so brings about total reflection. So, for these "extraordinary waves" or "*x*-waves," a given N-value produces a greater augmentation of the wave-speed, and the critical N-value for total reflection is smaller, than for the ordinary waves. The signal composed of *x*-waves, mounting into the ionosphere, finds its appropriate mirror at a lesser altitude than does the signal composed of *o*-waves, and it gets earlier back to earth. It may indeed come

[6] In addition to the other advantages of sending the waves up vertically, there is the feature that the angle between their line of motion and the field is the same when they are going up and when they are returning.

back from E while the o-signal goes on to F or from F_1 while the o-signal goes on to F_2, or from F_2 while the other goes irretrievably forth into space.

When the northwest-southwest antenna is used, the ionosphere takes charge of the signals, and separates them into an o-component and an x-component. Each travels according to its proper law, and the x-component reaches its lower-down mirror earlier and beats the o-component back to earth. Two echoes return instead of one. The earlier is plane-polarized with electric vector east-and-west; the laggard is plane-polarized with electric vector north-and-south. Suppose a long straight horizontal antenna is used to respond to the returning signals. It will respond to both, if pointed north-west-southeast; only to the earlier, if pointed east-and-west; only to the later, if pointed north-and-south.

All the foregoing were statements of theory at first, but thanks to the experiments of Wells and Berkner at Huancayo, they now are statements of data as well.[7] Figure 13 exhibits a small selection from the data.

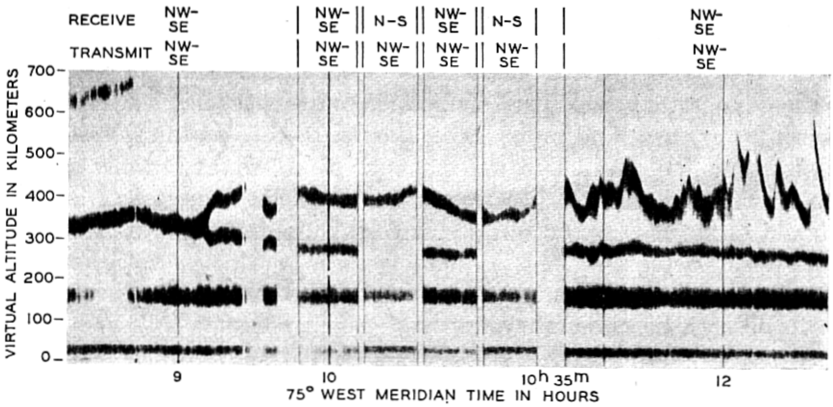

Fig. 13—Echoes of plane-polarized signals near the geomagnetic equator. (Wells and Berkner.)

Now look again at Fig. 4: formerly I asked the reader to ignore one of the curves, but now we will compare the two. The circles and the crosses indicate the o-wave and the x-wave respectively. One is

[7] This is a good place to speak of a question which may already have occurred to many readers, viz. the question why we assume the charged particles in the ionosphere to be free electrons rather than charged atoms or molecules. Were they of atomic or molecular mass, the separation of the o and x echoes would be inappreciable, and the "gyro-frequency" later to be mentioned (page 482) would be quite outside of the radio range. It is not, however, excluded that among the free electrons there may be a great multitude of charged atoms, perhaps even many times more numerous than they, though much less influential.

a fairly close copy of the other, shifted a certain distance along the horizontal axis. From the magnitude of the shift it is possible to compute the strength H of the earth's magnetic field; or let me rather say, it is possible to compute a numerical value which must agree with H, or else the theory will be vitiated. When the computation is made, the value turns out to be just a few per cent less then the field strength at ground-level. The action of the field through the electrons on the waves is exercised only in the ionosphere, which is hundreds of kilometers up in the sky; and it is quite reasonable to believe that these few per cent are actually the falling-off in the field strength from the ground up to that level. Such is the present belief, and many of those who work in terrestrial magnetism are happy over the prospect of measuring thus the field in regions where there seems to be no greater hope of anyone ever actually going, than of going to the moon.

Actually Fig. 4 shows data obtained in England, which is far from the equator; I point this out in order to mention that even when the waves are traveling obliquely to the earth's magnetic field, there is a separation of the signals into pairs of echoes, and these are still amenable to theory. In this general case of oblique transmission, the waves are polarized elliptically—a feature difficult to visualize without a certain amount of specialized knowledge, but lending itself to some very neat and pretty experimental tests.[8] In the special case of transmission parallel to the magnetic field, waves initially plane-polarized should remain of this character but their plane of polarization should rotate as they proceed. There are indeed so many curious and interesting details of the influence of the field through the electrons on the waves, that a writer must be ruthless in ignoring them if he is to observe decent limits of space. I will mention only in closing that the "gyro-frequency" $eH/2\pi mc$—which in our latitudes is around $1.3 \cdot 10^6$ mc.—plays the part of a resonance-frequency. Waves too close to this frequency are liable to great, not to say distressing, anomalies in transmission. Theoretical statements about waves in general are likely to assume two different forms, one appropriate to those of frequency higher and the other to those of frequency lower than the gyro-frequency; it is the former which appears in this paper.

In Fig. 14 there appears something which, if the war had begun before its discovery, would perhaps have been called a "blackout."

[8] It may perhaps be regarded as obvious that the ellipse of polarization should be described in opposite senses in the northern and southern hemispheres, since in one hemisphere the magnetic lines of force are coming out of the ground, in the other they are diving in. This inference was tested by a special experiment in Australia, and the result was taken as establishing the "magneto-ionic theory," as this general theory is often called.

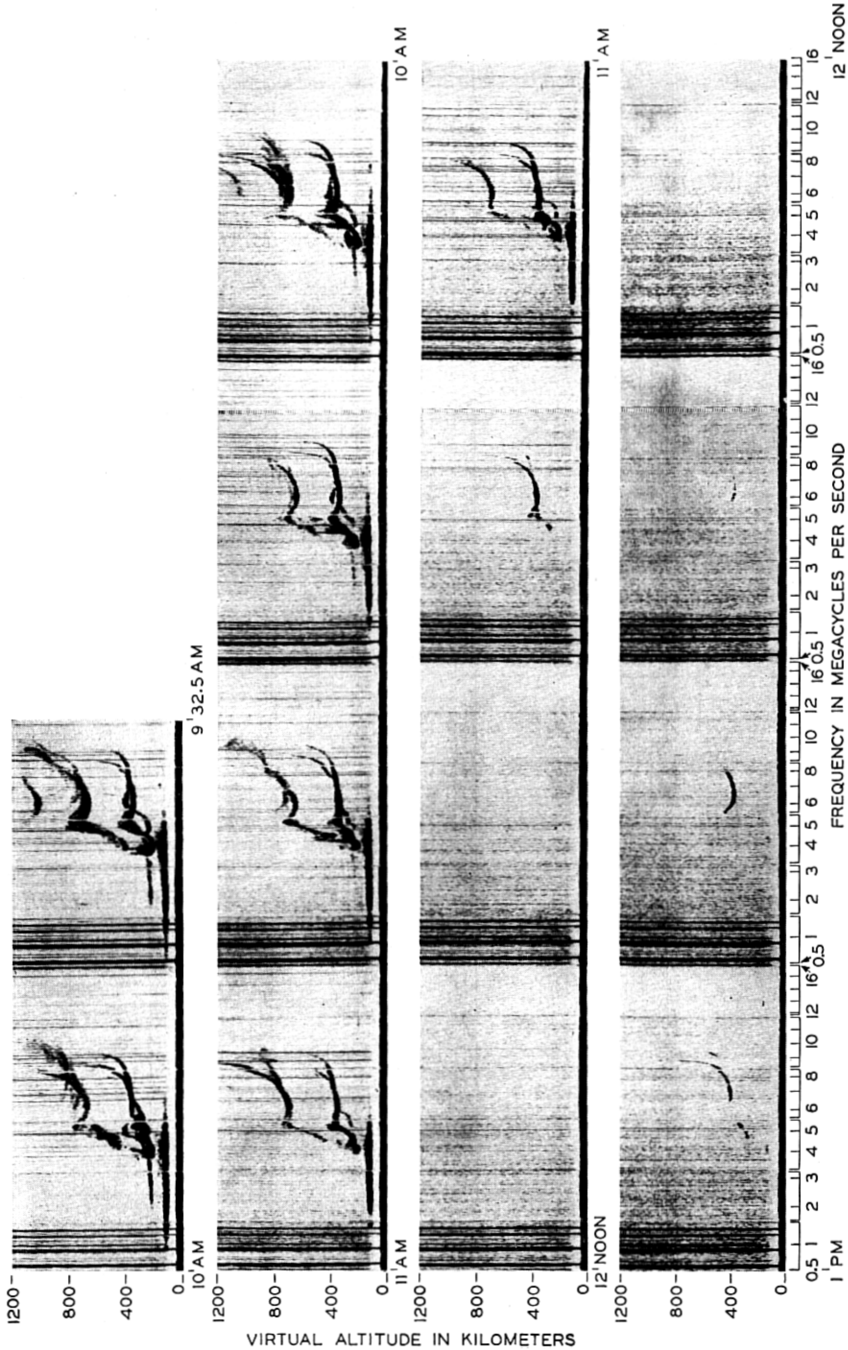

Fig. 14—Progression and recession of a fadeout. (Carnegie Institution of Washington.)

Being discovered however some years before the war, it was and is called a "fadeout." The apparatus was that which I mentioned on page 471; accordingly each of the pictures was traced in fifteen minutes, and as soon as each was finished the next was begun. See how the pattern of (P', f) curves, familiar and sharp in the earlier pictures, dissolves into fragments and then is completely wiped out! Later on it begins to come back piecemeal, and finally is restored as good as ever.

Since attention was focussed on such events in 1935, they have been reported by the scores in every year, varying in duration and in severity. It requires no (h', f) curve to show them, since ordinarily they cut off communication by radio, and with the sharpness of a knife. Many an engineer, to quote from Dellinger, has "dissected his receiving equipment in the vain effort to determine why it suddenly went dead." Over broad areas the extinction is sudden and simultaneous in many fadeouts, more gradual in others; the restoration is as a rule more gradual.

Shall we interpret this strange and striking effect as a sudden vanishing of the ionosphere and all the reflecting layers thereof, or as a swallowing-up of the signals by something which is suddenly created underneath the ionosphere? Against the first suggestion it is to be said, that no one can image anything which might so suddenly frighten all the electrons of the ionosphere back under cover, so to say—drive them into the arms of their parent molecules in a few seconds or minutes—when all day and even at night they manage to hold their freedom. Such a graph as Fig. 15 speaks also against it forcibly. Here in the upper part of the figure we see the critical frequencies [9] of F_2, F_1 and E as located every fifteen minutes on (P', f) curves such as those of Fig. 14. Each flock of data lies along a curve which, intercepted though it is by the fadeout, resumes so nearly at the level where it left off that one can hardly believe that the ionosphere totally vanished in between.[10]

As for the curve marked "f_{MIN}" in Fig. 15, it represents the lowest frequency at which echoes are observed. I have said nothing as yet about there being such a minimum-frequency. How indeed can there be one, and why should signals of any frequency however low fail to be echoed, since the mirror-density for any higher frequency is *a fortiori* more than a mirror-density for any lower?

[9] For the ordinary waves, as indicated by the superscript in symbols such as f^o_E.

[10] In violent magnetic storms the ionosphere is so convulsed that the echoes lose their sharpness entirely, and (h', f) curves like those of Fig. 7 are replaced by broad smudges; or echoes may vanish altogether. These are quite different from fadeouts.

Attempting to supply an answer to this question, I point out that in such part of the theory as I have thus far given, there is nothing corresponding to *absorption*. This is because the electron-current

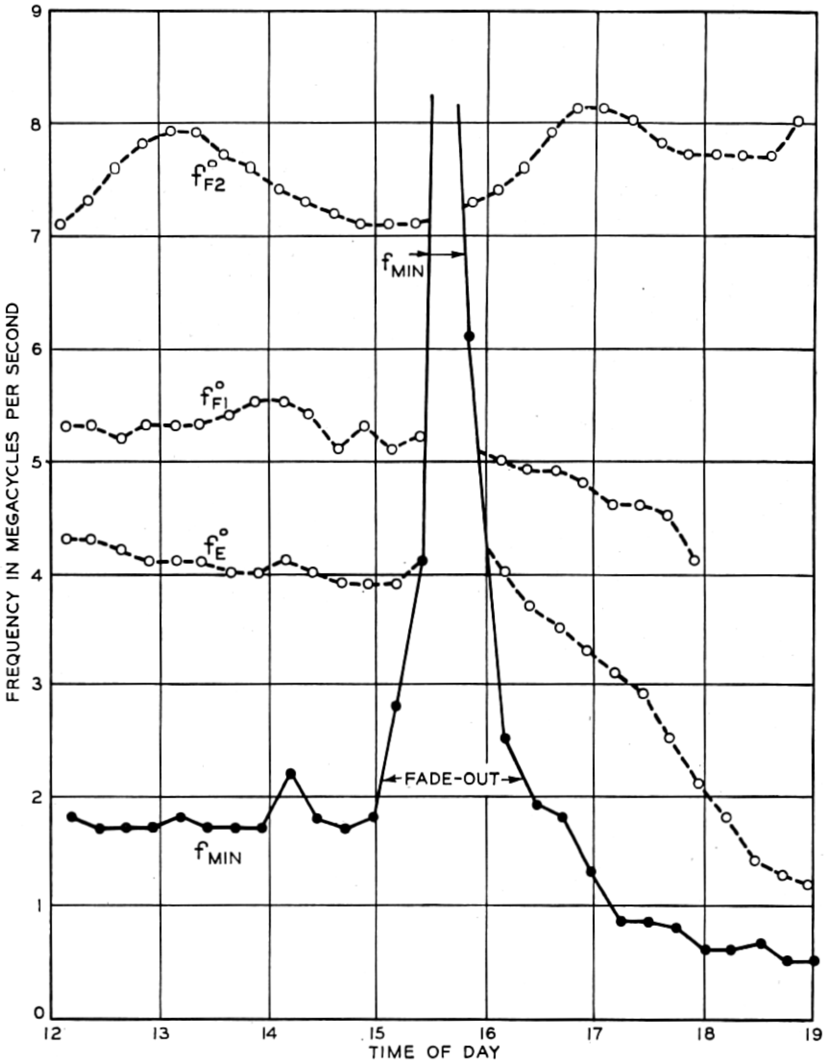

Fig. 15—Trend of the critical frequencies and of the minimum echoed frequency as a fadeout proceeds. (Berkner.)

(I_e in the former notation) is in quadrature with the electric vector in the waves, being thus a "wattless current." This in turn is because we have assumed each electron to be entirely free, oscillating in the wave-swept aether as though there were nothing else in the world but itself and the waves. If, however, the electron were occasionally to strike and bounce off a molecule it would leave some (even though but a small part) of its kinetic energy behind, and this would have to be replenished by the waves. Indeed as soon as collisions are taken into account, the algebra tells us that the electron-current is no longer in perfect quadrature with the electric vector. Dissipation of energy occurs together with reflection, and if it is sufficiently great —if, that is to say, the electrons collide often enough with molecules—it takes the place of reflection. In the terms of my acoustic simile on an earlier page, the signals are no longer echoed back from the hard slopes of the mountain-range of Fig. 3*A*, but are swallowed up and lost as if in something soft and woolly.

One would expect absorption to occur in the lower reaches of the ionosphere rather than in the upper, since the air is denser there and the electrons suffer many more collisions. The absorbing layer, that is to say, must be situated just where it is able to cut us off from the reflecting layers by dissipating the signals which we send. Why should it do so occasionally with such completeness, and most of the time not do so at all, for any except the lowest frequencies?

For an answer to this query, one looks again to the sun. Ordinarily, we will suppose, the ionizing agent coming from the sun penetrates deep enough into the air to form the reflecting layers high overhead, but is nearly consumed in so doing. Occasionally, though, the sun sends forth a quite abnormal transitory burst of radiation, so strangely constituted that it passes the reflecting layers without contributing to them or weakening itself, and continues so far down that at the level where it at last engenders free electrons they constitute a layer absorptive and not reflective. This would be no more than an *ad hoc* assumption, were it not that brilliant eruptions are frequently seen on the face of the sun at the moments when fadeouts are commencing. To some extent it is still an *ad hoc* assumption, for the light whereby the eruption is seen is certainly not ionizing light, and we must assume that the visible light is attended by rays of other wave-lengths having just the properties desired. Coincidence of fadeout and eruption is, however, so frequent, that it would now take a very sceptical mind to reject the assumption. The trend of the curve "f_{MIN}" in Fig. 15 sustains it.

And so I have now come back to the theory that it is radiation from the sun which makes the upper atmosphere into an ionosphere, by detaching electrons incessantly from the molecules thereof. (The detaching must be incessant, for the electrons are always liable to recapture by the molecules.) Of this theory it may be said that the major facts confirm it, though at night the E-ionization persists so tenaciously that we are obliged to seek for a separate agent (meteors, perhaps?); while numerous minor discrepancies can be explained away by making special assumptions which can neither be confirmed nor refuted because there is so little independent knowledge of the upper atmosphere. Not a very satisfactory situation for the present, but at any rate one which offers endless promise!

Thus it can readily be seen that as the ionizing light descends from heaven through the upper air, the ionization per unit volume should at first increase (because the air is getting denser) and then decrease (because the light is getting to be used up). This offers an explanation for a layer; and the mathematical working-out of the idea—due in the main to Chapman—shows that not only the existence but the shape of either peak in the curve of Fig. 3A is compatible with the theory. But there are several layers and peaks, not just one; how does this come about? Well, the atmosphere is a mixture of several gases, differently susceptible to the ionizing light; one can attribute a peak to each gas (indeed more than one to a gas, by invoking different states of the molecules). The height of a peak, the N-value at the crown of a layer, should rise and fall as the sun rises and sinks in the sky. This is true of the layers E and F_1, as we saw from Fig. 11, and again there is a quantitative theory by Chapman, which is borne out in some though not in full detail. Of F_2 it is not always true, as Fig. 11 proclaims; there is a minimum at noon in summer, and the highest N-values of all are attained in winter! One tries to cope with the discrepancy by assuming that as the sun climbs higher in the sky, the F_2 region expands so much in the heat that although the total number in the region is properly increasing, the number in unit volume suffers a decline. The layers do not disappear at night, though the N-values shrink. There seems to be plenty of time for recapture of all the electrons between sunset and sunrise, and one is driven to hunt for other causes of ionization which emerge when the sunlight is gone. These remain mysterious. Inrush of meteors into the high atmosphere has been suggested as one of the causes, and also incessant streams of charged particles similar to those which become intense during magnetic storms.

Sunlight is therefore not the only, yet apparently the major factor in maintaining the ionosphere. Not, however, any sunlight that we ever feel! This portion of the sun's outpourings is so thoroughly consumed above that it never reaches down to the levels where we live. Were it not so consumed, we should not be able to communicate by radio very far over the earth. The reader may think that this is not very important: our ancestors lived without radio, why should we worry about lacking it? Well, it is probably quite true that if the ionosphere were not overhead, we should not be worrying about the lack of radio. We should in fact probably not be worrying about anything at all, for we should not be here to worry. The ultra-violet light of the sun, pouring down upon the surface of the globe unhindered, would work changes so severe on organisms as we know them that life would have to be very different, and perhaps impossible. This lethal light is like an enemy, which in attacking a city spends itself in throwing up a barrier against itself; and the barrier not only keeps the enemy out, but is serviceable otherwise to the dwellers in the city.