SHORT-WAVE RADIO AND THE IONOSPHERE

OTHER BOOKS OF RADIO INTEREST

WIRELESS DIRECTION FINDING
R. Keen, M.B.E., B.Eng.

TECHNICAL INSTRUCTION FOR MARINE WIRELESS OPERATORS
H. M. Dowsett, M.I.E.E., F.Inst.P., and L. E. Q. Walker, A.R.C.S.

FOUNDATIONS OF WIRELESS
M. G. Scroggie, B.Sc., M.I.E.E.

BASIC MATHEMATICS FOR RADIO STUDENTS
F. M. Colebrook, B.Sc., D.I.C. A.C.G.I.

RADIO LABORATORY HANDBOOK
M. G. Scroggie, B.Sc., M.I.E.E.

GUIDE TO BROADCASTING STATIONS
Compiled by *Wireless World*

GREAT CIRCLE PROJECTION MAP OF THE WORLD

ETC., ETC.

SHORT-WAVE RADIO AND THE IONOSPHERE

With special reference to everyday professional and amateur problems of short-wave transmission and reception

By

T. W. BENNINGTON
ENGINEERING DIVISION, BRITISH
BROADCASTING CORPORATION

With 61 illustrations

Second Edition

ASSOCIATED
ILIFFE
TECHNICAL BOOKS

Published for
" WIRELESS WORLD "
by
ILIFFE & SONS, LTD.
LONDON, BIRMINGHAM, COVENTRY, MANCHESTER AND GLASGOW

First published 1944
(*as "Radio Waves and The Ionosphere"*)
Second Edition 1950
(*with amended title*)

Published for "Wireless World" by Iliffe & Sons, Ltd.,
Dorset House, Stamford Street, London, S.E.1

Made and printed in Great Britain
by W. & J. Mackay & Co., Ltd., Chatham, Kent.

(*Bks.* 525)

Contents

Preface

THE ROLE OF THE IONOSPHERE in long-distance short-wave communication is one of paramount importance ; indeed, without it such communication would be impossible. All, therefore, who carry on radio communication over long distances, whether as part of their professional duties or merely as a hobby, are continual " users " of the ionosphere, and, as such, would wish to know something about it, and about its effects upon their radio waves.

The aim of this book is to present such information in an essentially simple form, so that it will be of use even to those with only a limited technical knowledge. In this way it is hoped to meet the needs of all those " users " of the ionosphere, whether of amateur or professional standing, who have not the time to make a special detailed study of ionospheric propagation, and of the part it plays in the maintenance of short-wave communication.

The use of mathematics has been avoided, and the physical processes involved explained, so far as this is possible, in simple descriptive language. Whilst such explanations must, in some cases, fall somewhat short of the complete truth, this is felt to be justified by the above considerations. The practical side of the subject has always been kept in mind, and it is shown how the scientific ionospheric data may be applied to everyday problems of short-wave transmission and reception.

The subject is introduced in such a way as to make it comprehensible to the beginner : the formation and structure of the ionosphere are discussed, and its effects upon a radio wave briefly explained. The technique of ionospheric measurement is dealt with, and the nature of the continual variations that occur within the medium are discussed. The methods for applying the ionospheric information to short-wave transmission and reception are dealt with at some length, and some of the phenomena which particularly affect amateur radio transmission are specially mentioned. Finally the cause and nature of ionospheric disturbances and of certain other phenomena are gone into.

Throughout the book some liberty is taken in the use of certain terms, as, for example, " refraction " and " reflection," " rays " and " waves," but it is hoped that this will not result in any confusion.

In writing the book the author has sometimes made use of information which has been published before, and he wishes to acknowledge the debt he owes to those scientific workers whose published papers constitute the standard literature on ionospheric matters. He also wishes to thank the British Broadcasting Corporation for permission to publish information which has been made use of in the development of the Corporation's overseas services.

T. W. BENNINGTON

January, 1950.

ABBREVIATIONS USED

°	degree
cm	centimetre
db	decibel
EHF	extremely high frequency
G.M.T	Greenwich Mean Time
HF	high frequency
kc/s	kilocycles per second
km	kilometre
LF	low frequency
LUHF	lowest useful high frequency
Mc/s	megacycles per second
MF	medium frequency
MUF	maximum usable frequency
OWF	optimum working frequency
SHF	super high frequency
UHF	ultra high frequency
VHF	very high frequency
VLF	very low frequency

CHAPTER 1

Fundamentals of Long-distance Radio Communication

INTRODUCTION

RADIO COMMUNICATION OVER very long distances—whether by broadcasting, telegraphy, telephony, or by any other means—is nowadays carried out almost exclusively on the short waves, that is to say on waves from about 100 metres to 10 metres in wavelength. It is not possible to carry out such long-distance communication on the medium waves (those between about 100 metres and 1,000 metres in wavelength) nor upon any of the wavelength ranges shorter than 10 metres in wavelength, at least upon a regular and dependable basis. It *is* possible to do so upon the very long and upon some of the long waves (those over about 3,000 metres in wavelength), but the cost of the plant and of its operation upon these wavelengths is so high compared with that of installations with similar transmission ranges operating upon the short waves, that present-day long-distance communication is confined very largely to these latter wavelengths.

The short waves are therefore of immense importance in modern communication systems, as is also the medium by which they are enabled to travel to such great distances. This is the body of ionised air existing in the upper atmosphere—the *ionosphere*, as it is called. Without this medium, long-distance radio communication round the spherically-shaped earth would be impossible : it forms an essential part of every short-wave transmission path. Thus the short waves, together with their conducting medium, the ionosphere, constitute a most important adjunct to modern everyday life, and, besides this, form a very fascinating subject for amateur interest and experimentation.

It is the aim of this book to explain the essential facts about short-wave transmission and the ionosphere in language as simple as it is possible to use. Though the whole book should be intelligible to anyone, whether with previous knowledge of radio or not, this first chapter is intended for the veriest beginner, and may well be skipped by those with an elementary radio knowledge.

RADIO WAVE RANGE VARIES WITH WAVELENGTH

It has already been said that the only radio waves suitable for really long-distance communication are those in the short-wave range (apart from certain long waves whose use is more costly), and from this it will be gathered that the short waves have some characteristic which is not shared by other radio waves. This is true, though the beginner should clearly understand that it is not because of any fundamental difference in the *nature* of the short and of other radio waves. Radio waves are all electromagnetic disturbances in space, and are of the same essential character whether their wavelength is measured in thousands of metres or in fractions of a metre. But their actual behaviour during transmission varies very greatly according to their wavelength, because of the fact that phenomena which are of little importance when the wavelength is long become of great significance when it is reduced ; and, conversely, phenomena which have a big effect upon the long waves are of little consequence on those of shorter wavelength. This difference in behaviour greatly affects the distance range to which the waves can travel, and it is this which makes the short waves so suitable for long-distance transmission.

It should not be imagined, however, that there is any sharp dividing line between waves which behave or travel in one way and those which travel in another. As a matter of fact the boundaries between the different classes of waves are not only largely arbitrary, but also vary very considerably from time to time. Nevertheless it is expedient to classify the waves according to the characteristics which they usually and primarily exhibit, and to divide them into wavebands, each of which possesses, in general, these particular characteristics.

CLASSIFICATION OF RADIO WAVES

If there is one thing upon which radio engineers seem to find it difficult to make up their minds it is upon this question of the classification and nomenclature of the waves of differing wavelength. There is at practically any time a dispute going on in one or other of the world's technical journals about the validity of the latest classification, and usually the reasons given by the disputants are not easy to contradict. We cannot do better here, however, than to give what is, in its main essentials, the classification laid down by the Atlantic City Telecommunications Conference of 1947, which is shown in Table I.

As will later be explained, we can speak of radio waves in terms either of their " frequency " or of their " wavelength."

NATURE OF RADIO WAVES

It is not intended, in a book like this, to go into the highly complicated details about the production of a radio wave, or about its exact nature when produced. But a few fundamental points about it will be mentioned as being helpful later on to an understanding of the special behaviour of the short waves, in which we are mainly interested.

TABLE 1

Class	*Wavelength range*	*Frequency subdivision*	*Frequency range*	*Metric subdivision*
Very long waves	Above 10,000 metres	VLF (Very low frequency)	Below 30 kc/s	Myriametric waves
Long waves	10,000–1,000 metres	LF (Low frequency)	30–300 kc/s	Kilometric waves
Medium waves	1,000–100 metres	MF (Medium frequency)	300–3,000 kc/s	Hectometric waves
Short waves	100–10 metres	HF (High frequency)	3,000–30,000 kc/s	Decametric waves
Very short waves	10–1 metres	VHF (Very high frequency)	30,000 kc/s–300 Mc/s	Metric waves
Ultra short waves	1–0·1 metres	UHF (Ultra high frequency)	300–3,000 Mc/s	Decimetric waves
Super short waves	0·1–0·01 metres	SHF (Super high frequency)	3,000–30,000 Mc/s	Centimetric waves
Extremely short waves	0·01–0·001 metres	EHF (Extremely high frequency)	30,000–300,000 Mc/s	Millimetric waves

Imagine a radio transmitting aerial being fed with high-frequency electrical energy from a transmitter, in which case there will be an electric charge oscillating back and forth along the wire. This charge sets up " lines " of electric strain in the space surrounding the aerial, and, since it would do so equally well if the aerial were in a vacuum, it is evident that the air surrounding it has nothing to do with the matter. The electric strain lines do not, as a matter of fact, require any material medium to support them : they can exist in " free space "—that is, in space which contains no material medium whatsoever.

It is a principle in electricity that when an electric strain acts across a conductor it sets up an electric current along it. Similarly, a *changing* electric strain—like the one we are considering—sets up what is known as a " displacement " current, even in an insulator, like the air surrounding our aerial. An electric current, whether of the " displacement " or " conduction " type, will, in turn, set up a *magnetic* strain, acting at right angles to the electric strain which was its original cause. And a *changing* magnetic strain will, again, set up another electric strain, and so on. Thus, in the space surrounding our aerial a disturbance consisting of alternate electric and magnetic strains is set up,

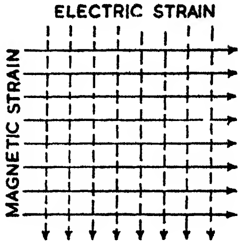


Fig. 1—Imagine this to be the wave front of a wave which is coming towards you. The directions of the " strains " are shown for a certain instant of time : a half-cycle later the directions would *both* be reversed

and, since each successive strain may be regarded as occurring farther and farther from the aerial, the electrical energy actually *leaves* the aerial, and is " radiated " into space.

Such a disturbance in space—consisting of co-existent electric and magnetic strains—constitutes an electromagnetic wave, and this travels through space at its own natural velocity of 300,000,000 metres (186,000 miles) per second. Its velocity through ordinary air is substantially the same as through " free space."

The outer edge of such an advancing disturbance is called the wave front, and if we could examine things there we should find that the direction of the electric strain, the direction of the magnetic strain and the direction of advance of the wave were all at right angles to each other. Thus if, looking straight in front of you, you imagine a wave advancing past you from left to right, the direction of its electric strains might be up and down from ceiling to floor and that of its magnetic strains horizontally towards and away from you. Fig. 1 may also help to explain the situation.

It is important to remember, however, that a radio wave travels not only outwards from the transmitting aerial in all horizontal directions (that is, over the surface of the ground), but also, provided it is not suppressed in any direction, in all upward directions from the horizontal as well.

VELOCITY, FREQUENCY AND WAVELENGTH

In order to make sure that we are clear on all fundamental points it may be as well to explain the relation between the velocity at which the wave travels, its frequency, and its wavelength.

As the electric charge continues to oscillate in the aerial, electromagnetic waves continue to be radiated from it. The rate at which they are emitted will depend on the rate or " frequency " at which the electric oscillation is set up by the radio transmitter. For each complete oscillation of the charge in the aerial one complete wave is emitted.

The velocity at which the wave travels in the air surrounding the aerial always being the same, the wavelength and frequency are connected by the relation—

$$\text{Wavelength (in metres)} = \frac{300{,}000{,}000}{\text{frequency (in cycles per second)}}.$$

A visual picture will help us to understand this. Suppose that the radiated waves are visible, and that we are able to watch them coming from the aerial and to follow them as they travel outward from it. We know the velocity at which the waves will travel, namely 300 million metres per second. Suppose the electric charge to be oscillating up and down the aerial 300 million times per second, in which case we say that it has a frequency of 300 million cycles per second. For each complete oscillation (or cycle) of the charge one complete wave is radiated. At the end of a second the wave front will be 300 million metres away, whilst the last wave will just be leaving the aerial, and the distance between the wave front and the aerial will be occupied by the 300 million waves which have been emitted. The distance occupied by each wave will therefore be 1 metre—that is, the wavelength is 1 metre.

Now let us decrease the frequency of the oscillating charge and cause it to oscillate a hundred times more slowly than before, namely at a frequency of 3 million cycles per second. At the end of a second there will now be only 3 million waves occupying the 300 million metres between wave front and aerial. The wavelength will therefore be—

$$\frac{300{,}000{,}000}{3{,}000{,}000} = 100 \text{ metres.}$$

We thus see that a wave of low frequency has a long wavelength, whilst a high-frequency wave has a short wavelength. We can refer to a wave either in terms of its frequency or of its wavelength, though frequency is perhaps the better designation.

In order to avoid getting confused by the large figures involved when dealing with frequency we can express this quantity in more convenient terms than that of cycles, namely—

1 kilocycle per second (kc/s) = 1,000 cycles per second
1 megacycle per second (Mc/s) = 1,000,000 cycles per second.

THE GROUND WAVE

It has already been mentioned that radio waves can be emitted from the aerial so as to travel outwards in all directions, both horizontal and vertical. We should, perhaps, require a rather special form of aerial in order to radiate equal amounts of energy in all of these directions, but let us imagine, for the moment, that that is being done.

If we could stop the radiation an instant after it had started and hold the waves still in space whilst we examined the situation, we should

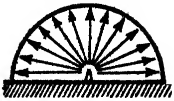


Fig. 2—Direction of travel of some of the "rays" of radio energy

get the picture of a huge hemisphere of radiated energy surrounding the aerial and supported on the ground, with the aerial at the centre of the circle described on the ground by the bottom of the hemisphere of energy.

It is rather difficult to show this in a diagram drawn on a plane surface, but it will suffice for our purpose to show a section cut through the hemisphere, as in Fig. 2. The arrows indicate the direction of travel of some of the radiated waves or " rays " of radio energy—some, we see, are travelling outwards over the surface of the ground, whilst others are going up towards the sky.

For the moment we will consider only those waves which are travelling outwards in directions parallel to the surface of the ground. This part of the radiated disturbance is called the " ground wave " since it remains in contact with the ground throughout its journey. As the wave travels along it sets up electric currents in the earth itself, and these cause a weakening or " attenuation " of the wave, for energy is taken from it in order to maintain them. This loss of energy is said to be due to " ground absorption," and its amount will vary according to the nature of the terrain over which the wave is travelling. It is less over sea than over any other type of terrain, and thus the " distance range " of the ground wave over sea is relatively great.

Part of the energy lost in the ground is replaced by that in the part of the wave immediately overhead, so there is a certain flow of energy downwards towards the ground. But the energy at the foot of the wave is only partly replaced in this way, and, since it goes on inducing currents in the earth and losing more and more of its energy the farther it goes, it eventually becomes so greatly attenuated that—to all intents and purposes—it dies away altogether.

In considering the propagation of the ground wave we must remember that a radio wave *normally* travels in a perfectly straight line, and, since the earth's surface is curved, it cannot, if it is to be receivable at any great distance, continue to do this, but must " bend " around so as to follow the earth's curvature. The ground wave, when it meets an intervening object—such as the " bulge " in the earth's surface—in its path, does, to some extent, " bend " around it by the process known as " diffraction." Also the wave, as it goes along, acquires a forward " tilt," so that some of the energy overhead tends to flow downwards towards the ground, and this assists the wave to bend around the curved earth. But these processes absorb energy from the wave and so weaken it that, for this reason alone, it is necessary to employ some other principle than that of ground-wave transmission in order to effect really long-distance communication.

Variation in Ground Losses with Wavelength

But there is another point which we should now consider—an important one in helping us to understand the reasons for the different behaviour of waves of different wavelength or frequency.

The losses to which the ground wave is subject, because of the earth currents at its foot, besides varying with the nature of the soil or water over which it is travelling, vary also with the wavelength. The shorter the wavelength the thinner is the layer of earth which it affects, and so the greater is the resistance of this. Thus the greater is the amount of energy lost in the earth, and the sooner does the wave become completely attenuated and die away. In other words, the longer the wavelength (the lower the frequency) the less are the earth losses, and the longer does the wave persist, so that with a given amount of energy radiated the greater is the *range* of the transmitting station.

On the longer wavelengths the ground wave can travel to quite considerable distances, though it should be stated that the great distance ranges obtained on the long wavelengths are not obtained by means of the ground wave alone. However, the main point is that as we reduce the wavelength—by increasing the frequency at the transmitter—the ground losses increase, and the range of the station gets less and less. Thus in Fig. 3, at a wavelength of about 1,000 metres, the ground absorption is relatively slight and so the range is relatively great. But

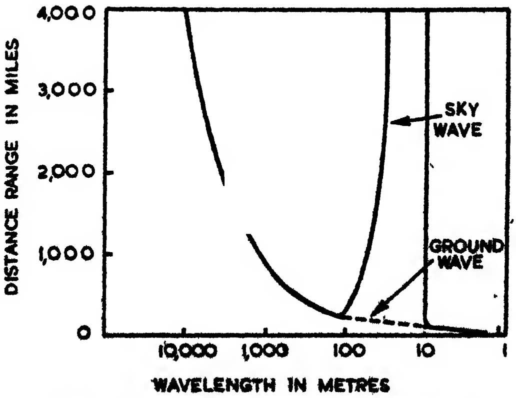


Fig. 3—Variation in range with wavelength

throughout the medium-wave range (1,000 to 100 metres) the losses are increasing rapidly, and the ground-wave range is so severely restricted that these waves are only of use for short-distance communication. From Fig. 3 it will be seen that this reduction in range continues—so far as the ground wave is concerned—right down to the shortest wavelengths. Below about 100 metres, therefore, the ground wave is of little consequence, and is not usually relied upon to provide communication.

THE SKY WAVE

But the *actual* range, as shown by the full-line curve of Fig. 3, is seen to increase very rapidly at about 100 metres, so that on waves between about 80 metres and about 10 metres the greatest ranges are obtainable.

Furthermore, these ranges are obtainable at a fraction of the cost in material and power of similar ranges on the long waves. This is because of the large dimensions and high operating costs of long-wave installations. For, considering only the aerial arrangements and bearing in mind that to be a good radiator its dimensions must approach in magnitude the order of a wavelength, it is easily seen how difficult and costly this is to achieve where the wavelength is long, and how simple and cheap in the case of the short waves.

It will be noted that this greatly enhanced range can occur over the whole range of waves which we call the " short " waves (from roughly 100 to 10 metres) and, as this is the subject with which the rest of this book will mainly deal, we had better examine the reason for its occurrence straight away.

In considering Fig. 2 we agreed to deal for the time being only with those waves which were going out at small angles to the earth's surface. We will now consider that other portion of the radiated energy which is contained in the top portion of our hemisphere of radiated energy, and which represents the upward-going waves, shown by the sloping arrows in Fig. 2. These waves, on leaving the aerial, commence to travel up towards the sky, and hence this portion of the radiated energy is known as the " sky " wave. In short-wave work we rely entirely on this sky wave to provide the energy which will actuate the receiver at the far-distant point.

But observing the direction of the arrows and remembering that the waves will normally advance in straight lines, it is at once seen that after they have been travelling for a second or so they will be many thousands of miles away from the earth and will never be in a position to operate a radio receiver located upon it. That would indeed be so if they were travelling all the time in normal air—such as exists at the earth's surface. This acts as an electric insulator. But fortunately for us—in more respects than one—the air surrounding the earth is not all in this "normal" state. For high in the atmosphere, and surrounding the earth like a shell, is a region where the air has been turned into an electrical "conductor," and air in such a state acts upon radio waves very differently from that at the earth's surface. This "shell" of conducting air is said to be "ionised," and hence the whole region it occupies has been given the name "ionosphere." It extends from about 50 miles to about 300 miles or more above the earth's surface.

WHAT HAPPENS TO THE SKY WAVE IN THE IONOSPHERE

Referring again to Fig. 2, let us consider that part of the radiated energy going out in the upward directions such as are depicted by any of the obliquely-sloping arrows. When we consider the energy going upwards in one single direction only, such as is shown by one of these arrows—the energy contained within an extremely thin sector of the radiated hemisphere—then we may look upon it as a "ray" of radio energy, and regard it in a similar way to that in which we regard a ray of light. Whilst referring to this "ray" of radio energy we should do well to remember that in actual fact there will always be, not one, but a great many such rays travelling upwards side by side.

The ray we are considering travels onward and upward in a straight line and at its normal velocity of 300 million metres per second. When it enters the ionosphere, with the latter's layers of conducting air, its behaviour alters. It begins to travel at a different velocity and, as a result of this, it is deflected from its straight path and commences to curve round so that it is travelling at a smaller angle to the earth's surface than when it left the ground. This curving process continues as it penetrates deeper into the ionised region, until eventually the ray is bent completely round, and is travelling towards the lower edge of the ionosphere again. It emerges into the ordinary air travelling at its original velocity and in a similar oblique path to that of its upward journey, and so it continues straight on towards the ground. This curving process is known as refraction, and is illustrated in Fig. 4,

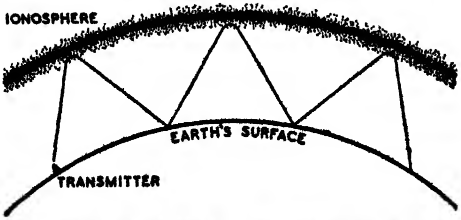


Fig. 4—Short-wave transmission round the curvature of the earth

where, for the sake of clarity, some of the dimensions have been greatly exaggerated.

Having lost very little of its energy during its journey—it *will* lose some in the ionosphere, particularly under certain conditions—the ray is able to actuate a radio receiver at the point on earth where it returns, which will be many miles distant from the transmitting station. Furthermore, on reaching the earth's surface it is reflected like a light ray from a mirror, and is sent off upwards again at the same angle at which it started. Reaching the ionosphere once more, it is again refracted and returned to earth at a point twice as far from the transmitter as that at which it first came down.

These processes are repeated again and again so that the wave travels to the greatest distances on earth in a series of hops, as is pictured in Fig. 4. This is very convenient and fortunate for us, for when the spherical shape of the earth is considered it is impossible to see how we should ever get a radio wave to Australia, for example, if it persisted in travelling in a straight line.

We have, in our examination, considered only one ray of radio energy, but if we bear in mind that all the adjacent rays are being affected by the ionosphere in a similar way we shall see that a considerable portion of the earth's surface at the distant points will be covered by the downcoming rays, so that many thousands of radio receivers will be able to pick up the signals sent out.

We shall go into this matter of ionospheric refraction in greater detail later on. Let us return for a moment to a consideration of Fig. 3.

PROPAGATION CHARACTERISTICS FOR DIFFERENT WAVELENGTHS

When we were discussing the ground wave we saw that the attenuation due to earth losses was least on the long waves and gradually increased as wavelength was reduced, thus leading to very small ranges on the shorter wavelengths.

How does attenuation vary in the case of the sky wave ? Well, the sky wave is subject to absorption whilst it is travelling in the ionosphere, and this varies with wavelength in the opposite way to the ground absorption, being least on the shortest waves and increasing as wavelength is increased. Thus the wave is, generally speaking, more greatly attenuated by the ionosphere the longer its wavelength.

But the whole truth is somewhat more complicated than this, so let us examine Fig. 3 again, starting at a wavelength of about 300 metres. At this wavelength the ionospheric absorption is so high that all the upgoing energy is lost in the ionosphere—the sky wave is *completely* attenuated, and there is only the ground wave on which to effect communication. At a wavelength of about 100 metres the ionospheric absorption is reduced sufficiently for energy to be returned from the

TABLE 2

Classes	*Frequency sub-divisions*	*Main characteristics*	*Principal uses*
Very long waves and long waves	VLF and LF	Wave travels to considerable distances over earth's surface and to great distances by reflection from lower edge of ionosphere	Medium and long distance point to point communication. Long-wave medium-distance broadcasting
Medium waves	MF	Wave travels over earth's surface to relatively short distances during day, at night some energy comes from ionosphere and range increases	Local broadcasting, marine and aircraft communication, direction finding
Short waves	HF	Wave travels up to ionosphere whence it is reflected back to earth. Conditions vary greatly with time of day and season, but great ranges obtained if conditions favourable	Long distance broadcasting, point to point communication. Amateur communication, et cetera
Very short, ultra short, super short and extremely short waves	VHF, UHF, SHF and EHF	No ionospheric reflection. Wave travels directly through lower atmosphere from transmitter to receiver	Short distance communication, FM broadcasting, television, radar, aircraft guidance systems. Amateur communication

upper atmosphere, and so we begin to get increased ranges due to the sky wave. Decreasing wavelength still further results in the ionospheric absorption being rapidly reduced, so the sky wave travels to greater and greater distances. This continues until a wavelength of about 10 metres is reached, when (as is seen from Fig. 3) the range is suddenly reduced to that of the ground wave again. This is not due to any increase in ionospheric absorption, but to the fact that, at this wavelength, the ionosphere is no longer capable of turning the wave around sufficiently for it to return to earth—its *reflecting* properties are reduced, and so the sky wave penetrates through it and is lost in outer space.

Returning now to the wavelength of 300 metres and proceeding towards the long-wave end of Fig. 3, we see that on the longer wavelengths great ranges are again achieved. As has been said, the ground wave on the long wavelengths is capable of travelling to considerable distances. But we must not fall into the error of supposing that the great ranges on the long waves are *entirely* due to good propagation of the ground wave. As we have seen, the ionospheric absorption increases with wavelength but it only does so up to a wavelength of about 200 metres, and, on the long waves, begins to decrease again. Also, on these wavelengths the sky wave does not *penetrate* into the ionosphere, but is reflected from its lower edge. Thus it does not encounter much ionospheric absorption, and so is not greatly attenuated. Consequently propagation of both ground and sky wave is relatively good on the long waves, and by the latter the greatest ranges can be obtained, at least on the longer of these waves.

It should be added that there are many factors affecting this matter which will introduce modifications to Fig. 3, and to the explanation of it just given. In particular, because the state of the ionosphere varies from hour to hour, season to season, and year to year, the limiting wavelengths given above for each type of propagation are not to be taken rigidly, but rather as indicating in a general sort of way how propagation varies with wavelength.

We may now summarise the propagation characteristics of the wavelengths or frequencies classified in Table 1 and indicate the principal uses to which they are put. This is done in Table 2 (page 19).

From now on we shall concentrate almost entirely on the short radio waves, and in the next chapter it will be well to examine briefly the way in which their transmission medium, the ionosphere, is formed, and to learn something of its structure.

CHAPTER 2

Formation and Structure of the Ionosphere

DISCOVERY OF THE IONOSPHERE

WHEN, IN 1901, MARCONI SUCCEEDED in picking up radio signals in Newfoundland from his station at Poldhu in Cornwall, he created a problem that required considerable explanation. For it had previously been mathematically proved that the ground wave could not curve around the earth's surface sufficiently to reach such a distant point, and furthermore it was known that radio waves travelling through the atmosphere itself—a medium then supposed to possess constant electrical properties—must travel in perfectly straight lines. **HERTZ** had shown that they could be made to diverge from such straight paths only by interposing within the path a material having different properties from that of air, such, for example, as a metal sheet. By what means, then, did they manage to reach a point which necessitated their curving round an obstacle about 200 miles high—the " bulge " in the earth's surface between transmitter and receiver ?

It was then independently suggested by **HEAVISIDE**, and by **KENNELLY**, that the air in the upper atmosphere was not the same as that at ground level—that the earth was, in fact, surrounded by a region of *conducting* air, capable of reflecting radio waves and so of preventing their escape from the atmosphere, and of guiding them round the earth's curved surface.

It was soon seen that the air, which had always been regarded as a good insulator, might, in the high atmosphere, be converted into a conductor, by the process of ionisation. For in that region there would exist the atoms of gas capable of being made into ions and free electrons, and also the downcoming energy from the sun, capable of doing the work. But the theories of **HEAVISIDE** and **KENNELLY** were to remain as theories for many more years—years during which the art of radio communication underwent tremendous expansion and development, and during which the ionised region itself was more and more utilised.

It was not until 1924 that the first experiments which directly *proved* the existence of the **HEAVISIDE** layer were actually made, by Sir **EDWARD APPLETON** and his co-workers, using the Bournemouth transmitter of the B.B.C. In that year, and in 1925, they made many such experiments, and found that the actual structure of the ionospheric region was not so simple as might have been supposed. There were, in fact, not one, but several ionised layers of gas, lying one above the other in the atmosphere. Furthermore the ionisation, as might have been expected, was found to vary constantly and considerably, from hour to hour, from season to season, and over a still longer period.

Since then the technique of ionospheric measurement has been considerably expanded and improved, and, as a result, a great deal has been learnt about the ionosphere, enabling us to understand more about the reasons for the existence of such a region, to learn the details of its structure and of the continual variations which occur within it, and to develop methods for the application of this knowledge to our practical problems of short-wave transmission and reception. Not all is known, however, about the region even yet—there remain many points on which only incomplete and partial knowledge is yet available. So, in what follows, we will attempt to give an idea of the main details of the causation and structure of the ionosphere, with the reservation that, on some points, future knowledge may necessitate modifications of what is said.

FORMATION BY THE SUN

The ionosphere is brought into existence by energy which is radiated from the sun. This fact very quickly becomes evident when we study some of the variations in the conductivity of the air which occur within it. Perhaps the most striking evidence of its dependence on the sun's radiations is its behaviour during a total eclipse of the sun. During such an event, when the sun's rays begin to be cut off by the moon, the conductivity or degree of "ionisation" of the air in the ionosphere begins to decrease, and this decrease continues as the eclipse proceeds towards totality. A minimum in the ionisation occurs at about the centre of the eclipse, and when it is over the ionisation increases again and soon returns to a normal state. This shows that the agent responsible for the ionisation of the air has been prevented from reaching it during the eclipse, its path having been blocked, as it were, by the presence of the moon in between the sun and earth, in the same way that the sun's light and heat have been cut off from the earth's surface during the eclipse. Furthermore, the fall in the ionisation is observed to start at the same time that the amount of light and heat reaching the earth starts to decrease, thus proving that all three radiations—light, heat and ionising radiations—are travelling towards the earth at the same speed.

Again, there are very many variations in the degree of ionisation of the air as between night and day, and also as between summer and winter, as would be expected if the sun were responsible for it. All the evidence goes to show, therefore, that the agent causing the ionisation is indeed a part of the energy emitted by the sun. Fig. 5 is a plot of some ionospheric observations made before and during an eclipse of

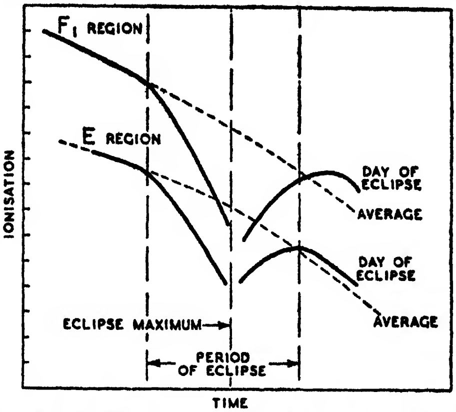


Fig. 5—Effect of an eclipse of the sun on the ionisation of the F_1 and E layer

the sun, and illustrates the effects just mentioned, so far as they affected two of the ionospheric layers.

The Sun's Radiations

The sun pours out into space an enormous amount of energy in the form of heat, light and other types of waves. A small part of the total energy emitted travels in the direction of this planet, and again, a small part of this reaches the ground. This we are very well aware of, because our eyes and other organs are sensitive to certain ranges of solar waves which we know as light and heat. But a considerable part of the sun's energy never reaches the ground. It is expended in ionising the gases of the upper atmosphere.

These radiations of the sun are electromagnetic waves, and, as such, have much in common with radio waves. They are propagated according to the same general laws, and at the same velocity. But, compared with radio waves, they are of tremendously short wavelength (high frequency). The longest waves to which our eyes are sensitive—that of red light—is only about 0·0008 cm in wavelength, and the shortest, that of violet light, about 0·0004 cm wavelength. Shorter in wavelength than these, and, of course, invisible to us, is a whole range of radiations extending down to very much shorter wavelengths still, and known as the " ultra-violet " radiation.

We have already seen that when electromagnetic waves of " radio " wavelengths travel through bodies of un-ionised gas—such as the air at the earth's surface—there is no appreciable effect either upon the gas or upon the wave. But when we come to waves so short in wavelength (or so high in frequency) as the ultra-violet waves from the sun, the position is quite different, as we shall shortly see.

THE ATMOSPHERIC GASES

The gases constituting the earth's atmosphere—the principal of which are oxygen and nitrogen—are not uniformly distributed throughout the whole thickness of the atmosphere. In the lower six miles or so —called the troposphere—they are kept uniformly mixed by the action of the weather, but at great heights, where weather phenomena do not exist, the constituent gases become separated. The lighter ones tend to rise and float at higher levels than the heavier ones, so that at certain levels there will be a preponderance of one particular gas.

The details of the actual distribution of the gases in the high atmosphere are not yet definitely known, though new information is being acquired. It seems likely that the distribution of the gases themselves is to some extent affected by the rays which pass into the atmosphere from the sun, for these can cause the gases to exist in different forms, and this fact may account, in part, for the different structure of the ionosphere at different times of day and at different seasons of the year. We may take it, in any case, that the stratification of the ionosphere into " layers " of ionisation is due, to some extent, to the predominance of different types of gas at different heights.

IONISATION OF THE GASES

Gases, like all material bodies, are composed elementally of atoms, and atoms, so far as they concern us, are made up of positively-charged nuclei around which revolve negatively-charged electrons. The whole atomic system is, so far as external points are concerned, in a state of electrical equilibrium, the negatively-charged electrons just counterbalancing the positive charge of the nucleus. A neutral atom of this kind exerts no electrical force outside its own structure, and its structure is so small that the atom itself is not affected by a passing radio wave.

But the wavelength of an ultra-violet wave is so short that interaction can take place between it and the atomic structure itself, and, the sun's ultra-violet emissions embracing a whole band of these very short wavelengths, if it happens to be at the right wavelength for an atom of a particular type—say of one particular gas—the wave can exert a tremendous force upon the atomic structure. In fact it can set the atom into such a state of agitation that electrons are dislodged from their orbits and may even be knocked out of the atomic structure altogether. An electron which is dissociated from its parent atom in this way is known as a " free " electron, and exists as a separate negatively-charged unit. The atom which has lost an electron is no longer electrically neutral but acquires a positive charge, and is known as an " ion."

The ultra-violet wave, in performing this operation, expends a considerable amount of its energy, and, if the number of atoms on which to operate is sufficient, it eventually expends the whole of it—it is " absorbed " by the gas, and dies away. Atoms of different gases respond most readily to ultra-violet rays of different wavelengths, according to the " frequency " of the waves and the nature of the gas atoms. Thus waves of one wavelength may affect the atoms of nitrogen, whilst it requires a different wavelength entirely to liberate electrons from the atoms of oxygen in a similar way. The amount of " energy " contained in the ultra-violet waves may also have something to do with this matter, in determining whether or not a particular class of radiation is capable of ionising the atoms of a particular gas. But, as has been said, the sun's ultra-violet radiations appear to contain all the wavelengths necessary for the ionisation of the different atmospheric gases and, furthermore, the radiations are intense enough to set up a large amount of " ionisation," which is the term by which this electron-liberating process is known.

Of course, if the free electrons get near enough to an atom which has lost an electron they rapidly become attached to it, thus forming an electrically neutral structure again. This is the process known as " recombination " and if it went on rapidly enough the ionisation would soon disappear. But the chance of a free electron finding an ion in the high atmosphere is very much less than at ground level, because of the comparative rarity of the gas at those heights, and thus of the relative sparsity of the atoms. Also, as fast as—often faster than —the free electrons become re-attached to ions others are set free by the continually arriving ultra-violet rays, so that, until the rays are cut off, there is a more or less steady supply of them.

It will be remembered that the essential feature of an electrical insulator is that its electrons are all held fast within their parent atoms, and cannot be made to move outside them. A gas which is in an ionised state possesses the properties of an electrical conductor—because its free electrons are capable of independent movement—and acts upon radio waves in a similar way to that in which a metallic conductor would act. It is therefore the splitting up of the neutral gas atoms into ions and free electrons which changes the electrical nature of the upper atmosphere and so renders it impervious to radio waves coming up from the earth. When the sun's rays are cut off from the atmosphere—as at night or during a total eclipse—then the molecules and free electrons *do* recombine so as to cause the density of the free electrons to diminish. The *rate* of recombination will depend upon the density of the gas, being low at the outer part of the atmosphere where the gas is very rare, and greater at lower levels where more gas atoms per unit of volume are to be found.

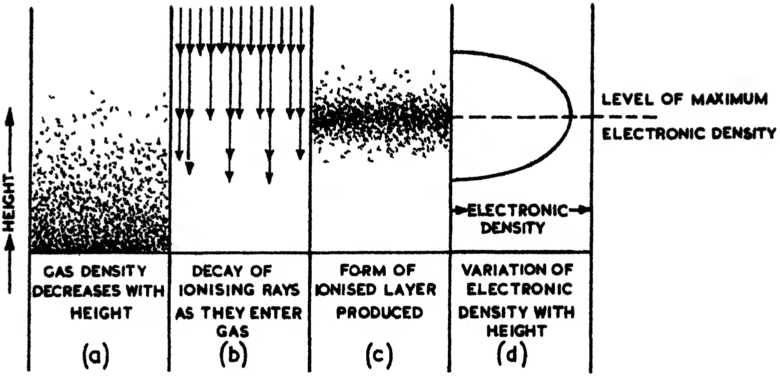


Fig. 6—Details of the formation of an ionised layer

DISTRIBUTION OF THE IONISATION

Thus we have the ultra-violet rays coming down from the sun and encountering the gases of the high atmosphere. As they meet a concentration of any particular gas the waves of a certain wavelength are absorbed, and their energy is expended in ionising the gas. At first, though the rays are tremendously strong, the gas itself is very rarefied (that is, the atoms are few and far between) and so little ionisation is produced. As they travel on the gas gets denser, and so more and more ionisation is produced because there are now plenty of atoms on which the rays can expend their energy. But this expenditure of energy means that the rays themselves become weaker, and so lower down still they are unable to ionise so many atoms. Finally the energy in the rays is completely used up and so the ionisation ceases to be produced.

Fig. 6 is an attempt to show *(a)* how the gas density increases with decreasing height above the ground, *(b)* how the ionising rays expend themselves as they pass into the gas, *(c)* the general form of the ionised layer produced, and *(d)* how the resultant electronic density varies with height above ground.

The density of the free electrons is therefore not constant throughout the ionised gas, but is at a minimum at the top and bottom edges of the region where a particular gas is ionised, and increases to a maximum at the centre. This is seen to be the case in *(d)* of Fig. 6, where electronic density is plotted against height above ground.

"LAYERS" OF IONISED GAS

This, then, is the sort of thing which happens at different levels in the atmosphere, ultra-violet rays of different frequencies penetrating down to different heights before expending their energy in setting up ionisation, and different concentrations of gas predominating at these different heights. The result is that throughout the ionosphere are formed several *layers* of ionised gas, each having ionisation densities similar to that depicted in (*d*) of Fig. 6. It is important to appreciate that the ionisation is not uniformly distributed with altitude, but is stratified into these layers, each of which varies considerably in its characteristics, however, from time to time.

The condition of the air in between the well-defined ionised layers is not definitely known, but certainly the ionisation density there would appear to be less than in the layers themselves.

Full information about the reasons for the heights taken up by the different layers is still lacking, so we had better refrain from discussing the matter further. We may, however, give an idea as to the positions in the atmosphere of the various layers, and list the names by which they are known.

Somewhere near what is the outer fringe of the earth's atmosphere lies what is (so far as we know) the uppermost layer of the ionosphere, at a height varying between about 150 and 300 miles above the ground. This layer, named—after its discoverer, Sir EDWARD APPLETON—the " Appleton Layer," is also called the F layer of the ionosphere. This is the most important layer so far as short-wave propagation is concerned, for it acts as the principal refracting layer in long-distance communication.

It may here be remarked that the allocation of letters to designate the various layers was initiated by Sir EDWARD APPLETON, who, when he discovered the upper layer in 1925, called it the F, and the other layer then known the E, whilst a still lower layer which he located was called the D. This, as he says, left several letters at the disposal of future workers for allocation to other layers which they might discover above or below these three. So far, no other letters have had to be brought into use, and it does not now seem likely that they ever will.

The F layer has been found to divide into two distinct layers during the daytime, the lower one taking up a position about 120 miles above the ground. When this happens the upper part is called the F2 layer and is the most important short-wave refracting layer, whilst the lower part is the F1, and is capable of propagating radio waves of certain frequencies.

During the daytime there is also present, at a height of about 70 miles above ground, the original KENNELLY-HEAVISIDE layer, now known as the E layer. This is also capable of refracting certain short waves during the day, though at night its ionisation decreases to a very low value, whilst not disappearing altogether. Slightly below the E there is found, during the daytime only, the D layer of the ionosphere, a region in which occurs the greater part of the absorption to which short waves are subject. There does not appear to be any appreciable ionisation lower in the atmosphere than this : at least, no ionised layer *regularly* exists lower down.

STRUCTURE OF THE IONOSPHERE

Having listed the ionospheric layers and indicated their main effects in short-wave propagation, let us now look at a diagram which will help us to form an idea of the main details of the ionospheric structure as it exists on some significant occasions. As has been mentioned, owing to the varying intensity of the solar rays with time, the ionisation density will not always be the same within the layers ; nor, at least in the case of the F layer, will the height always be the same. Both height and density will vary, seasonally, diurnally, and also over a long period.

Then, again, the ionospheric structure will vary considerably at different points on the earth's surface, for the sun's action upon the outer atmosphere will obviously vary with geographical latitude. Let us, however, picture the various layers as they might exist over this country during the winter and during the summer, bearing in mind that such a picture must be subject to considerable modification in the light of further knowledge.

Fig. 7*a* shows the structure prevailing during the winter day. With its bottom edge at a height of about 150 miles lies the F2 layer, whilst immediately below it—hardly separated from it in fact—lies the F1. Then, with its bottom edge at a height of about 70 miles, we have the E layer, and immediately below that the D.

At night there is not much difference in the general ionospheric structure as between winter and summer—though there *is* a considerable difference in the ionisation density prevailing—so Fig. 7*b* will serve to give an idea of the night-time structure throughout the year. We have, in fact, simply the F layer at a height of about 180 miles

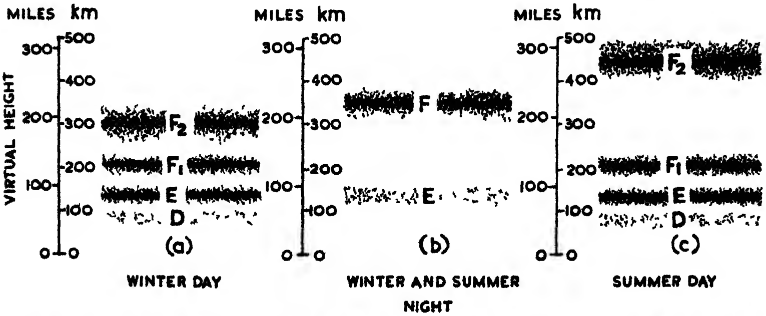


Fig. 7—How the structure of the ionosphere changes from day to night and from winter to summer. (Note : The layer positions shown are based on *minimum virtual light measurements*, and the heights of maximum ionisation may, in fact, be considerably different from those shown)

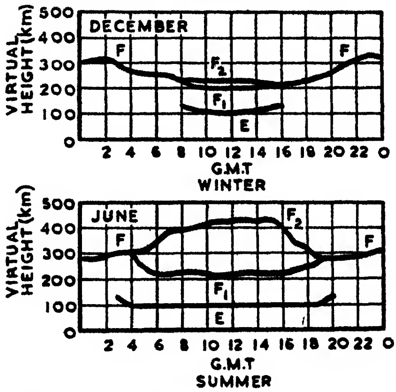


Fig. 8—Monthly average virtual heights of principal ionosphere layers. From observations made at Slough

and no other ionisation present with the exception of a very weakly ionised E layer.

Fig. 7*c* depicts conditions during the summer day. The three lower layers occupy approximately the same positions which they did during the winter day, but the F2 layer now appears at a much greater height. It will be observed that, whilst the lower layers occupy the same positions throughout the year, it is the F layer whose position varies so considerably, its height being less during the winter day than at night, but greater during the summer day than at night. Also that the division of the daytime F into two layers is much more pronounced during the summer than during the winter day. We shall have more to say about these variations—and more particularly about those in the ionisation density—later on.

It may help, at this stage, if we give some typical measured values (of the lower edge) of the F, F1, F2 and E layers as obtained throughout the 24 hours of a winter and a summer day. These are shown in Fig. 8, where however the D layer is not shown and the E layer only during those hours when its ionisation is sufficient to be of any significance as far as the propagation of *short* waves is concerned.

The reader need not attempt to memorise all these different heights for every hour of day. It will be sufficient for our purpose at the moment to picture the daytime ionosphere as a region comprising three reflecting layers (the F2, F1 and E) and one absorbing layer (the D), and the night-time ionosphere as consisting almost solely of the F layer.

This will simplify matters somewhat, and we can now go on to examine in more detail the behaviour of a radio wave in the ionised region.

CHAPTER 3

Radio Waves in the Ionosphere

REFRACTION IN THE IONOSPHERE

WE HAVE ALREADY SEEN that a radio wave travels through "free space"—and through ordinary un-ionised air—with a velocity of 300 million metres per second, but that when it enters the ionised gas of the ionosphere its velocity is altered, and its direction of travel changed. We have seen also that whilst ordinary air acts as an electrical insulator, ionised air acts as a conductor. Its conductivity is not so high, of course, as is that of a metal sheet, but it has this property in common with a conducting metal—it possesses electrons which are capable of being set into independent motion, and its degree of conductivity is determined by the number of such free electrons existing per unit of volume.

Electrons, when set in motion, constitute an electric current, and it is the setting up of such currents by the radio wave which gives the ionosphere its refractive properties, and so enables it to bend the wave from its straight path, and cause its return to earth again. This refraction or reflection of the wave can happen, under the right conditions, whether it is sent upwards obliquely or whether it is sent vertically upwards, even though the latter means that the wave must be brought to a complete stop somewhere in the ionosphere, and then sent travelling downwards again.

The full explanation of the behaviour of a radio wave in the ionosphere is very involved, and is beyond the scope of this book. Since it is necessary, however, to know *something* about the matter in order to understand the principles of short-wave transmission, we must be content with the simplified explanation given below.

WAVE VELOCITY

It will be remembered that a radio wave has been said to consist of oscillating electric and magnetic fields (or strains, as we called them), the energy in the wave manifesting itself in both of these ways and being measurable by either electric or magnetic means. Now the velocity at which the wave travels depends upon the nature of the *current* set up by these oscillating fields. A current is, in reality, a movement of electrons, and such movement may be caused in a conductor by the action of an electric field, but in an insulator, movement of the electrons *through* the material is not possible. Ordinary air is an insulator, and thus the electric field of a radio wave travelling through it does not set up any actual current (though it does set up a " displacement " current), and the wave travels at the velocity of 300 million metres per second.

It may be asked why it travels at this particular velocity, and the answer is that this is determined by the rate of change of the electric and magnetic fields in the wave—if these effects are to move at all they can only do so at this speed. It will be remembered, from Chapter I, that it is the " displacement " current which sets up the magnetic field of the wave, and that it is the rate of change of the electric and magnetic fields which governs the velocity of propagation. This velocity of 300 million metres per second—designated " c "—is the natural velocity of light, and of all other electromagnetic waves, in " free " space.

When the wave travels in " ionised " air, however, the situation is different, because the electric field is now able to set the free electrons into motion, thus setting up an *actual* or " conduction " electric current, which of course affects the behaviour of the wave. The " conduction " current which is now set up is not in the same *phase* as the displacement current, and therefore cancels the latter out—at least in part—so that the rate of change of the ensuing magnetic and electric fields of the wave is altered. This alteration results in the *phase* of the wave at a given point in the new medium being shifted, so that the wave behaves as though it had been speeded up. As soon, therefore, as the wave sets up a conduction current its velocity increases. As we have already seen the farther we go up into the ionised region the greater does the density of the free electrons become ; and so, as the wave penetrates farther into the region, the conduction current will become more and more effective, and the wave velocity get increasingly greater.

Fig. 9 illustrates what will happen under these conditions. Suppose a b to represent the wave front of a wave advancing towards the ionosphere by the oblique path indicated by the arrow. Whilst the wave is travelling in un-ionised air a and b will travel at the same speed, and the wave front at successive instants of time will be as indicated by the next few lines perpendicular to the direction of travel, which is constant. When the wave enters the ionised layer, however, the top part of the

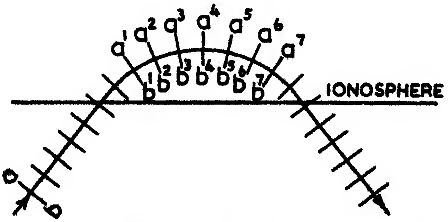


Fig. 9—" Bending " of the wave path within the refracting layer

wave front will be in a region of greater electronic density than the bottom part, and will therefore travel at a greater velocity. Thus a^1, a^2, et cetera, in the figure will be travelling at a greater velocity than b^1, b^2, et cetera. If this is the case it will be seen that the wave path cannot continue in a straight line, but will bend away from regions of high electronic density to those where it is low, and so describe the trajectory depicted in Fig. 9. Thus the wave is diverted from its oblique upward path by the effect of the ionised air, and sent downward towards the earth again.

REFRACTION OF A LIGHT RAY

We may find this matter easier to understand, however, if we look upon the upgoing radio energy not as a wave disturbance but as a single " ray " of energy, possessing, as it does, similar characteristics to those of a ray of light. This being so, we can apply to it the laws of optics.

Fig. 10*a* illustrates what happens to a ray of light when passing from a rare into a dense medium, such as from air into glass. The shaded area represents the dense medium, A B being the surface of separation between the two media, and C D, a line drawn at right angles to the surface, is called the " normal " to this surface or boundary. I O is the light ray approaching the surface so as to reach point O—the point of " incidence "—at an angle i_1 to the normal. This angle i_1 is called the " angle of incidence." Instead of continuing its course in a straight line in the new medium, the ray is refracted or bent in towards the normal, so as to travel along the line OQ, the angle i_2 made with the normal in the new medium being called the " angle of refraction." It is because the new medium is denser than the old, and possesses a greater " refractive index," that the direction of bending of the ray is in towards the normal, the *amount* of bending depending on the relation between the refractive indices of the two media. According to the laws of refraction the sine of the angle of refraction times the refractive index (μ) of the second medium is equal to the sine of the angle of incidence times the refractive index of the first medium, that is—

$$\sin i_1, \mu_1 = \sin i_2 \, \mu_2.$$

When the light ray passes from a medium of greater to one of smaller refractive index (as in b, Fig. 10) it will be seen that the ray must bend *away* from the normal in order to keep the product of sin i and refractive index μ constant.

If we increase the angle of incidence—by making the ray path more oblique—then the angle of refraction will increase also, until we get an angle of refraction of 90°, when the path of the refracted ray is exactly along the surface of separation. A *further* increase in the

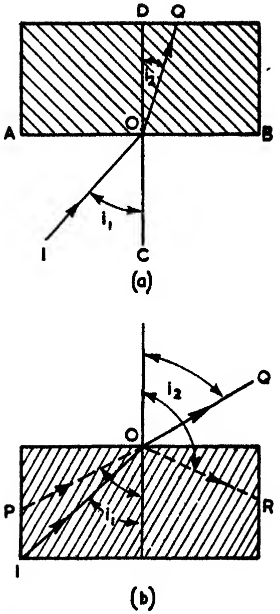


Fig. 10—Behaviour of a light ray in passing into media of differing refractive indices

angle of incidence leads to the result shown by the dashed lines in Fig. 10*b*. The line P O is the direction of the incident ray and O R that of the refracted ray—the angle of refraction is greater than 90° and the light ray cannot penetrate into the new medium at all, but undergoes " total reflection " at its surface. This total reflection will occur after the product of the sine of the angle of incidence and the refractive index of the first medium is equal to the refractive index of the second medium, the sine of an angle of refraction of 90° being, of course, 1·0.

REFRACTION OF THE RADIO RAY

Now we can return to our ray of radio energy, which is approaching an ionosphere layer where it will undergo refraction in a similar manner to that in which the light ray was deflected from its course. In order to compare the two cases we must first of all make a number of assumptions, the principal of which are that the earth and ionosphere are flat, and that the earth's magnetic field has no effect upon the wave. Whilst neither of these is true they will enable us so to simplify things that we can obtain a picture of what goes on in the ionosphere—a picture which may later be somewhat modified.

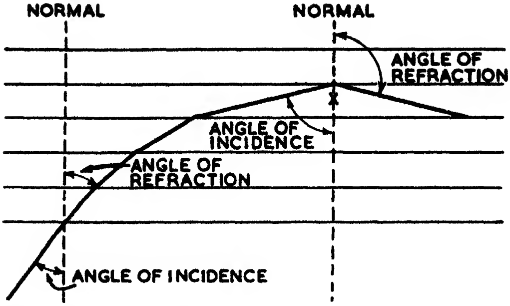


Fig. 11—Showing successive refractions at imaginary thin layers of constant electronic density, each layer having greater electronic density than that next below it

Since the radio ray obeys the laws of optical refraction, and since the refractive index of ordinary air is 1·0, whilst that of the ionosphere gets increasingly smaller than 1·0, it follows, from what was said at the end of the last section, that the ray must penetrate into the region until the refractive index μ_2 is reduced to the value sin i, for, when this has occurred, the angle of refraction will be 90° and the ray will be travelling parallel to the earth's surface.

Let us see in a little more detail what happens.

The decrease in refractive index in the ionosphere is due to the increase in electron density within it, and, as we saw in Chapter 2, this increases continuously as we proceed from the bottom edge upwards to the point of maximum electronic density. Suppose, however, that instead of increasing constantly with height, the electron density were itself in a series of thin "layers," as shown in Fig. 11, each thin "layer" having a constant electron density which was greater than that of the "layer" next beneath it, and therefore a smaller refractive index. Then we should get the effect shown in the figure. The radio ray would be refracted away from the normal each time it came to the boundary of one of the thin layers, and it is seen that its path would become more and more oblique until eventually the angle of refraction would exceed 90°, and the ray would then start travelling downwards again, as at X in Fig. 11. It will be seen that the smaller the angle of incidence in the first place (that is, the more nearly vertical the approach of the ray towards the layer), the greater would be the number of thin layers it would enter before the angle of refraction exceeded 90°.

With the continuously increasing electron density that does in fact exist it means that the ray is *continuously* refracted, and instead of its path being in a series of oblique straight lines it becomes the curving sweep with which we are already familiar. Thus, due to the increasing electron density in the ionospheric layer, the radio ray is bent around in a curve and sent downwards to the earth again. As we increase the distance over which we wish our ray to travel, by making its approach to the ionosphere more and more oblique, so does the ray penetrate less and less deeply into the ionospheric layer and require a smaller electron density to effect its return to earth.

REFRACTIVE INDEX VARIES WITH FREQUENCY

Now we must leave our " ray " picture for a moment and go back to our consideration of the radio energy as a " wave " disturbance, in order to explain some very important points. We saw, earlier on in this chapter, that it was the magnitude of the electronic current set up by the wave which determined its change of velocity in the ionosphere, and hence the amount of " bending " of its path.

When the wave sets the free electrons into motion the impetus given to them will vary according to the rate at which the electric field of the wave is changing. The velocity attained by the electrons will be determined by the time during which the field continues to act in one direction, which is tantamount to saying that the *magnitude* of the conduction current will be greater the slower the changes in direction of the field. Thus the lower the frequency of the wave the greater will be the effect of the moving electrons—in altering the velocity and direction of travel—upon it. It varies, in fact, inversely as the square of the frequency or directly as the square of the wavelength. And so we see that a wave of high frequency (short wavelength) will penetrate farther into an ionised layer before the electron density is sufficient to ensure its reflection than will a wave of low frequency (long wavelength).

We have thus learnt two fundamental points about short-wave transmission, which we may now summarise. They are—

(1) The greater the angle of incidence on the ionosphere—that is, the greater the distance we wish to cover in one hop—the less will the wave penetrate into the layer and the smaller will be the electronic density required to return it to earth.

(2) The lower the frequency of the wave the less far will it penetrate into the layer and the smaller will be the electronic density required to return it to earth. Also, there is an upper limit to the frequencies which *can* be returned, depending on the maximum electron density existing within the ionised layer.

Let us be sure that we fully appreciate the implications of this second point. It means that the decrease in refractive index within the ionosphere depends not only upon the electronic density existing, but also upon the frequency of the wave. The higher the frequency (the shorter the wavelength) the less rapidly will the refractive index decrease, the less will the wave path be bent, and the farther will it penetrate into the ionised layer. And if we continue to increase the frequency the wave will continue to penetrate farther into the layer until it reaches the point of critical electronic density somewhere at the centre of the layer. A *further* frequency increase means that the wave does not encounter a sufficient electronic density (or a sufficiently small refractive index) for its path to be bent through an angle of 90° to the normal, and at this frequency the wave escapes through the layer and is not returned to earth at all.

EFFECT OF IONOSPHERE CURVATURE

So far we have considered the ionosphere to be flat, and have pictured it so in several of our explanatory diagrams. But in reality the ionospheric layers have curved surfaces like that of the earth itself, and, as has been mentioned, this curvature introduces some modifying effects upon the behaviour of the radio wave within the layers. Though we cannot consider these in detail we may see what their main implications are.

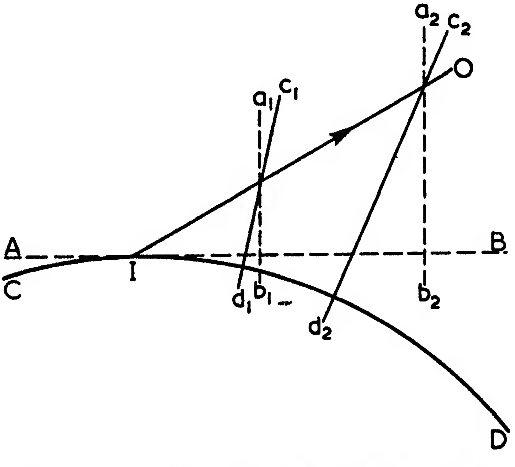


Fig. 12—Illustrating the effects of ionosphere curvature

Reverting back to Fig. 11 we can see that if each of the thin " layers " were curved instead of being flat, the angles of incidence made upon them by the advancing ray would be smaller than those shown in the figure, and thus the angles of refraction would be smaller. We may perhaps explain the effect by means of Fig. 12, in which A B is the lower boundary of the flat ionosphere and C D that of the curved one. We may imagine the electron density to be uniform throughout the region, in which case the radio ray will proceed in the constant direction I O. It is seen that, whilst in the flat case the angles to the normals a_1 b_1, a_2 b_2 remain the same, in the curved case those to the normals c_1 d_1, c_2 d_2, get progressively smaller.

Due to the curvature of the ionosphere, therefore, the ray path would become more nearly vertical as it proceeded upwards, and it would need a definitely increasing electron density in order for the angles to the normals to be kept the same as in the flat case. It follows that, in an ionosphere where the electron density increases constantly with height, it will need an increased electron density compared with the flat case in order to overcome the curvature effect and bend the ray so that it is travelling in a horizontal direction, and then in a downward direction again. So the result of the ionosphere curvature is that the apex of the wave path occurs where the refractive index is lower—and the electron density higher—than would be the case if the ionosphere were flat.

Summing up the results of the curvature we may say that—

(1) It results in somewhat lower frequencies being returned for any angle of incidence at the lower boundary than if the ionosphere were flat.

(2) The modification to the flat case which it introduces is least when there is a thin, sharp reflecting layer and greater when the electron density increases only slowly with height.

GROUP VELOCITY

Now we come to what is rather a difficult phenomenon to explain, though, with the help of Fig. 13, we may attempt it. Earlier on we said that when the wave entered the ionosphere its velocity increased owing to the phase shift given to it by the effect of the " conduction " current which was created by the setting into motion of the free electrons. But

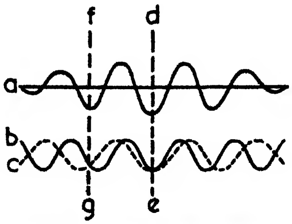


Fig. 13—How the point where the constituent waves reinforce each other varies with time, causing the " group " to travel slower than the constituent waves

the velocity we were speaking of—which exceeded that of light in free space—is the " wave velocity " and not the velocity of the signal as a whole. It is, in fact, impossible to transmit any *energy* at a greater velocity than that of 300 million metres per second, and so no *signal* can exceed this velocity.

The energy contained in the signal has its velocity reduced, not increased, whilst within the ionosphere. The signal is thus retarded, and, if it is travelling vertically, brought to a complete stop before being returned to earth. What affects it is not the speed of individual waves but a quantity known as the " group velocity," which refers to the speed of the signal as a whole—the speed at which the *energy* travels.

We have seen that, because the refractive index depends upon the frequency of the wave, the ionosphere is what is called a " dispersive " medium, that is, different frequencies travel within it at different speeds. If our signal consists of a wave group like that depicted at a in Fig. 13—and all signals must consist of some such group of waves of varying amplitude—then it is really made up of a number of sine waves of different frequencies, as at b and c in Fig. 13. As shown in the sketch, these two constituent waves reinforce each other at point d e, thus causing the peak of the signal group to occur at that point. But the fact that they are travelling at different speeds means that the point of reinforcement will occur later and later as the two frequencies travel along—as, for example, at f g after a short interval of time. Thus the peak amplitude of the group is moving slower than either of the constituent waves, and this means that the group as a whole is travelling at a slower speed than either. And so the group velocity in the ionosphere is always lower than it is in ordinary air.

The group velocity U depends, in fact, upon the refractive index, and is equal to $c\mu$, where, as before, c is the velocity of light in free space and μ is the refractive index. As the refractive index of ordinary air is 1·0, whilst in the ionosphere it gets increasingly less than 1·0, it can be seen that the group velocity gradually decreases, and, furthermore, the greater the wave velocity the smaller will the group velocity be. So the greater the electron density the slower is the signal as a whole propagated.

Whilst this retardation of the signal occurs at all angles of incidence we are now in a position to see more clearly how it is that a signal sent *vertically* upwards towards the ionosphere can be returned to earth. On entering the ionosphere the signal ascends at an ever-decreasing speed as it passes into regions of increasing electron density and decreasing refractive index. Finally, when the refractive index is reduced to zero it is seen (from $U = c\mu$) that the group velocity will also become zero—the signal will stop in its tracks—and then commence to travel downwards again. Gathering speed as it goes, it will emerge from the layer and travel with the velocity of light towards the ground —there to be received like an echo coming back from the sky.

As we shall see in the next chapter, it is by the use of the echoes from such vertically ascending signals that the heights and other characteristics of the ionospheric layers are measured and recorded.

CHAPTER 4

Measurement of the Ionospheric Characteristics

PULSES

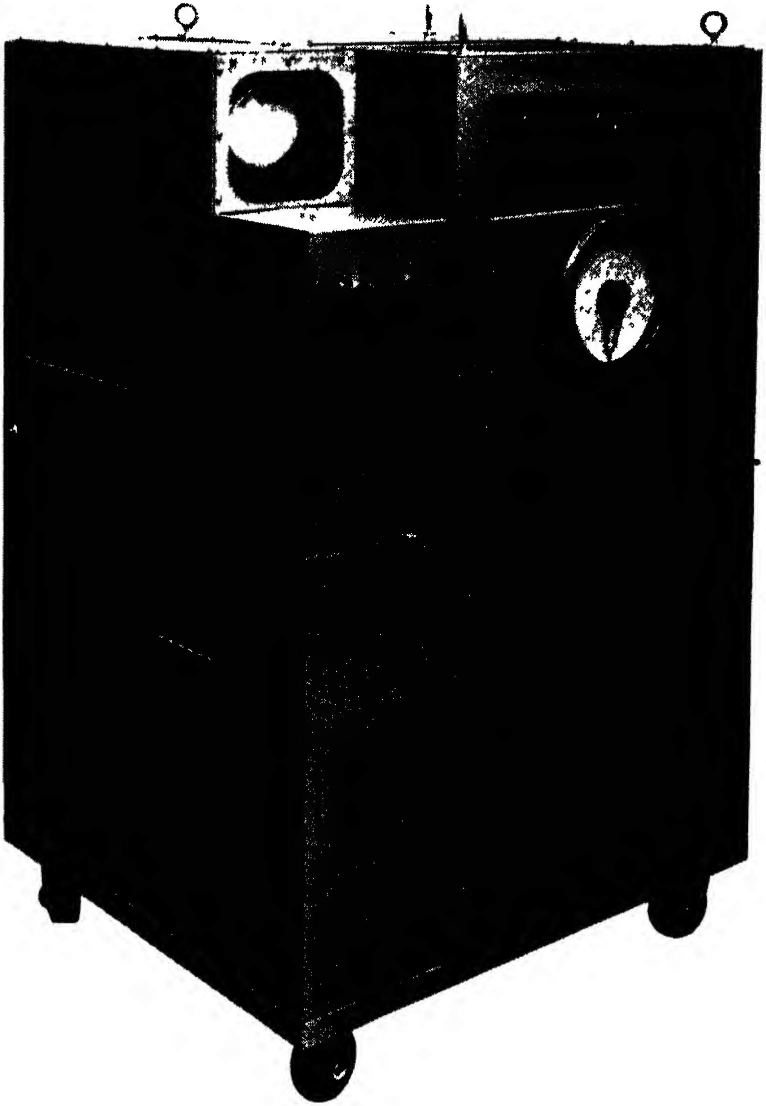
IT IS TRUE TO SAY that the methods developed for measuring the heights and other characteristics of the ionospheric layers are the foundation upon which the art of " radar " is built, and that the first object whose position was ever located by radar methods was the E layer of the ionosphere. In 1924 and 1925 Sir EDWARD APPLETON used a method for locating the layer in which, at a distance from the transmitter, he observed the interference effects between the ground wave and the downcoming wave when he altered the frequency, and then, by triangulation, determined the point in the atmosphere from whence the down-coming wave came. Soon afterwards a method was developed—by BREIT and TUVE—whereby the operation could be done from one point on the earth's surface, and this is now generally used.

In this the type of signal employed is a very short, sharp burst of energy, like the " dot " in the Morse code, though much shorter even than this. In point of fact it usually lasts for only 100 millionths of a second, and is called a " pulse." These pulses are repeated at intervals of about one-fiftieth of a second, this relatively large interval allowing plenty of time for one pulse to travel to the ionosphere and back before the next is sent off. By a suitable arrangement of the aerials at the measuring transmitter the pulse is made to travel vertically upwards, and a receiver located near the transmitter picks up the echo when it returns from the ionosphere.

Short as it is, the pulse consists of a small group of waves such as is depicted in Fig. 13*a*, and it therefore ascends with the velocity of light, slows down and eventually stops in the ionosphere, reverses its direction and returns to earth as an echo, in the manner described at the end of the last chapter.

OBSERVATION OF THE ECHO

The pulse and its echo are observed on a cathode-ray tube which is connected to the output of the receiver. The trace on the screen of the tube is actuated by a time base which is synchronised to the same repetition rate as that of the pulses, so that when a pulse is sent off the trace commences its sweep from left to right of the screen at a known speed. It is arranged that the received pulses will cause an upward deflection of the trace of the tube, and the sort of thing which appears



AN AUTOMATIC IONOSPHERIC RECORDER

The range of the apparatus is from 550 kc/s to 22 Mc/s, and a timed and dated record covering this range, with frequency marks at 100 kc/s intervals and height marks for every 50 km is completed in 5 minutes. The records are similar to that shown in Fig. 20. The equipment illustrated has been developed at the National Physical Laboratory of the Department of Scientific and Industrial Research. (Crown copyright reserved)

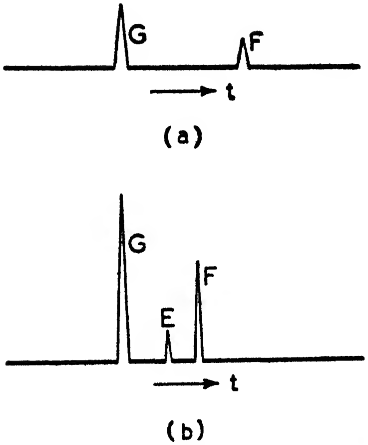


Fig. 14—Echoes returning (a) from the F layer, (b) from the E and F layers. G is the original pulse signal, received at the moment it is sent

on the screen is illustrated in Fig. 14. In (*a*) G is the pulse picked up by the receiver immediately it was sent off, and F is the echo received a moment later. As we know the speed at which the trace is moving from left to right we can measure the time to which the distance between G and F corresponds. This is the time taken for the pulse to go up to the ionosphere and back again. If we multiply this by the velocity of light we have—we will assume for the moment—the distance it has travelled, or *equivalent* path (p′). Half of this gives the height above ground at which the pulse was reflected, known as the *equivalent* or *virtual* height (h′). Fig. 14 (*b*) shows a case where a certain amount of energy was returned from the E layer, though the majority was reflected by the F, as shown by the longer time delay.

In practice the cathode-ray tube can be calibrated directly in terms of height rather than in terms of time delays, so we can read off the height of reflection of the pulse without having to make any calculation. What we are measuring, however, is the virtual height h′, and not the true height h, to which the pulse has reached.

We had better explain the difference between these two quantities straight away.

VIRTUAL HEIGHT

In converting the time delay between the transmitted pulse and the reception of its echo as observed upon the screen into terms of height, or in calibrating the cathode-ray tube directly in terms of height, we make the assumption that the pulse is travelling throughout the whole of its journey at the velocity of light c. Whilst this is true during its journey through the un-ionised air we have already seen that whilst in the ionised region it is retarded, that is, its group velocity is less than c. But, since we have no means of knowing exactly how the electron density varies with height in the ionosphere we do not know exactly *how* or in what *degree* it has been retarded, and so are unable to measure the actual height h to which it has ascended. We measure instead its *virtual* height h′.

In Fig. 15 we have illustrated the case for a pulse signal sent up somewhat more obliquely than is usual for measurement work, in order to make clear the difference between h and h′. A is the transmitter and F the receiver, and the trajectory described by the pulse is the full line ABCDF. It is whilst the pulse is travelling over BCD that its velocity is reduced, but if it had continued with its original velocity and had followed the path BED it would have arrived at F at exactly the same moment as it does in fact arrive there after following the curved path BCD. When therefore we take the delay between the reception of the signal and its echo, and multiply this by the velocity of light, then dividing the answer by 2 will give the height h′, and not the height of point C, the top of the actual trajectory. h′ will therefore always be greater than the true height h—or at least never less—and the difference between the two will depend upon the electronic gradient within the layer.

It is possible, by comparing the results obtained on a number of different frequencies, to calculate, from the virtual height, the actual heights to which the wave penetrates, but this is an extremely involved affair, and quite beyond the scope of this book.

We do know this, however, and should remember it when studying curves of h′ against frequency—that, whilst over a number of frequencies the difference between virtual and true height will not be great, near the frequency at which the wave escapes through any layer the large increase in *measured* height is partly due to retardation within the layer, and the difference between virtual and true height is at a maximum.

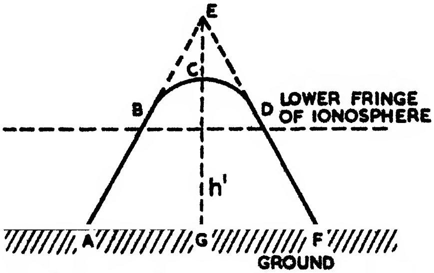


Fig. 15—Showing the difference between virtual and true height

PLOTTING THE IONOSPHERIC CHARACTERISTICS

Suppose now that we are about to start sending a whole series of pulses up to the ionosphere with the object of finding out what the prevailing conditions are. We arrange to start sending the pulses upon a relatively low frequency (long wavelength) and gradually to increase the frequency—in steps of, say, 2 kc/s—as we proceed. We will read off from the screen of the cathode-ray tube the height at which the pulses are reflected, and plot this in a graph of virtual height against frequency, known as an h′ - *f* curve. This is shown in Fig. 16, whilst the sort of display we should get upon the tube is shown, for several significant frequencies, in Fig. 17.

At first, upon the lower frequencies, we obtain no echo at all, because, on these frequencies, the upgoing energy is all absorbed in the D layer, as explained earlier on. At about 1·6 Mc/s we commence to get an echo from a height of 115 km, and this continues for a time, so that at a frequency of 2 Mc/s the display looks somewhat as is shown in Fig. 17 (1), where G is the " ground " pulse received at the moment it is sent off and E is the echo. We know, from the height indicated, that this echo is coming from the lower part of the E layer. Soon this echo begins to move to the right on the trace—indicating penetration into the layer—and then to decrease in size, whilst another echo appears further to the right, so that at about 3·1 Mc/s the display appears as in Fig. 17 (2). The E echo then disappears and that from what is evidently the F1 layer moves to the *left* of the tube instead of, as we might have expected, to the right, so that at about 3·7 Mc/s we have the display shown in Fig. 17 (3). As we plot this in Fig 16 we get an apparent *decrease* in height of the F1 from an increase in frequency. This is due to the effect we have already explained—the retardation of the pulse near the

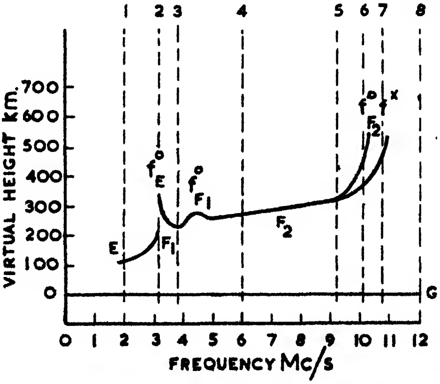


Fig. 16—An h′ - *f* curve, showing the virtual heights for the echoes obtained on a typical winter day

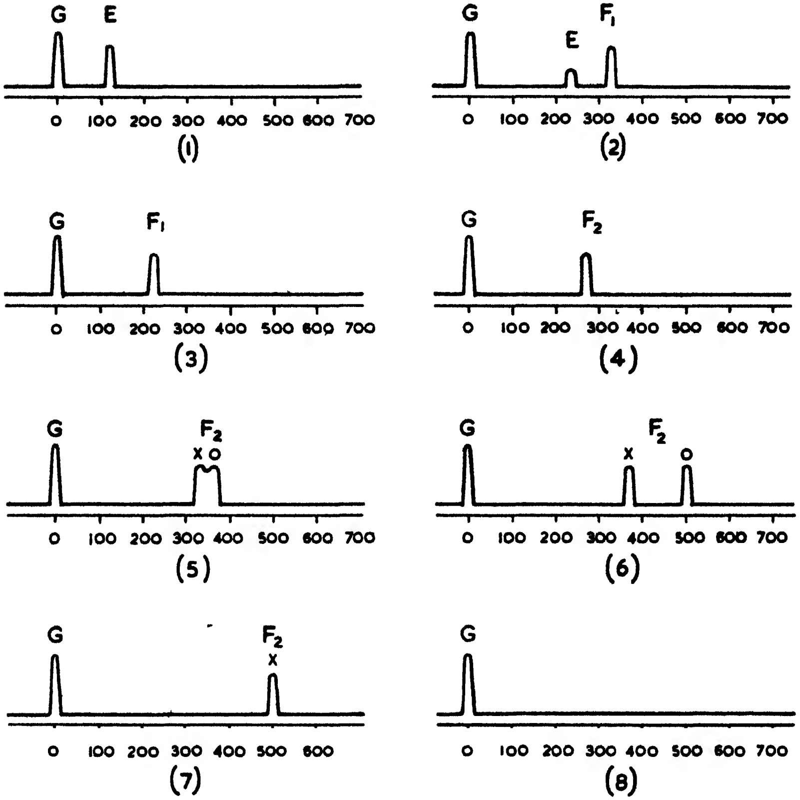


Fig. 17—Cathode-ray tube displays for several frequency points, as marked in Fig. 16

escape frequency of the E layer. We must remember that this retardation in the E will occur not only at frequencies just below, but also at those just above the E layer escape frequency, so the big upward curl of the E curve and the downward curl of the F1 curve are both due to this effect. We are, in fact, reading a falsely great height for the F1 layer at first, and it is not till we get away from the E layer escape frequency, and the retardation ceases, that the F1 layer measured height becomes nearer the true one. At 3·7 Mc/s, however, the measured height is nearer the true one for the F1 layer.

Continuing to increase the frequency and to plot the results in Fig. 16, we first of all get a slight increase in height—penetration into the layer—and then at about 4·2 Mc/s we get a decided kink in the curve. As we are supposed to be making our measurements during the winter day—see Figs. 7 and 8—there is very little difference in height between the F1 and F2, so that the wave goes from one layer to the other almost imperceptibly. Nevertheless, the kink in the curve has this definite meaning, and does enable the change from F1 to F2 to be detected. It is due to a repetition of the retardation phenomena which we have just described for the first reflections from the F1.

The pulses are now going up to the under side of the F2, the highest layer in the ionosphere. As we continue to increase the frequency the wave begins to penetrate slightly into the layer, so that at 6·0 Mc/s we have a display like that shown in Fig. 17 (4). The F2 echo continues moving slowly to the right of the screen until at about 9·2 Mc/s—Fig. 17 (5)—we notice a peculiar effect—the echo pulse is beginning to split in the centre so as to form two separate echoes for the one pulse being sent off. The separation between the two then increases, the right-hand echo moving rapidly to the right—indicating deeper penetration and more retardation—and the left-hand echo much more slowly.

We shall endeavour to explain this splitting of the ray into two later, but let us at once learn the names by which the two rays are known. That to the right of the trace is called the " ordinary ray " whilst that to the left is known as the " extraordinary ray," and at 10·1 Mc/s we have a display like that shown in Fig. 17 (6). At 10·2 Mc/s the ordinary ray echo disappears, indicating that it has penetrated right through the layer, and our plot of its behaviour is shown by the left-hand " prong " of the forked curve in Fig. 16. We are left with only the ground-ray pulse and the extraordinary ray echo, and a further increase in frequency causes that too to move rapidly to the right, so that at about 10·9 Mc/s the display is like that shown in Fig. 17 (7). The extraordinary ray echo now also disappears, indicating that the whole wave has passed the point of maximum electron density, and has penetrated through the layer.

Any further increase in frequency—such as that to 12·0 Mc/s shown in Fig. 17 (8)—does not produce a further echo, and we are left with the ground-ray pulse only. Our pulses have penetrated through the highest layer of the ionosphere and there is nothing further to return the energy to us.

EFFECT OF THE EARTH'S MAGNETIC FIELD

Now we come to a rather difficult matter—one whose complexities are such that we cannot hope to deal with them fully here. We will merely attempt to give an idea as to the reason for its occurrence. This matter is the splitting of the ray into ordinary and extraordinary components, which is seen to commence at about 9·2 Mc/s—Fig. 17 (5)—and which leads to the forking of the $h'-f$ curve of Fig. 16. It is due to the action of the earth's magnetic field.

When the wave is travelling in ordinary air, and is not setting up any electronic motion, the earth's magnetic field has no effect upon it.

But as soon as the wave sets up such movements in the ionosphere it begins to be affected by the field. For the field exercises a force upon the moving electrons, producing a sort of twisting effect upon the paths in which they vibrate ; and, because of its dependence upon the nature of the electronic motion, the wave itself is affected. The ray is split into two rays, each travelling with different speeds, requiring different electronic densities for their reflection and therefore reaching to different heights before reflection occurs. One of the rays is therefore delayed on the other, as will be apparent from Fig. 17 (6). Actually this effect occurs at all frequencies, but on the lower ones where the ray does not penetrate far into the layer the delay is so small that it cannot be seen upon the cathode-ray tube, and only becomes apparent at those frequencies where the ray penetrates deeply into the layer.

It can be shown that the earth's field causes the direction of the electric field of the wave to change in a very complicated way, this " polarisation " being different for the two components into which the wave is split. But perhaps we can best explain their different behaviour in another way.

ORDINARY AND EXTRAORDINARY RAYS

Suppose, in the case of our pulse wave, sent vertically upwards, that when it enters the ionosphere the electric field is acting so that the electrons are set vibrating in a direction exactly parallel to that of the earth's magnetic field. The field in such a case will have no effect upon them, and consequently its effect will not be apparent in the behaviour of the wave itself. The pulse will ascend until the magnitude of the electronic current is sufficient to reduce the group velocity to zero, and then it will commence to descend.

Suppose, now, that the electric field is acting so as to set the electrons vibrating in a direction *transverse* to that of the magnetic field. The earth's field will now have the maximum effect upon them—its twisting effect upon their paths will be at its greatest. And this twisting effect is equivalent to an increase in the strength of the electronic current itself, so that the wave is more affected than before. Its group velocity is reduced by a greater amount, it deviates more from its original path and it is completely reflected with a lesser density of electrons than before. It is therefore reflected lower down in the ionosphere than is the wave we first considered.

In practical cases—when the wave approaches the ionosphere with the direction of its electric field at an angle to that of the earth's field—the wave is resolved by the ionosphere into two separate components, each behaving differently and according to the general cases stated above. They become differently polarised, travel with different velocities, follow different paths and require different electronic densities to ensure their reflection. That behaving according to the first case is called the " ordinary " ray and its performance is represented by the left-hand " prong " of the fork in Fig. 16. That behaving according to the second case is the " extraordinary " ray and its behaviour is recorded in the right-hand " prong " of the fork. As will be seen, because the extraordinary ray requires a smaller electronic density for its reflection, after a frequency is reached at which the ordinary ray has penetrated the layer, echoes of the extraordinary are still received. As the frequency is further increased its behaviour follows closely that of the ordinary ray at lower frequencies, until it, too, penetrates the ionosphere layer. The difference in the frequencies at which the two rays penetrate through the layer depends upon the strength of the earth's magnetic field, and therefore varies somewhat at different locations on the earth's surface, being about 0·65 Mc/s in this country.

It should perhaps be remarked that in these practical cases the ordinary ray is not *truly* ordinary in the sense that it is entirely unaffected by the earth's magnetic field. Nevertheless, so far as its depth of penetration before reflection is concerned, it does behave as if there were no magnetic field and is therefore called the " ordinary " ray. And when we come to relate the frequencies for long-distance transmission to those observed by vertical incidence measurements we shall see that it is the ordinary ray escape frequency which is of most importance.

THE CRITICAL FREQUENCIES

Because the refractive index within the ionospheric layers depends upon both the frequency of the exploring wave and upon the electron density, waves of different frequencies ascend to different heights before being reflected. And as the frequency is gradually increased the wave ascends higher and higher until it reaches the region of maximum electron density, after which it penetrates through the layer because there is an insufficient electron density to ensure reflection of a wave of its particular frequency. The highest frequency returned from any layer is called the " critical " frequency for that layer, and is thus a measure of the maximum electron density existing within it, the maximum electron density being, in fact, proportional to the square of the critical frequency. When, after penetrating through a lower layer, we obtain reflections from a layer higher up in the atmosphere, it is only because the electron density in the higher layer (and thus its critical frequency) is higher than in the lower layer. If this were not so we should be unable to detect the presence of the higher layer by these means, because any frequency which penetrated the lower layer would automatically be too high to be reflected from the higher layer. As we shall see later, these critical frequencies are of the utmost importance, for they enable us to tell, from the results of our ionospheric measurements, what are the useful frequencies for long-distance short-wave transmission, and what are the frequencies which will be useless owing to penetration of the layers at oblique incidence. Thus we are enabled to put the data obtained from the measurements to practical use in short-wave communication.

The critical frequency of a layer is denoted by the symbol f, with a subscript such as $_E$ denoting the layer referred to, and a superscript denoting whether the ordinary or extraordinary ray is indicated. Thus we commonly have f_E, f^O_{F1}, f^X_{F1}, f^O_{F2} and f^X_{F2} for indicating the critical frequencies for the different layers and rays.

DATA OBTAINED FROM THE h′ - f CURVES

When the h′ - f curve is analysed, therefore, quite a lot of useful information is obtained from it, both from the practical point of view in its application to short-wave transmission problems, and also from the scientific point of view in giving information about the ionosphere itself and about its causative agency. It is from the former point of view that we shall look at things.

As has been indicated, the state of the ionosphere and therefore the form of the h′ - f curves will vary greatly, not only with time of day, season of the year, etc., but also with geographical latitude and longitude. So, in order to obtain a knowledge of the world-wide structure of the ionosphere—and it is obvious that that is what we require when transmitting over great distances—what we really need is a series of h′ - f curves obtained at different times of day and at different locations on the earth's surface. As a matter of fact there are a number of stations located in different parts of the world regularly engaged in making these curves, and it is from an analysis of their combined results that we obtain the world-wide information that we require in short-wave work.

Referring to the curve of Fig. 16—a curve such as might be obtained at a mid latitude in the northern hemisphere on a winter day—we may indicate some of the principal data which may be deduced from an inspection of it. These are—

Minimum frequency returned (f min)	1·6 Mc/s
Minimum virtual height E layer (h′E min)	115 km
Critical frequency E layer (f^O_E)	3·1 Mc/s
Minimum virtual height F1 layer (h′F1 min)	220 km
Critical frequency F1 layer (f^O_{F1})	4·2 Mc/s
Minimum virtual height F2 layer (h′F2 min)	250 km
Critical frequency F2 layer (ordinary ray) (f^O_{F2})	10·25 Mc/s
Critical frequency F2 layer (extraordinary ray (f^X_{F2})	10·9 Mc/s

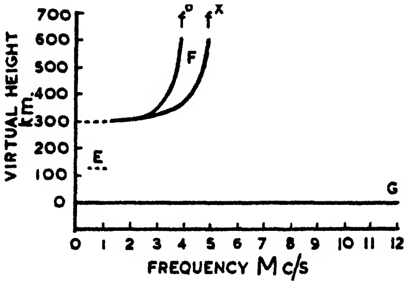


Fig. 18—An $h' - f$ curve obtained during the winter night

Night-time and Summer-time Curves

Fig. 18 depicts the sort of $h' - f$ curve we might have obtained had we made our measurements during the winter night. These $h' - f$ curves should be compared with the pictorial diagram of Fig. 7, and also with Fig. 8, in order to help explain some of the heights recorded.

In Fig. 18, although we can detect the presence of the E layer, most of the energy penetrates it and goes up to the F layer. There is, at night, no kink in the F layer curve, which shows that the F exists as a single layer instead of two separate ones as during the day. Furthermore, the F critical frequency—for the ordinary ray—is now only 3·7 Mc/s as against 10·25 Mc/s during the day. There is thus a large decrease in the highest frequency returned from the ionosphere as between day and night, which is what we might expect, observing that when the sun's rays are removed from the layer the free electron production ceases, and, the recombination process continuing, the number of electrons per unit volume in the layer is gradually reduced.

Had we made our measurements during the summer night the curve would have assumed much the same form as that shown in Fig. 18, except that the F critical frequencies would have been somewhat higher, because of the later occurrence of sunset in the summer than in the winter.

Fig. 19 shows the form the $h' - f$ curve would take on a summer day. Although this looks very complicated it really is not so, and a glance at Figs. 7 and 8 should soon help to make it clear. The E and F1 layers are seen to lie at about the same heights as shown by the winter day-time curve, but their critical frequencies are considerably higher than during winter, as we would expect from the fact that the sun is more nearly overhead in summer and its rays are therefore stronger. Furthermore, because the ray can penetrate deeply into the F1 before going up to the F2, both ordinary and extraordinary rays for the former layer are clearly seen. The F2 layer is at a considerably

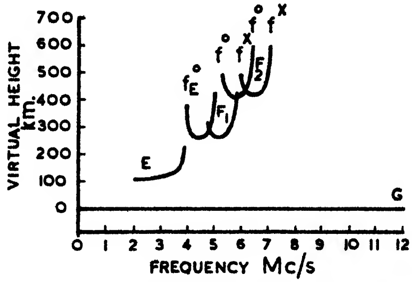


Fig. 19—An h′ - *f* curve for a typical summer day

greater height than during the winter day, and, *contrary* to what we should have expected, its critical frequency is only 6·5 Mc/s as against 10·25 Mc/s during the winter day. So far as its electron density is concerned it therefore behaves in an anomalous manner when the summer day increase in the sun's radiations is considered, and in a manner opposite to that of the two lower layers. But we shall hear more of this matter later on.

Modern Measuring Methods

Although the ionospheric measurements are often made manually, as we have hitherto assumed (and it will be seen that with suitable synchronisation and timing of the pulse rate and the sweep of the cathode-ray tube trace, a standing pattern will appear upon the screen), at most of the regular ionosphere observatories the transmission and reception is arranged to take place automatically, and the curves themselves to be automatically plotted by photographic means. An apparatus is arranged at the transmitter to send out the pulses at the correct rate, whilst the frequency is automatically increased over the necessary frequency range. The photographic film for reading the height is at the same time moved in suitable steps corresponding to the steps in the frequency increase. Thus the h′ - *f* curve is traced out photographically, the whole frequency range being covered in a few minutes or even, sometimes, in a few seconds. The apparatus is arranged so that it is automatically switched on and off, and thus repeats the whole process at suitable intervals, so that practically continuous recording is going on. An example of a curve obtained in this way is given in Fig. 20.

In the next chapter we shall examine the nature of the continual variations which occur in the virtual heights and critical frequencies of the ionospheric layers, and endeavour to explain the reasons for some of these.

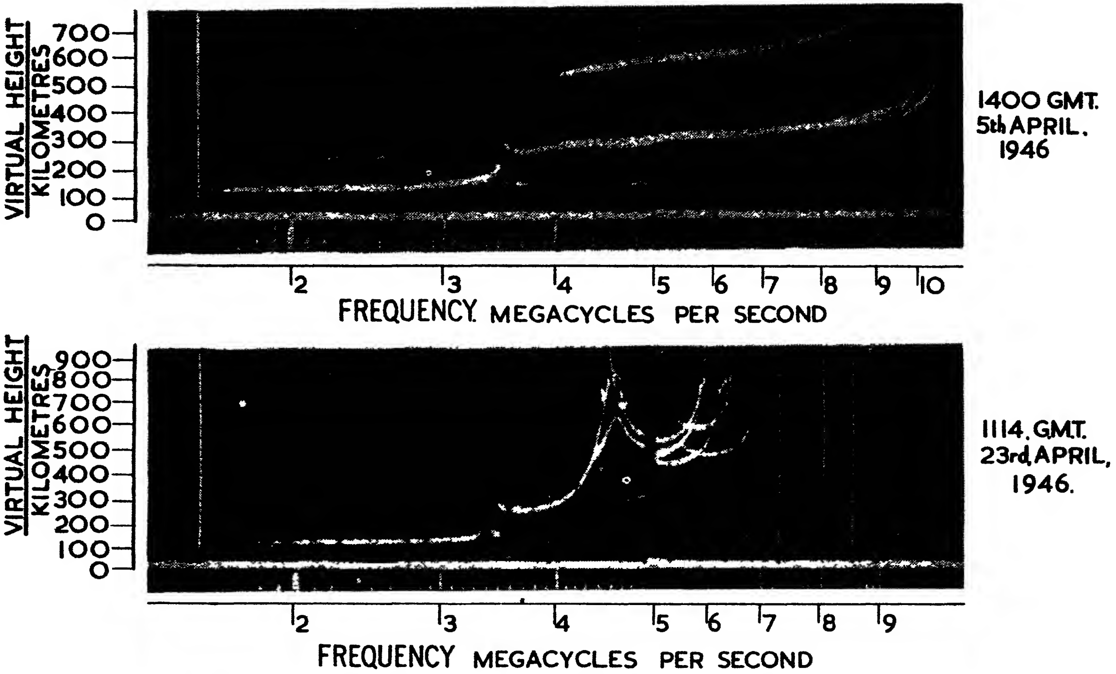


Fig. 20—Specimens of h′ - *f* curves obtained by photographic means. The upper photograph shows conditions on a "quiet" day and the lower while an ionospheric storm was in progress. (Photographs reproduced by courtesy of the Director, Radio Research Station, Slough)

CHAPTER 5

Ionospheric Variations

CAUSES OF VARIATION IN THE IONISATION

AS WE SAW FROM the h′ - *f* curves given in the last chapter, there are large changes in the critical frequencies of all the ionospheric layers as between day and night, and between summer and winter. We should, of course, expect large diurnal and seasonal changes in the ionisation of the upper atmosphere, because of the variations in the amount of solar radiation acting upon it. Thus, because of the constant movement of the earth in relation to the sun, the sun's zenithal angle at any point on earth varies with the time of day and with the seasons, and so the power of the ultra-violet radiation, in ionising the atmospheric gases, varies likewise. Thus we have more or less regular diurnal and seasonal components of variation in the critical frequencies, and in some cases in the virtual heights, of the layers.

If this were the only cause of variations in the ionisation these would be much more simple than they actually are, for at a given time on a given day in any one year we might expect the ionisation to be roughly similar at any terrestrial point to what it was at the same time and day of the previous year. But this is not the case—there is, in fact, a long period variation in the ionisation superimposed on the diurnal and seasonal variations. This variation, which is spread over about eleven years, is caused by a change in the activity of the sun itself, which lasts about that time. It had been known for many years that there *was* this periodic change in the sun's activity, for certain solar phenomena—notably the sunspots which appear on the sun's surface—had been found to vary in cycles lasting approximately eleven years, and certain terrestrial phenomena had been found to exhibit a marked tendency to vary in sympathy. There is, however, no marked variation in the intensity of the sun's *visible* light which arrives at the earth's surface, for this has been found to remain more or less constant throughout the sunspot cycles. It is evident, though, that the *ultra-violet* rays emitted from the sun must undergo very large variations in accordance with the cycles of solar activity, for the electronic density in all the ionospheric layers, which they produce, has been found to exhibit a remarkable correlation with the degree of sunspot activity.

When the ionospheric structure is viewed on a world-wide basis the effect of the diurnal, seasonal and sunspot cycle changes taken together becomes extremely complex, since when it is high summer in one hemisphere it is deep winter in the other, and when it is noon in one part of the world it is midnight in another. However, if we consider all these variations separately we shall, perhaps, obtain an idea as to their main effects in short-wave propagation—for, in short-wave communication, we are vitally interested in these changes.

The greater the strength of the ultra-violet radiation the greater is the free electron density in the layers, and the higher is their critical frequency. The higher the critical frequency the higher are the frequencies which the layers will refract when the radio wave strikes them at oblique incidence, as it will do when we send it out so as to communicate over a distance. If we ignore these variations and attempt to operate our short-wave stations on frequencies chosen haphazardly, and without regard to the ionosphere variations, the chances of our being able to maintain good communication will be very remote, for part of the time our waves will fail to arrive at the receiving point because they have penetrated through the ionosphere altogether, while at other times they will fail to do so because of complete absorption in the lower ionosphere. If, on the other hand, we carefully choose our working frequency to suit the conditions of ionisation prevailing at any time, then the wave will be properly refracted, and will travel to great distances with extraordinarily little loss of energy.

We can best study these variations in terms of the critical frequency, because it is easiest to relate this to the actual working frequencies for various distances. We might remember that the actual ionisation in the layer is proportional to the square of the critical frequency.

DIURNAL VARIATIONS

Turn again to Figs. 16, 18 and 19, and note how the critical frequencies and virtual heights varied as between day and night and between summer and winter. Suppose we have available one of these $h' - f$ curves for *every* hour of day, during a winter and a summer day. We could then plot the critical frequencies and virtual heights against time of day, and so obtain a diurnal characteristic curve of the ionosphere for each of these seasons. Actually we should obtain a more representative curve for these seasons if we took, not the measurements for a single day, but the mean of those obtained at each hour of day over a whole month. Such a set of curves is given in Fig. 21, the values shown being for the months of June and December during a year of minimum and one of maximum sunspot activity. We will concentrate upon the diurnal variations in critical frequency first.

Examining first the variations in E layer critical frequency over the day—all critical frequencies are those for the ordinary ray—we see that it varies in direct accordance with the altitude of the sun, increasing from a very low value at sunrise, reaching a diurnal maximum at noon, and then decreasing towards sunset. Since very low values of critical

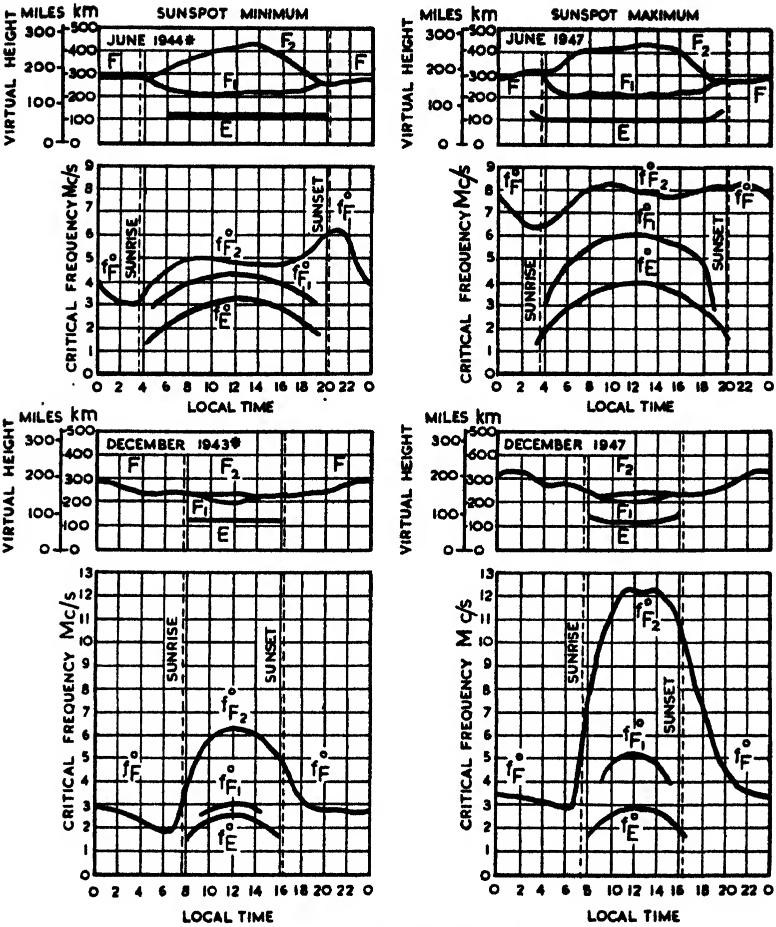


Fig. 21—Monthly graphs of critical frequencies and virtual heights observed at Slough during summer and winter of years of minimum and maximum sunspot activity

frequency are difficult to measure, that of the E is not shown in Fig. 21 when it is below about 1·5 Mc/s, though it is known to persist at night with lower critical frequency than this.

We must here digress for a moment to remind the reader that the free electron density in a layer at any time will depend, not only upon the rate of ion production, but also upon the rate of recombination of the electrons and ions. The density will only increase when the production rate is greater than the recombination rate. The recombination rate will depend, in the main, upon the density of the air itself, that is upon the number of atoms of gas per unit of space. For it will be clear that the more numerous the atoms are (that is, the less the distance between them), the greater will be the chance that a free electron will come into contact with an ionised particle and so recombine with it so as to re-establish the electrical neutrality of the structure. In the lower part of the ionosphere a high atom density does exist, and thus when the sun's rays diminish or cease to affect the layers, the ionisation rapidly decreases.

The layer therefore behaves, in general, according to the simple theory of ionisation by ultra-violet radiation by the sun, the ionisation being at a maximum when the ultra-violet radiation affecting it is greatest, as at noon, and falling away on either side of this time to a very low value. The critical frequency is, as a matter of fact, proportional to the cosine of the sun's zenithal angle, being given by—

$$f_E = \mathrm{K}\cos^{\frac{1}{4}} \chi$$

where χ is the zenithal angle of the sun and K is a factor depending on the intensity of the solar radiation.

It should be noted that according to this theory of ultra-violet ionisation the E layer should disappear entirely soon after sunset. In fact the critical frequency, after falling to about 0·6 Mc/s, seems to remain at that value throughout the night—there appears to be an agency other than ultra-violet light which maintains a low value of ionisation when the sun's rays are cut off. Sir EDWARD APPLETON has suggested that this agency is the small meteors which are continually arriving in the earth's atmosphere, and that the ionisation which these swiftly moving particles set up may maintain the " residual " night-time ionisation of the layer. It should be noted that this ionisation is so low as not to affect appreciably the propagation of short radio waves, though on long and medium wavelengths it would be of significance.

Examining the diurnal variations in the critical frequency of the F1 layer in Fig. 21, we find that it exhibits very similar characteristics to those of the E. It does, in fact, like the E layer, behave according to the theory of ionisation by ultra-violet light from the sun, and it closely obeys the $\cos^{\frac{1}{4}}\chi$ law. At night, of course, the F1 layer is non-existent, having merged with the F2 to form the single night-time F layer. As to the D layer, though it does not reflect waves at vertical incidence, it may be taken that its diurnal variations are similar to those of the E.

When we examine the diurnal variations of the F2 layer we see that it does not conform to the above law at all, for its critical frequency, particularly in summer, goes on increasing long after noon. Since this is the most important layer in short-wave propagation, and since its behaviour is more complex than that of the other two, we had better leave it for separate consideration in a few moments.

SEASONAL VARIATIONS

The seasonal variations in the critical frequency of the E and F1 layers are again simple and straightforward, the critical frequencies being higher in the summer than in the winter of any one year, and the diurnal maxima being greater in summer than in winter. This again is consistent with the ultra-violet ionisation theory, the critical frequency being proportional to the cosine of the sun's zenithal angle and thus reaching a peak in summer.

It may here be remarked that this rise in the critical frequencies of the E and F1 layers during the summer day is such that they become of much more importance in short-wave transmission at this time than they are at any other time during the year, often being of *prime* importance for transmission over certain distances. It will be noted also that the layers remain in existence for much longer periods in summer than in winter, as would be expected from the longer duration of the summer day.

The seasonal variations in the F2 layer are again seen to be anomalous in character, so we had now better consider that layer in more detail.

ANOMALOUS BEHAVIOUR OF THE F2 LAYER

Examining first the diurnal variations we see that the critical frequency increases from sunrise, but that it is usually higher during the hours after noon than at noon. In fact, during the summer there is an actual fall in the critical frequency around noon, and then it goes on *increasing* during the afternoon to reach a diurnal maximum around sunset There is a decrease after sunset which is rapid in winter but more gradual in summer, so that in the latter season the midnight value is but little lower than that for noon. Generally speaking there is a continued gradual decrease after midnight, though sometimes, during winter, there is a slight increase a considerable time *before* sunrise.

Turning to the seasonal variations, we notice at once a remarkable thing. Although the sun's radiations are much stronger in summer than in winter day-time, the critical frequency of the F2 is much lower during the summer day than during the winter day. During the winter night, however, it is lower than during the summer night, as would be expected. Though all these variations may seem complicated we may sum up their main features by saying—

(1) there is a diurnal variation lagging considerably behind the variation in the sun's altitude ; and

(2) there is a day-time seasonal variation of the opposite character to that of the other layers, and in opposite sense to that of the sun's altitude.

Though we cannot say for sure what are the reasons for this anomalous behaviour of the F2 the following is considered to be the most satisfactory explanation : In the first place, owing to the rarity of the gas atoms at this height and thus to the low recombination rate, the diurnal variation lags behind the altitude of the sun, and so the tendency is for maximum critical frequency not to occur till some time after noon. Then there is a temperature effect, according to which the uppermost part of the atmosphere expands greatly when heated by the sun, so that the effects are most apparent during the summer day. In an expanded atmosphere there are less atoms per unit of space than in a contracted atmosphere, and so the ionising radiation is unable to produce so many free electrons per unit of space, and thus the critical frequency is lower. During the summer day the F2 layer is more expanded than during the winter day and so its critical frequency is lower, in spite of the increase in ionising radiation. Again, during the summer day itself the critical frequency rises after sunrise as the ionising rays become stronger, but towards noon the gas expands, due to the heating effect, so that the critical frequency begins to fall. In the afternoon the gas begins to cool and to contract, so that the critical frequency rises to reach a peak just before sunset, after which the night-time decrease due to recombination sets in.

The pre-sunrise increase of critical frequency which often occurs in winter, but which is not apparent in Fig. 21, is accounted for on a similar basis. During the long winter night the gas cools very considerably and contracts during the process, so that the number of ions per unit of space is increased. Then, before dawn, there is a definite rise in the critical frequency, due, not to the beginning of free-electron production, but to the crowding of the free electrons already there into a smaller space. Apart from this pre-sunrise increase in the critical frequency the night-time critical frequencies are lower in winter than in summer. This is due to the shorter period of time during which the ionising radiation is operative in winter, the earlier onset of darkness—during which only recombination takes place—resulting in the critical frequency falling to very low values during the winter night.

Combining the diurnal and seasonal variations in F layer critical frequency, we see that it reaches its highest values during the winter day and its lowest during the winter night. The variation between day and night critical frequencies is therefore great in winter and relatively small in summer. The greatest rate of variation between day and night frequencies thus occurs in winter just after dawn and just after sunset.

It should be noted that in the ionosphere there is no lag in the seasonal effects such as occurs in the seasons of weather, which follow the sun's seasonal position a month or two later.

THE SUNSPOT CYCLE

As has been said, the activity of the sun itself varies over a long period, and one indication of this is in the size and number of the sunspots which appear upon it. These sunspots are only one of the indications of the variation in solar activity, though they are the most convenient of solar phenomena to observe. Such observations are regularly made at many astronomical observatories, and the information published in the form of " relative sunspot numbers." These are arrived at by taking the sum of the total number of sunspots observed plus ten times the number of spot groups, this sum being multiplied by a factor depending upon the telescope used and the viewing conditions. The observations from the different observatories are correlated by that at Zurich, and the final " number " published from there. It might be thought that this is a somewhat arbitrary method of assessing the sun's activity, but, in fact, it has proved a very faithful one, and the sunspot " number " has been found to possess a remarkable degree of correlation with certain terrestrial phenomena.

Records of this index of the solar activity go back for very many years ; in fact, they are continuous as far back as 1749. It has been found that there is a mean period of 11·1 years in the degree of solar activity as evidenced in this way—that is to say, a mean period of 11·1 years is occupied by a complete cycle of activity from minimum, through maximum, to minimum activity again. By this it is not meant that the sun's activity increases in a smooth and regular manner from minimum to maximum—there are in fact large and erratic variations from day to day. Nor are the cycles themselves at all regular, either in

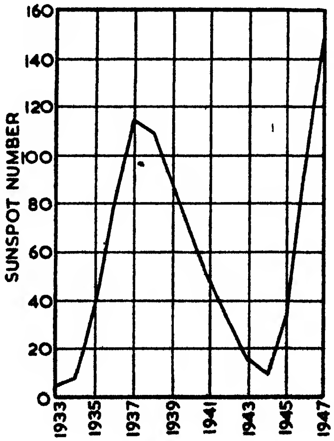


Fig. 22—Annual means of relative sunspot numbers

amplitude or in frequency—though the cyclic change in the average activity is markedly apparent. This is seen from Fig. 22, in which the annual means of the relative sunspot numbers are given for each year from 1933 to 1947, during which period roughly 2½ cycles of solar activity were completed, the minimum years being 1933 and 1944 and those of maximum activity 1937 and 1947. It is seen that the year-to-year changes in the sunspot activity have been markedly different in the two cycles and that the 1947 maximum was much higher than that of 1937. It is, in fact, impossible to predict from past cycles what changes will occur in the activity a very long way ahead, though predictions of the *general* activity several *months* ahead may be made with good accuracy.

Long Period Variations in the Ionisation

In Fig. 23 are given some (idealised) h′-*f* curves obtained at noon in December during consecutive years during which the sunspot activity was increasing. From these it is seen that the critical frequency of all the ionospheric layers increased in accordance with the degree of sunspot activity, though the amount of increase was different in the different layers, being greatest for the F2 and least for the E layer. Curves taken at night or at other seasons of the year will also show this increase, though again it has been found to be different at different times of day and seasons.

It is evident, therefore, that the variations in the sun's radiations of ultra-violet light which occur in accordance with the variation in sunspot activity give rise to large variations in the ionisation of the upper atmosphere. Ionospheric observations over two sunspot cycles show, in fact, that the electron density in the F2 layer is about four times as great at sunspot maximum as at sunspot minimum, and it is evident that such large changes will have to be taken account of in choosing the frequencies for use in short-wave communication.

Turning to Fig. 21 again, if we compare the two left-hand sets of curves (sunspot minimum) with the two right-hand sets (sunspot maximum) we can see the nature and order of the critical frequency changes which took place as between sunspot minimum and maximum. We see that the increase in critical frequency was greater for the F than for the other two layers, that it was greater during the day than during the night, and greater during winter day than during summer day.

The month-by-month variations in the relative sunspot numbers for the "increasing" phase of the current sunspot cycle are shown in Fig. 24, together with the monthly means of the noon F2 critical frequency as measured at Slough. It is seen that the sunspot numbers vary erratically from month to month but that there is a steep general

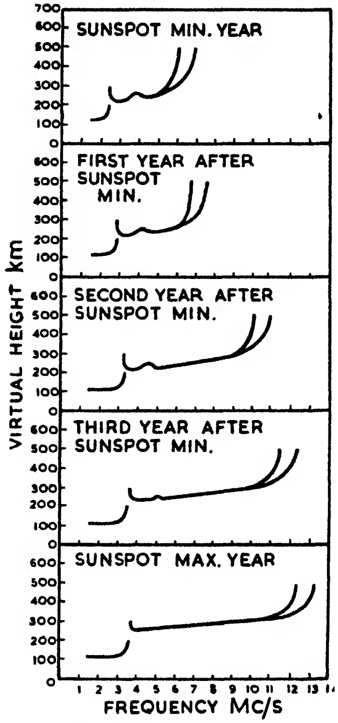


Fig. 23—h′ - *f* curves for noon in December during consecutive years between sunspot minimum and maximum

rise in their values. In the critical frequency curves we have, of course, the large seasonal changes superimposed on these due to the changes in solar activity. These are such as to give high values of critical frequency in winter and low values in summer, with the exception that there is usually, at the extreme mid-winter period, a secondary decrease, so that the highest values actually occur about November and February. Nevertheless the general rise in critical frequency in sympathy with the increase in sunspot number is clearly seen, showing how the ionosphere responded to the changes in the activity of its producing agent, the sun.

This is still more clearly seen, however, in Fig. 25, which gives twelve-month running averages of the monthly sunspot numbers, together with twelve-month running averages of the noon and midnight F, and of the E, critical frequency, as measured in England. The object of taking twelve-month running averages is to smooth out the temporary fluctuations in the sunspot numbers and the seasonal effects in the critical frequency values, so that the long-period effects in both quantities may be more clearly seen. It is done by taking for the mean for the epoch at the centre of any month the average of the twelve monthly means having that month as the centre.

It is seen that there is exceptionally good correlation between the curves for sunspot activity and those for critical frequency when considered in this way. From the curves it will be seen that the twelve-month running averages of the E critical frequency followed those of the sunspot numbers very closely indeed, though there was a slight delay in reversal tendency at the maximum. Those for the F and F2 layers also followed the sunspot numbers closely, except that towards the maximum the critical frequency failed to increase so rapidly as the sunspot numbers, as if there were some saturation effect in the ionisation of the F layer. Also the times of reversal of the trends in the critical frequency values for the F layer do not correlate so well with those for the sunspot numbers, that at the maximum lagging about three months behind the sunspot number. Nevertheless over most of the cycle the correlation between the curves for sunspot numbers and those for critical frequency of all the layers was remarkably good.

It is interesting to note that over the period shown the noon E layer critical frequency increased by 1·26 times and that for the noon F2 layer by 2·04 times, implying increases in the actual ionisation of the E layer of 1·59 times and in that of the F2 layer of 4·16 times, as between sunspot minimum and maximum. It is also very interesting to note that although the increase in sunspot numbers was considerably greater in the present cycle than in the past one, the increase in the critical frequency of the F layer was not much different in the two cycles.

It will be seen that if the latest trends in the twelve-month running average values are taken into account it is possible to forecast what the value will be several months ahead with good accuracy. Knowing this it is then possible to exterpolate the running average critical frequency curves out to a similar time. If then the known diurnal and seasonal factors in the critical frequency variations are taken into account it is possible to predict the average critical frequency values for each layer for each hour of day for the month in question, and so to forecast the conditions for short-wave communication.

GEOGRAPHICAL VARIATIONS

It is obvious that the ionisation of the ionospheric layers will vary considerably with geographic latitude and longitude, and that in conducting world-wide short-wave transmissions we shall have to take this into account. At any one time of day the sun's zenithal angle will, of course, vary with geographic latitude, and the amount of ionising radiation entering the atmosphere will vary accordingly. As we approach the equator, where the sun is more directly overhead, we

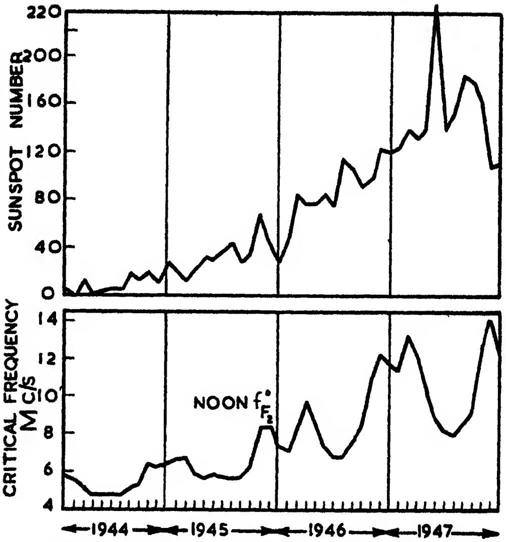


Fig. 24—Monthly means of relative sunspot numbers and of noon F_2 critical frequency at Slough during the "increasing" phase of a sunspot cycle. (Sunspot numbers for 1947 are provisional only)

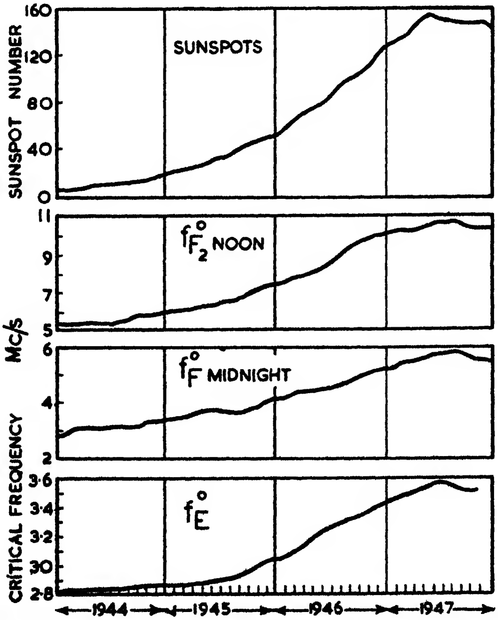


Fig. 25—Twelve-month running averages of critical frequencies and sunspot numbers

should expect the critical frequency of the layers to increase, and then to decrease again in the southern hemisphere. So far as the E and F1 layers are concerned this expectation is well borne out, and their critical frequencies do follow the sun's angle. At noon at the equinoxes, for example, they are highest at the equator and have a similar value for similar latitudes in the northern and southern hemispheres. At noon during other seasons they are highest at that point north or south of the equator where the sun is overhead at noon, and decrease proportionately as one proceeds north or south from this point. The seasonal variations are, therefore, similar for similar latitudes in the northern and southern hemispheres, but six months " out of phase " in the two hemispheres.

As to the variations in the E and F1 critical frequencies with geographic longitude, they are simply those diurnal variations which we have already discussed. In other words, as the earth turns upon its axis, the diurnal variations which occur at any terrestrial point are repeated all along the same parallel of latitude, so that along this parallel similar values of critical frequency occur at similar values of local time.

The variations in F2 critical frequency with latitude and longitude are much more complex, and it seems that the reason for this may be that, whilst the larger part of the ionisation is undoubtedly due to the sun's ultra-violet radiation, some of it may be due to bombardment of the gas atoms by corpuscles, some of which may arrive from the sun, and some, possibly, from points outside the solar system altogether, such, for example, as the Galaxy. However that may be, there seems to be a component in the ionisation which varies independently of the sun's zenithal angle, and which, moreover, is affected to some extent by the earth's magnetic field, particularly so in the auroral regions.

In order to examine the way in which the critical frequency varies with latitude one should examine the observations made by stations located on one meridian of longitude but at different latitudes, their observations being made at similar values of local time. It is difficult however, to do this, because the observing stations are very rarely located along a given meridian, but Fig. 26 gives the observations made for June and December at three stations as near as possible to one meridian of longitude, two being at similar latitudes in the northern and southern hemispheres, and one being near the geographic equator. During June the diurnal characteristic for the northern hemisphere station is the typical summer one which we have already discussed. Near the equator the critical frequency, as would be expected, is much higher, both by day and night. Moreover the diurnal characteristic assumes a somewhat different shape, as is usual for stations in equatorial latitudes, there being a tendency for the critical frequency to remain

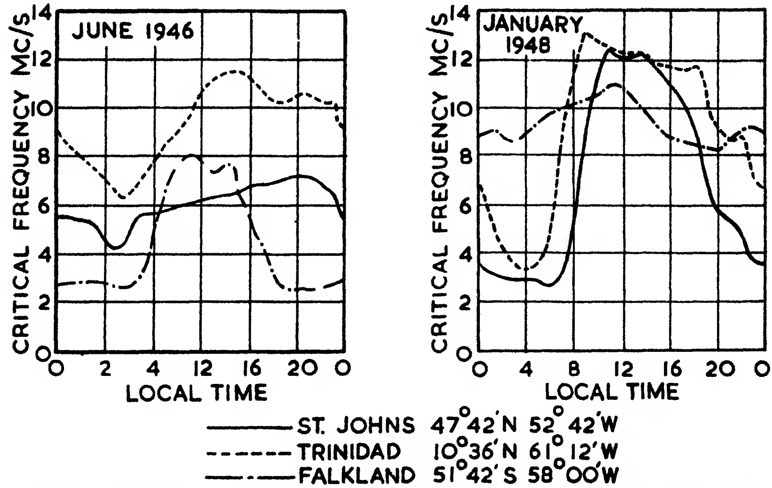


Fig. 26—Variation in F, F_2 critical frequency with latitude at similar times of day

at a relatively high level, or even to increase, for some time after sunset, and then to decrease sharply from midnight to just before sunrise.

At the southern hemisphere station the diurnal characteristic is a typical winter one, giving lower night values and higher midday values than for the northern hemisphere station. The midday values are, however, not so high as is usual during winter in the northern hemisphere. During December the diurnal characteristic for the northern hemisphere station is a typical winter one, having low night-time and high day-time values of critical frequency. Near the equator the characteristic is such that, generally speaking, higher values occur at all times except during the midday period, when the critical frequency is lower than for the higher latitude station. This, it is thought, is due to the temperature effect which we have already discussed, the upper atmosphere near the equator being in a more expanded state than in higher northern latitudes.

As to the December diurnal characteristic for the southern hemisphere station, it is of such a peculiar shape and so unlike that for the northern hemisphere station in local summer that its variations are difficult to account for. It will be noted, however, that the night-time values are much higher and those for day-time considerably lower than is the case for the northern hemisphere station.

From the above it will be gathered that the variations in F, F2 critical frequency with latitude are, in general, rather complex. In order to obtain a world-wide picture of the ionospheric variations with latitude, curves like Fig. 26A are drawn, which show (idealised) critical

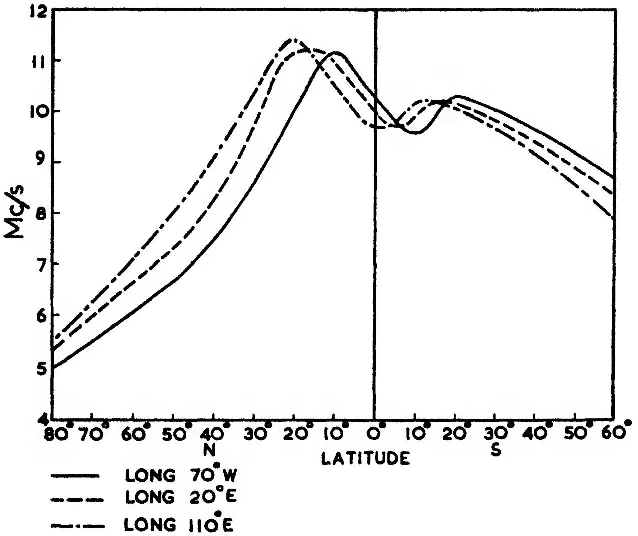


Fig. 26A—Variation in F_2 critical frequency with latitude at noon in June

frequency variations with latitude at noon in June along different meridians of longitude.

The Longitude Effect

Since the sun's zenithal angle will have similar values along any parallel of latitude at similar values of local time, as the earth turns upon its axis, it would be logical to assume that the ionospheric critical frequencies in any one latitude would be similar for similar values of local time. As has been said, this is the case for the E and F1 layers, but ionospheric observations have shown that it is not so in the case of the F and F2. In Fig. 27 are given the monthly mean values of F, F2 critical frequencies for June, 1947, for two pairs of stations, the stations of each pair being located in very similar latitudes but in widely different longitudes. It is seen that though the general form of the curves in each pair is similar, there are large differences in the values of critical frequency recorded at similar values of local time. It will be noticed that the lower values in each pair were obtained at stations in the western zone of the northern hemisphere (Portage la Prairie and Baton Rouge) and the higher values at stations in the eastern zone (Slough and Delhi). Now although the geographic latitudes of Slough and Portage la Prairie and of Delhi and Baton Rouge are very similar, their geomagnetic latitudes are widely different, and it has been found that this is the reason for the difference in their critical frequency values. The general effect is seen in Fig. 26A.

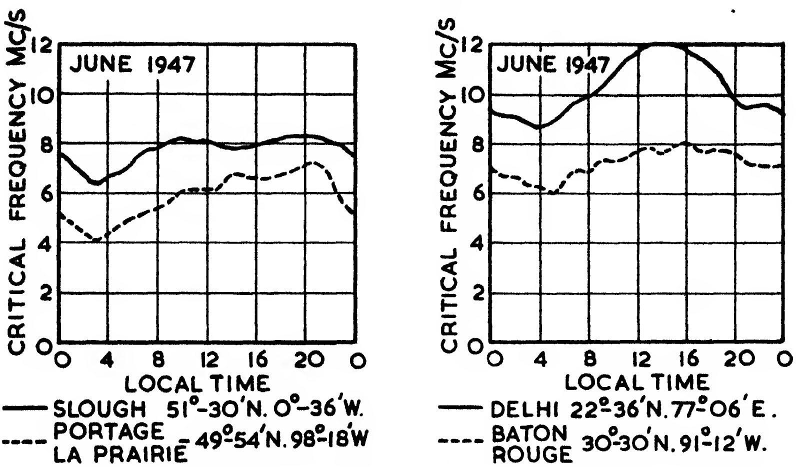


Fig. 27—Variation in F, F_2 critical frequency with longitude at similar times of day

In fact, the F and F2 critical frequencies vary, not only according to geographic latitude and longitude but according to magnetic latitude and longitude as well. Thus exactly similar values of critical frequency along a parallel of geographic latitude would only occur at similar values of local time where the geomagnetic latitudes were also similar. Also, whilst the critical frequency does increase towards the low latitudes, the highest values do not occur in equatorial regions, for there appears to be a trough which runs along the geomagnetic equator. Whether or not these conditions are due to the effect of the earth's magnetic field upon the corpuscular ionising agency, or whether they are due to the effect of the field upon the free electrons already existing, is not known, but this " longitude effect " in the F layer ionisation is itself quite marked and well established. It is such as to produce relatively high values of critical frequency in relatively low magnetic latitudes and vice versa, so that places in Europe would, for example, have higher values than those in Eastern America with a similar geographic latitude, and places in South Africa lower values than those in eastern South America with a similar geographic latitude.

As we shall see later, this longitude effect in the ionisation has to be taken into account in world-wide short-wave transmission operations.

CHANGES IN VIRTUAL HEIGHT

Fig. 21 shows that the minimum virtual height of the E and F1 layers is practically constant at all hours of day during which they exist, and this is so also at all seasons of the year and epochs of the sunspot cycle.

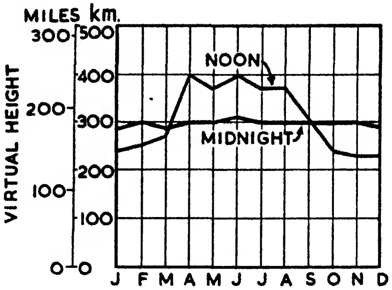


Fig. 28—Monthly mean of noon and midnight virtual height for F and F_2 layers

The day-time F2, on the other hand, lies at a much greater height during the summer day than during the winter day. This variation in height is probably connected with the expansion of the gas during the summer day because of heating. As will be seen from Fig. 21 the night virtual height is usually in the vicinity of 300 km both in winter and in summer. In winter there is a considerable decrease before sunrise, the layer remaining during the day at a lower level than during the night. In summer, however, there is a large increase in virtual height after sunrise, and the layer during the day is at a far greater height than at night. The layer returns to its night-time level in summer before, and in winter after, sunset. Fig. 28 gives the minimum virtual heights for each month of a year, and thus shows the seasonal variation in this quantity The geographical variation in virtual height is such that lower values occur in low than in high latitudes.

As we shall see later, changes in the virtual height affect the frequencies to be used for long-distance short-wave working, as well as do the variations of critical frequency. It may here be mentioned that, in general, small virtual heights and high critical frequencies lead to higher working frequencies, whilst great virtual heights and low critical frequencies necessitate the use of low working frequencies.

Summary of the Ionospheric Variations

Now let us summarise these rather complicated ionospheric variations, in order to clarify the information we have obtained—

(1) The critical frequency of the E and F1 layers varies diurnally, seasonally and geographically according to the variations in the zenithal angle of the sun.

(2) The critical frequency of the F2 layer has a diurnal variation lagging somewhat behind that of the sun's zenithal angle. It is also subject, both diurnally and seasonally, to a temperature effect which tends to produce lower values of critical frequency

around noon during the summer day than during the hours before and after noon, and much lower values during the summer day than during the winter day. Its night-time critical frequency is lower in winter than in summer.

(3) Probably owing to a corpuscular component in its ionisation the F2 critical frequency varies with geographic latitude in a complex manner, and with longitude according to the geomagnetic, as well as the geographic, values.

(4) The critical frequencies of all the layers, both by day and night, summer and winter, vary directly with the 11·1 year cycle in sunspot activity.

(5) The virtual heights of the E and F1 layers are more or less constant during the daily period when they exist, at all seasons and epochs in the sunspot cycle. This applies also to the night-time F layer. The virtual height of the F2 layer is subject to variations, being lower than the night-time F during the winter day and much higher during the summer day.

In the next chapter we shall go on to see how the critical frequency values, and their variations, affect the working frequencies of use for short-wave communication.

SHORT-WAVE AERIAL ARRAYS

Aerial arrays at the B.B.C. short-wave station at Woofferton, Shropshire. The array on the left is for 17 metres and that on the right for 14 metres. Each array consists of four sets of 4 horizontal half-wave radiators stacked one above the other at half-wave intervals. The bottom horizontal elements are one wavelength above ground. Behind each radiating curtain is a parasitically excited reflecting curtain. Either curtain may be used as radiator or reflector. The direction of maximum radiation is at right-angles to that in which the horizontal elements lie, i.e., either towards or away from the observer, depending on which curtain is being fed with power. The transmission lines conveying power from transmitter to aerial array may be seen towards the left of the photograph. (*Photograph by courtesy of the B B C*)

CHAPTER 6

Short-wave Transmission

OBLIQUELY INCIDENT RAYS

IN CHAPTER 3 WE DISCUSSED the behaviour of a radio wave in the ionosphere, and saw that with a given electronic density the refracting power of a layer varies inversely as the square of the frequency (or directly as the square of the wavelength). Also that the greater the angle of incidence the less is the amount of refraction required, the less far does a wave of given frequency penetrate into the layer, and the lower is the electronic density required to return it to earth.

In Chapter 4 we saw how the ionosphere characteristics were measured for vertical incidence, and in Chapter 5 we described some of the changes in the ionisation which have been disclosed by such systematic measurements.

We now wish to know how we can put all this information to use in practical short-wave communication—how we can utilise it in ensuring the efficiency of our communications, and in the planning of our future operations. For it is obvious that to use to the best advantage such a complex structure as the world-enveloping ionosphere we must arrange to be kept constantly conversant with the nature and order of the changes taking place, and have ready a relatively easy technique for applying this information to our short-wave operations. Furthermore, if we are to plan our operations some way ahead—as is practically always necessary—we must be able to *anticipate* the ionospheric variations, and so to obtain a world-wide picture of the structure as it will be at some future date. Such techniques and prediction methods have now been developed and tested, with very good results, and in fact large-scale short-wave operations are nowadays based upon them, though this is not to say that they are yet perfect. Apart from such large-scale operations it is obvious that *all* users of the ionosphere—whether of professional or amateur standing—will do much better if they conduct their operations in the light of the latest ionospheric information than if they perform them in a completely haphazard manner.

In practical short-wave transmission we do not radiate equal amounts of energy at all horizontal and vertical angles, as was mentioned in Chapter 1 and pictured in Fig. 2 (page 14), but, by a suitable arrangement of the aerial system, we concentrate the radiated energy in the form of a " beam," which can be arranged to cover only the most useful angles, both horizontal and vertical. The advantages of this, so far as the horizontal angle is concerned, are immediately apparent, for the available energy may all be sent out in those azimuthal directions in which the distant receivers lie, and none of it wasted in undesired directions. It should also be apparent that, since the wave will only reach the receiving location after reflection from the ionosphere, it will be advantageous to concentrate the energy in the vertical plane as well, so as to direct it upon that point in the ionosphere where a ray must undergo reflection if it is to reach the earth again at the receiving location. This, in the case of single-hop transmission, is a point in the ionosphere half-way between transmitter and receiver.

It is not intended here to go into the theory of aerial directivity, but we will merely say that so far as vertical directivity is concerned it may be achieved by taking advantage of the energy which is reflected upwards from the ground beneath the transmitting aerial. By suitably positioning the aerial as to its height above ground, this reflected energy can be arranged to reinforce that being radiated by the aerial itself, so that all sorts of vertical radiation patterns can be produced.

In order to illustrate the principle (and for comparison with Fig. 2), the radiation patterns of Fig. 29 are given, which are for a horizontal aerial half a wavelength long, placed at different heights above the ground. By arranging additional half-wave aerials one above the other so that their radiated energies combine in a certain vertical direction greater directivity still can be achieved, and the energy may be concentrated in a relatively narrow " beam " at any angle to the horizontal. Also, by placing a reflecting curtain behind the aerial the " backward " radiation may be almost eliminated and the energy concentrated in one " forward " direction. There also exist special forms of aerial comprising only a few half-wave elements and suitable for amateur use, which give a good " gain " in both horizontal and vertical planes. The angle at which the radiated energy is sent out, measured from the horizontal, is called the " elevation " angle.

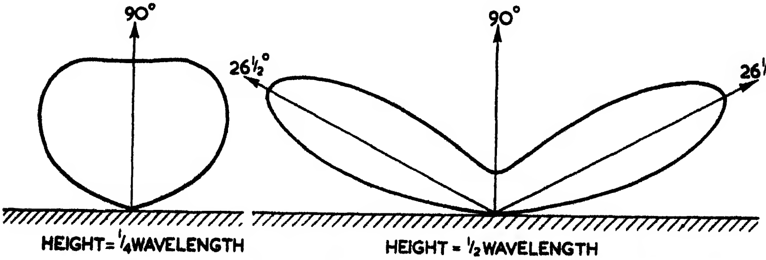


Fig. 29—Sky wave radiation patterns of a horizontal half-wave aerial placed at different heights above ground

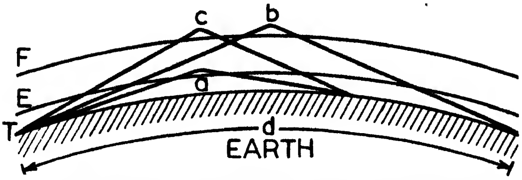


Fig. 30—Illustrating the geometry of one-hop ionospheric reflections

Fig. 30 will help to make clear the geometry of one-hop transmission via the ionosphere. In it we will suppose that E is the E layer and F the F layer of the ionosphere, whilst T is the transmitting station. As the elevation angle is reduced from the vertical (90°) the distance d (over which the ray travels) increases, but it will be clear that because of the curvature of the earth there is a definite limit to the distance that can be covered in one hop, depending on the height at which the reflecting layer lies. This limiting distance corresponds to an elevation angle of 0° (horizontal transmission), though in practice it is almost useless to transmit at a smaller elevation angle than about 2½°, because if this is done most of the energy is absorbed in the ground near the transmitter. The maximum distance for one-hop transmission via the E layer (corresponding to 0° elevation angle) is about 1,400 miles, via the F1 layer about 1,900 miles and for the F, F2 layer about 2,500 miles, depending on its height.

From Fig. 30 two other important facts should be noted—

(1) for a given layer height the greater the distance ; and

(2) for a given distance the lower the layer, the greater will be the angle of incidence on the layer.

Thus, for two rays travelling via the same layer, that of which the apex of the trajectory is at *b* has a greater angle of incidence than that of which the apex of the trajectory is at *c* in Fig. 30. Also, for two rays traversing the same distance, that which travels via the lower layer as at *a* has a greater angle of incidence than that which travels via the higher layer as at *c*.

RELATION BETWEEN VERTICALLY AND OBLIQUELY INCIDENT RAYS

We saw in Chapter 3 that a ray of radio energy incident on an ionospheric layer will penetrate into the layer until the refractive index is reduced to a value equal to the sine of the angle of incidence, and that it will then start travelling downwards again. Thus it will penetrate furthest into the layer when the angle of incidence is 0° (vertical incidence), and less and less as the angle of incidence is progressively increased, that is, as the distance *d* is made larger. If, however, the frequency is altered so that it is allowed to penetrate in each case to the point of critical electron density (minimum refractive index), then it means that higher and higher frequencies can be used as the angle of incidence is increased, and the distance over which it is desired to communicate made larger. Thus the highest frequency which can be used for any distance is related to the critical frequency at vertical incidence and to the angle of incidence.

The highest frequency for any distance is called the Maximum Usable Frequency (MUF) for that distance, and for a flat ionosphere this would be equal to the critical frequency multiplied by the secant of the angle of incidence (the secant is 1·0 at 0° and about 5·0 at the greatest angle of incidence it is possible to make in practice).

The curvature of the earth and ionosphere introduces modifications of a complex nature into this law, and we cannot go into them here. We need only note that a layer of given electron density returns frequencies which are higher the greater the angle of incidence on the layer, and thus, so far as the curvature permits, higher frequencies will be returned the greater the distance of transmission and the lower the reflecting layer lies. The relation between the critical frequency and the MUF for any distance may be described in terms of factors by which the critical frequency must be multiplied in order to obtain the MUF. These are called the MUF factors.

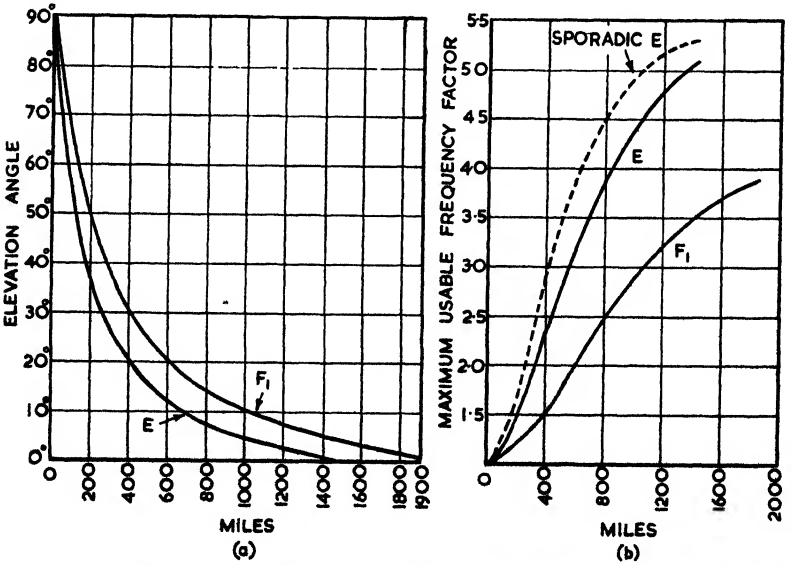


Fig. 31—Elevation angles and MUF factors for transmission over various distances by way of the E and F_1 layers, with (b) MUF factors for transmission by way of Sporadic E

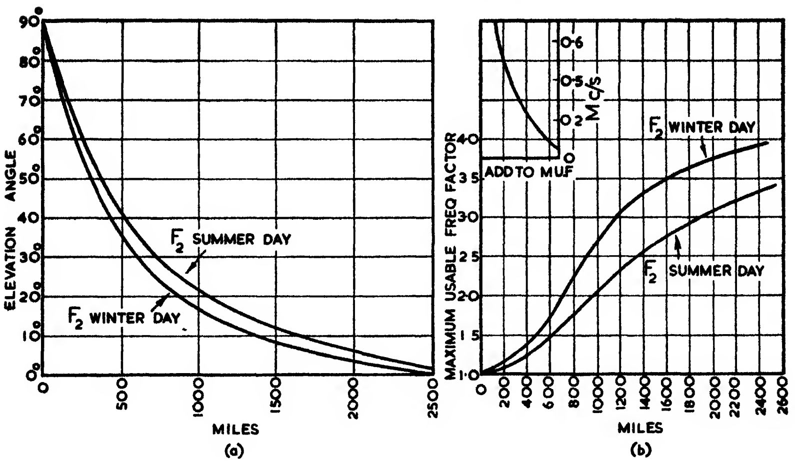


Fig. 32—Elevation angles and MUF factors for transmission over various distances by way of the day-time F_2 layer, during winter and summer

In Fig. 31 (*a*) are given the elevation angles for transmission over various distances out to the maximum distance it is possible to cover in one hop, for transmission by way of the E or F1 layers, while in Fig. 31 (*b*) are given the MUF factors for transmission by way of these layers, by which the vertical incidence critical frequency should be multiplied in order to obtain the MUF appropriate to the distance. Since these layers always lie at an approximately constant height the elevation angles and MUF factors are valid for all times and seasons during which the layers exist. Because the E layer lies at a smaller height than the F1 the elevation angles for transmission over any distance by way of it are smaller, and the MUF factors greater, than is the case for transmission by way of the F1. The factors for transmission by "sporadic E" shown in Fig. 31 (*b*) will be explained later.

In Fig. 32 (*a*) are given the elevation angles for transmission by way of the daytime F2, during winter and summer, while in Fig. 32 (*b*) are given the MUF factors appropriate to this layer during these two seasons. Since the layer is at a smaller height during the winter than during the summer day the elevation angles are smaller and the MUF factors are greater during winter than during summer. The elevation angles and MUF factors appropriate to other seasons may be found by interpolation, it being assumed that the cases shown are for the months of June and December respectively.

The critical frequencies to be used in conjunction with the MUF factors in order to obtain the MUF for any distance are those for the ordinary ray. For very short distance transmission via the F, F2 layers it is, however, usual to make a correction to the MUF so obtained to allow for the effect of the extraordinary ray, which will provide reception at these short distances. As the distance over which communication is effected is increased, that is, as the angle of incidence is increased, the separation between the ordinary and extraordinary rays decreases, until, when the distance is about 600 miles, it becomes so small as to be of no account in practice. The curve at the top left corner of Fig. 32 (*b*) gives the value in Mc/s which should be added to the MUF obtained from the ordinary ray critical frequency, in order to take account of the effect of the extraordinary ray at these short distances.

In Fig. 33(*a*) are given the elevation angles for transmission by way of the night-time F layer, throughout the year, while in Fig. 32 (*b*) are given the MUF factors appropriate to this layer. It will be remembered that the layer remains at an approximately constant height throughout the year. The values given in Figs. 32 and 33 for the F2 and F layers are respectively for noon and for midnight. For other hours of day the appropriate figures may be found by interpolation between these two values, and a study of the variation in virtual height with time of day shown in Fig. 21 should help in this matter.

The MUF factors for the E and F1 layers remain the same throughout the sunspot cycle, but in the case of the F and F2 layers there appears to be some increase in the thickness of the layers towards sunspot maximum. The values given in Figs. 32 and 33 are those for sunspot minimum, and the corresponding values for sunspot maximum are about 10 per cent smaller.

MAXIMUM USABLE FREQUENCIES

Although it is obviously not possible here to give MUF factors which will take into account all the variations in layer height and thickness, those given in Figs. 31, 32 and 33 are near enough for most practical purposes, if interpolation for times of day and seasons not given is made. If therefore we have available curves like those of Fig. 21, showing the average critical frequency at all times of day, and we wish to ascertain the MUF for transmission over any distance, then we read off the critical frequency for the appropriate time *at the centre of the transmission path*, and multiply it by the appropriate factor obtained from the curves. This will give the MUF for any distance up to 2,500 miles, beyond which the transmission will be by multiple hops, which case will be dealt with later.

It is useful to have available a set of curves from which one can read off the MUF directly, and these can be prepared by calculating the MUF for each hour of day for various distances. Fig. 34 shows such a set of MUF curves, obtained from the critical frequency values given in Fig. 21, and by the use of factors obtained from Figs. 31, 32 and 33.

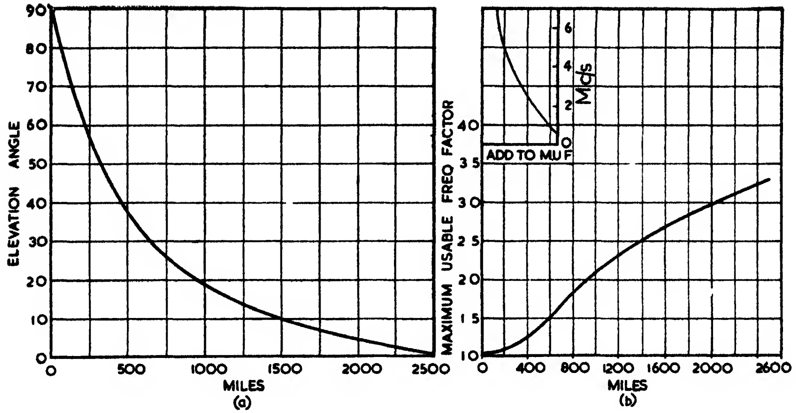


Fig. 33—Elevation angles and MUF factors for transmission over various distances by way of the night-time F layer, throughout the year

From these we can see at a glance what is the highest frequency for one-hop transmission over any distance out to the limits for such one-hop transmission.

It is interesting to study Fig. 34 with a view to ascertaining the *range* of frequencies usable for short-wave transmission at a mid-latitude in the northern hemisphere, remembering that the MUFs shown, since they are calculated from monthly averages of critical frequency, represent the *average* values usable during winter and summer of sunspot minimum and maximum years. The lowest MUFs shown, that is, those for 0 miles, are, of course, the critical frequencies for the extraordinary ray. As the distance increases the MUFs get greater at all times of day. The lowest MUFs occur shortly before sunrise, and we see that for longest distance transmission these range from about 6 Mc/s in winter to about 10 Mc/s in summer of a sunspot minimum year, and from about 9 Mc/s in winter to about 18 Mc/s in summer of a sunspot maximum year. At noon the MUFs for longest distance transmission range from about 16 Mc/s in summer to about 24 Mc/s in winter during a sunspot minimum year, and from about 24 Mc/s in summer to no less than 44 Mc/s in winter during a sunspot maximum year. Thus the total range of usable frequencies for longest distance transmission, having regard to time of day, season of the year and epoch of the sunspot cycle, is from about 6 Mc/s to about 44 Mc/s.

As we shall later see, the MUF represents a frequency which is not only usable, but *near* which the best communication will be achieved, so that the frequencies shown in Fig. 34 for the various times of day, seasons of year and epochs of the sunspot cycle are of great significance in ensuring maintenance of short wave communication over a long period.

EFFECT OF THE LOWER LAYERS

Transmission on short waves is usually by way of the F or F2 layers, though, of course, if a frequency very *far* below the MUF is used transmission may be by way of one of the lower layers. As has been indicated, however, it is bad practice to use a frequency *very* far below the MUF for long-distance transmission, because such a frequency will be heavily absorbed. As has already been noted from Fig. 21, the critical frequency of the lower layers is always much below that of the F2. It should be appreciated, however, that, because of the smaller virtual height, the E layer may sometimes have a greater MUF than the F1 or F2, and the F1 than the F2 layer. Around noon in summer, when the F2 ionisation is exceptionally low and that of the E exceptionally high, over a certain range of elevation angles the E layer critical frequency and the angle of incidence the ray makes on the E *determine the MUF*, and any frequency which is high enough to penetrate the E at this elevation angle also penetrates the upper layers. It will not occur for the shortest distances because in these cases the angle of incidence at the E is small, and there is thus more tendency for the wave to penetrate the layer. Then at an angle corresponding to a certain distance it will occur, and will continue out to about 1,400 miles, the maximum distance possible for one-hop transmission by the E.

Similarly, beyond this distance, the F1 may control the MUF out to a distance of about 1,900 miles, the maximum distance possible for one-hop transmission by the F1. Beyond this distance the F2 layer MUF will still continue to rise because the angle of incidence at that layer still increases with increasing transmission distance, and so the MUF is again controlled by the F2, out to the limit of distance for transmission by that layer.

In Fig. 34 it will be noticed that in the MUF curves for June there occur " humps " in the curves for 500, 1,000 and 1,500 miles, which peak around noon. It is during the periods covered by these " humps " that the lower layers control the MUF. In the afternoon, it will be remembered, the ionisation of the E and F1 layer falls, whilst in the summer that of the F2 continues to rise, so that towards evening the F2 again controls the MUF for all distances. In calculating the MUF for any distance during these times, it is therefore necessary to multiply the critical frequencies of each layer by the MUF factor appropriate to the distance and layer. Then, whichever calculation yields the highest value, this is the MUF for that distance, and the calculation indicates which of the layers will control transmission over that distance.

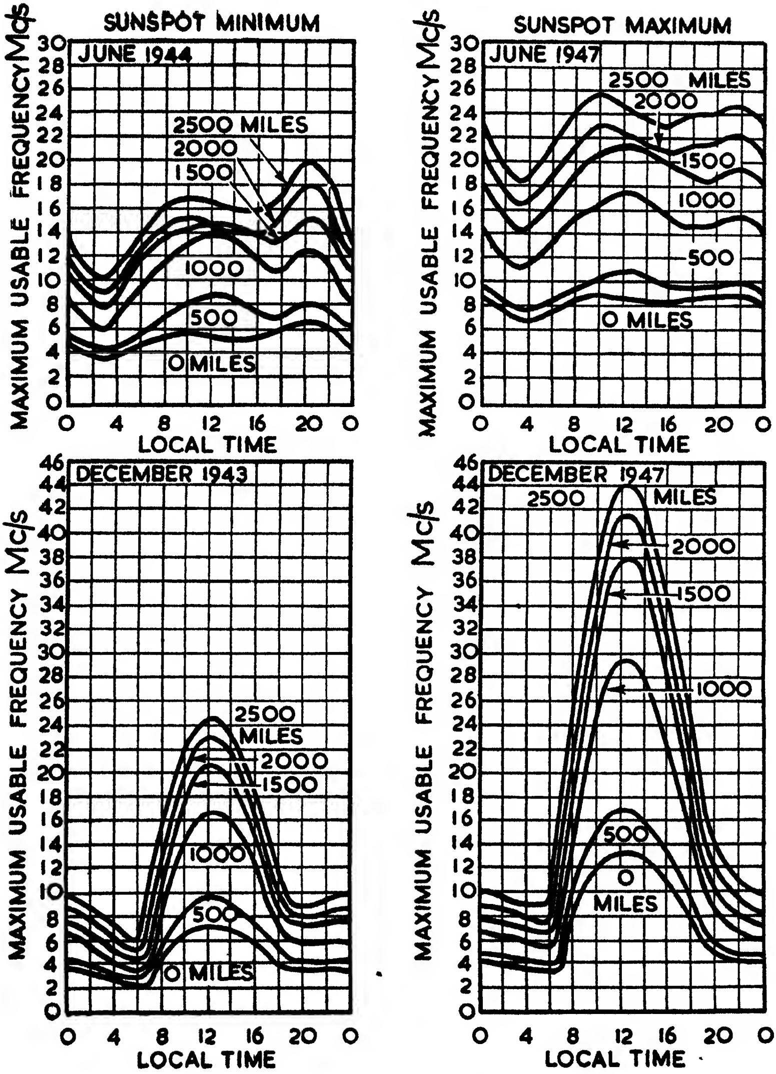


Fig. 34—Maximum usable frequencies for various distances, obtained from the critical frequency values of Fig. 21 by use of the MUF factors given in Figs. 31, 32 and 33

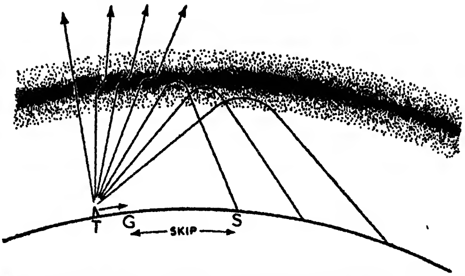


Fig. 35—Showing the reason for the existence of a skip zone

Skip Distance

It will be clear that if we worked on a frequency below the critical frequency, *all* the upgoing rays would be returned to earth, no matter at what angle they impinged upon the refracting layer. None of the rays would penetrate the layer at such a low frequency. But, as we have said, it is extremely bad practice to work on a frequency as low as this when we wish to cover long distances. We need to use a frequency which is near the MUF for the distance over which we are transmitting, and this is always higher than the critical frequency. If we use this frequency then it means that the higher angle rays will penetrate the ionosphere altogether. We have the situation shown in Fig. 35 ; the high-angle rays penetrating the layer and those at the angle corresponding to the distance for which our frequency is the MUF and all lower angles being returned. The result is that there is an area round about the transmitting station and beyond the limits of the ground wave in which there are no rays coming down from the ionosphere at all. It is not served by any of the waves radiated by the station. This area is called the skip zone, and the distance across it in any direction—that is, the distance between the transmitter and the point where the first refracted ray reaches the ground—is the skip distance, the waves being pictured as " skipping " over the area.

The dimensions of the skip zone and of the skip distance will depend entirely on the ionisation of the layer and on the frequency used, and they will thus vary for a given frequency with time of day, season of year and phase of the sunspot cycle. They will not vary with the amount of power radiated, since no increase in power makes any difference as to whether the wave penetrates the ionosphere or not—that depends simply on the frequency and the ionisation prevailing.

Now the distance at which a given frequency is the MUF is also the skip distance for that frequency, for at the angle of incidence appropriate to that distance all higher frequencies will penetrate the layer. The MUF and any *lower* frequency will be refracted so that the wave is receivable at the distance considered, though if the frequency is decreased *much* below the MUF the attenuation due to ionosphere absorption will increase, and signal strength will therefore be reduced. So from the MUF curves—such as those of Fig. 34—we can read off for any time of day (interpolating where necessary) the skip distance appropriate to any frequency. If, for example, we are interested in communication over a fixed distance—say, 1,000 miles—we can see what frequencies would skip at each time of day and thus which of our available frequencies we must use in order to avoid skipping. If, on the other hand, we are interested primarily in the performance of a single frequency—say 14 Mc/s—we can see to what distances it would be usable for each time of day, and over what distances it would be unusable owing to skip.

To make sure, then, that none of the waves intended to reach a certain location penetrate the ionosphere and cause the location to fall within the skip zone, we must be certain that we use a frequency not higher than the MUF appropriate to that distance.

OPTIMUM WORKING FREQUENCIES

It must be remembered that the MUF curves are compiled from critical frequency measurements which are the *average* of those obtained on every day of the month, so that the MUFs shown are themselves the average MUFs for the month. In the diagrams of Fig. 36 are given for the month of March, 1946, the distribution of the critical frequencies of the F and E layers about the monthly mean as observed at Great Baddow, Essex. The full line gives the monthly mean and the figures give the number of cases when the critical frequency fell between the indicated limits of frequency. It will be seen that in the case of the E layer the day-to-day deviation is extremely small, and the same thing is true of the F1 layer. In the case of the F and F2 layers, however, there is considerable day-to-day deviation, even on relatively undisturbed days. The cases shown represent the worst usually experienced, for it has been found that there is more deviation at the equinoctial periods than at other seasons of the year.

An analysis of such diagrams as those of Fig. 36 has disclosed the fact that, whilst the day-to-day deviation in E and F1 critical frequency is of negligible account, that of the F and F2 commonly varies up and down about the monthly average by about 15 per cent on undisturbed days. This means that if we are to operate a regular everyday shortwave service we cannot, in the case of the F and F2 layers, work on a frequency quite as high as the average monthly MUF. If we did our communication would fail on a number of days during the month when the MUF was below the monthy average, due to penetration of the layer. In order to ensure continuous transmission on all the

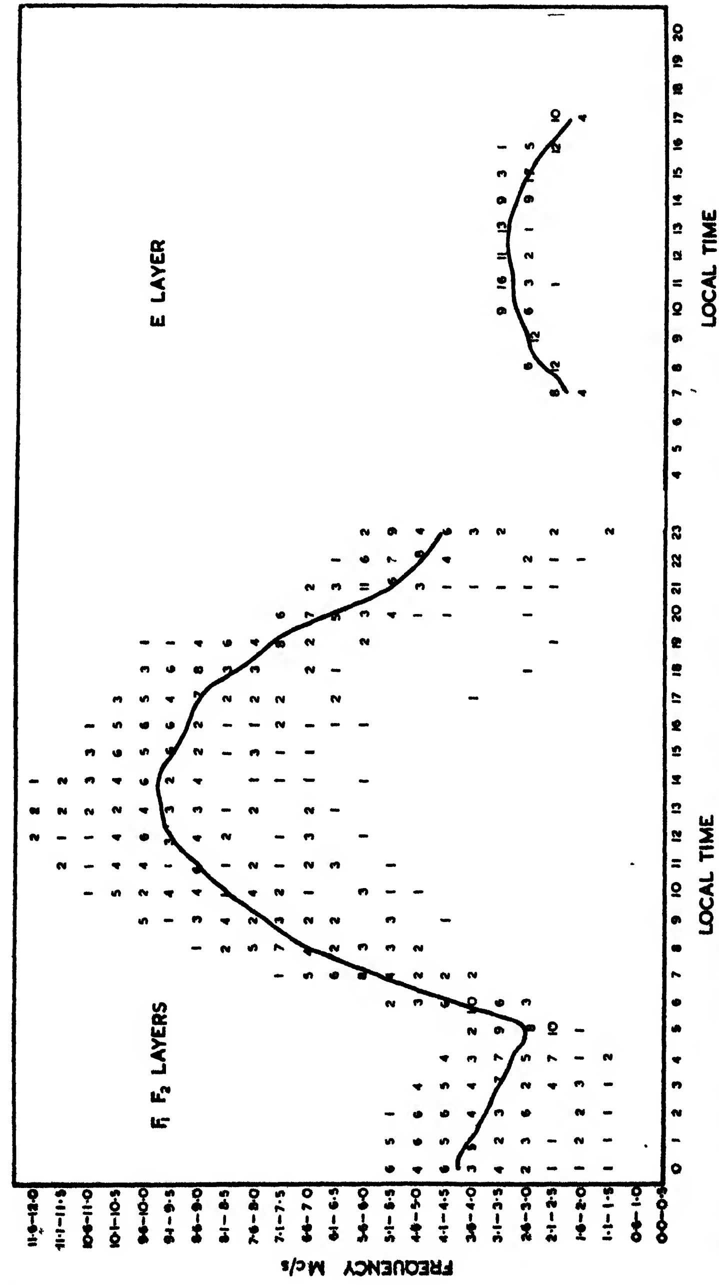


Fig. 36—Critical frequency observations made at Great Baddow, Essex, during March, 1946. The curves show the monthly mean values, and the figures indicate the number of observations which fell within the frequency bands

undisturbed days of a month it is therefore the practice to find the F or F2 layer MUF for the transmission path, and then to deduct 15 per cent from this in order to allow for the day-to-day deviation. The frequency thus found is called the Optimum Working Frequency (OWF) and the actual working frequency is kept as near below the OWF as is practicable for all hours of day. When the MUF is controlled by the E or F1 layers no such deduction is necessary, and the OWF for these layers coincides with the MUF. Examples of circuit curves plotted on this basis will be given later.

It will be noted that monthly predictions of OWF are designed to ensure reliable communication on every normal day of the month, but that this is not necessarily the sort of information required by amateurs and others whose main interest is the exploitation of certain frequency bands, usually of exceptionally high frequency. As will be seen from Fig. 36, since the MUF varies up as well as down about the monthly average, there will be occasions during any month when it is possible to achieve communication on frequencies not only above the OWF but also considerably above the monthly average MUF as well. For example, on ten days of any month the MUF for the F and F2 layers is likely to be 5 per cent to 15 per cent above the monthly mean, while on five days it may be from 10 per cent to 25 per cent above this. Thus, whilst the OWF indicates the high limit frequency for regular everyday communication, amateur contacts may be established on quite frequent occasions on frequencies up to 35 per cent above it.

ANGLES OF ELEVATION

It is obvious that good communication over a given distance will only result when the design of the aerial is such as to radiate a considerable proportion of the total energy at the correct elevation angle. This would be relatively easy if ionosphere conditions remained steady and constant, but, as we have seen, they vary continuously not only with time but also, at a given time, with geographical location. It is to be noted that variations in electronic density, as well as those of layer height, will have some effect upon the angle of elevation, because, for a given usable frequency, the nearer this is to the MUF the further will the wave penetrate into the layer and the greater will be the height from which it is returned. And, as will be seen from Fig. 30, the height of reflection is the controlling factor in determining the elevation angle necessary in order to return the ray to earth at a given distance from the transmitter. The higher the point from which the wave is returned the greater is the length of the " hop," and the height, in turn, depends upon the electronic density, or, with a given electronic density, upon the frequency.

It is evident that if we were to try and take account of all the variations of layer height with time of day, season and geographical location, as well as of the variations in height of reflection with changing electronic density, we should have to be continually altering the elevation angle. Except in certain special classes of service this is impracticable, and a compromise is usually effected by radiating energy at a number of elevation angles, all of which are likely to be useful angles at some time or another. In other words, the aerials for different directions and distances are arranged to radiate a " lobe " of energy which is wide enough to cover all useful vertical angles. At the same time the working frequency, which is never higher than the OWF, that is, 15 per cent below the MUF, is changed from time to time as the electronic density changes, so that it never falls too far below the OWF, and thus the height of reflection does not vary greatly from time to time. Nevertheless, in order to prevent wastage of energy the lobe width and the elevation angle of its centre must be related to the estimated heights of reflection.

For this purpose the elevation angles given in Figs. 31, 32 and 33 will be found useful in practice, so far as single-hop transmission is concerned. It is to be noted that it is over the first few thousand miles outwards from a transmitter that both the MUF and the elevation angle change most rapidly, so that the use of more than one frequency and more than one type of aerial is often more necessary for providing a good signal over an area of a given longitudinal extent at these medium distances than is the case at greater distances.

A method of approximation in selecting the appropriate elevation angles usually has to be followed, according to the following procedure: First find the elevation angles corresponding to the centre of the area to be served from Figs. 32 and 33 for F or F2 layer transmission—having regard to the diurnal and seasonal variations in virtual height. Then find those corresponding to the inner and outer boundaries of the area in the same way, and so obtain an idea as to the elevation angle for the centre of the lobe, and as to its necessary width in the vertical plane. Then the effect of the E and F1 layers in controlling transmission at certain times of day may have to be taken into account, and finally the elevation angle for the centre of the lobe may have to be slightly readjusted having regard to the types of aerial available, in order to ensure that no part of the area falls within the skip at any time.

An example will serve to make the matter clear, the relevant data from Figs. 31, 32 and 33 being shown in Table 3.

For amateur purposes an aerial with a narrow major lobe having its centre at about 8° would probably be most satisfactory for long-distance work, but such a low elevation angle is almost impossible to achieve with arrays such as are convenient for amateur use. However, there are several types of array which, when erected about one wavelength

TABLE 3

LONGITUDINAL EXTENT OF AREA TO BE COVERED—1,000 TO 2,000 MILES FROM TRANSMITTER

Elevation angle	*To inner boundary*	*To centre*	*To outer boundary*
By night-time F	18°	10°	4°
By winter day F2	16°	8°	3°
By summer day F2	22°	12°	6°
By E layer	5°	8° (2 hops)	5° (2 hops)
By F1 layer	10°	4°	8° (2 hops)

Width of lobe 19°, that is, 3° to 22°.
Centre of lobe—at say 12½°.

above ground, have a major lobe centred at between 10° and 20°, and also have a considerable forward gain, and their use on the higher amateur frequencies is usually practicable at amateur stations. For use on lower frequencies, where the erection of an array is usually impossible at an amateur station, a simpler type of aerial—like a dipole—is often employed, and, as will be seen from Fig. 29, this, if erected between ½ and 1 wavelength above ground, will radiate a considerable amount of energy at useful angles.

For professional communication services elaborate types of aerial arrays are usually employed, having azimuthal and vertical lobes to suit the transmission paths over which they are to be used. In the broadcast services, for example, different arrays are employed for serving different areas, each one being designed with regard to the distance, the latitudinal and longitudinal extent of the service area, and the average heights of the refracting layers. In many cases it is possible to " slew " the beam over several degrees of azimuth by electrical means, if required.

Since the earth is practically spherical it would be expected, for one-hop transmission at least, that the angle of arrival at the receiver would be the same as the elevation angle at the transmitter, and measurement has shown this to be substantially true. It has also shown that, under quiet ionospheric conditions, these angles are relatively stable and only vary by about 3°. During periods of ionosphere disturbance, periods which we shall discuss later on, the ionosphere is much less stable, and the angles may vary from time to time by as much as 9°.

It might also be expected that there would be considerable *lateral* deviation of the radio rays, due to lateral gradients in the ionisation. Experiments over different circuits have shown, however, that under quiet ionospheric conditions this lateral deviation is very small, and does not as a rule exceed about 2°. It is known, however, that during disturbed ionospheric conditions considerable lateral deviation does occur, deviations of 20° being quite common and those of 60° not unknown.

In this chapter we have dealt mainly with one-hop transmission, since this forms the basis for the MUF and OWF technique for transmission by multiple hops which we shall discuss later. But, so far as the elevation angles for multiple-hop transmission are concerned, we may deal with them now.

In multiple-hop transmission it has been found that the radiated energy becomes considerably diffused by the ionosphere, so that the energy arrives at the receiver at a number of different angles, and it is difficult to relate these to a specific number of hops. Nevertheless experiments—particularly pulse experiments—have shown that there is usually one predominant angle of arrival corresponding to the least number of hops possible having regard to the *average* virtual height of the refracting layer over the transmission path, but that energy sometimes arrives by a lesser and also by a greater number of hops. The predominant arrival angle may vary with time of day, and the proportion of the total energy arriving at a given angle, that is, by a given order of hops, may be modified by altering the elevation angle at the transmitter. However, the ionosphere itself *tends to control* the number of hops, that is, more energy tends to arrive at the angle corresponding to the least number of hops having regard to the average virtual height, provided a reasonable amount of energy is radiated at the appropriate elevation angle.

It will be appreciated that the virtual height changes over very long transmission paths are very complex. For example, the increase in virtual height from winter day to summer day in the northern hemisphere results in the useful elevation angle for medium-distance transmission becoming greater in summer than in winter. When the transmission path extends into the southern hemisphere, since the seasonal change is in opposite phase in the two hemispheres, the virtual height change will be in the opposite direction. Any increase in the elevation angle brought about by an increased virtual height in one hemisphere will tend to be cancelled by a decreased virtual height in the other hemisphere. The useful elevation angle will tend to remain the same, but the distance traversed per hop over various parts of the transmission path will change. The useful elevation angle will, in fact, be determined by the proportion of the path in each hemisphere.

It will be seen therefore that in multiple-hop transmission it is not usually possible to specify a precise elevation angle corresponding to a precise number of hops. It has, however, been found in practice—and this has been tentatively confirmed by experiment—that for all multiple-hop transmission where a relatively large area has to be covered it is generally satisfactory to radiate the energy over a range of vertical angles of from 2½° to 13½°, with the centre of the lobe at 8°. This takes account of the variations in virtual height and MUF with time of day and season, and also of the necessity to cater for different orders of hop. It appears to apply to the above cases irrespective of the length of the transmission path, the ionosphere itself controlling the exact mode of propagation. This is not to say that, in certain special types of service where the transmission is to a more precise geographical point, other elevation angles and lobe widths might not prove more efficacious.

In the next chapter we shall explain the modern MUF and OWF technique applicable to world-wide coverage by multiple-hop transmission, and also mention the methods used in the prediction of ionospheric conditions.

CHAPTER 7

Multiple-hop Transmission and Ionospheric Forecasting

MULTIPLE-HOP TRANSMISSION

IN FIGS. 31 TO 34 THE METHODS for finding the MUFs applicable to distances up to about 2,500 miles were illustrated. Beyond this distance transmission must be by multiple hops, that is, by successive reflections or refractions in the ionosphere and at the earth's surface. Although the theory relating the critical frequency and angle of incidence to the MUF is the same for multiple as for single-hop transmission, and although the methods described for applying the critical frequency data to single-hop transmission are those on which the methods used for multiple-hop transmission are based, there arise complexities in the latter case which render necessary a special technique.

Multiple-hop transmission cannot, in fact, be treated as a simple extension of the single-hop theory, because, after the first hop, the mode of transmission becomes so complex that it is not possible to divide the transmission path into a number of different hops and examine the ionosphere at the centre of each. Experience has shown, for example, that just beyond 2,500 miles the MUF does not sharply decrease, as might have been expected owing to the sudden decrease in the length of each hop, but, owing to diffusion of the rays by the ionosphere, remains approximately at the value corresponding to the smallest elevation angle one-hop case. Again, because of this, the skip zone is not, as might have been expected, repeated within the area assumed to be covered by the second or succeeding hops, that is, only one skip zone is at all clearly defined, and that is the zone from the limits of the ground wave to the point where the first downcoming ray arrives.

As to the reason for this "spreading" of the radiated energy, it will be appreciated that there is considerable "scattering" of the energy each time it passes through the E layer on its way up to the F, whilst there is also some lateral scattering due to gradients in the ionisation, or by particular "clouds" of ionisation. All this seems to result in transmission over a multiplicity of paths, so that, irrespective of what was said in the last chapter about elevation and arrival angles, when it comes to the calculation of the MUF for multiple-hop transmission, a simple division of the transmission path into separate hops is impossible. Fortunately, however, a method has been found—more by experience than by theory—which seems to fit the conditions very well. For multiple-hop transmission paths it has been found that it is only necessary to consider the ionospheric conditions at two points, each of these being one-half the length of a hop of maximum possible length from each end of the transmission path. So, in effect, we assume lowest elevation angle transmission for all multiple-hop paths, no matter what their length may be. In practice, therefore, we draw upon a Mercator map the Great Circle path between the transmitter and the receiving location, and then mark points upon this 1,250 miles from each end. These are called the " control points," and we note the MUF at each of these points at any particular time. Whichever of these MUFs has the lowest value is the MUF *for the transmission path* at this particular time, and in this way we take account of the geographical variation in MUF along the path.

As an example let us consider a case where the transmitting end of the circuit is in broad daylight, whilst the receiving end is in deep darkness. The MUF in such a case will decrease all along the path from transmitter to receiver, and the control point having the lowest MUF will be that at the far end of the circuit, viewed from the transmitter. This will be the MUF for the path, and will determine the working frequency to be used. Since the frequency thus indicated will be one which we know will be returned from a point on the path where the ionisation is at its lowest, this must also be returned from points nearer the transmitter where the ionisation is higher. On the other hand we are obliged to use a frequency lower than one which could be returned from the control point at the near end of the path, because if we did not it would penetrate the ionosphere at the far end.

World-wide Critical Frequency Data

In order to obtain information about ionospheric conditions all over the world there have been established, as was mentioned in Chapters 4 and 5, numbers of ionospheric " observatories," located at " strategic " points—from the point of view of obtaining the most comprehensive information—all over the world. At the present time there are over fifty of these stations operating in various British, Dominion and foreign countries throughout the world. Regularly throughout the twenty-four hours of every day they are engaged in making measurements upon the ionosphere immediately overhead, these being in the form of h′ - *f* curves like those described in Chapter 4, and illustrated in Figs. 16, 18, 19 and 20 (pages 44–52). The significant details as to critical frequencies and virtual heights obtained from these hourly h′ - *f* curves are transmitted to some authority in each of the countries concerned—in this country to the Radio Research Board of the Department of Scientific and Industrial Research—the information in its usual form being a monthly summary of the ionospheric characteristics observed. There is thus a steady stream of this

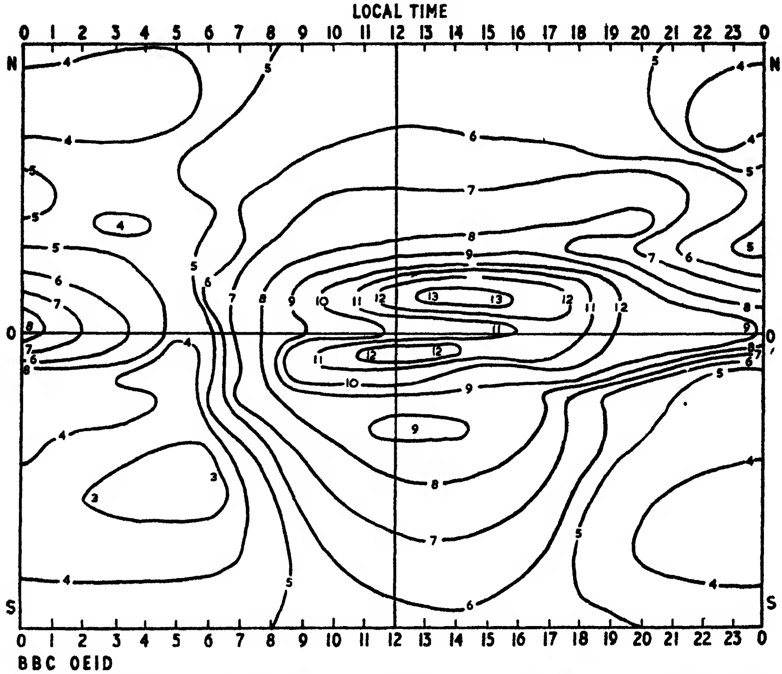


Fig. 37—Specimen contour chart of predicted critical frequencies. It is applicable to the " I " zone only and is for the month of May of a year about half-way between sunspot minimum and maximum

ionospheric information coming in, which can be analysed in order to build up a world-wide picture of the ionospheric conditions existing at any time or, more usually, a picture of the average conditions existing during any one month. The data is plotted on similar lines to those used for plotting weather conditions over a wide area—the familiar weather charts issued by the Meteorological Office.

IONOSPHERIC CONTOUR CHARTS

Let it be assumed for the moment that the " longitude effect " which we described in Chapter 5 does not exist, but that the same values of critical frequency and virtual height as are obtained at any one observing station would be obtained at all other points on the earth's surface lying in the same latitude at the same instant of local time. If, therefore, we take a Mercator projection of the world's surface (without the usual geographical features upon it) and along the parallel of latitude appropriate to each observing station enter the monthly mean of the critical frequencies obtained at that station, equally spacing the twenty-

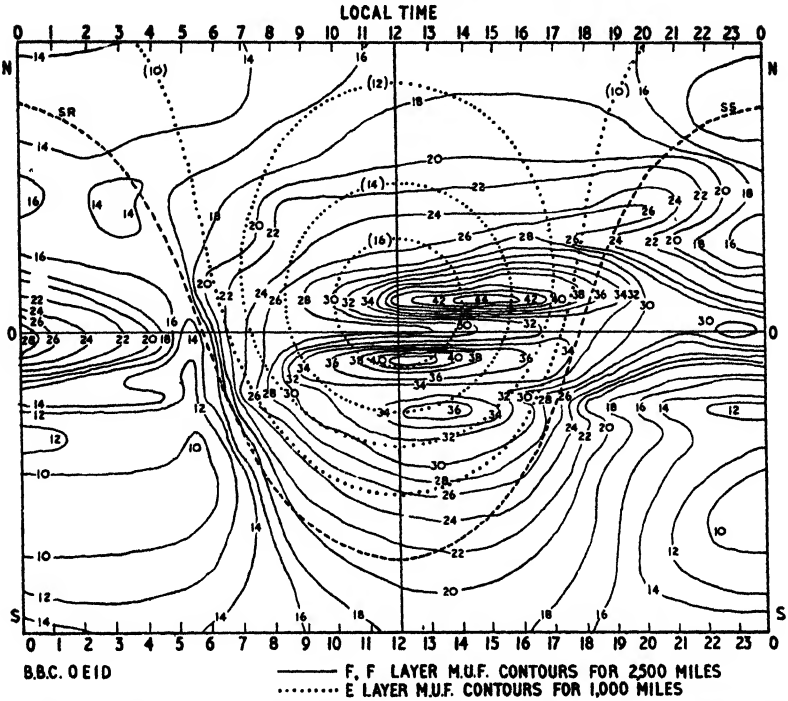


Fig. 38—Specimen contour chart of predicted maximum usable frequencies. It is applicable to the " I " zone only, and is for the month of May of a year about half-way between sunspot minimum and maximum

four-hourly measurements so as to cover the entire width of the projection, then we may assume that we have recorded mean conditions for that latitude for the twenty-four hours. We may enter the hours of local time along the top of the projection (as in Fig. 37) and when we have entered in the measurements of every one of the observing stations in its appropriate latitude, we may assume that we have recorded the monthly average critical frequency on a world-wide basis. Now, if we join up all the points of equal critical frequency, we produce a critical frequency contour chart (similar to Fig. 37) which depicts the world-wide variation in critical frequency. If we draw the chart upon transparent cloth we can, by laying it over a Mercator map of the world and by sliding it along to represent the rotation of the earth upon its axis, see the monthly mean critical frequency at any place at any value of local time.

Similarly, if we analyse the virtual height data from all the different stations we can obtain the MUF factors appropriate to each latitude and time of day, and by plotting these upon a Mercator projection and joining up the lines of equal factor, can produce a contour chart of world-wide MUF factors.

Finally we can produce a contour chart of MUFs, which is what we actually require in order to apply the ionospheric data directly to short-wave transmission. This can be done by reading off, from the critical frequency contour chart, the critical frequency values for each hour of day for, say, every 10 degrees of latitude, and, from the MUF factor chart, the MUF factor values appropriate to the same times and geographical positions. The critical frequencies are then multiplied by the appropriate factors and the resulting values plotted upon another chart in the correct latitudes and under the correct times of day. When points of equal value are joined by lines, the result is a contour chart showing the world-wide variation in MUF. It is most convenient in practice to draw such charts in terms of the F or F2 layer MUF for 2,500 miles and of E layer MUF for 1,000 miles (it is not necessary to show the F1 layer MUF at all) and a chart like this would be similar to that pictured in Fig. 38.

ZONAL METHOD OF ALLOWING FOR LONGITUDE EFFECT

As was mentioned in Chapter 5, it is incorrect to assume, as we have done in the previous section, that the critical frequency will be the same along a given parallel of latitude at equal values of local time, for the ionisation of the F and F2 layers varies, not only according to the geographical latitude and longitude, but also according to the magnetic latitude and longitude as well. This introduces complexities into the compilation of the contour charts, for, since the magnetic axis of the earth has no relation to local time, it means that we cannot transpose longitude for time, as was done in the explanation given in the previous section.

The present solution to this difficulty is to make a sort of compromise between the geographical and magnetic influences on the critical frequencies, by dividing the world into separate zones, and plotting separate contour charts for each zone. The delineation of the zones, according to their present definition, is as shown in the map of Fig. 39, there being an " east," a " west " and two " intermediate " zones. Since the zonal boundaries are certain magnetic meridians it will be seen that points within each zone of a given geographical latitude and longitude will have a *similar* magnetic latitude and longitude. All the ionospheric measurements obtained from stations within a particular zone are used in the construction of the contour chart for that zone only, and thus the geomagnetic influence on the critical frequencies is, to some extent, allowed for.

The chart will then not be quite correct for every place within a zone,

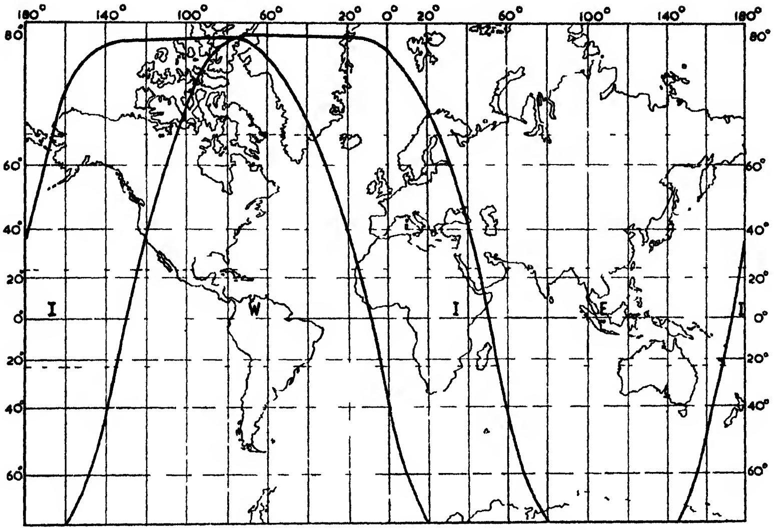


Fig. 39—Zones covered by a series of three contour charts

since the geomagnetic influence will tend to vary across it, but it will show the *average* conditions prevailing in that zone, which is good enough for most practical purposes. Thus the critical frequency contour chart of Fig. 37 and the MUF contour chart of Fig. 38 apply to conditions within the " intermediate " zones only.

It is not implied that the above are the exact and only methods employed in the construction of contour charts, and those used by every organisation engaged in this work, but the description given serves to illustrate the principles involved.

FORECASTING IONOSPHERIC CONDITIONS

Although contour charts compiled from the measured values of critical frequency and virtual height are useful for many purposes, what we usually require in the planning of short-wave transmission operations is not a chart showing conditions as they have existed, but one showing them as they will exist at some future date. This involves the forecasting of ionospheric conditions all over the world, and this is a process too complicated to describe in detail in this book. But we may, perhaps, briefly touch upon the methods by which it may be done.

In Fig. 25 (page 63), we illustrated how the twelve-month running average of the critical frequencies of the various layers followed that of the twelve-month running average of the sunspot numbers. By observing the current trends in the sunspot curve it will be seen that this may be exterpolated for several months ahead with good accuracy, and, having thus forecast the average value of sunspot activity likely to prevail, it is possible to go forward with the twelve-month running average critical frequency curves by maintaining the relation which the latter have with the former. Thus the general critical frequency variation over a period of time—at least over one not exceeding a few months—may be seen for the various locations.

It is then necessary to take account of the diurnal and seasonal components in the critical frequency variations, so as to obtain a diurnal curve of *predicted* critical frequency for each of the ionosphere observatory locations at any season. Fig. 40 shows the change in F and F2 critical frequency in England for various significant times of day during June and December over the " increasing " phase of the current sunspot cycle, as obtained from the critical frequency curves of Fig. 21. During June, it is seen, the change was more or less constant over the twenty-four hours, but in December it varied greatly as between day and night. As regards the changes in E and F1 critical frequency with changing sunspot activity, these are relatively simple and are more or less constant for all hours and seasons.

The changes shown in Fig. 40 represent those produced by an increase in twelve-month running average sunspot values of about 140, and it will be seen that such curves can be broken down so as to obtain the diurnal and seasonal components of critical frequency variation for any given change in sunspot numbers. If these are applied to the measured critical frequency values, for, say, the same month of the previous year, a predicted diurnal characteristic can be built up for the coming month. This may be done for every station, and then, for a given time of day, curves showing critical frequency variations with latitude along different meridians of longitude may be plotted. These, when smoothed, are like those of Fig. 26A, and show the predicted variation

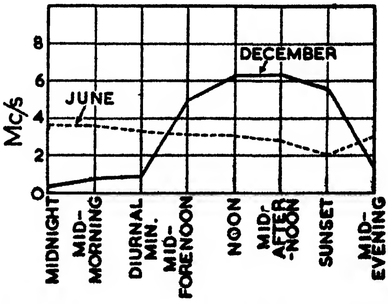


Fig. 40—Changes in F and F_2 critical frequency in England from sunspot minimum to maximum at various significant times of day

TABLE 4

DISTANCE FACTORS BY WHICH F LAYER 2,500-MILE CONTOURS OR E LAYER 1,000-MILE CONTOURS MUST BE MULTIPLIED TO OBTAIN MUF FOR SHORTER DISTANCES

Distance	*Factor*			
(Miles)	*For F or F2 layer*	*For E layer*	*For F1 layer*	*Add to MUF Mc/s*
0	0·35	0·22	—	0·7
250	0·36	0·39	—	0·4
500	0·42	0·64	—	0·2
750	0·52	0·85	—	—
1,000	0·64	1·00	—	—
1,250	0·74	1·10	—	—
1,500	0·83	1·14	—	—
1,750	0·90	—	1·15	—
2,000	0·95	—	1·16	—
2,250	0·98	—	—	—
2,500	1·00	—	—	—

in critical frequency for every latitude along a given meridian. Several of these are prepared, for different times of day, and for different layers.

These are then used in order to prepare contour charts of *predicted* critical frequency or MUF, for the different zones, showing the expected world-wide ionospheric conditions during the month yet to come. Figs. 37 and 38 show such contour charts for the " intermediate " zone.

Finding the Working Frequencies

Charts prepared in this way may be used to find the predicted *average* MUF—and from these the Optimum Working Frequency (or OWF)—for a transmission path of any distance, in any part of the world, for every hour of day for the month for which they are valid.

If the transmission path is of one hop, the relevant frequencies may be obtained in the manner described in the previous chapter, or alternatively they may be obtained from the MUF contour charts. The contours of this, as has been said, are in terms of the MUF for 2,500 miles, the maximum possible distance for one-hop transmission by the F or F2 layers, and in terms of 1,000-mile transmission by the E layer. If the path for transmission via either layer is exactly of these lengths then the MUF will be read off *directly* from the chart at the point at the centre of the path, and the OWF found by deducting 15 per cent from the MUF values. If, for one-hop transmission the path is of any other length, the MUF can be read off from the chart at the centre of the path and multiplied by a factor which is the ratio of the MUF for the given distance to that shown on the chart.

A set of " distance factors " suitable for this operation is given in Table 4 (page 95). Since the F1 layer contours are not given upon the chart its MUF is obtained, for the limited distances over which it can control transmission, by multiplying the E layer MUF by the Distance Factors given for F1 layer transmission. It should be added that, in the case of the F layer, the use of these Distance Factors tends to give inaccurate values of MUF at very short distances, and, for these distances, it is better to obtain the MUF from the predicted critical frequency, as described in the previous chapter.

The full procedure for finding the MUF for any transmission path is thus as follows : Using a Mercator map of the same size and co-ordinates as those used for the contour charts, the location of the transmitting and receiving points is first marked off, and the Great Circle path joining them is drawn in. If the path is 2,500 miles or less in length the centre of the path is then marked off, whilst, if it is of greater distance than this, the two control points 1,250 miles from each end of the path are similarly indicated. By reference to the map of Fig. 39 it is next ascertained in which zone the separate control points lie. If, as is the case for paths 2,500 miles or less in length, there is only one control point, or if where there are two they both lie in the same zone, it is only necessary to make use of one contour chart ; but if the control points lie in different zones, then the contour charts appropriate to both zones must be consulted.

Using the transparent contour chart, this is placed over the Mercator map so that the equators on each coincide, and the transparency is slid over the map until its 00 hours meridian coincides with the meridian on the map corresponding to the standard time in which it is desired to work. For instance, if it is desired to work in terms of G.M.T. we start operations with the 00 hours meridian coincident with the 0° longitude meridian on the map. Fig. 41 will help to make this clear. The 2,500-mile MUF at each control point is then read off, and by sliding the transparent contour chart along so that each hour in turn coincides with the Greenwich meridian, this is done for every hour of the day, the MUFs for each point being entered upon a work sheet. Having done this it is only necessary to strike out the *higher* of the two MUFs appearing on the work sheet to be left with a value which is the MUF for the whole path. A deduction of 15 per cent is then made from this value, the result being entered in another column of the work sheet which shows the OWF for every hour of day. Other columns on the work sheet are for the MUF by way of the E or F1 layers.

CIRCUIT CURVES

It is now advantageous to plot the data from the work sheet in the form of circuit curves, such a graph being shown in Fig. 42, where the

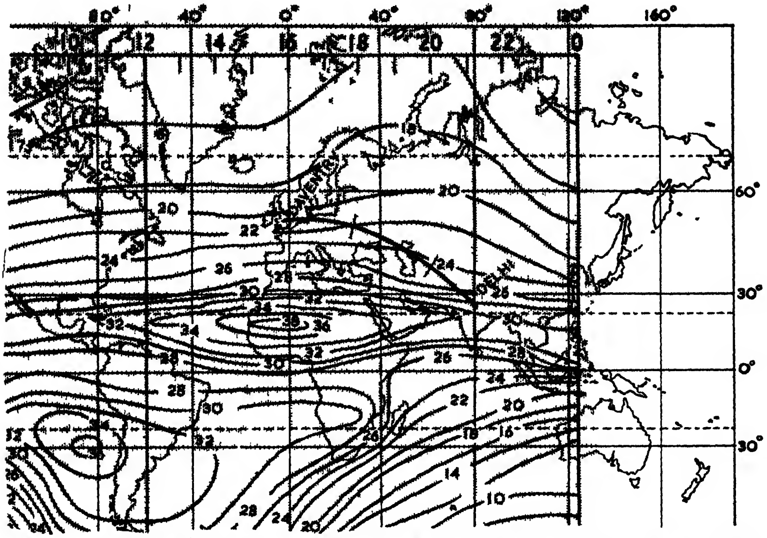
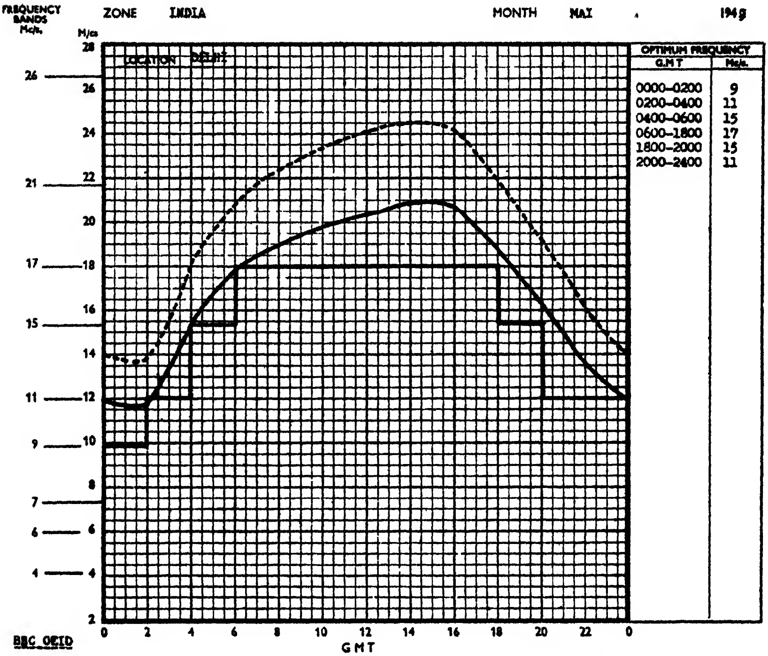


Fig. 41—Showing the method by which the MUF is read off for every hour of day. The transparency is shown over the Great Circle path to Delhi at 16 hours G.M.T.

conditions are shown for the multiple-hop London/Delhi circuit for May, 1949, the dashed line showing the MUF and the full line the OWF. We are now in a position to choose the actual working frequencies for every time of day, and these will depend upon the frequency allocations made for the particular services in which we are interested, that is, the highest frequency band, not above the OWF, at the various times of day, which we may have available.

In Fig. 42 are shown (as horizontal full lines) the actual working frequencies inserted into the curve in terms of the broadcast bands, while the vertical lines connecting them indicate the times when a frequency change is necessary. Of course, it does not follow that one would necessarily change frequency at all the times indicated, particularly in a broadcast service, in which continuity of transmission on a single frequency for as long as possible is a desirable feature. Nevertheless, these curves show at a glance the frequencies which *should* be used at any time of day, and, in general, the transmission schedules are compiled in conformity with them. Thus, having available, several months in advance, curves like this for every circuit over which it is desired to transmit, one is able to plan transmission operations with some facility.

Fig. 43 is a circuit curve for a single-hop circuit for a summer month, as obtained from a contour chart of predicted MUF, and, although we



OPTIMUM FREQUENCY	
G.M.T	Mc/s.
0000–0200	9
0200–0400	11
0400–0600	15
0600–1800	17
1800–2000	15
2000–2400	11

Fig. 42—Curves of frequencies for a specific route. Dotted line shows MUF for path, while the full lines refer to optimum working frequency; broadcasting bands are shown

have already described the principles by which the MUF is obtained in such cases, we may here indicate a few points of interest about this curve. It was found, in this case, that the E layer controlled the MUF between 0900 and 1500 G.M.T., its MUF for this relatively short distance being higher than that by way of the F2. Thus, when the dashed line MUF curve was plotted, a " hump " appeared in the curve between those times, the significance of which we have already described. In finding the OWF for the month the deduction of 15 per cent was made from the F layer MUF, but not from the MUF for the E layer, because the day-to-day variability of ionisation in this layer is so small as not to necessitate this. Thus in drawing in the full-line OWF curve this was taken right up to the MUF curve for the times during which the E layer controls, so that the OWF coincides with the MUF. The allocated frequencies are then fitted into the curve in the manner already described.

It is interesting to note that, whilst the day-time working frequency for a long circuit where transmission is by the F2 will decrease in

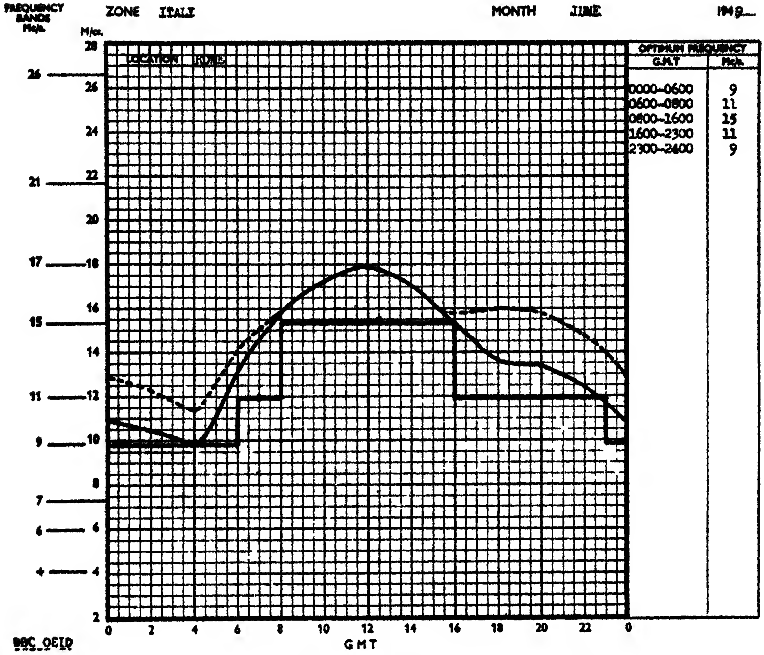


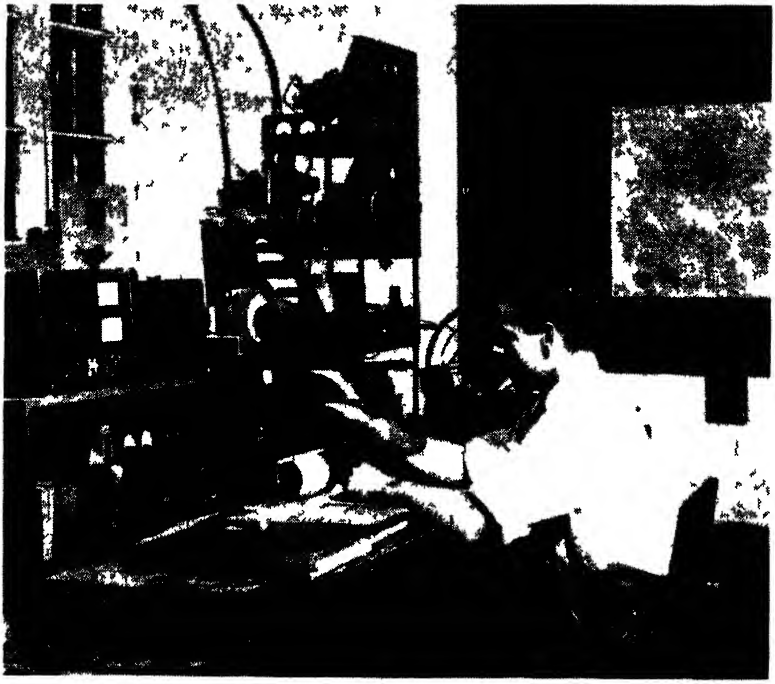
Fig. 43—Frequency curves for a relatively short route

summer, that for a short distance circuit will tend to increase, because the ionisation of the E layer is higher in summer day than in winter day.

This, then, is the modern technique in which, by the use of contour charts showing predicted conditions, the ionospheric data are put to use in practical engineering. The authority responsible for the preparation of such charts in this country is the Radio Research Board of the Department of Scientific and Industrial Research, and the charts are prepared, at present, six months in advance, at the Radio Research Station, Slough. These are available to all official bodies which maintain regular short-wave transmitting and receiving services in this country, and no doubt they would be found of great use, when employed according to the above-described techniques, by many other radio engineering concerns as well.

Sets of charts and other data are also supplied monthly to the Radio Society of Great Britain, so that they are always available for the amateur members of that society to borrow or consult. It is possible, though, that the methods described are somewhat too involved and elaborate for general amateur use, and, in any case, the problems involved in amateur transmission are somewhat different from those in the maintenance of a regular everyday short-wave service

In the next chapter we shall have something to say about these special "amateur" problems, though the phenomena discussed should, of course, be of interest to professional technicians as well



AN AMATEUR RADIO STATION

In front of and on the left of the amateur are the receivers, with "plug-in" coils in the box above one receiver and the loudspeaker above the other. To the right is the "rack-mounted" transmitter, with (from bottom to top) its power supplies and modulating circuits, oscillator circuits and final amplifier with co-axial cable connection to the aerial. On the desk are the microphone and log-books. Upon the wall is a map of the world on a projection which gives the true bearing of any point from the point where the string is attached. (*Photograph by courtesy of the R S G B*)

CHAPTER 8

Amateur Transmission on High Frequencies

DIFFERENCES BETWEEN AMATEUR AND PROFESSIONAL PRACTICE

WHEREAS THE TASK OF THE professional short-wave engineer engaged in the operation of long-distance radio services is to maintain steady and reliable communication on each and every day throughout the year, the interests of the amateur radio man usually lie along quite different lines. The spirit of adventure plays a large part in the pursuit of his hobby—he is usually out to do something "phenomenal" in the way of long-distance communication, such as to effect the maximum number of contacts over the greatest possible distances. This is not to say, of course, that the exploits of amateur radio men are not, in the long run, of a useful nature, for, as everyone knows, their operations have done a great deal to advance the knowledge of short-wave engineering in practically all of its branches.

However, the point is that the amateur's aim is usually that of the exploitation of certain frequency bands—often of very high frequency—in order to secure the maximum results in the way of " DX contacts," rather than in the more mundane task of maintaining regular communication with a certain area, and because of this his needs are different from those of the professional radio engineer. The latter is obliged to impose safeguards upon his operations in order to ensure that his working frequencies do not fail on certain days by reason of their being too high, and even with these he is sometimes defeated in his aims by the incidence of ionosphere storms and disturbances.

No such limitations bother the amateur—apart, of course, from the fact that his actual transmissions are subject to the same natural laws as are those of the professional engineer—he is prepared to work upon a frequency far above the OWF, and even above the mean monthly MUF—on the mere chance of effecting transmission with some distant colleague, or of learning something new about conditions on such high frequencies. And it is to be noted that if ionospheric conditions are such as to favour his experiments, then his use of a very high frequency is likely to give rise to better results than if he had worked upon a lower frequency. For, with the low power which is usually at his disposal, the higher the frequency upon which he can work the greater is his chance of putting down a workable signal at the distant point, because generally speaking the higher the frequency the less is the amount of energy lost by ionospheric absorption.

This applies, of course, to professional as well as to amateur transmissions, but the professional engineer has usually available enough power to overcome a considerable amount of ionospheric absorption, whereas the amateur has not.

Another interesting point about amateur operations is that, since they may be conducted upon " unorthodox " frequencies, the amateur is often in a position to observe new and interesting phenomena which may be of great significance in explaining points about the formation and structure of the ionosphere, about short-wave transmission generally, and about other matters of scientific interest. Many contributions of real importance in the field of scientific knowledge have, in fact, come from amateur radio men, and will, no doubt, continue to do so in the future.

AMATEUR F LAYER PROPAGATION

So far as transmission by way of the normal E and F1 layers of the ionosphere is concerned there are few, if any, points of difference between professional and amateur practice. It is to be noted in this connection that the amount of power radiated makes no difference to the MUF of a layer—a layer either returns or does not return a certain frequency according entirely to the electron density existing, and irrespective of the power radiated. That is so far as a true refracted wave is concerned, and does not apply to the " scatter " of energy from an ionised region, which we shall allude to later on.

So far as F or F2 layer propagation is concerned, the OWF, as has been said, has little significance for the amateur, who will often obtain his best results on frequencies up to 35 per cent above it, which is well above the monthly mean MUF as well. An analysis of diagrams like Fig. 36 indicates that during about ten days in any month the F or F2 layer MUF may be from 5 to 15 per cent above the monthly mean value, and that on five days it may be from 10 to 25 per cent above this.

As an example of what this may mean to the amateur there has been plotted in Fig. 44 a monthly mean MUF curve for 2,500 miles, for a certain circuit. This is considered to be the frequency which would be propagated during fifteen days of the month, the OWF also being shown, and considered to be the frequency which would be propagated during all undisturbed days during the month. The two upper curves, which are respectively for frequencies 10 per cent and 20 per cent above the monthly mean MUF, are considered to show frequencies which would be propagated, respectively, on ten and on five days during the month. The significance of all the curves to amateur operators is shown by the insertion, as horizontal shaded areas, of some of the amateur wavebands at present in use in this country.

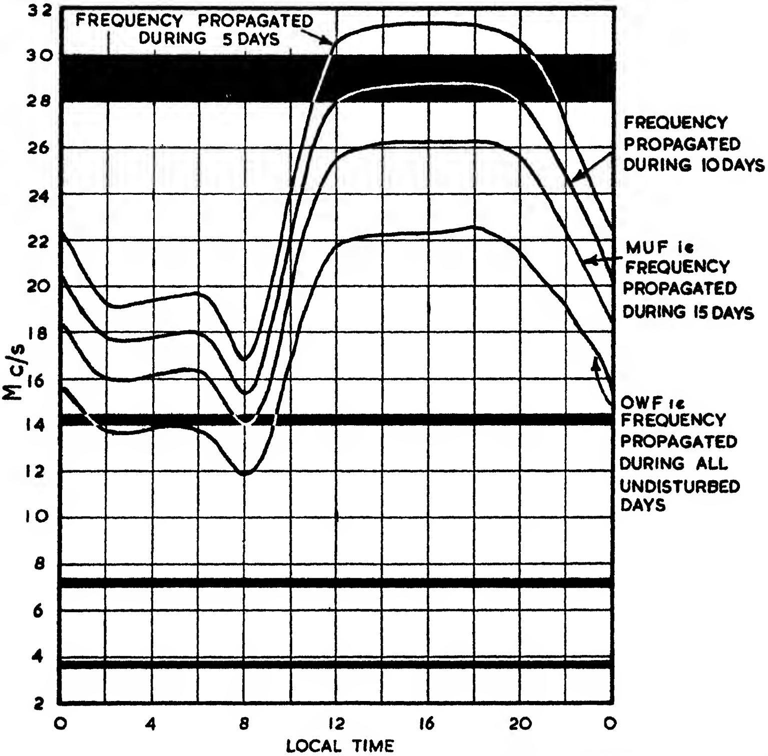


Fig. 44—Range of usable frequencies for F or F_2 layer transmission during one month

Range of MUFs for a Sunspot Cycle

In the hope of giving some information which may serve as a rough guide for amateurs, in place of the more elaborate methods previously described, the curves of Figs. 45 to 48 are presented, showing the MUFs which might be expected for certain circuits from this country at the winter, equinox and summer periods of sunspot minimum and maximum years. Fig. 45 shows these for a circuit running westwards from this country ; Fig. 46 for one in a south-westerly direction ; Fig. 47 for one in a south-easterly direction ; and Fig. 48 for one running in an easterly direction. If the area between each pair of curves is considered to represent the range of MUF for a complete sunspot cycle, then it ought to be possible, from a study of current sunspot trends, to interpolate the corresponding MUF curve for any of these seasons during any year during the remainder of the sunspot

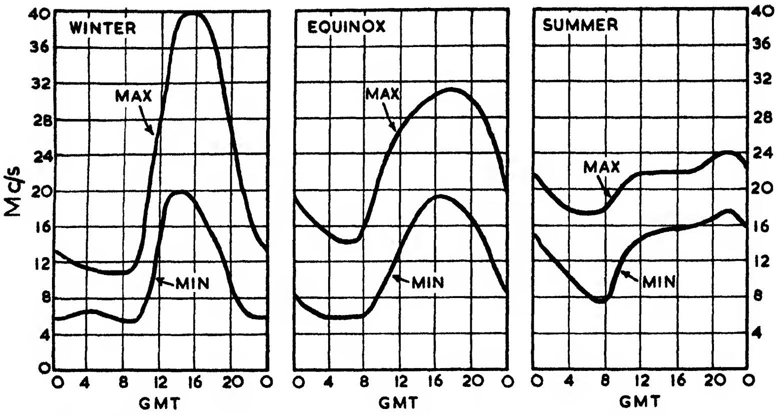


Fig. 45—MUFs for the London-New York circuit at sunspot maximum and minimum

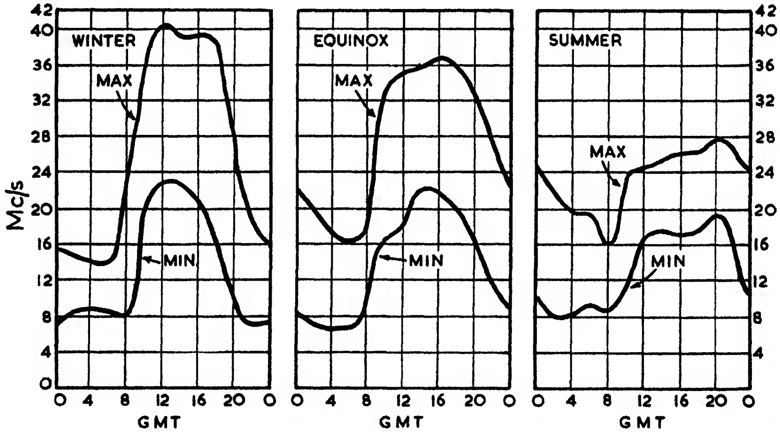


Fig. 46—MUFs for the London-Buenos Aires circuit at sunspot maximum and minimum

cycle. From this it should then be possible to estimate the frequency capable of propagation over these circuits for any proportion of the total time, and also to estimate roughly the frequencies capable of being propagated over other circuits running in similar directions.

Sporadic E Ionisation

Frequently during the summer months "patches" of ionisation appear within the E layer having a critical frequency far greater than that of the layer itself, and therefore capable of propagating radio waves

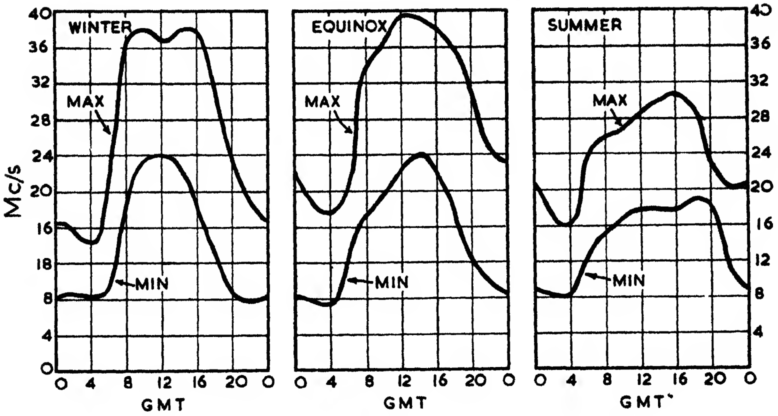


Fig. 47—MUFs for the London-Cape Town circuit at sunspot maximum and minimum

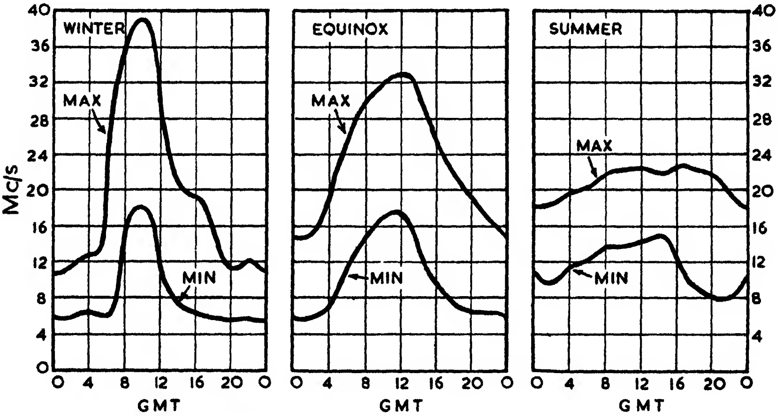


Fig. 48—MUFs for the London-Chungking circuit at sunspot maximum and minimum

of much higher frequency over medium distances. This type of ionisation is known as " sporadic " or " abnormal " E ionisation. Because it is of a random and intermittent character, and therefore not susceptible to prediction, it is practically useless in the operation of regular short-wave services, because one can never tell whether one can obtain communication by way of it at a given time or not. Amateur radio men, however, are often able to make use of it in the establishment of communication over these medium distances, and, by its use, are able to work over these distances on exceptionally high frequencies.

In the vertical incidence h′ - *f* curves the presence of sporadic E ionisation is often indicated in the manner shown in Fig. 49. The critical frequency of the normal E is discernible by the upward curl of the E layer curve, but, on frequencies higher than this, reflections are still received from the height of the E layer, and the higher layer reflections may or may not be observable at the same time. The frequency range over which such reflections may extend is very variable, but they have been known to extend up to 20 Mc/s. There is no upward curl at the frequency at which the sporadic E ceases to be observed, because of the " thinness " of this ionisation.

Sporadic E seems usually not to cover wide areas at any one time and to be relatively short-lived at any one ionospheric location. Thus *long-distance* propagation does not often occur by way of it. because of the remote possibility of its being present simultaneously at the widely separated ionospheric points necessary to effect multiple-hop propagation. But single-hop propagation (up to about 1,400 miles) often occurs by means of it, and, as is indicated by critical frequency measurements, this may occur on frequencies up to about 100 Mc/s.

An analysis of the measurements of sporadic E made by the Department of Scientific and Industrial Research at their Radio Research Station at Slough during 1946 has been made, and in Fig. 50 are given the results for each month in terms of percentage of total time at each hour of day during which sporadic E of critical frequency greater than 4 Mc/s was present. These measurements are made at hourly intervals, and, because the sporadic E may vary quite rapidly, such hourly measurements may not give a complete picture of its characteristics. Nevertheless they do give a general idea as to its rate of occurrence.

We see that sporadic E was recorded at all hours of the day and night, though it tended to be more prevalent during the day than during the night. The general diurnal distribution was, in fact, such as to produce a broad peak around noon and then a subsidiary peak around sunset. This is well brought out by the curve of Fig. 51, which

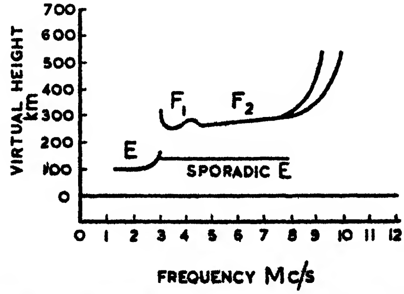


Fig. 49—An h′ - *f* curve in which the presence of sporadic E ionisation is indicated

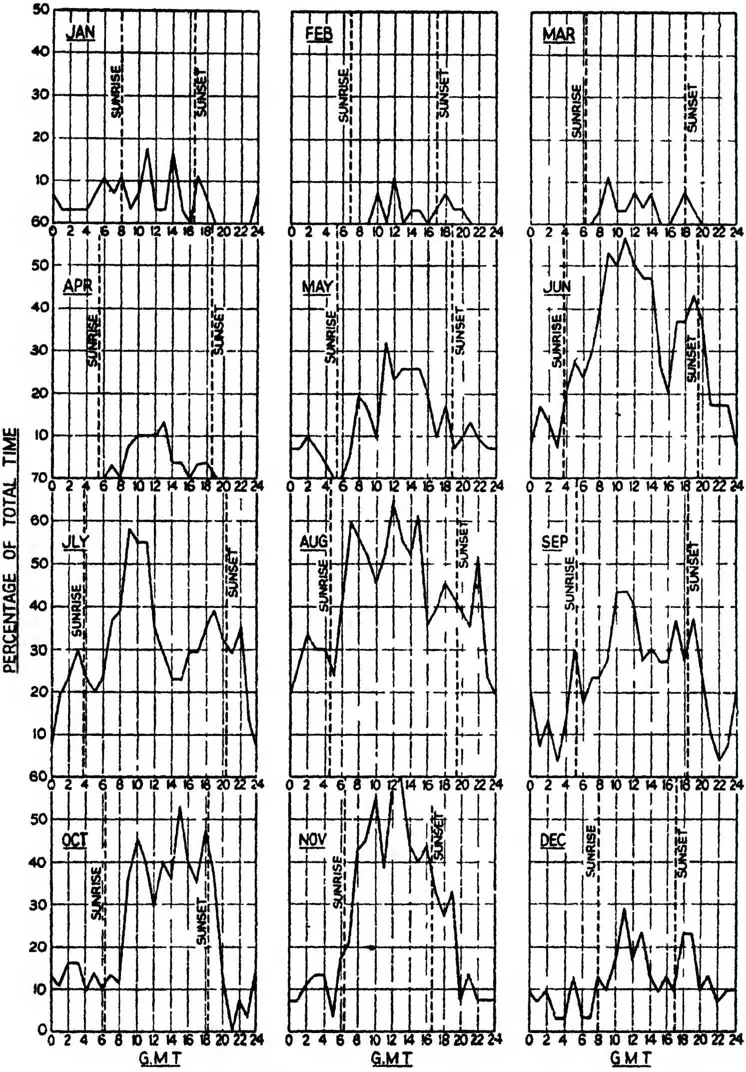


Fig. 50—24-hour distribution of sporadic E, 1946 ($f^c > 4$ Mc/s)

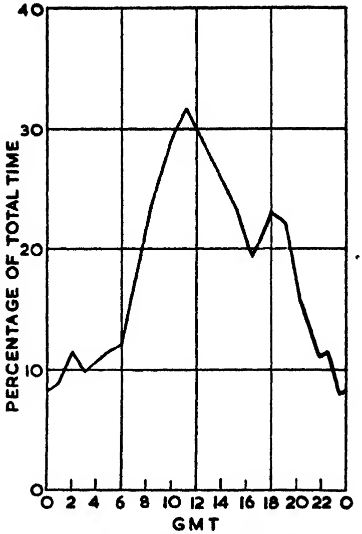


Fig. 51—Mean annual 24-hour distribution of sporadic E, 1946 ($f^c > 4$ Mc/s)

gives the annual distribution for 1946 for each hour, in terms of the percentage of the total time.

But the important thing to note from Fig. 50 is that sporadic E was much more prevalent during certain months than during others: 1946 was a rather unusual year in this respect, so in Fig. 52 is given the monthly distribution in terms of percentage of the total time, the curve being the mean of the measurements made over three years. This curve shows that it is only during the summer months when the phenomenon is present for a sufficient proportion of the total time to be of much significance in amateur communication.

Although, therefore, we cannot predict the occurrence of sporadic E we can say this much about it—

(1) It is subject to erratic and rapid variations, such that it may completely change its character within a matter of minutes.

(2) In general it has a diurnal variation such as to produce two broad peaks in its rate of incidence, the major one around noon and the subsidiary around sunset.

(3) It has a relatively well defined monthly distribution, such that, in the northern hemisphere, a minimum rate of incidence occurs during the period February–April. During the months of May, June, July, August and September it is present for a considerable proportion of the total time, and falls to low

> values of incidence again about October, to remain so during the remainder of the year. In the southern hemisphere this monthly distribution is six months out of phase with that in the northern hemisphere, so that the maximum occurrence is in local summer.

It may be added that measurements made in other parts of the world indicate a definite variation with geomagnetic latitude, and it is especially prevalent in the auroral zones—particularly during ionospheric disturbances.

The actual cause of the phenomenon is not yet known. Since it occurs quite frequently at night it would seem not to be caused by the sun's ultra-violet radiations, though this does not necessarily mean that it has nothing to do with the sun. As the aurora is almost certainly caused by corpuscles arriving from the sun, and as sporadic E is usually present in coincidence with displays of the aurora, this would seem to indicate that solar corpuscles may have something to do with the formation of sporadic E. However, there is some evidence that there may be two distinct types of sporadic E, for the aurora never spreads as far south as the geomagnetic equator, and only rarely as far south as Great Britain, yet sporadic E is frequently present in both these localities. Possibly, therefore, the sporadic E of mid and low latitudes may be produced by a different agent from that which causes the auroral type.

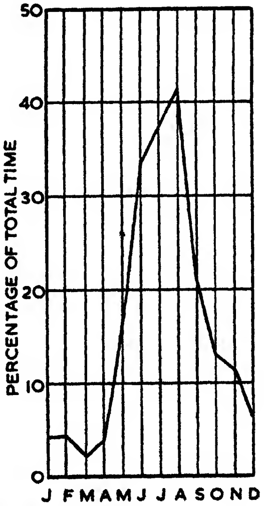


Fig. 52—Monthly distribution of sporadic E ($f^{c} > 4$ Mc/s)—mean for 1945, 1946 and 1947

Sporadic E in England was observed to increase during a meteor shower of marked intensity, so it is possible that the sporadic E of mid latitudes may have some connection with the arrival in the earth's atmosphere of showers of meteoric dust. Amateur experiments in the United States have indicated that a " cloud " of sporadic E appears to " drift " over the earth's surface from east to west, varying in intensity as it moves.

PROPAGATION ON HIGH FREQUENCIES BY SPORADIC E

In Fig. 31 (*b*) is given the MUF factor curve appropriate to transmission by way of the sporadic E. Although the phenomenon occurs at E layer heights it will be noticed that the factors are somewhat greater than those for transmission by way of the normal E, and this is because sporadic E is a thin, highly ionised region, whereas the normal E has a more gradual increase in ionisation with height. Since the ionisation of the sporadic E is often higher than that of the F2 layer it is, at times, responsible for propagation on *much* higher frequencies than are possible by way of the F2, as will be evident from the relatively smaller height at which it lies.

It is extremely difficult, however, from an examination of the measured sporadic E critical frequency data above, to obtain a clear idea of the possibilities of this medium for high-frequency propagation, because of the rapid changes which occur. Critical frequencies of 20 Mc/s are occasionally measured, indicating propagation over a distance of 1,400 miles on 106 Mc/s. On the other hand, critical frequencies as high as 10·9 Mc/s—which would be necessary to sustain propagation on 58 Mc/s—are not by any means a daily occurrence, even during the summer months when sporadic E occurs most often. Yet it would appear from practical results—such as amateur transmissions to Europe on 58 Mc/s and the frequent reception in this country of harmonics of European stations on very high frequencies—that sporadic E of 10·9 Mc/s or higher critical frequency must be an almost daily occurrence during those months, and, furthermore, must occur fairly often during the day.

However, some of this high-frequency communication may not take place by way of sporadic E at all, but by means of tropospheric refraction, and, again, due to the rapid variations of the sporadic E, both in time and space, it may not be observed by an ionospheric measuring station during its hourly observations, even though it is occurring quite often at other places in the vicinity.

However, an analysis of reports of amateur contacts with places in Europe during the summer of 1947 did indicate that communication on 58 Mc/s to the further parts of Europe occurred mostly on days when a large amount of sporadic E or that with high critical frequency appeared in the ionosphere measurements. In fact, so far as communication to places between about 600 and 1,400 miles distant is concerned, there appeared no doubt that transmissions on 58 Mc/s did take place by way of sporadic E.

It seems entirely feasible, therefore, for amateur operators to make deliberate use of this phenomenon during the months May to September inclusive for the establishment of contacts over medium distances, though, whilst sporadic E also *affects* the operation of regular radio services in more ways than one, it does not seem that it could be put to any definite use on a regular basis in the operation of such services.

TROPOSPHERIC REFRACTION

Communication on very high frequencies to places less distant than those mentioned above is often achieved by means of " tropospheric refraction," and although this is not an ionospheric phenomenon and therefore might seem to have no place in this book, a few words about it may be relevant in this chapter.

In Chapter 1 the propagation of the ground wave was described, but, in actual fact, the ground wave is made up of more than one component, and that which was there described is more truly known as the " surface " wave. On very high frequencies this surface wave is of little importance, because it is only effective up to a few feet above the earth's surface. What *is* important is another component of the ground wave called the " space " wave—a direct wave passing from the transmitting to the receiving aerial—and on these frequencies this provides the normal means of communication. The space wave does not, however, travel in a perfectly straight line through the lower atmosphere but, owing to the *normal* variations in the atmosphere with height, curves slightly downwards. This causes the wave to be propagated to distances somewhat beyond the " optical " horizon, and thus the " radio horizon " under *normal* conditions is considerably further away than the optical horizon, for a given height above ground of the transmitting aerial. Under abnormal conditions, however, other effects occur.

The temperature of the lower atmosphere normally decreases steadily with height, but frequently in fine weather the lapse-rate alters, the temperature falls less rapidly than normal and, under some conditions, may even increase with height. Also, the atmospheric water-vapour content may, under certain conditions, decrease more rapidly with height than it does in a " normal" atmosphere. When one or other of these conditions exists, and especially when they both exist simultaneously, a radio wave, if of high enough frequency, may be bent downwards *in the lower atmosphere* sufficiently for it to reach the earth again at distances far beyond the normal radio horizon, and, under some conditions, it may repeat this process again and again, and so travel to considerable distances by these repeated refractions.

It is stressed that the ionosphere plays no part in this process, the wave usually never getting more than a few thousand feet above the surface. The controlling factor in such propagation is the alteration of the dielectric constant of the air with height.

When the weather is fine and settled, horizontal stratifications of refracting air are most liable to exist—when rough and stormy weather prevails the atmosphere becomes more mixed, and the normal lapse-rate in temperature and the normal variation in water-vapour content are likely to prevail. Thus the conditions for the good long-distance propagation of very high-frequency radio waves are that the air above the surface should be exceptionally dry or exceptionally warm—after allowing for the normal decrease in temperature with height—compared with that at the surface.

It has been well established that it is under these fine weather conditions that abnormally long ranges on very high frequencies do occur, and that when the fine weather " breaks up " the radio conditions return to normal. It is to be noted that it is only for very high-frequency waves that this tropospheric refraction is effective—hence the commonest and most effective guiding takes place on wavelengths of the centimetre class—only moderate guiding would be expected on the highest of the short-wave frequencies, and then comparatively rarely.

Nevertheless, in the summer months—when the conditions described above are more likely to exist—communication on some of the higher-frequency amateur bands does occur in this way over distances of a few hundred miles. Because the conditions are, no doubt, bound up with the heating and cooling of the air during the day, such communication is much more liable to occur during the night than during the day, and there appear to be peaks in the signal strengths of such signals around sunset and sunrise. Though definite information as to the maximum ranges likely to be achieved by tropospheric refraction does not seem to exist at present, it is known that distances of 400 miles have been covered on 58 Mc/s in temperate latitudes, and in tropical countries radar ranges on 200 Mc/s have often attained 700 miles, and sometimes as much as 1,500 miles.

Although it is therefore difficult to be certain—because of the difficulty of separating tropospheric communication from that by way of sporadic E, it would seem that many amateur contacts from this country with places in Belgium and Holland, and possibly some with places in Denmark, France and Switzerland, may occur by means of tropospheric refraction. In the summer months, therefore, we might expect such communication sometimes to occur in the higher short-wave frequencies up to distances of about 400 miles.

SCATTER

Another phenomenon by which the amateur sometimes effects transmissions—it is not unknown for the professional engineer to do so—is that of " scatter." It is caused by the fact that the ionisation in the E layer is often of an irregular or " patchy " type, sometimes being in the form of " clouds " of ionisation, and from these some of the energy in the radio wave is " scattered " in all directions, so that it eventually reaches the earth again.

Anyone who has operated a short-wave receiver within the skip zone of any particular station will realise that it is quite untrue to say that *no* signals at all are *normally* obtainable within that zone. Signals of a kind are *normally* obtainable, though they are weak and erratic in character, and are generally not suitable for reliable communication. They are due to this " scattering " of energy from the radio wave as it passes through the E layer. It is usually only when a high-powered transmitter is used that the amount of such scattered energy returned to earth is at all useful in providing, for example, a broadcast signal of programme value, and even then the signal is not, of course, of a character corresponding to that due to the true refracted wave. However, as has been said, signals suitable for amateur contacts are sometimes obtained in this way.

There are, in general, two distinct types of such scatter, known respectively as "short" and "long" scatter. Short scatter occurs from " clouds " in the E layer, where the wave first passes through it on its upward journey to the F. It is to be noted that there is no particular MUF for this type of scattered signal, energy being partially reflected from a wave of any frequency, according to the intensity of the " cloud " : also that the scattered energy may arrive from directions off the Great Circle path between the transmitter and the receiver,

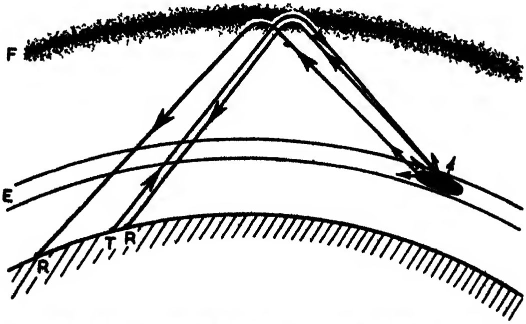


Fig. 53—The mechanism of long scatter. T is the transmitter and RR are receivers within its skip zone

the energy being scattered back obliquely from clouds lying well off the path.

Long scatter occurs by a mechanism which is illustrated in Fig. 53. Here again the energy may arrive at the receiver from points off the Great Circle path, but in this case there is a definite high limiting frequency on which the scatter can occur or, alternatively, a minimum distance for a given frequency from which the scatter can arrive. This is because the energy is only scattered in the E layer after it has travelled to the scatter point by normal refraction in the F layer, and its journey in the reverse direction is made by a similar route. It is necessary, therefore, for the frequency to be lower than the MUF for F layer propagation for the distance between transmitter (or receiver) and the scatter point in the E layer. Later information indicates that some of this long scatter may occur at the point where the ray first strikes the ground, and not at a point in the E layer.

Although, as has been said, it is only when high power is used that *good* signals can sometimes be obtained in this way, amateur operators do obtain contacts by making use of the scattering process, particularly that of the " long " type. The way in which they do so is to direct their transmitted energy upon a distant scatter point for the purpose of obtaining communication with a station which lies within the skip zone of their station. The finding of such a scatter point is, of course, largely a fortuitous matter, but is sometimes achieved by directing the radiated energy towards a point where it is thought that abnormal E layer conditions may exist, such, for example, as towards the auroral zones.

In the next chapter we shall examine some other factors which affect short-wave communication, particularly in the determination of the *lowest* frequency upon which communication may be effected at a given time.

CHAPTER 9

Radio Noise, Ionospheric Absorption and the Low Limiting Frequency

USEFUL FREQUENCY BANDS

AS WAS STATED IN CHAPTER 7, the working frequency for any transmission path at any time of day should be chosen so as to be as near below the OWF—or predicted OWF—as possible, and that when the OWF rises or falls by a considerable amount then a change in the working frequency is indicated. The use of a frequency higher than the OWF would result in there being a danger of the wave penetrating the ionosphere altogether, whilst the use of a frequency too far below the OWF would result in weak reception because of the heavy losses due to ionospheric absorption.

Of course, a certain amount of scope is allowable as to how far below the OWF it is permissible to work, depending on the power available, but the guiding principle should be : " Work on as high a frequency which is below the OWF as is possible, for on this the losses due to ionospheric absorption will be at a minimum, and the received signal strength will therefore be at a maximum."

Thus as we decrease the frequency from the OWF--which represents the high limiting frequency for regular day-to-day operation—the wave will become more and more attenuated because of the losses due to absorption, until finally we reach a " low limiting frequency " for the transmission path, on frequencies below which the signal strength at the receiving end will be below the minimum necessary for good reception. This low limiting frequency is sometimes called the Lowest Useful High Frequency (LUHF) and its determination is more complicated, and the techniques for doing so less reliable, than is the case with the MUF and OWF. For this reason we shall not here attempt to lay down any precise rules for its determination, particularly as some of the data involved in its calculation are not yet completely agreed upon as being valid, and are not, in any case, widely available. We may, however, briefly discuss some of the factors which enter into the matter, so far as they are known.

For any transmission path we have, then, a band of frequencies which are useful for communication between particular transmitting and receiving stations—a band bordered on the one hand by the OWF, and on the other hand by the less easily definable LUHF. It is possible that, on certain circuits at certain times of day, the LUHF may exceed the OWF, in which case the useful band of frequencies ceases to exist, and the circuit becomes unworkable.

The following are the factors which determine the lowest frequency on which effective communication can be achieved—

(1) At every geographical location there is a certain radio noise field strength occasioned by the production of naturally produced radio waves in the atmosphere (mainly from lightning flashes), and their propagation from their source to the location in question. There may also be a certain level of artificial or man-made noise. In order to ensure good reception the signal field strength must exceed that of the noise by a certain value, that is, the signal/noise ratio must have a certain value. The minimum signal strength for tolerable reception is called the " required field strength " and this will vary geographically according to the field strength of the noise at the receiving location.

(2) The required field strength will also vary considerably according to the aerial and type of receiver in use, and the class of radio service being given.

(3) As the working frequency is decreased below the OWF for the transmission path the losses due to ionospheric absorption increase, and the signal field strength at the receiving location consequently decreases. Finally a frequency is reached on which the signal field strength only just equals the required field strength for the service being given, and at the geographical point where the receiver is located. This frequency represents the low limiting frequency, or LUHF.

RADIO NOISE

At the output end of a radio receiver there is always present, in addition to the wanted signal, a certain volume of unwanted sound, called the " noise." The signal will only be of an intelligible—or, alternatively, agreeable—quality when it has a strength relatively great compared with that of the noise. Thus we have the situation where a very weak received field with a low noise may provide a more intelligible signal—or a more agreeable programme—than will result from a very strong signal field which is accompanied by a high noise level. In other words, the signal/noise ratio is of more importance than the strength of the signal itself.

Speaking very broadly, there are two categories of noise : that produced within the receiver itself—called the " set noise "—and that picked up by the receiving aerial together with the wanted signal—called the " radio noise." The set noise sets the absolute low limit to the value of signal field strength necessary to produce a workable signal with a given receiver, since, in the absence of all external noise, it will still be present, and so limit the working capabilities of the receiver. But, since the set noise is largely a matter of receiver design we are not concerned with it here, but will go on to consider the phenomenon of radio noise.

The radio noise, it will be appreciated, is an important quantity in a communication system, affecting the design of both transmitting and receiving systems. It may be considered to consist of three separate components—

(1) Interfering waves produced by man-made devices, whether such devices are for the purpose of emitting signals or not.

(2) Interfering waves produced by natural phenomena in the earth's atmosphere.

(3) Interfering waves produced by phenomena outside the earth's atmosphere.

The problem of man-made noise exists mainly in urban and industrial areas, and in these localities may reach such levels as to necessitate very high signal fields to ensure good reception. But perhaps the only solution to this problem is that of State legislation against the radiation of electrical interference, so we shall not discuss the matter here. The third category of radio noise mentioned above is briefly discussed in the last chapter of this book, which leaves us now to deal only with the problem of noise produced in the atmosphere.

Most of this naturally-produced radio noise is caused by lightning discharges occurring in the troposphere, which radiate electro-magnetic waves that are picked up by the receiving aerial, and so give rise to the interfering sounds at the output of the receiver. Thunderstorms are occurring in some part of the world at practically all times of day and night, and at each lightning stroke an enormous amount of power is radiated, and the stroke can be regarded as a radio transmitter radiating power on all frequencies. The radio noise intensity actually produced is, broadly speaking, extremely high on the lowest radio frequencies and, being inversely proportional to frequency, reaches almost negligible proportions on the ultra-high frequencies.

PROPAGATION OF THE NOISE

" Local " thunderstorms, that is, those occurring within a few hundred miles of the receiving station, affect the receiver by reason of the " ground " wave from the lightning strokes, and give rise to the well-known " crashing " noises which often mar medium-wave reception in this country, particularly during the summer months. But the waves produced by lightning strokes are propagated, not only as ground waves, but as sky waves as well, so that the greater part of the radio noise present on short waves in the temperate regions arrives from far distant sources.

The principal thunderstorm-producing areas of the world are certain land areas in the tropical regions, and it is thought that the distant noise which affects the temperate zones comes from these areas. The exact location of these " noise zones " varies with the seasons, and, in general, they move north and south over several degrees of latitude with the sun. The frequency of the lightning flashes occurring in these tropical thunderstorms is far greater than in those which occur in temperate zones, and, at a distance, the noise emanating from them is more in the nature of a continuous " crackling " than of an occasional " crash." The thunderstorms occur much more frequently during the local afternoon and evening than during other times of day, and there is a diurnal variation in the intensity of the noise of about five to one, the maximum occurring in late afternoon and the minimum in early morning. Near the noise zones themselves the noise received is principally that due to the ground waves from the lightning strokes, the sky waves " skipping " over the area. Its intensity is very high on the low frequencies and decreases to low values on the ultra-high frequencies but, generally speaking, is higher on all frequencies than at a long distance from the noise source.

At a long distance from the noise zones the noise varies in a complex way with frequency, that on the lower frequencies being largely due to

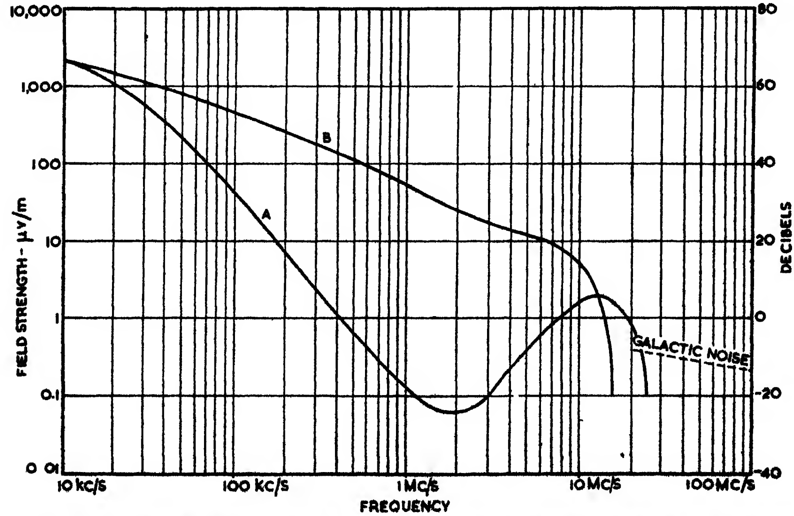
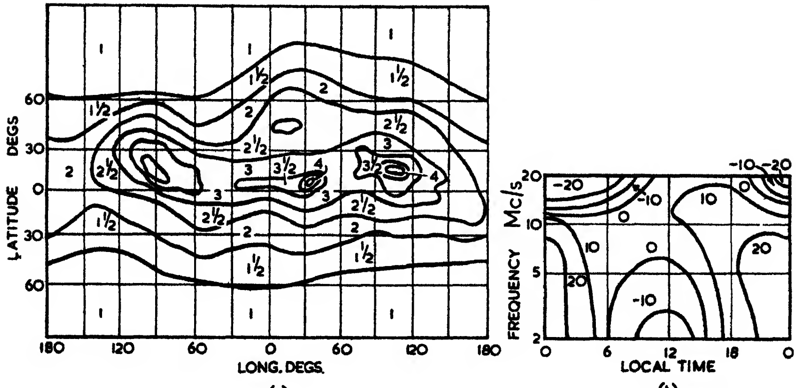


Fig. 54—Required field strength for good reception of radiotelephony in the presence of noise in the Northern Temperate Zone—A at noon and B at midnight



(a) (b)

Fig. 55—(a) Noise-chart of the world for the months of June, July and August, and (b) noise curves for noise grade 2½ during summer, showing required field strength for radiotelephony for various frequencies and times of day. *(From information given by Messrs. K. W. Tremellen and J. W. Cox, M.A.)*

ground wave, and on the higher ones entirely due to sky wave, propagation of the noise. As one gets further and further from the zones the general level of noise decreases, so that in Arctic regions its intensity is lower than anywhere else in the world.

Fig. 54 gives an idea of the nature of the noise variations with frequency in the northern temperate zone at two times of day. Curve A is for midday, and shows a gradual decrease in noise intensity with frequency towards about 2 Mc/s, due to poorer propagation of the noise. Then there is a considerable increase towards the short waves due to better sky wave propagation, and finally a big decrease towards about 25 Mc/s due to penetration of the ionosphere by the noise sky waves. Curve B—for midnight—shows different characteristics, because sky wave propagation of the noise now occurs even on frequencies in the medium wave range. The noise reaches low values on lower frequencies than during the day because at night the ionosphere cannot sustain sky wave propagation on the higher frequencies. Curves for other times of day exhibit characteristics partly of the midday and partly of the midnight type. It is to be noted that, on frequencies above about 20 Mc/s, the extra-terrestrial noise (discussed in the last chapter) is often of a greater intensity than is the atmospheric noise.

NOISE CHARTS

From a study of the noise producing areas and of the diurnal and seasonal variations in the intensity of noise produced, and in its propagation, it has been found possible to construct noise charts showing the geographical distribution of the radio noise over the world's surface during different months of the year. As an example of such a chart Fig. 55 (*a*) is given, which shows conditions during the months of June, July and August, the noise intensity being shown according to a number of noise " grades," from grade 1 to grade 4½. It will be seen that the areas of highest noise grade are in tropical regions, albeit they lie, during the months in question, somewhat north of the equator.

Having ascertained, from such a chart, the noise situation at a particular receiving location, it is then necessary to consult the noise curves for the particular grade in which the receiver lies, in order to find the signal strength necessary to provide a workable signal in the presence of the noise. Fig 55(*b*) gives the curves for grade 2½ in summer, showing the field strength for various frequencies and times of day.

It should be added that the details of the distribution of radio noise over the world's surface are not yet fully known, and more data is required before the above methods can be regarded as really reliable.

REQUIRED FIELD STRENGTH FOR VARIOUS CLASSES OF SERVICE

The required field strength for satisfactory reception varies widely according to the class of radio service being given. So far as the field required in the presence of radio noise is concerned the reference level is that for intelligible radiotelephony, the required field strength being shown in the curves of Fig. 55(*b*) relative to one microvolt per metre. The corresponding fields for some other classes of radio service are given in Table 5.

IONOSPHERIC ABSORPTION

The absorption to which a wave is subject in the ionosphere will determine the distance at which the required field strength can be put down with a given radiated power. As we have already seen, the free electrons of the ionosphere are set in motion by a radio wave, and so long as their motion is unrestricted what energy is taken from the wave is given back by the electrons when in motion, so that no energy is lost. But their motion is not always unrestricted, for, from time to time, they collide with molecules of gas, and, during collision, lose the energy which they possess at the moment of impact. This energy comes from the radio wave, which thus loses it and so becomes attenuated.

The energy lost by the wave in this way will depend on the number of collisions made per second, and on the frequency of the wave. The number of collisions per second is, of course, much greater relatively low down in the atmosphere than at the height of the F layer, because lower down the air is much denser and therefore the chance of collision with a molecule much higher. The greater part of the absorption, therefore, occurs in the D layer and the lower part of the

TABLE 5

Class of Service	*Signal strength for intelligible reception, in dbs above that required for radiotelephony*
Double-sideband radiotelephony	0 (reference level)
Single-sideband radiotelephony	6
Standard broadcast service	26
International short-wave broadcast	15
Manual C.W. telegraphy	− 17
High-speed automatic C.W. telegraphy	− 11
Double-sideband telephoto system	0

E, though some also occurs in the F layer when the wave penetrates a long way into it and is greatly retarded, as near the critical frequency of the layer. In practical transmission, however, the D and E layers can be regarded as the absorbing regions. The absorption will be greater the greater the time spent in traversing these regions, that is, it will, in general, increase as the length of the transmission path is increased. It is at a maximum on a frequency of about 1·5 Mc/s.

On *short waves* this absorption is inversely proportional to the square of the frequency, apart from that occurring in narrow bands near the critical frequencies of the different layers. It will be seen that this is so because the motion imparted to the electrons is less the higher the frequency, and thus the higher the frequency the less are the number of collisions occurring per second.

The absorption varies greatly with time of day and season of the year, the electronic content of the absorbing layers being proportional to the sun's altitude. Thus it varies from an extremely low value at night to a maximum around noon, and it is much greater in the tropics, where the sun is more directly overhead, than in the lower geographical latitudes. The seasonal variation is, however, not at all clear, for during the winter months there sometimes occur days of high absorption, when, according to the sun's altitude, it should be at a low value. In the auroral zones a special type of absorption occurs, which appears to be connected with the degree of ionospheric disturbance prevailing in those zones.

It is possible to construct world charts from which the absorption over any transmission path may be calculated. These charts are based upon the sun's zenithal angle at all points on earth, and the absorption per unit distance of the path is considered to be proportional to the electronic content of the E and D layers. However, world-wide experimental data on this subject is lacking at this time, and the validity of such charts cannot be considered yet to be proven.

THE LOW LIMITING FREQUENCY

It will be seen that if the noise level prevailing at every point on earth at any time is known, and if the signal/noise ratio for each class of radio service is known, we can then state the required field strength for any class of service at any receiving location at any time. Then, from a consideration of the power radiated and the absorption to which the wave is subject over the transmission path, we can find the actual field strength at the receiver on every frequency. The frequency on which the field strength put down just equals the required field strength is the low limiting frequency or LUHF, and on this and higher frequencies up to the OWF regular service should be possible.

It should be noted that the LUHF, unlike the MUF, is not a fixed and definite quantity set by conditions in the ionosphere itself, but will vary with the nature of the service being conducted. For example, ionospheric absorption can be overcome by an increase in power radiated, or by the use of an aerial with greater directivity, and thus the LUHF may be lowered, whereas no such alteration of the MUF is possible by similar means.

As has been said, in the absence of reliable world-wide data from which to calculate the LUHF, the best procedure is to work as near to the OWF as possible, for there the field strength at the receiving location will be at its greatest. As the frequency is reduced the absorption increases until finally the LUHF is reached, and on frequencies below this the field strength will be insufficient to give the service required.

In the next chapter we shall consider some of the abnormal ionospheric conditions which from time to time occur, and which have very undesirable effects upon short-wave communication.

CHAPTER 10

Ionospheric Storms and Other Phenomena

General and Particular Effects of Sunspot Activity

Since the ionisation of the upper atmosphere is brought about mainly by the action of the sun's ultra-violet radiation, and since the latter is, generally speaking, increased when the sunspots are numerous, the general effect of increased sunspot activity is to raise the critical frequencies of the ionospheric layers. Fig. 25 showed us that the critical frequency of all the layers does, in fact, increase in conformity with the increase in sunspot activity.

The net result of the increased ionising radiation at the solar maximum is that short-wave propagation is improved, for with its increased electron population the F layer is able to refract much higher frequencies than is the case at the minimum, and we are thus able to extend upwards our range of useful frequencies for short-wave communication. Our high-limit frequency is much increased. At the same time the ionospheric absorption—whilst it is certainly greater at the maximum than at the minimum, due to increased ionisation in the D layer—is not increased in the same *degree* as is that of the refracting power of the higher layers, and so the low-limit frequency—set largely by the absorption—is not unduly raised. The result is that the band of useful frequencies is broadened with increased solar activity, and beneficial results to short-wave communication ensue.

But the solar activity produces other effects in the ionosphere, for in connection with it there occur from time to time great upheavals on the sun, from which *extra* radiation is emitted, such as cause abrupt changes in the ionisation in various parts of the atmosphere. This gives rise to abnormal behaviour on the part of radio waves, and to a condition often leading to their complete failure as a means of communication. These disturbances are of two distinct kinds, having, so far as their effect upon the short waves is concerned, characteristics of a very different nature. One of these is the " Dellinger fade-out " (after Dr. J. H. Dellinger, of U.S.A.), or simply " radio fade-out," and the other is the " ionospheric storm." It is important, therefore, to distinguish between the effect of the *general* increase in the sun's activity—which tends to *improve* short-wave conditions—and the effect of local disturbances which occur on and within it.

Though the sunspots may themselves be regarded as centres of " disturbance " in the sun, near which the upheavals occur, and from which the excessive radiations of various kinds emanate, it is extremely difficult to obtain definite correlation, in a large number of instances, between particular sunspots and terrestrial disturbances. For instance, the mere size of a sunspot (apart from that of a few " giant " sunspots) is no criterion as to its potentialities for causing ionospheric and other terrestrial disturbances. The sunspots occur in different solar latitudes and, since they move across the sun's surface as the sun itself rotates, there seems to be a critical area near the sun's central meridian from where their disturbing radiations are more likely to have terrestrial effects, than when they are in other solar longitudes.

If a sunspot is found to be spectroscopically active, then, irrespective of its size, it has greater potentialities as an ionosphere " disturber " when in the central area than has a sunspot which is not so active.

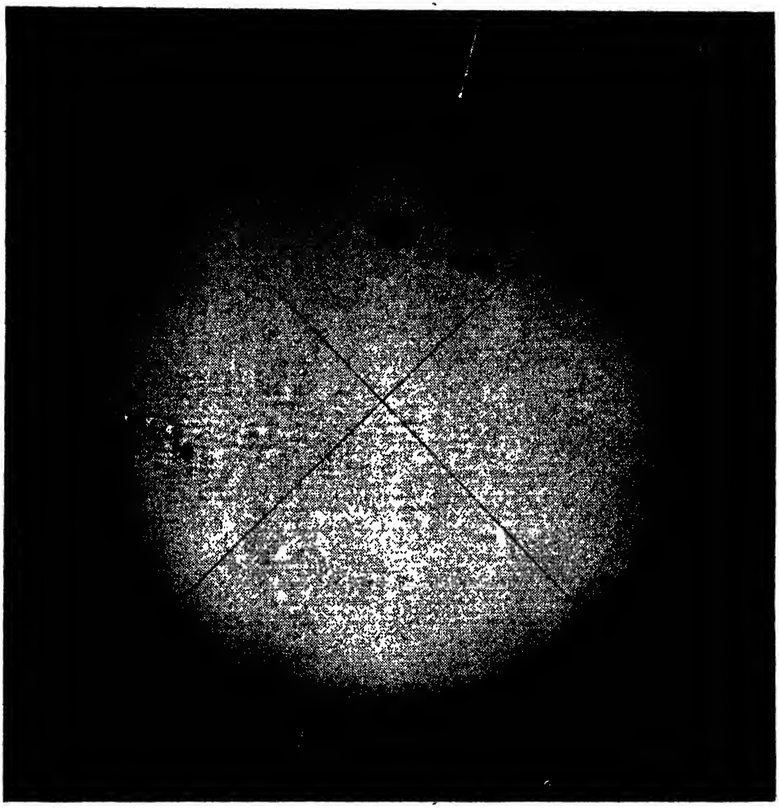
Another feature of the sunspots is their magnetic polarity, for it has been found that there is a magnetic field associated with them, and it has been suggested that this magnetic polarity has something to do with the incidence of storms in the two terrestrial hemispheres.

But it cannot be said that this latter point is by any means clearly proven, and there remains much to be learnt about this subject. So let us go on to consider something about the disturbances themselves and we may then mention the supposed relation with *particular* sunspots.

SOLAR FLARES

Fairly frequently there occur eruptive disturbances on the sun, from which are emitted not only huge quantities of gaseous matter, but also a large amount of energy in the form of radiations of various wavelengths, including those of visible light. The disturbances are thus observed from the earth as " flares " of bright light. They often take place in the vicinity of sunspots, and if they occur on that side of the sun which is facing the earth and are of sufficient intensity, they produce certain terrestrial effects which occur only on the earth's sun-lit hemisphere. The effects may be produced no matter on what part of the visible disc of the sun the flares occur. The radiations from the disturbances would appear to be sent out in all directions and, as the terrestrial effects start at the same time as the flare is observed, they must be due to waves travelling at the speed of light. The flares usually last but a few minutes, though their terrestrial effects are still felt some time after they have died down.

The radiations of ultra-violet wavelength which are emitted seem to have the property of penetrating the higher ionospheric layers and of producing a layer of intense ionisation at about the height of the D layer—just below the E. The intensity of the ionisation there produced continues to grow so long as the waves continue to arrive, but when the flare dies down and the wave radiation ceases, the ions and electrons



A GIANT SUNSPOT

Photograph of the sun taken on 5th February, 1946, showing, in solar latitude 27° N., a sunspot of area nearly 5,000 millionths of the sun's hemisphere, equivalent to about 100 times the cross-sectional area of the earth. An intense solar flare occurred over this sunspot region on 6th February, and 18½ hours later, that is, at about 10.18 on 7th February, a severe geomagnetic storm began. This was accompanied by a severe ionospheric storm and radio disturbance. (*Royal Greenwich Observatory photograph; published by courtesy of the Astronomer-Royal*)

start to recombine. Though they do this at a very rapid rate owing to the great density of the gas, an hour or more may elapse before normality is restored, owing to the great intensity of the ionisation which the waves produced in the first place.

Coincident with the time of observation of a bright solar flare there is often a sudden cessation of all short-wave signals. This result of the flare is the "Dellinger fade-out," and the correlation between the two phenomena is now conclusive.

DELLINGER FADE-OUTS

The effect on short-wave reception is very impressive, for everything appears to go " dead," even the background noise disappearing. All frequencies within the short-wave range may be affected, though the disturbance is usually more intense and lasts longer on the lower frequencies. These fade out before the high frequencies, and do not come in again until after the latter have done so. On the frequencies affected, signals usually fade away entirely within a minute or two, and may not return until anything up to two hours later, signal strength increasing from zero in a more gradual manner than that in which it was reduced. Only the sky waves are affected, and only those over transmission paths passing through the sunlit hemisphere. It has been found that on transmission paths which pass through low latitudes—where the sun's rays are more perpendicular—the disturbances are more intense than on paths which pass entirely through high latitudes.

The cause of the short-wave failure is the sudden production of greatly increased ionisation in the D layer, in which the radio wave causes collisions between the electrons and the neutral atoms. In so dense a region the number of collisions occurring is very great, and as at each collision energy taken from the wave is expended the result is such a heavy absorption of energy from the wave that it is often completely lost. The absorption is greater on low than on high frequencies, whilst it is greater in low than in high latitudes, because the ionisation produced is greater where the sun's radiations are most intense. For this reason transmission paths in the dark hemisphere are unaffected by the disturbance. When the ionising radiation ceases to arrive—after the flare has died down—the electrons and ions start to recombine, and the absorption of the wave decreases. Owing to the density of the gas, the recombination rate is high, and, as the ionisation disappears, the short-wave signals come in again, the higher frequencies first.

Such disturbances are accompanied by a sudden brief fluctuation of the earth's magnetic field, caused by the sudden presence within it of large numbers of moving ions, whose movement is in reality a vast electric current carrying an associated magnetic field. This geomagnetic disturbance, like the ionosphere disturbance, is confined to the sunlit hemisphere, and is more intense in low than in high latitudes, and in these and other respects it is quite different in character from the " magnetic storm." A radio " fade-out " of this kind, with its accompanying magnetic disturbance, is illustrated by the field strength and magnetic intensity records shown in Fig. 56. It should be noted that since the solar flares are random occurrences there seems to be no possibility of making forecasts of *their* happening, and therefore none of predicting the onset of a Dellinger fade-out.

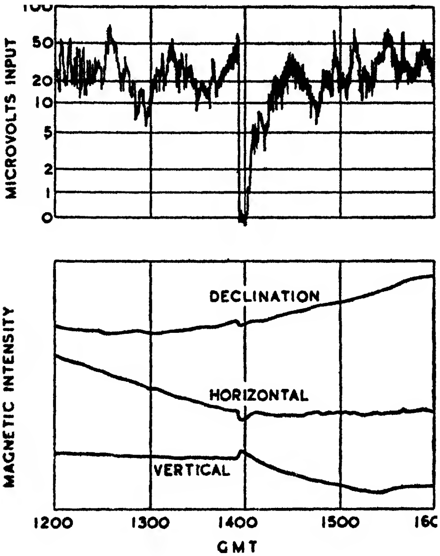


Fig. 56—Sudden disturbance of the ionosphere on 6th April, 1936, as shown by fade-out of 13·525 Mc/s transmission from GLH, Dorchester, observed at Riverhead, New York, and coincidental terrestrial magnetic perturbation recorded at Cheltenham, Maryland

IONOSPHERIC STORMS

The ionospheric storm constitutes the major form of disturbance to short-wave communication because, while usually its effect is not so intense as that of the Dellinger fade-out, it is of much greater duration. During ionospheric storms short-wave signals on wavelengths normally well received drop to a very low level, and often disappear entirely. There is almost always a great increase in the amount of fading experienced, and a prevalence of the particular type known as " flutter " fading. The higher frequencies are most affected, and there is no discrimination between the sunlit and dark hemispheres. Transmission paths in low latitudes are less affected, however, than those passing through high latitudes, the paths most severely disturbed being those which pass through certain zones centred on the geomagnetic poles.

Ionosphere measurements indicate that the principal effect is in the upper part of the ionosphere, and that the F layer is usually at an abnormally great height during the disturbance and has an abnormally low ionisation density. The storms usually last for several days. The critical frequency and virtual height graphs of Fig. 57 show how the values for these quantities depart from the normal on days of ionosphere storminess.

It is thought that disturbances of this type are caused by the emission of streams of particles from the sun, and that these corpuscles—which may be particles of ionised calcium—are shot out from disturbances which occur in the vicinity of sunspots. There is strong evidence that the corpuscular stream is emitted at the time of occurrence of a bright solar flare. Although part of the emission is known to fall back upon the sun, some of it appears to be forced outwards by radiation pressure until it escapes from the sun's atmosphere. It would thus leave the sun at the same time as the wave radiation which produces the Dellinger fade-out, but the corpuscles would travel much slower than the ultra-violet electromagnetic waves, and so would not reach the earth until some time later.

Furthermore, while the wave radiation is emitted in all directions, and so reaches the earth irrespective of the position of the flare on the visible disc of the sun, the corpuscular radiation seems to be in the form of a cone-shaped jet, with the flare at its apex. So that, unless it is emitted from a position on the sun which is " pointing " more or less towards the earth, it misses this planet altogether. It has been noticed, for instance, that when a Dellinger fade-out occurs as the result of a bright solar flare *near the central meridian* of the sun, it is often followed —from 17½ to 36 hours later—by an ionospheric storm, but that disturbances due to flares occurring in other parts of the sun are not followed by a storm.

Apart from this, it is often noticed that when an ionospheric storm starts there is a sunspot in a position about ½ to 1½ days past the central meridian in the direction of rotation of the sun. So it would appear that if there is a disturbance on the sun near its central meridian—whether visible as a bright solar flare or indicated by a sunspot—the corpuscles which are shot out do encounter the earth and produce an ionospheric storm possibly about 30 hours after leaving. This would indicate that they had travelled through space at a speed of about 900 miles per second.

It should be added that the correlation between the start of ionospheric storms and the solar phenomena is not by any means perfect, some storms occurring when no flare has been seen and when there is no spot near the central meridian, though there may be spots in other parts of the sun. This may be because the solar disturbance has not yet produced a visible sunspot, or because the corpuscles are not always emitted in a direction normal to the sun's surface. On the other hand, there is a definite tendency for some storms to recur at intervals of about twenty-seven days, corresponding roughly to the average rotation period of the sun, which would indicate that the same disturbance had produced an ionospheric storm during its successive passages across the sun's central meridian.

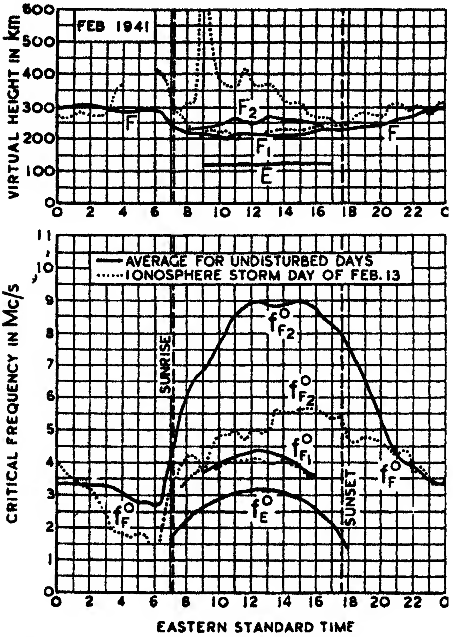


Fig. 57—Virtual heights and critical frequencies of the ionospheric layers, observed at Washington in February, 1941, by the U.S. Bureau of Standards

On reaching the earth's atmosphere the corpuscles are affected by the geomagnetic field, which carries them in the direction of the magnetic poles. Consequently, their effects, both upon the ionosphere and in other ways, are most intense in zones around the poles. A state of turbulence is set up in the ionosphere, particularly in the upper layer, leading to erratic conditions for the refraction of radio waves, with consequent abnormal fading. The F layer then appears to expand and to rise, and the stratification of the ions is upset. The ionisation per unit of volume is thus reduced, so that waves which are normally refracted begin to penetrate the layer. At the same time there appears to be an increase in absorption in the lower layers.

According to one observer of an exceptionally severe storm, the F layer continued to decrease in ionisation and to rise in height until it finally disappeared altogether, when an entirely new layer appeared lower down. This, in turn, behaved in a similar manner until it, too, disappeared, to be followed again by another layer, the time between the appearance and disappearance of a layer being about three hours, and the phenomenon becoming less and less evident until eventually the turbulence subsided.

Ionospheric storms have caused the *entire disappearance* of signals on certain frequencies in a few cases for as long as two days. Though not usually so intense as this, the average time for which conditions remain *subnormal* is between one and two days, whilst they have been known to remain so for nearly a fortnight. During the storms the highest frequencies which the F layer will refract may be reduced by as much as 50 per cent below normal.

In the prediction of ionospheric storms a *moderate* amount of success has so far been achieved. Long term indications that a storm is likely to occur may be given by the twenty-seven day recurrence tendency mentioned above (although this often fails), or by the observation of giant sunspots or of solar flares in the critical zone near the centre of the sun's disc. Short term indications may be obtained from the observation of fluctuations in the field strengths or bearings of distant radio stations whose signals pass through the auroral zones, and by noting the occurrence of abnormal fluctuations in the geomagnetic field. But a great deal of work remains to be done before the predictions can be considered very reliable.

MAGNETIC STORMS AND POLAR AURORAE

Ionospheric storms are almost always accompanied by "magnetic storms," that is, by violent fluctuations in the geomagnetic field. The intensity of the storms is often great, that is, the range of variation in the magnetic elements is of large magnitude. Like the ionospheric storm, the magnetic storm is most intense in the (geomagnetic) polar regions. Though the two phenomena are clearly connected, the start of the ionospheric and magnetic storms does not appear to be always exactly simultaneous, and the ionospheric storms nearly always persist for some time after the geomagnetic field has returned to a "quiet" state. Magnetic storms are caused by the abnormal movements of ions and electrons within the geomagnetic field, due to the action of the solar corpuscles.

In a recent analysis of sunspot and geomagnetic activity (international magnetic character figures) H. W. NEWTON confirmed the existence of a definite relationship between the larger sunspots (particularly the giant spots) and the peak value of magnetic activity during the solar disc passage of the sunspots. This is a statistical result only and is not operative in every case. The peak occurred about 1½ or 2 days after the sunspots crossed the sun's central meridian. This relationship rapidly diminished with decreasing size of the sunspot, and for the smaller sunspots it practically disappeared, which, however, does not mean that such sunspots are *never* associated with magnetic storms. The correlation was greatly strengthened if only those sunspots which were associated with solar flares were considered. Very intense flares showed a remarkable association with the more intense magnetic storms. Flares of lesser intensity, and even those of minor intensity when unusually frequent, near a particular sunspot, were also clearly connected with a rise in magnetic activity after the sunspot had crossed the central meridian. It was found that the " great " magnetic storms did not tend to recur after an interval of about twenty-seven days, but that small storms did display this tendency. The relationship of many of the smaller magnetic storms to sunspots and solar flares still remains very obscure. These results are of great interest in the radio field in view of the close connection between geomagnetic and ionospheric storms.

Another effect of the solar corpuscles is the production of the polar aurorae, which are intimately associated with variations in the geomagnetic field, and which occur frequently in the zones surrounding the magnetic poles. In these zones, incidentally, ionospheric and magnetic storms are of greater intensity and occur with greater frequency than elsewhere, so that transmission paths which pass through them are much more liable to disturbance than those traversing lower latitudes. The coloured light of the aurorae is due to the emission of visible rays by atoms of atmospheric gas when subjected to bombardment by the solar corpuscles. The height at which the aurorae occur has been measured, and it is thought that the lower limit of about fifty-five miles above the earth's surface represents the furthest distance that the corpuscles penetrate into the earth's atmosphere. Although the aurorae are usually confined to the zones around the magnetic poles, there are occasions when they are observed over much greater areas. The implication is that, under these conditions, the stream of solar corpuscles entering the atmosphere is of exceptional intensity, and such occurrences are almost always accompanied by ionospheric and magnetic storms of very great severity.

It should be noted that it is not the magnetic storm which causes the ionosphere disturbance—in fact rather the reverse seems to be the case, the abnormal electronic currents in the ionosphere upsetting the normal geomagnetic field. Also that radio failures which accompany displays of the aurorae are not, as is so often stated, caused by the aurorae themselves. The aurorae are rather an additional and secondary effect produced by the same solar corpuscles which cause the ionospheric storm and consequent radio failure.

SOLAR AND GALACTIC NOISE

Some time before the 1939–1945 war it had been noticed by certain amateur operators that, at the beginning of a radio fade-out of the " Dellinger " kind, a peculiar " hissing " noise was audible in their receivers. This was later shown—by Sir EDWARD APPLETON—to be due to the reception of radio waves emitted from the sun. It had long been thought that, since the sun radiated visible and thermal radiations corresponding to a black body at 6,000° K, its radiations should include waves of " radio " frequency corresponding to the same temperature, though these would be so weak as to be undetectable with ordinary radio receivers. It appears, however, that during the eruptions which produce solar flares, the sun's radiations in this part of the spectrum are enormously increased, and that they then become audible in ordinary receivers even on frequencies in the ordinary short-wave bands, as had been noticed by amateurs. Solar radio noise on these wavelengths does not, however, last very long, because, on the building up of the ionised layer which absorbs the terrestrial short-wave communication, the solar radio waves are also absorbed, and so cannot get through to the earth's surface.

There is another form of radio noise of extra-terrestrial origin, which was first observed by K. G. JANSKY in 1932, and which is sometimes known as " star static " by reason of its origin in the stars. JANSKY, by the use of a sharply directional aerial system, found that it varied diurnally in a systematic way in its direction of arrival. He found that its direction changed nearly round the compass every twenty-four hours, except that, in the middle of the night, when its direction had reached north-west, it also began to come in from the north-east, and so commenced another cycle of variation.

At first he thought that it must be coming from the sun, but later disproved this, and found that it was coming from the direction of the Galaxy, a fact which has since been confirmed by G. REBER and other workers, by the use of improved aerial arrangements for reception of the noise. The means by which this galactic radio noise is produced is not yet fully understood, though theories have been advanced to account for its origin. On frequencies above about 20 Mc/s, this noise is often the principal source of interference to reception (See Fig. 54).

FADING

Fading is a phenomenon which is almost always present, to a greater or lesser extent, in short-wave reception, and, for this reason, indirect reception by way of the ionosphere is almost never of the same steady quality as is that given by direct reception of the ground wave. In the broadcast services, for example, indirect reception is classed as " second grade" service irrespective of the distance over which it is given, as compared with " first grade " service given over a considerable portion of the service area by direct reception. The term " fading " as used above refers to the random variation of signal strength at the receiving point, and should not be confused with the " fade-out " which we discussed earlier in this chapter. Fading of the short-wave signal is due to the fact that the signal is comprised of a number of different " rays " of radio energy which have reached the receiver after travelling *via* the ionosphere over paths of different lengths. As the lengths of the different paths are constantly changing due to varying ionospheric conditions, and as the changes in path length for the various rays are not always the same, but vary for the different rays, the energy in the arriving rays " adds up " in a random manner.

At one instant, for example, the path length for two of these rays may be such that they arrive at the receiver so that the energy in one reinforces that in the other, whilst at the next instant the path lengths may have changed relative one to the other so that the energy in the second ray may completely cancel out that in the first. It is the " phase " of the different rays which determines the resulting signal strength, and the interference between the different down-coming rays will cause the signal to vary in strength in a random manner.

In addition to this there is the fact that during its passage through the ionosphere the " polarisation " of a wave, that is, the direction and magnitude of its electric and magnetic forces, changes in a very complicated manner, so that the polarisation of the arriving wave may be constantly changing with time in a random manner. Most receiving aerials absorb more energy from a wave which is polarised in one particular plane than they do from those polarised in any other, so it will be clear that this constantly changing polarisation will contribute something to the fading.

The sky wave signal is, therefore—unlike that due to the ground wave—almost never of constant strength. The character of the fading varies considerably with varying ionosphere conditions. It exhibits different characteristics in respect of time—" rapid " or " slow " fading—or in respect of intensity—" deep " or " shallow " fading—all depending, as has been said, upon the prevailing ionosphere conditions. Deep fading is especially prevalent near the outer edge of the skip zone, for there, at one instant there may be many rays coming down at the receiver location whilst at the next instant the paths may have so changed that all the rays pass over the receiver and reach the earth at a point more distant from the transmitter.

There is one type of fading which is particularly serious in the case of broadcast reception, since it leads to bad distortion of the received programme. This is known as " selective " fading and is due to the fact that, because the path length in the ionosphere varies with frequency, the fading may be different for frequencies which differ by only a few hundred cycles. A speech or music modulated carrier is made up of a large number of different frequencies, and the " quality " of the received programme is dependent upon these having the same relations as were present in the transmitted programme. If, therefore, the different frequencies are propagated in a different and changing fashion in the ionosphere, the received programme will be distorted.

During ionospheric storms a peculiar form of fading known as " flutter " fading is often experienced. In this the variation in signal intensity takes the form of a fast rhythmic beat, almost of the nature of a low-frequency oscillation superimposed on the modulated carrier. It is well described by the term " flutter " fading but, apart from the fact that it is due to disturbed conditions in the F layer, not enough is known about its cause to permit of explanation of the specific reasons for its peculiar character.

IONISATION BY METEORS

Meteors, or " shooting stars " as they are more familiarly called, are small particles of matter—ranging from those a few inches long to those of microscopic size—which are continually entering the earth's atmosphere from outer space. Broadly speaking they are of two kinds, those which arrive from random directions, and those which arrive in the form of " showers " at certain times, and which are really the remains of disintegrated comets. It has been computed that hundreds of millions of the former kind enter the atmosphere every 24 hours.

Some years ago it was noticed by two engineers of All India Radio that, whilst receiving the unmodulated carrier wave of a high-powered transmitter at a location a few miles distant, heterodyne whistles of an unusual character could be heard. These were of short duration, and commenced as a high note which fell rapidly in pitch, and usually died away before reaching zero frequency. A similar phenomenon had also been observed at a B.B.C. receiving station in England. The Indian engineers concluded that the whistles must be due to a Doppler effect arising from the interference of the direct ground waves from the nearby transmitter, by waves which were being reflected from a rapidly moving reflecting surface, and later were able to establish, by direct observation, that the whistles were produced by meteors.

It was then seen that a meteor, by reason of its very high velocity, must produce a local area of ionisation in its track, and that radio waves were reflected from this area so as to beat with those received direct from the local transmitting station. At first it was thought that the fall in the pitch of the whistles was due to the fact that the meteors were very rapidly retarded by the earth's atmosphere, but it was later shown that the pitch of the whistle would fall even if the meteor approached at a constant velocity. The character of the whistle depends, in fact, upon the track of the meteor in relation to the position of the observer. Many more meteors can be heard in this way than can ever be seen, and probably only a small proportion of those entering the atmosphere are heard.

Having established the fact that ionisation of the atmosphere can be caused by meteors, some interesting speculations—and experiments—have been made as to their effect upon the ionosphere, and in causing other short-wave phenomena. In the U.S.A., for example, it was noticed that F.M. stations working on very high frequencies would sometimes burst in upon the signals of other stations working upon a similar frequency, but far beyond their normal range. It has been concluded that these " bursts " are due to the reflection of the radiated energy by areas of ionisation produced by meteors, thus causing the waves to reach the earth at points very far beyond the normal very short-wave range.

During recent years radar methods have been employed for the observation of meteors. By this means transient echoes are obtained from the region of the meteor trails, indicating that an area of ionisation exists there, and that it persists for a short time. Observations made on meteors belonging to some of the well-known showers have indicated that while the duration of the echoes does not, in the majority of cases, exceed one second there are some which last up to 20 seconds.

The heights from which the transient echoes come have also been measured, and one set of such measurements indicates that the virtual height ranges from about 80 to 160 km (50 to 100 miles). The majority, however, came from a height of from 110 to 120 km (69 to 75 miles), which would correspond to the height of the ionospheric E layer.

From the observations it has been possible to compute the velocities of the meteors, and it appears that these vary widely. Some recent observations gave velocities ranging from about 17 to 46 km (11 to 29 miles) per second, with a maximum in the region 35 to 40 km (22 to 25 miles) per second. It seems probable that the meteors having the greater velocities are those travelling at the greater heights, the velocity being affected by the atmospheric gas density at different levels.

The fact that the majority of meteors seem to penetrate the atmosphere to about the height of the E layer has led to speculation as to whether they may not have something to do with some of the abnormalities which occur in that region. Two distinct possibilities emerge :—(1) The low night-time " residual " ionisation of the E layer—which cannot be accounted for on the basis of solar ultra-violet ionisation—may be maintained by the continual arrival of showers of meteoric dust. (2) The phenomenon of sporadic E, consisting of highly ionised " clouds " in the E layer, may be due to ionisation by meteors.

Neither of these possibilities has yet been proven, but the radar technique is proving very useful, if only because it enables the meteors to be kept under observation by day as well as by night, a thing not hitherto possible.

Conclusion

HAVING REGARD TO ALL that has been said in the last chapter, the reader may begin to wonder, not that a received short-wave signal is sometimes weak and unreliable, sometimes fluctuating and distorted ; but that a signal travelling via the ionosphere is ever receivable at a far distant point at all. The fact remains that the short waves have become the main—and certainly the most economical—medium for long-distance radio communication, and that, without the ionosphere, no such communication would be possible at all. Not that that would necessarily be a disaster for mankind, for, without the ionosphere, human life on the earth would be impossible, so it is doubtful whether such communication would be missed.

It is true that the signals received by short waves are often inferior in quality to those provided by short range direct-wave systems, and it is probable that this will always be so, for it is difficult to see how each of the ionospheric vagaries is to be completely overcome. Nevertheless great improvements in equipment and in operating techniques have taken place in recent years, and are still going forward, and there is no doubt that the present-day short-wave signal is more reliable than ever before.

But this book has done nothing if it has not shown that there is a great deal yet to be learnt about the ionosphere, and about its causation and its effects. Such knowledge is, however, steadily, if slowly, being acquired and as it advances there follow improvements in the techniques for applying it to short-wave radio communication.

But knowledge of the upper atmosphere and of the physical processes enacted there covers a far wider field than that of radio communication, and much that is learnt is of great value in many different branches of science. And so it may be that in this region of the upper atmosphere, and in the solar rays and other radiations with which it is permeated, there may be hidden much that will eventually prove to be of great and far-reaching benefit to mankind.

Index

★ ★ ★ ASSOCIATED ILIFFE TECHNICAL BOOKS ★ ★ ★

The books listed below are usually obtainable at all booksellers. In case of difficulty, please write to : **The Publishing Dept., Dorset House, Stamford Street, London, S.E.1.** *Complete list of radio and television books sent on application.*

FOUNDATIONS OF WIRELESS. *4th Edition.* By M. G. Scroggie, B.SC., M.I.E.E. Explains the principles on which the design and operation of modern receivers are based, and also covers the elementary principles of transmitters. 358 *pp.* 7*s.* 6*d. net. By post,* 7*s.* 10*d.*

GUIDE TO BROADCASTING STATIONS. *5th Edition.* Compiled by *Wireless World.* Gives details, arranged for easy reference of over 1,300 short-wave broadcasting stations of the world, and over 400 European long- and medium-wave stations. 1*s.* 6*d. net. By post,* 1*s.* 7*d.*

RADIO LABORATORY HANDBOOK. *4th Edition.* By M. G. Scroggie, B.SC., M.I.E.E. A practical handbook describing laboratory equipment and methods of working. 430 *pp.* 12*s.* 6*d. net. By post,* 12*s.* 10*d.*

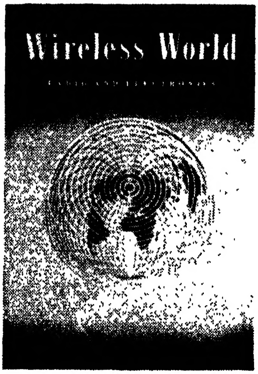
RADIO DATA CHARTS. *5th Edition.* By R. T. Beatty, M.A., B.E., D.SC., revised by J. McG. Sowerby, B.A., GRAD. I.E.E. A collection of 43 Abacs, or nomograms, providing most of the data required for the design of radio receivers. 91*pp.* 7*s.* 6*d. net. By post,* 7*s.* 11*d.*

" WIRELESS WORLD " GREAT CIRCLE PROJECTION MAP. With this map—first drawn by the late J. St. Vincent Pletts, the true distance and bearing from London " as the radio wave travels " to any part of the world can be ascertained simply and quickly. All measurements, either of distance or direction, can be obtained with a foot rule with the inches divided decimally. Scale : 1 in. = 1,000 miles.
34 *in. deep* × 27 *in.* 2*s.* 6*d. net. By post,* 3*s.*

TELEVISION RECEIVING EQUIPMENT. By W. T. Cocking, M.I.E.E. Comprehensive information about modern receiving equipment, giving many practical details and design data. *Third Edition in preparation.*

BASIC MATHEMATICS FOR RADIO STUDENTS. *2nd Edition.* By F. M. Colebrook, B.SC., D.I.C., A.C.G.I. Presents the whole subject in an easily understandable and interesting form. The author deals first in a general manner with the fundamental mathematical processes that are the foundation of all applied science, and then discusses the application of mathematical ideas to radio problems.
298 *pp.* 10*s.* 6*d. net. By post,* 10*s.* 11*d.*

WIRELESS DIRECTION FINDING. *4th Edition.* By R. Keen, B. ENG., (HONS.) A.M.I.E.E. This manual treats the subject of Direction Finding historically, descriptively and with a wide range of bibliographical references. A completely new chapter deals with systems of navigation using the hyperbolic grid, and an outline is also given of the application of the Principle of Least Squares to the plotting of a wireless fix. Aircraft Approach and Landing Systems are included and a description of the system developed by the U.S. Army Air Force as " SCS 51."
1072 *pp.* 45*s. By post,* 45*s.* 9*d.*


Wireless World