Ionospheric Radio Propagation

U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS
CIRCULAR 462

ERRATA TO ACCOMPANY NATIONAL BUREAU OF STANDARDS CIRCULAR 462, IONOSPHERIC RADIO PROPAGATION

1. Figure 7.30 on page 133 is reproduced to the wrong scale. It should be the same scale as figures 7.11 to 7.29, pages 123 to 132.

2. The formula on page 113, column 1, fourth line should read: $Q=1+0.005R$.

June 23, 1958
USCOMM-NBS-DC

U. S. DEPARTMENT OF COMMERCE
Charles Sawyer, Secretary
NATIONAL BUREAU OF STANDARDS
E. U. Condon, Director

Ionospheric Radio Propagation

NATIONAL BUREAU OF STANDARDS CIRCULAR 462
Issued June 25, 1948

UNITED STATES
GOVERNMENT PRINTING OFFICE
WASHINGTON : 1949

For sale by the Superintendent of Documents, U. S. Government Printing Office, Washington 25, D. C.
Price $1.25

Foreword

This book has been prepared by the Central Radio Propagation Laboratory of the National Bureau of Standards in order to meet the need for an elementary presentation of the theory and practical use of radio-wave propagation involving the ionosphere.

Preparation of the volume represents one of the completed projects of the Central Radio Propagation Laboratory of the National Bureau of Standards. This laboratory is charged with the function of centralizing within the United States all information on radio propagation. The laboratory operates, either directly or by cooperative arrangement, a large number of ionospheric stations that furnish many of the data on which our knowledge of propagation conditions is based. It is responsible for the reduction, analysis, and interpretation of the data from these stations. It is also charged with responsibility for observational and theoretical research on the propagation of radio waves.

Ionospheric Radio Propagation is in part a revision and expansion of the IRPL Radio Propagation Handbook that was prepared during the war at the request of the Armed Services. Current knowledge and technics of making radio-propagation calculations are summarized, and much of the material that has been included has appeared heretofore only in the form of reports that have had a limited circulation. The theoretical treatment is by no means comprehensive but is intended to explain the basic facts and principles of electromagnetic-wave propagation and the ionosphere for persons who have not had advanced courses in electrodynamics. In the effort to present basic theory in a simple manner, it has been necessary at times to use a mathematical treatment which would not be appropriate in a work prepared for more advanced students. Frequent references to the literature are given for further study, and the reader is encouraged as far as possible to consult them.

This book does not give all the necessary charts, graphs, and computing aids required for radio-propagation calculations in different parts of the world at different times, because ionospheric conditions vary widely from month to month and gradual improvements and changes in calculating technics are made. In general, therefore, a series of prediction charts, such as appear in the monthly CRPL series D of reports available from the Superintendent of Documents, Government Printing Office, Washington 25, D. C., is needed for practical use. From time to time such charts and computing aids as are necessary will be published and issued by the National Bureau of Standards.

The sciences, and indeed all fields of human activity, depend on the contributions of many individuals, and the National Bureau of Standards is indebted to many investigators and agencies for permission to use their results in this book. In particular, the excellent work that has been done by the Radio Propagation Unit of the United States Army Signal Corps is acknowledged, especially that on atmospheric radio noise. Finally, acknowledgment is made to members of the Bureau's staff who have been responsible, under the direction of J. H. Dellinger and Newbern Smith, chief and assistant chief, respectively, of the Central Radio Propagation Laboratory, for various chapters of this book: chapters 2, 4, and 5, A. G. McNish; chapter 3, R. Bateman, H. V. Cottony, H. P. Hutchinson, and A. H. Morgan; chapter 6, W. B. Chadwick and R. Silberstein; chapter 7, T. N. Gautier; chapter 8, J. W. Herbstreit, K. A. Norton, and Edna L. Schultz; chapter 9, T. N. Gautier and R. Silberstein.

E. U. CONDON, *Director.*

CONTENTS

IONOSPHERIC RADIO PROPAGATION

CHAPTER 1

INTRODUCTION

1.1. Purpose

The purpose of this book is to present the elementary principles of sky-wave or ionospheric radio-wave propagation at high frequencies and their practical application to the problems of radio communication. Its scope is limited to those waves which are propagated over long distances by reflection from the ionosphere.

The frequencies involved range ordinarily from 0.5 to 30 Mc/s or so. Low frequencies, below 0.5 Mc/s, might also be considered as belonging to this category, but consideration of them is excluded from this volume because the principles by which they are propagated are different. Owing to variations in the state of the ionosphere, much higher frequencies—up to 70 Mc/s—have been known to be propagated by the ionosphere, but except insofar as these frequencies are propagated by the ionosphere they will not be considered here. Ground-wave propagation will also not be considered, although it is an important mode of propagation for short distances within the spectrum range given above. In other words, the scope of this volume is determined entirely by whether or not the waves are reflected by the ionosphere.

The volume is not intended to be either a comprehensive treatise on the theory of wave propagation or a strictly practical handbook replete with charts and nomograms to solve by rule-of-thumb all types of practical problems of radio communication. It is intended rather to set forth in simple form the physical and mathematical theory underlying the principles of radio communication by reflection from the ionosphere and to bring these principles into understandable relation with the practical problems of radio communication.

1.2. Outline of Subject Matter

The ionosphere ordinarily consists, during daylight hours, of three main layers of ionization capable of reflecting radio waves at certain frequencies. These are the E layer, the $F1$ layer, and the $F2$, centered at heights of about 100, 200, and 300 km, respectively. The density of ionization, and hence the ability to reflect higher frequencies, is greatest for the $F2$ layer and least for the E layer. Below the E layer is a region, called the D region, which strongly absorbs radio waves of certain frequencies. Absorption in the D region is greatest for the lowest frequencies.

During night hours the ion density in all of the reflecting layers decreases, and hence the frequency which each will reflect decreases. During night hours, moreover, the $F1$ and $F2$ layers merge into one, and the ability of the D region to absorb radio waves also decreases. Thus, during night hours the ionosphere may be thought of as consisting almost solely of a single F layer, which is, nevertheless, still designated as $F2$.

The characteristics of the ionosphere are different in different geographic locations. They change with season and with the sunspot cycle; the various factors which are responsible for the changes are not independent in their effects. There are also random fluctuations which are not yet completely explained. Thus the variations of the ionosphere present a complicated geophysical phenomenon, the principles of which, however, need to be understood in order to achieve the best use of radio.

A typical practical problem of radio communication may consist in selecting a suitable frequency for a given service over a given path at a given time, within the limitations of the transmitting and receiving equipment. The frequency

must be low enough so that it will be reflected from the upper layers to the place at which reception is desired, for if the frequency chosen is too high it will penetrate the ionosphere. On the other hand, if too low a frequency is selected, it will be subject to great absorption in passing through the *D* region, and then the radiated power requirements for satisfactory reception will be excessive. Under certain conditions this selection is easily accomplished, while under others no practical radio communication can be established. Consideration also needs to be given to conditions at the transmitting location and at the receiving location, including the limitations imposed by radio "noise" (either natural or man-made), fading, interference, etc. There are many variations of this basic practical problem, for each of which a standard method of solution exists.

1.3. Outline of Chapters

The chapters of this volume are developed along a plan to yield a basic understanding and satisfactory solution of the usual problems of radio communication. Throughout the volume, the mks (meter-kilogram-second) system of units has been used because it has the advantages of simplicity and practicality. In accordance with present radio engineering practice, ratios of attenuation, voltage, etc., are frequently given logarithmically in decibels, although the primary mode of expression is kept as the unit itself rather than in decibels above or below a given reference level.

Chapter 2 presents in simple form the mathematical theory underlying the propagation of radio waves by the ionosphere, starting with the elementary laws of Ampere and Faraday and deriving therefrom Maxwell's equations and the wave equation. The physical significance of the equations is brought out and illustrated on occasion by simple analogy. The mathematics has been made as simple as possible, although at some sacrifice of rigor. Use of the imaginaries and other similar convenient notations has been avoided in order to hold the treatment to the simplest types of manipulations.

Methods of measuring the characteristics of the ionosphere and other factors affecting radio propagation are described in chapter 3. The descriptions are not given in sufficient detail for the contents of the chapter to be used as operating instructions for the making of ionospheric measurements; the chapter serves to point out the types and principles of measurements on which our knowledge of the ionosphere is based. The pulse-echo method of determining virtual heights and critical frequencies, described in this chapter, is being used at about 60 different ionosphere stations scattered over the earth. These stations are operated by various local governments and private research organizations in a vast international cooperative effort to delineate the characteristics of the ionosphere. Field intensities of various remote transmitting stations are measured continuously at many additional places. Measurements of radio noise and other auxiliary observations are also in progress, but on a less comprehensive scale, in various parts of the world.

The data acquired by the observational methods of chapter 3, interpreted in terms of the physical and mathematical principles outlined in chapter 2, are the basis for the physical concepts of the ionosphere described in chapters 4 and 5. These are the geophysical phenomena which determine the ability of the ionosphere to transmit waves of various frequencies. The main features and the geographic distribution of the ionosphere are set forth in chapter 4, together with a theoretical explanation of why it is as it is. Time variations of the ionosphere are described in chapter 5, together with an explanation of the causes of the variations. An elementary discussion of solar phenomena is included because of their intimate relation with ionospheric effects. With the conclusion of chapter 5, most of the salient scientific facts and principles necessary for application to the practical problems of radio communication have been presented. The methods of applying these facts and principles are covered in the chapters which ensue.

Chapter 6 covers in detail the methods for determining the maximum usable frequency for any path at any time. It takes into account the various possible modes of propagation and applies principles which have been found to work in practice, although the reasons for the success of some of these principles have not been established from theory. The application of these principles is illustrated by suitable examples. The procedures used in predicting maximum usable frequencies in advance, allowing for effects of changing solar activity, are set forth in detail.

The practical problems of ionospheric absorption are treated in chapter 7. These differ fundamentally from the problems involving the maximum usable frequency. If a frequency intended for a given purpose exceeds the maximum usable frequency, good reception of the signal cannot be obtained regardless of how much power is put out by the transmitter, but in calculating limitations imposed by absorption account must be taken of the amount of power radiated and the field intensity required for intelligible reception in the presence of the several interfering factors, such as atmospheric radio noise. The theory upon which these calculations are based is covered in the first five chapters, but the necessary parameters and the methods by which the calculations may be accomplished most easily constitute a separate body of material.

Because intelligible radio reception is limited,

among other factors, by the intensity of radio noise, chapter 8, on radio noise of atmospheric, solar, and cosmic origin has been included. Because investigation in this field is relatively recent, involving meteorological factors which are not clearly delineated on a world-wide basis, severe limitations are imposed on our present state of knowledge concerning radio noise. Chapter 8 couples, in the best way possible at present, the observed facts of radio noise and the theoretical generalizations that may be drawn from them and from related facts of radio propagation.

The considerations brought out in chapters 7 and 8 are employed practically in chapter 9 to enable the calculation of the lowest useful high frequency that may be used for communication over certain paths at certain times employing a given amount of power, and also the calculation of the inverse quantity, i. e., the lowest required radiated power to permit communication for specified conditions and frequencies. Several examples of the calculation of the quantities involved are given, and the necessary charts and data for these calculations are presented.

1.4. Practical Aids to Computation

The necessary aids required to carry out the calculations described in these chapters are not completely presented in this volume, but only such samples as are required for solution of the illustrative problems. There are several reasons for this. Some of the computational material is of current significance and must be changed from month to month. Other of the computational material, while still following the principles set forth in this book, will be changed from time to time as new evaluation of the parameters involved becomes possible. Furthermore, the inclusion of all the necessary tables and charts within this volume would involve presentation of them in a form inconvenient for regular use. It must be emphasized that the charts in this volume with which an overlay is to be used are to a smaller scale than the corresponding ones in the practical aids mentioned in the following paragraphs.

The National Bureau of Standards supplies material for solving radio-propagation problems. This includes the CRPL–D series "Basic Radio Propagation Predictions," which present monthly contour charts of predicted values of $F2$-zero-muf, $F2$-4,000-muf, E-layer-2,000-muf, median fEs, issued three months in advance. This publication is available on a purchase basis from the Superintendent of Documents, United States Government Printing Office, Washington, 25, D. C. The cost of an annual subscription is $1.00 (foreign, $1.25). Each issue contains charts of extraordinary-wave critical frequency for the $F2$ layer and charts of maximum usable frequency under average conditions for a transmission distance of 4,000 km. These charts are provided for each of the three zones into which the world is divided, for the purpose of taking into consideration the variation of the characteristics of the $F2$ layer with longitude, that is, the geomagnetic effect. There is a chart of maximum usable frequency for E-layer transmission over a path length of 2,000 km, and charts showing the highest frequency of sporadic-E reflections as well as percentage of time occurrence for sporadic-E in excess of 15 Mc are included.

The tables and charts referred to above, for use in practical calculations of ionospheric radio propagation, together with information and recommended procedures, will be available as a Circular of the National Bureau of Standards. This will include all the most recent material available for the radio-propagation calculations.

The National Bureau of Standards also presents as a weekly service the CRPL–J series, which gives an estimate of periods of radio propagation disturbances a month in advance. The disturbances considered here are of the ionosphere-storm type, which is most severe in polar regions, and tapers off toward the equator. Sudden ionosphere disturbances (SID), characterized by simultaneous fadeouts in the entire high-frequency spectrum, on paths in the daylight side of the world, are not included in this forecast.

This weekly service is supplemented by semihourly warnings of radio disturbances broadcast by the National Bureau of Standards broadcast station WWV, Washington, D. C., at 19 and 49 min. after each hour. WWV operates continuously on 2. 5, 5, 10, and 15 Mc with carrier power outputs of 700 w, 8, 9, and 9 kw, respectively; on 20 Mc with carrier output of 8. 5 kw, except for the first four workdays following the first Sunday in each month when the carrier output is 100 w; and on 25, 30, and 35 Mc with carrier output of 100 w. A series of W's signifies that disturbed conditions over North Atlantic paths exist or are expected within 12 hr; a series of N's signifies that conditions are expected to be quiet.

Tables and graphs of ionospheric characteristics observed at various stations are available from the National Bureau of Standards. From time to time, additional material of interest in connection with radio propagation is presented in occasional reports, a list of which is available upon request.

CHAPTER 2

THEORY OF RADIO WAVE PROPAGATION

2.1. Maxwell's Equations

a. Laws of Ampere and Faraday

The principles of the propagation of radio waves rest upon a few fundamental laws of electricity and magnetism, most of which are taught in every high-school physics course. Among these are the laws of Ampere and Faraday relating magnetic and electric phenomena. Ampere's and Faraday's laws, as such, apply only to voltage and currents in electric circuits. In order to understand and study the phenomena of electric and magnetic fields in space, it is necessary to generalize the laws so that they apply not only to well-defined circuits but also to any point or region in space. As so generalized, they become equations which are valid at any point in space. From the "field equations" so written, we can derive all the properties of electric and magnetic fields, and their interaction with charges; we can, in fact, by applying the equations to electric circuits, derive all of electric circuit theory from them also.

Faraday's law states that the total electromotive force, V, around a circuit is proportional to the rate of change of the magnetic flux, ϕ, threading through the circuit. This is clearly seen in the case of a loop of wire threaded by a magnetic flux of density B. If the magnetic flux $\phi = \int\int BdS$, indicated in figure 2.1, increases, there will be an electromotive force in the wire given by $V = (\partial/\partial t)\int\int BdS = \int\int(\partial B/\partial t)dS$ integrated over the total area of the loop. We have said nothing about the resistance of the wire—if the wire were of infinite resistance the electromotive force would be the same. Thus it is clear that if no wire is present the electric force or electric intensity

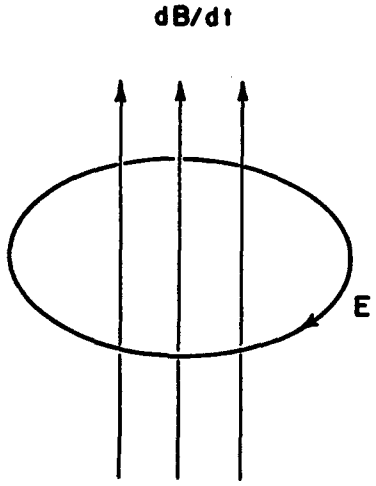


FIGURE 2.1. *Illustration of Faraday's law.*

E at every point in the same path in air (or in a vacuum) will be the same, and so will the total emf around the circuit, given by $V = \oint Eds$, integrated around the circuit. This emf is the summation of all the emf induced in the wire, and is equal to the electric intensity E integrated around the circuit. If the magnetic flux increases at the rate of 1 weber/sec and the area of the loop is in square meters the relationship is given by

$$-\int\int(\partial B/\partial t)dS = \oint Eds, \qquad (2.1)$$

E being expressed in volts per meter and the loop integral being taken in a clockwise direction looking in the direction of B. The left-hand side states that the flux increases are added together over the total area of the loop S, and the right-hand side states that the electric force in the instantaneous direction of travel is added around the circuit s. The right hand side of the equation is also the work done on a unit charge in traveling around the circuit while the field is changing. No limitations are placed upon the shape of the loop.

b. Expression of the Laws in Differential Form

To express this law in a form valid at any point in space, let us imagine the changing magnetic field as extending over a considerable region and concentrate our attention on a small rectangular area within that region shown in figure 2.2. We shall conduct our integration around the rectangle $dxdz$, starting from the origin of the coordinate system, going in the z-direction along the z-axis, then in the x-direction, then in the minus z-direction, and finally in the minus x-direction back to our starting point. As we do so we shall add together all the values of E, the electric intensity; that is, we shall compute the work done on a unit charge carrying it around the circuit. Going out in the z-direction the work done is $E_z dz$, where E_z is the electric intensity along that leg of the circuit, and dz is the length of that leg. But coming back on the opposite leg, the work done is $-E_z dz - (\partial E_z/\partial x)dxdz$, where $(\partial E_z/\partial x)dx$ represents the change in E_z due to the fact that our position in the z-direction has been shifted by an amount dx. Similarly going outward along the x-direction, the work done is $+E_x dx + (\partial E_x/\partial z)dzdx$, and coming back it is $-E_x dx$. Adding these together, we have for the total work done, i. e., for the integral of Eds around the loop:

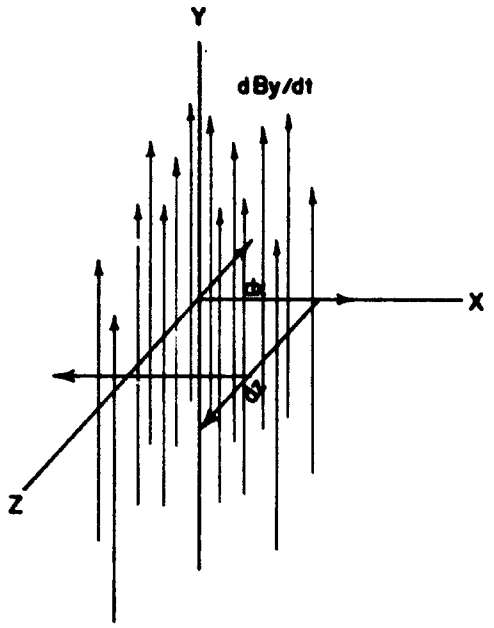


FIGURE 2.2. *Illustration for derivation of Faraday's law in differential form.*

$$\oint Eds = (\partial E_x/\partial z)dxdz - (\partial E_z/\partial x)dxdz.$$

We identify this expression as the right-hand side of eq 2.1. But Faraday's law says that this is numerically equal to minus the rate of flux increase within the circuit. The area of the circuit is $dxdz$ so that the flux increase within the circuit is $(\partial B_y/\partial t)dxdz$, which we now identify as the quantity appearing on the left-hand side of eq 2.1. We now write

$$-(\partial B_y/\partial t)dxdz = (\partial E_x/\partial z)dxdz - (\partial E_z/\partial x)dxdz.$$

or

$$-\partial B_y/\partial t = \partial E_x/\partial z - \partial E_z/\partial x. \qquad (2.2a)$$

Two more equations may be derived for the components of the magnetic field in the x-direction and in the z-direction, namely:

$$-(\partial B_x/\partial t) = \partial E_z/\partial y - \partial E_y/\partial z. \qquad (2.2b)$$

$$-(\partial B_z/\partial t) = \partial E_y/\partial x - \partial E_x/\partial y. \qquad (2.2c)$$

Ampere's law states a reciprocal principle regarding the production of a magnetic field by an electric current. If we replace the magnetic flux, B, in figure 2.1 by an electric current of density I, we have a magnetomotive force, H, around the loop instead of the electric force, E. This circuit equation may be written

$$\int IdS = \oint Hds, \qquad (2.3)$$

where I is in amperes per square meter and H is in ampere turns per meter. I is sometimes written as ρu (ρ=charge in coulombs and u=velocity in meters per second) and sometimes as σE (σ= conductivity in mhos per meter and E=electric intensity in volts per meter).

Following the procedure outlined, we may generalize Ampere's law to a point relationship in, terms of differential equations analogous to 2.2a, b, and c, which express Faraday's law, namely,

$$I_y = \partial H_x/\partial z - \partial H_z/\partial x. \qquad (2.4a)$$

$$I_x = \partial H_z/\partial y - \partial H_y/\partial z. \qquad (2.4b)$$

$$I_z = \partial H_y/\partial x - \partial H_x/\partial y. \qquad (2.4c)$$

But Ampere's law in its simplest form is strictly true only for steady electric currents flowing in conductors; in order to embrace all electrical phenomena, another term must be added. Consider a circuit containing a battery and a condenser as shown in figure 2.3. While the condenser is charging, a current will flow in the circuit, and a magnetic force would be created around any path circling the wire. The same magnetic force would be observed even around a path enclosing the space between the plates of the condenser. This is due to the "displacement current" flowing in the space between the plates indicated in figure 2.3 by the dotted lines. The displacement current between the two plates is continuous with the charging current and is

$$dD/dt = \kappa dE/dt. \qquad (2.5)$$

If there is conductivity or "leakage" in the condenser, the magnetic field around the loop encircling the space between the plates of the condenser will be the sum of the effects of the displacement and the conduction currents; the conduction current density between the plates of the con-

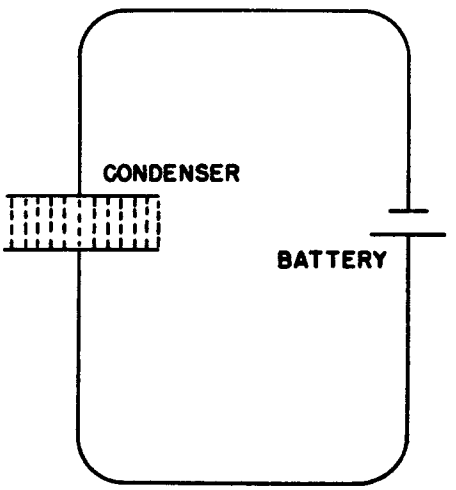


FIGURE 2.3. *Electric circuit illustrating displacement current between plates of condenser.*

denser can be written σE, where σ is the total conductivity of the space between the plates. Thus we must include displacement current along with conduction current in each of the eq 2.4a, b, and c, by writing $I+\kappa(dE/dt)$ in place of I, using the appropriate subscript for each component. If we write $B=\mu H$, in eq 2.2a, b, and c, where μ is the permeability of the medium, the six field equations may be written:

$$(I_x+\kappa\partial E_x/\partial t)=\partial H_z/\partial y-\partial H_y/\partial z \quad (2.6a)$$

$$(I_y+\kappa\partial E_y/\partial t)=\partial H_x/\partial z-\partial H_z/\partial x \quad (2.6b)$$

$$(I_z+\kappa\partial E_z/\partial t)=\partial H_y/\partial x-\partial H_x/\partial y \quad (2.6c)$$

$$-\mu\partial H_x/\partial t=\partial E_z/\partial y-\partial E_y/\partial z \quad (2.7a)$$

$$-\mu\partial H_y/\partial t=\partial E_x/\partial z-\partial E_z/\partial x \quad (2.7b)$$

$$-\mu\partial H_z/\partial t=\partial E_y/\partial x-\partial E_x/\partial y \quad (2.7c)$$

These equations simply state Ampere's and Faraday's laws; that an electric current or its equivalent, a changing electric field, produces a magnetic field along a path encircling that current or changing electric field (2.6a, b, and c), and a changing magnetic flux produces an electric field along a path encircling the changing magnetic flux. They state these relationships quantitatively for all points or regions in space, not limited to electric circuits, and in a convenient form for mathematical manipulation.

c. Derivation of the Wave Equation

The equations express the relationships for the components of I, E, and H directed along each of the three axes of our coordinate system. But there is nothing sacred about our coordinate system. Assuming that we are dealing with the above-described electromagnetic relationships in a portion of space which is nonconducting and where therefore no I can exist, let us select our coordinate system so that E is in the direction of the x-axis. Equations 2.6b, 2.6c, and 2.7a may be disregarded because one side of each contains only terms which are nonexistent. There is no $\partial H_x/\partial t$, and any H present must be perpendicular to the x-axis. Let us rotate our coordinate system about the x-axis until the direction of H coincides with the y-axis, there being then no H_z. We have now reduced our equations to

$$\kappa\partial E_x/\partial t=-\partial H_y/\partial z. \quad (2.8a)$$

$$-\mu\partial H_y/\partial t=\partial E_x/\partial z. \quad (2.8b)$$

Let us differentiate eq 2.8a with respect to t and 2.8b with respect to z. We obtain

$$\kappa\partial^2 E_x/\partial t^2=-\partial^2 H_y/\partial z\partial t \quad (2.9)$$

$$-\mu\partial^2 H_y/\partial z\partial t=\partial^2 E_x/\partial z^2, \quad (2.10)$$

which, on combination, yield

$$\kappa\mu\partial^2 E_x/\partial t^2=\partial^2 E_x/\partial z^2. \quad (2.11)$$

This is a familiar equation in physics. It represents a transverse-wave propagated in the z-direction with a phase velocity $v=1/\sqrt{\kappa\mu}$; for free space $\kappa_0=(1/36\pi)\times 10^{-9}$

$$\mu_0=4\pi\times 10^{-7}$$

and thus the phase velocity in free space, c, is equal to 3×10^8m/sec. It is satisfied by functions of the form $E_x=E_{x_0}\cos\omega(t\pm z/v)$, where $\omega=2\pi$ times the frequency f. The wavelength $\lambda=v/f$. At time $t=0$, E_x will have the value E_{x_0} at points along the z-axis distant λ, 2λ, 3λ, etc., from the origin (see fig. 2.4). Similarly, for $z=0$, that is,

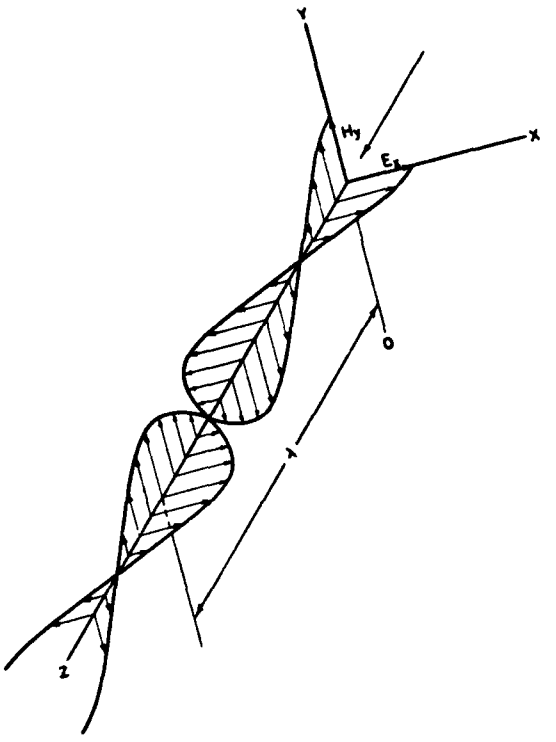


FIGURE 2.4 *Electric and magnetic vectors in plane polarized wave.*

at the origin, E_x will have the value E_{x_0} when $t=1/f$, $2/f$, $3/f$, etc. It may be seen from the foregoing that E and H are perpendicular to each other and also to the direction in which the wave is being propagated. The $-$ sign represents a wave traveling in the $+z$-direction, and vice versa. The equation for the wave may be written in

several different forms, each of which has its own advantages for certain purposes

$$E_x = E_{x_0} \cos \omega(t - z/v) \tag{2.12a}$$

$$E_x = E_{x_0} \cos \beta_0(ct - \eta z) \tag{2.12b}$$

$$E_x = E_{x_0} \cos (\omega t - \beta z) \tag{2.12c}$$

$$E_x = E_{x_0} e^{j(\omega t - \beta z)}, \tag{2.12d}$$

etc., where

$\beta_0 = 2\pi/\lambda_0$ (phase constant)

v = phase velocity $= \omega/\beta$

λ_0 = vacuum wavelength

η = refractive index $(= \sqrt{\mu\kappa/\mu_0\kappa_0})$

$\beta = 2\pi/\lambda$

λ = actual wavelength

$j = \sqrt{-1}$.

The exponential form is symbolic, like the use of j in alternating-current theory, and is convenient for mathematical manipulation, only the real part of the result being used.

In the lower atmosphere both κ and μ have essentially their free space values κ_0 and μ_0, respectively, so that the velocity of propagation is c, the velocity of light. If this condition applied throughout the entire atmosphere, the problem of radio-wave propagation over long distances would be simple—there would be none.

2.2. Motion of Ions in Electric and Magnetic Fields

Propagation of radio waves over long distances is rendered possible, though slightly more complicated, by the presence of free charges (ions or free electrons) in the upper atmosphere. These ions render the atmosphere electrically conductive, and consequently the term in I of the Maxwellian equations cannot be ignored in treating propagation in those regions. To apply the wave equations to an ionized region it is necessary to consider the current or polarization P that arises from the behavior of ions in the presence of electric and magnetic fields.

An ion in free space in the presence of an electric field experiences a force: $f = eE$, where e is the electric charge on the ion in coulombs, E is the electric intensity in volts per meter, and the force f, is given in newtons. This force imparts a velocity, u, to the ion in accordance with the familiar Newton's law of motion

$$du/dt = eE/m, \tag{2.13}$$

where m is the mass of the ion. (For our purposes a free electron will be considered as an ion also). The velocity of the ion in the direction of the electric intensity increases so long as the field is applied, approaching the velocity of light; naturally negative ions or electrons, which may be regarded as ions, will be accelerated in the direction opposite to the intensity. So long as the velocity is well below the velocity of light, it is given by integrating eq 2.13, so that

$$u = (e/m)Et, \tag{2.14}$$

assuming that the ion starts from rest.

As an ion carries an electric charge this motion of an ion corresponds to an electric current given by

$$I = ue = (e^2/m)Et. \tag{2.15}$$

It may be noted that the longer the electric intensity is applied the stronger the current becomes, and the current will not be reduced to zero unless an electric intensity is applied in the opposite direction for a sufficient time that the product Et is equal to that which created the motion. Of course, if the ion strikes a gas molecule the momentum is largely lost and the current must be built up again.

A constant magnetic field exerts no force on an ion at rest, but if the ion is in motion it experiences a force given by

$$\boldsymbol{f} = \mu e \boldsymbol{u} \times \boldsymbol{H}. \tag{2.16}$$

This is a vector equation; it denotes that the force is perpendicular to the velocity and to the magnetic intensity and is proportional to the component of the velocity, which is perpendicular to the magnetic intensity. Thus, an ion moving parallel to the magnetic intensity experiences no force. If the ion is moving in the direction of the x-axis and the magnetic intensity is in the direction of the y-axis, then, for a positive ion and a right-handed coordinate system, the force is in the direction of the z-axis.

As the force, and hence the acceleration of the ion are always perpendicular to the component of velocity perpendicular to the magnetic intensity that component velocity is continually changing, so that the ion's motion in the plane perpendicular to the magnetic intensity is a circle, the diameter of which is determined by balancing the centrifugal force due to the circular motion by the centripetal force due to the magnetic intensity. The radius of the circle is given by

$$r = mu/e\mu H. \tag{2.17}$$

The number of rotations per second is given by

$$f_H = u/2\pi r = e\mu H/2\pi m \tag{2.18}$$

and is independent of the velocity of the ion. For electrons

$$f_H = 2.84 \times 10^{10} B = 3.57 \times 10^4 H.$$

In the atmosphere, where the earth's magnetism gives rise to a magnetic intensity of the order of 40 amp turns per meter, the gyromagnetic frequency, f_H, is in the neighborhood of 1.5 Mc/sec for electrons and about 50 c/sec for ions of atomic oxygen. Thus ions in the upper atmosphere in a state of thermal agitation are not free to move like uncharged gas molecules but must encircle the lines of magnetic intensity. As random motions will have velocity components parallel to the lines of intensity, which are unaffected by the foregoing considerations, the paths of the ions will be helixes about the lines of the earth's magnetic intensity as shown in figure 2.5.

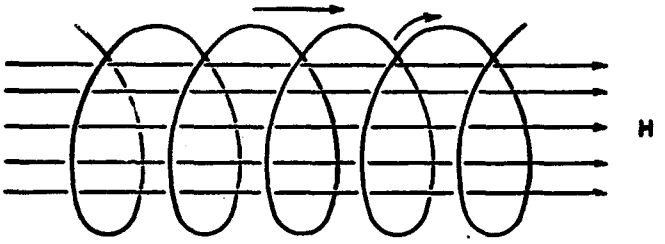


FIGURE 2.5. *Helical motion of ions due to thermal agitation in presence of uniform magnetic field.*

When magnetic and electric intensities are both present the motion of an ion is complicated. Naturally, when the electric intensity is parallel with the magnetic intensity there is no interaction between the magnetic intensity and the ion moving under the influence of the electric intensity. The motion of the ion then is exactly as given in eq 2.13.

But when the intensities are crossed, for example, when E is in the x-direction, and H is in the y-direction, the motion produced by the electric intensity is altered by the magnetic intensity. The motion in the direction of the electric intensity does not increase indefinitely, but the ion swings around and eventually moves against the electric intensity, the resultant average motion being perpendicular to both the electric and magnetic intensities as shown in figure 2.6, drawn for ions starting from rest. This average motion is independent of the mass and charge of the ion, and even of the sign of the charge, negative and positive ions both moving in the same direction. Therefore, where there are the same number of positive and negative ions no conduction currents can flow in a direction perpendicular to any magnetic intensity which may be present, except insofar as these conditions are altered by collisions of ions with each other or with other objects. The motion of an ion under these conditions may be thought of as first an acceleration

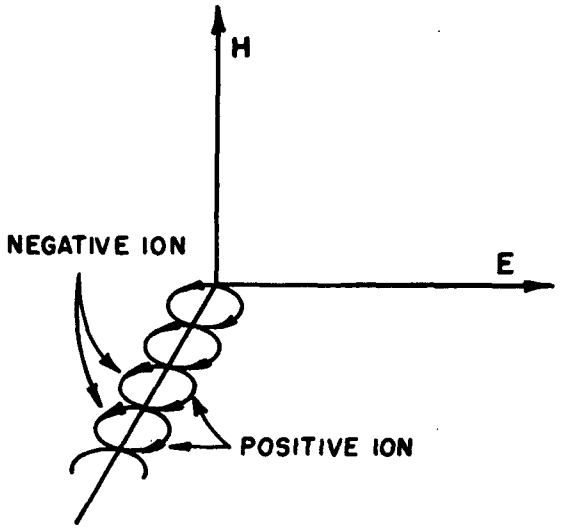


FIGURE 2.6. *Motion of positive ion in presence of crossed electric and magnetic fields.*

in the direction of the electric intensity, and, as this results in a velocity, the ion is deflected by the magnetic field and it finally moves contrary to the electric intensity until its momentum is spent, at which time the motion starts over again. In performing this motion, the ion has slid off in a direction at right angles to the electric and magnetic intensities.

2.3. Propagation of Radio Waves in an Ionized Medium

a. Absence of Constant Magnetic Field

Let us suppose that an alternating electric intensity representable by $E=E_0 \sin \omega t$, such as exists in an electromagnetic wave, is applied to a region containing ions. In the absence of a constant magnetic intensity, the only force acting on the ions will be the electric force. We may then write

$$md^2x/dt^2 = eE = eE_0 \sin \omega t, \qquad (2.19)$$

where m is the mass of the ion, and d^2x/dt^2 is its acceleration. Integrating once and shifting m to the other side, we get

$$dx/dt = -(e/m\omega)E_0 \cos \omega t. \qquad (2.20)$$

This represents the velocity of an ion. If there are N ions per cubic centimeter in the region, and if each carries an electric charge, e, the current, I, represented by the motion of all these ions is

$$I = Nedx/dt = -(Ne^2/m\omega)E_0 \cos \omega t. \qquad (2.21)$$

This we may substitute for I in eq 2.6a, assuming as before that the electric intensity of the wave is in the x-direction only and that the magnetic

intensity of the wave is in the y-direction only. We should also perform the differentiation with respect to time called for in the equation.

Equation 2.6a becomes

$$(-Ne^2/m\omega)E_0 \cos \omega t + \omega\kappa E_0 \cos \omega t = -\partial H/\partial z$$

or

$$[\kappa - Ne^2/m\omega^2]\omega E_0 \cos \omega t = -\partial H/\partial z. \quad (2.22)$$

In the atmosphere, where $\kappa=\kappa_0$ and $\mu=\mu_0$, this is equivalent to the existence of a dielectric constant

$$\kappa_i = \kappa_0 - Ne^2/m\omega^2, \quad (2.23)$$

so that the phase velocity in this ionized medium is

$$v = 1/\sqrt{\mu_0\kappa_i} = c/\sqrt{\kappa_i/\kappa_0} = c/\sqrt{1 - Ne^2/\kappa_0 m\omega^2} = c/\sqrt{1 - f_N^2/f^2}, \quad (2.24)$$

where $f_N^2 = 80.5N$, N is in electrons per cubic meter, e is the electronic charge (1.59×10^{-19} coulomb), e/m is the specific electronic charge (1.767×10^{11} coulombs/kg), and c is the velocity of light in a vacuum. The quantity $\sqrt{1 - Ne^2/\kappa_0 m\omega^2}$ is called the refractive index of the medium, designated by η. So long as $Ne^2/\kappa_0 m\omega^2 < 1$, the wave continues to be propagated in the medium, but when N is great enough that

$$Ne^2/\kappa_0 m\omega^2 = 1, \quad (2.25)$$

the left-hand side of eq 2.22, and, hence, the right-hand side, as well, reduce to zero. Under this condition, an electromagnetic wave can no longer be propagated. As both e and m may be regarded as constant, this condition depends upon the values of N and ω. Solving eq 2.25 for ω in terms of N, we get

$$\omega = 2\pi f = \sqrt{Ne^2/\kappa_0 m} = 2\pi f_N. \quad (2.26)$$

A wave of frequency, f given by eq. 2.26, will meet a refractive index of zero at the level where the ion density has the given value of N, and will be reflected at this level; this is the level where $f=f_N$. If N is the maximum density of ions in a given ionospheric region, all frequencies less than the corresponding value of f_N are reflected back to the earth by that region while all frequencies greater than the value f_N continue to be propagated and pass through that region. This value of frequency is called the "critical frequency" for the region and is designated "f_c" or "f_0."

As the quantity m appears in the denominator in the expressions for f_N and for κ_i, ions with the least mass have the greatest effect on propagation. In particular, electrons are most effective in reflecting radio waves because of their small mass, one electron being the equivalent of over 50,000 ions of molecular oxygen or nitrogen. The terms electron density and ion density are frequently used interchangeably. Except in specific cases, the term ion is assumed to include electrons as well, as a special class of ions.

b. Presence of Constant Magnetic Field

The presence of a constant magnetic intensity in the atmosphere due to the earth's magnetism alters the motions of the ions when an alternating electric intensity is applied. If there is a component of the constant magnetic intensity perpendicular to the electric intensity the ions are deviated in accordance with eq 2.16. The ions then describe elliptical paths, as shown in figure 2.7. This introduces an electric component of the wave in the y-direction. The ions acquire momentum from the electric intensity of the wave, which they do not lose when the intensity oscillates in the opposite direction. This momentum "helps" them in their motion in response to the alternating phases, so that their motions are greater than they would be in the absence of a constant magnetic intensity; thus the presence of a constant magnetic intensity tends to increase the apparent electric conductivity of the gas for waves of certain frequencies.

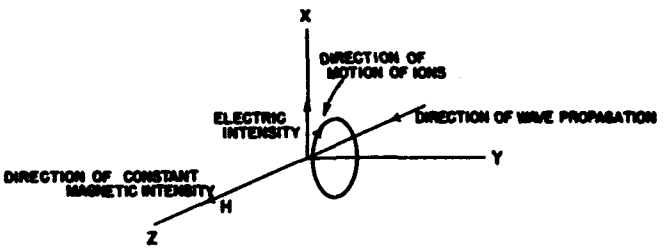


FIGURE 2.7. *Motion of ions caused by radio wave propagated in direction of a constant magnetic field.*

The mathematical development for this case is beyond the scope of this volume. It leads to an effective dielectric constant given by

$$\kappa_i = \kappa_0 \left[1 - \frac{x}{1 - \frac{y_T^2}{2(1-x)} \pm \sqrt{\frac{y_T^4}{4(1-x)^2} + y_L^2}}\right], \quad (2.27)$$

where

$x = Ne^2/\kappa_0 m\omega^2 (= f_N^2/f^2)$; $f_N = \sqrt{Ne^2/4\pi^2\kappa_0 m}$

$y_T = \mu_0 e H_T/m\omega (= f_T/f)$; $f_T = \mu_0 e H_T/2\pi m$

$y_L = \mu_0 e H_L/m\omega (= f_L/f)$; $f_L = \mu_0 e H_L/2\pi m$

H_T = magnetic intensity transverse to direction of propagation

H_L = magnetic intensity along direction of propagation.

The $\pm$ sign in the equation for κ_0 means that the

wave is split into two components because of the magnetic field, and that the dielectric constant is different for each component. The + sign refers to the so-called ordinary, the − sign to the extraordinary wave. The conditions for reflection ($\kappa_t=0$) are

$$x=1$$
$$x=1\pm y,$$

where

$$y=\sqrt{y_T^2+y_L^2}=\mu_0 eH/m\omega=f_H/f$$

H=total magnetic intensity

$f_H=\mu_0 eH/2\pi m$ (gyrofrequency).

The case $x=1$ corresponds to the ordinary wave:

$$f_N^2=Ne^2/4\pi^2\kappa_0 m=f^2. \quad (2.28a)$$

The case $x=1\pm y$ corresponds to the extraordinary wave:

$$f_N^2=Ne^2/4\pi^2\kappa_0 m=f(f\pm f_H). \quad (2.28b)$$

There are thus two levels at which the extraordinary wave can be reflected; ordinarily most of the reflection takes place at the lower level, where $f_N^2=f(f-f_H)$. Note that where $y_T=0$, however, (and in practice, where y_T is very small, i. e., for vertical transmission near the magnetic pole of the earth):

$$f_N^2=f(f+f_H) \text{ ordinary wave} \quad (2.29a)$$

$$f_N^2=f(f-f_H) \text{ extraordinary wave,} \quad (2.29b)$$

and the level of reflection for the ordinary wave is different, corresponding to the higher level for the extraordinary wave in the general case. For the extraordinary wave, with the — sign in eq 2.28b, f is greater than f_N, which means that higher radio frequencies are reflected in the presence of a magnetic field than would be reflected in its absence. The relationship between the frequencies of the ordinary and extraordinary waves reflected at a given level is the same as that for the respective critical frequencies for a given ionospheric region. Designating f_e for the ordinary wave by the symbol f_o and for the extraordinary wave by the symbol f_x

$$f_x(f_x-f_H)=(f_o)^2. \quad (2.30)$$

As there is usually a component of the electric intensity of the wave parallel to the constant magnetic intensity as well as perpendicular to it, two different frequencies are usually returned by the same ion density, that of the ordinary wave and that of the extraordinary wave. If f_o is sufficiently large with respect to f_H the relationship is approximated by

$$f_o=f_x-f_H/2. \quad (2.31)$$

2.4. Phase, Group and Signal Velocity

In section 2.3 it is shown that an electromagnetic wave is propagated in an ionized medium with a phase velocity given by eq 2.24 $v=c/\sqrt{1-Ne^2/\kappa_0 m\omega^2}$. This indicates that the phase velocity is in excess of the velocity of light. This is no contradiction of the theory of relativity because it signifies nothing more than that if a wave of constant frequency is being propagated in an ionized medium the distance from one wave crest to another is greater than it would be in free space. It does not mean that an electromagnetic signal is propagated in excess of the velocity of light.

When a signal is emitted from an electromagnetic oscillator it consists of a group of waves of various frequencies, that is to say, a carrier frequency modulated by side bands. Owing to the fact that the refractive index, and hence the phase velocity, is different for different frequencies, these waves which make up the signal will all be propagated in the ionosphere at different velocities. If two waves at nearly the same frequency are emitted, they appear as a continuous frequency modulated at a frequency equal to one-half the difference of the two frequencies. The wave groups so produced are propagated not at the phase velocity but at the group velocity, which, in the ionosphere, is always less than the velocity of light. In the simple case where there is no magnetic field in the ionosphere the group velocity w is given by

$$w=\eta c, \quad (2.32)$$

where η is the refractive index for the average of the two frequencies. It may be noted that since the phase velocity is given, by eq 2.24 as $v=c/\eta$, consequently $wv=c^2$.

A pulse, or packet, of radio waves consists of a large number of frequencies so that the above simple theory is not applicable. This pulse is propagated at the signal velocity, which also never exceeds the velocity of light. The signal velocity is not susceptible of exact definition as is the group velocity, but for practical purposes it may be taken as approximately equivalent to the group velocity.

The theory underlying this idea is complicated, but an appreciation of the physical significance may be derived from everyday experience. If an impulse is started in a pool of water by the dropping of a pebble, or by the passage of a boat, a group of waves is set into motion. As these waves are propagated it may be noticed that the leading crests die out and new crests are formed to the rear. Thus the pulse, as a whole, is propagated less rapidly than the individual waves of which it is composed. The velocity of an

individual wave in the pulse corresponds to the phase velocity and the velocity of the pulse as a whole corresponds to the signal velocity. Because of the difference of the refractive index for different frequencies the characteristics of the pulse change as the propagation continues to greater and greater distances.

When a pulse of electromagnetic waves enters the ionosphere it behaves in the same manner. As the pulse approaches the height of reflection its velocity of propagation becomes less and less, reaching zero instantaneously at the height of reflection. Because the pulse must pass through lower ion densities before it is reflected the length of time required for a signal to be returned from a given height in the ionosphere is greater than would be required if it traversed nonionized regions until reflected. This phenomenon is known as retardation.

We may calculate the height of reflection by dividing the length of time required for the pulse to return by the velocity of propagation. If the velocity of propagation is taken as the velocity of light, the height so calculated is greater than the actual height. This calculated height is called the virtual height and is given by

$$h' = (t/2)c, \tag{2.33}$$

in which t is the time for the echo to return. The virtual height has more significance for radio propagation than the true height.

If the ion density has a maximum (N_{max}) at some height, then a wave with angular frequency given by $\omega^2 = N_{max}e^2/\kappa_0 m$ encounters a refractive index $\eta = 0$ at this height. This frequency is the critical frequency, and a wave of this frequency is infinitely retarded on reaching the height. Waves of slightly lower frequencies, or slightly higher frequencies if they are reflected from a layer above this region, show very great retardation.

Although we can measure directly only the virtual heights at which radio waves are reflected, we can compute the actual height of ionospheric layers, provided we assume a particular form of ion distribution.

To compute the true height of reflection of a wave we must correct the virtual height for retardation in the portion of the highest layer through which it has passed, as well as all lower layers. The correction for retardation is the difference between the virtual distance and the true distance the wave has traveled in each layer.

Assuming a parabolic distribution of ion density the length of the virtual path traversed by a wave within a layer is

$$2\int_0^{y_m} ds/\eta = y_m(f/f_o)\log_e[(f+f_o)/(f-f_o)] \tag{2.34}$$

if $f > f_o$, and

$$\int_0^{y} ds/\eta = (1/2)y_m(f/f_o)\log_e[(f_o+f)/(f_o-f)] \tag{2.35}$$

if $f < f_o$, where f is the frequency of the wave, f_o is the critical frequency of the layer, y is the height of reflection, and y_m is the height of maximum ionization, the value $y = 0$ being taken as the bottom of the layer.

In the lowest layer from which we obtain reflections, the virtual heights satisfy the relation

$$h' = h_0 + (1/2)y_m(f/f_o)\log_e[(f_o+f)/(f_o-f)] \tag{2.36}$$

which can be solved for h_0, the height of the bottom of the layer, and y_m, the half-thickness, if we have the virtual heights corresponding to at least two frequencies. These values can then be substituted into the expression for the true height of reflection,

$$h = h_0 + y_m[1 - \sqrt{1-(f/f_o)^2}]. \tag{2.37}$$

For practical reasons it is often convenient to determine the true height of maximum electron density from the relationship

$$h_{max} = h'_{0.834}, \tag{2.38}$$

where $h'_{0.834}$ is the virtual height for the frequency

$$f = 0.834 f_0, \tag{2.39},$$

since this is the value of f which exactly reduces the right-hand side of eq 2.36 to $h_0 + y_m$. In case a layer whose true heights are desired lies above another layer, corrections for retardation in the lower layer must be made.

The method of calculation outlined above is valid within the limits set forth only if the distribution of ion density as a function of height is representable by a parabola. For other distributions, other methods of calculation must be employed.

2.5. Oblique Incidence

In the foregoing discussion attention has been directed to the case of a radio wave propagated vertically upward into the ionosphere. Although this case is basic to the understanding of radio propagation, it is a case that is never encountered in practical communication. The wave which is received at a distance from the transmitter impinges on the ionosphere obliquely. It is not "reflected" from the ionosphere in the sense that a ray of light is reflected from a mirror; the wave enters the ionosphere, is deviated by refraction, and emerges from the ionosphere in much the same manner as a ray of light is deviated on passing through a layer of hot air near the ground to produce the familiar optical phenomenon known as a

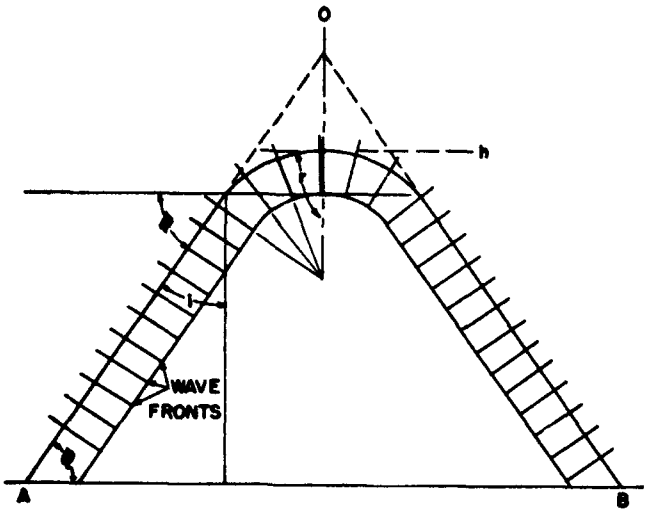


FIGURE 2.8. *Refraction of waves by the ionosphere.*

mirage. The optical analogy is a very close one.

We may now consider a narrow beam propagated from A so that it enters the ionosphere at an angle ϕ (see fig. 2.8). As the upper part of each wave front enters the ionosphere before the lower part, the upper part feels the effect of lowered index of refraction first. Consequently, the upper part of each wave front has an increased phase velocity so that the entire wave front as it enters the ionosphere wheels about like a column of soldiers obeying the command "column right." As the higher parts of the ionosphere have a greater ion density, the bending effect on the upper part of the wave front is greatest, so that the wheeling process continues and the waves are directed to the point B.

As propagation over long distances occurs if the direction of propagation becomes horizontal at some point, it is not necessary in the case of propagation over any considerable distance for the wave to be bent through as large an angle as is necessary for reflection at vertical or nearly vertical incidence. If we designate by r the angle which the direction of propagation (or ray) makes with the vertical after refraction and by i the angle which the ray makes with the vertical on entering the ionosphere, we may apply Snell's law directly from optics:

$$\eta=\sin i/\sin r. \tag{2.40}$$

For reflection, as explained above, it is necessary that $r=90°$ or $\sin r=1$. Then $\eta=\sin i$. Substituting for η the value from eq 2.24, we have

$$\sqrt{1-f_N^2/f^2}=\sin i, \tag{2.41}$$

which leads by simple trigonometric transformation to

$$f=f_N/\cos i=f_N \sec i \tag{2.42}$$

Remembering that f_N is the critical frequency for an ion density of N, eq 2.42 shows that if a certain frequency is returned at vertical incidence by the ion density at a height h then the same ion density will return at oblique incidence, i, a frequency sec i times as great.

Thus, the significance of the study of vertical-incidence phenomena to the problems of radio propagation is apparent. If the critical frequency for the ionosphere over a certain point on the earth is known, the frequencies which will be reflected over that point can be calculated. The angle of incidence can be calculated from geometrical considerations, knowing the separation of the two points between which propagation is to be established and the height at which reflection occurs. The height in this consideration is the virtual height, not the actual height, for we are dealing with the apparent paths of the radio waves.

In the practical problem of determining suitable frequencies for communication over long distances additional considerations must be taken into account, including the curvature of the earth, because, for large angles of incidence, the radio waves impinge on the ionosphere at smaller angles than they make with the vertical at the point of departure from the earth's surface.

2.6. Absorption of Radio Waves in the Ionosphere

a. Nondeviative Absorption

As was shown in section 2.3, a plane radio wave entering the ionosphere sets the ions into motion. Thus the electromagnetic energy of the radio waves is in part transformed into kinetic energy represented by the motion of the ions. If the ions do not collide with gas molecules or other ions, all of this energy is reconverted to electromagnetic energy and the wave continues to be propagated (although not necessarily in its original direction) with undiminished intensity. But if the ions engage in collisions, they dissipate the energy which they have acquired from the wave and consequently the wave is attenuated. This attenuation or absorption is, as might be expected, proportional to the number of ions N and to the frequency of collisions. Therefore, for a given wave, attenuation is ordinarily greatest in the region where the product of ion density and collisional frequency is greatest. If the ion density is sufficiently below the critical value so that the wave is not greatly deviated but the collisional frequency is so high that absorption is appreciable, the absorption is known as nondeviative absorption. Nondeviative absorption is very important in most radio-propagation problems.

To examine the process of absorption we shall consider the simplest case, that in which a mag-

netic field is absent. To do this we shall rewrite eq 2.13

$$eE_x = m d^2x/dt^2 + \nu m dx/dt, \qquad (2.43)$$

in which acceleration du/dt has been replaced by d^2x/dt^2, and a second term, $\nu m dx/dt$, has been added to take into account the effect of collisions, this term having the significance of a frictional term. (It may be noted that all three terms in eq 2.43 have the dimensions of a force, MLT^{-2}). This equation is physically valid for the *average* displacement of a single ion if E is constant or if $(dE/dt)/E$ is small with respect to ν. It is also valid for the *average displacement of a large number of ions* if $(dE/dt)/E$ is small with respect to $N\nu$. We shall also assume that E varies sinusoidally as in a wave; for this purpose it will be convenient to adopt the form for E given in eq 2.12c, namely,

$$E_x = E_{x_0} \cos(\omega t - \beta z).$$

By differentiation and substitution, it may be seen that

$$x = A \cos(\omega t - \beta z) + B \sin(\omega t - \beta z) \qquad (2.44)$$

is a solution of eq 2.43, in which $B = +(eE_{x_0}/m)(\nu/\omega)/(\omega^2+\nu^2)$ and $A = -(eE_{x_0}/m)/(\omega^2+\nu^2)$. Remembering from eq 2.21 that $I = Ne dx/dt$, we may now write the value for I in a region where there are collisions between the ions and the molecules.

$$I = Ne dx/dt = Ne\omega[-A \sin(\omega t - \beta z) + B \cos(\omega t - \beta z)]$$

$$= [Ne^2 E_{x_0}\omega/m(\omega^2+\nu^2)][\sin(\omega t - \beta z) + (\nu/\omega)\cos(\omega t - \beta z)]. \qquad (2.45)$$

Substituting this value of I in eq 2.6*a*, we have

$$\{[Ne^2\nu/m(\omega^2+\nu^2)]\cos(\omega t - \beta z) -$$

$$[\kappa - Ne^2/m(\omega^2+\nu^2)]\omega \sin(\omega t - \beta z)\}E_{x_0} =$$

$$-\partial H_y/\partial z. \qquad (2.46)$$

Substituting the expression for E_x in eq 2.7b, we have

$$-\mu dH_y/dt = (\partial/\partial z)E_{x_0} \cos(\omega t - \beta z). \qquad (2.47)$$

Proceeding as with eq 2.8*a* and 2.8*b*, that is, differentiating eq 2.46 with respect to t and eq 2.47 with respect to z, and eliminating the term $\partial(\partial H)/\partial z \partial t$ by substitution, we have

$$\mu\omega^2 E_{x_0}\{[Ne^2/m(\omega^2+\nu^2) - \kappa]\cos(\omega t - \beta z) -$$

$$[Ne^2\nu/m\omega(\omega^2+\nu^2)]\sin(\omega t - \beta z)\} =$$

$$(\partial^2/\partial z^2)E_{x_0} \cos(\omega t - \beta z), \qquad (2.48)$$

which is the equation satisfied for the simple propagation of a wave in an ionized medium if collisions are appreciable.

A solution of this equation is

$$E_{x_0} = E'_{x_0}\epsilon^{-kz} \cos(\omega t - \beta z), \qquad (2.49)$$

in which

$$2k\beta = -\mu Ne^2\nu\omega/m(\omega^2+\nu^2)$$

and

$$k^2 - \beta^2 = \mu\omega^2[Ne^2/m(\omega^2+\nu^2) - \kappa]$$

If

$$\omega^2 \gg \nu^2,\ \kappa = \kappa_0,\ \mu = \mu_0,\ c = 1/\sqrt{\kappa_0\mu_0},$$

$$\beta = (\omega/c)\sqrt{1 - Ne^2/m\kappa_0\omega} = -\omega\eta/c$$

$$k = (\mu_0 Ne^2\nu/m\omega)(c/2\omega\eta) = \nu Ne^2/2c\eta\kappa_0 m\omega^2 = \nu f_N^2/2c\eta f^2.$$

Equation 2.49 expresses the fact that a radio wave propagated in an ionized medium, where the collisional frequency is appreciable, decreases in intensity by absorption. The decrease in intensity in passing through 1 m of the region is determined by the absorption coefficient k, that is, the intensity is decreased by the factor ϵ^{-k}. Since k usually varies with height, the loss in intensity is proportional to $\epsilon^{-\int k dz}$. As was anticipated earlier in this section, k is proportional to N and ν. It is inversely proportional to the square of ω, the angular frequency of the wave. It is inversely proportional to the refractive index η, and hence directly proportional to the time spent in traversing a unit distance in the medium. The quantity $\alpha = \int k dz$ is called the absorption index.

b. Deviative Absorption

When η does not differ greatly from unity the absorption is called nondeviative, but when η is small, resulting in considerable deviation of the wave from its original direction of propagation, the absorption is called deviative. This case occurs when the ion density is close to the critical value. Under the circumstance, η approaches 0 and k approaches ∞. At the critical frequency $\eta = 0$, so that a wave propagated vertically into the ionosphere at this frequency is neither reflected nor propagated; it is completely absorbed

c. Absorption in the Presence of a Magnetic Field

In the presence of a magnetic field the process of absorption is more complicated. In particular, if the wave frequency is close to the gyrofrequency, the absorption becomes very great, for when this condition obtains each ion can absorb a large amount of energy from the wave, which is completely dissipated when it engages in a collision. The formula for the nondeviative absorption coefficient in the presence of a magnetic field is

935333 O - 51 - 2

$$k_m = \frac{k_0}{f^2}\left[p + \frac{2(p-1)^2 p}{(p^2-1)-p^2\tau^2}\right], \qquad (2.50)$$

where k_0 is the absorption coefficient for a frequency of 1 Mc in the absence of a magnetic field, and p is a solution of

$$p^2(1-\tau^2) - p(2-\tau^2 \sin^2\theta) + 1 = 0,$$

$\tau = f_H/f$, and θ is angle between the direction of propagation and the geomagnetic field.

2.7 References

E. V. Appleton, Wireless studies of the ionosphere, J. Inst. Elec. Engrs. (London, **71**, 642 (1932).

Mary Taylor, Appleton-Hartree formula and wave propagation, Proc. Phys. Soc. **46**, 408 (May 1, 1934).

H. Booker, General properties of formulae of the magneto-ionic theory, Proc. Roy. Soc. (London) [A] **147**, 352 (1934).

S. K. Mitra, Report on the present state of our knowledge of the ionosphere, Proc. Ntl. Inst. Sci. India **1**, 131 (1935).

H. G. Booker, Application of the magneto-ionic theory to the ionosphere, Proc. Roy. Soc. (London) [A] **150**, 267 (1935).

H. G. Booker, Oblique propagation of electromagnetic waves in a non-isotropic medium, Proc. Roy. Soc. (London) [A] **155**, 235 (1936).

N. Smith, Extension of normal-incidence ionosphere measurements to oblique-incidence radio transmission, J. Research NBS **19**, 89 (1937) RP1013.

H. R. Mimno, The physics of the ionosphere, Rev. Modern Phys. **9**, 1 (1937).

N. Smith, Application of vertical-incidence ionosphere measurements to oblique-incidence transmission, J. Research NBS **20**, 683 (1938) RP1100.

H. G. Booker, Propagation of wave pockets incident obliquely upon a stratified doubly refracting ionosphere, Trans. Roy. Soc. (London) [A] **237**, 411 (1938).

O. Rydbeck, Propagation of electromagnetic waves in an ionized medium, Phil. Mag. **30**, 282 (1940).

CHAPTER 3

MEASUREMENT TECHNICS

3.1. Purpose

The basis of our knowledge of the characteristics and behavior of the ionosphere and of some of the other factors, such as radio noise, which influence radio transmission and reception at ionospheric frequencies lies in the measurement of these factors. Accordingly, in this chapter are described the fundamentals of those measurement technics which apply to the specialized subject of radio propagation, in order to provide a background of information on how the various factors which are dealt with in other parts of this book are actually measured. In general, only the methods most widely used at present are given; there is no attempt to treat the subject of radio propagation measurements exhaustively.

3.2. Measurement of Virtual Height

a. Group Retardation or Pulse Method

The pulse method, originally devised by Breit and Tuve, is the most straightforward way of measuring the virtual height (h') of the ionosphere layers. A succession of pulses of short duration (30 to 100 μ sec) is sent vertically upward at regular intervals (1/30 to 1/120 sec) by a transmitter and at the receiver the time delay of the echo is measured on an oscilloscope, usually a cathode-ray tube, with some form of fast time base. The virtual height h' in kilometers, corresponding to a delay time t_s, in microseconds, is then, for each echo:

$$h' = 0.15\, t_s.$$

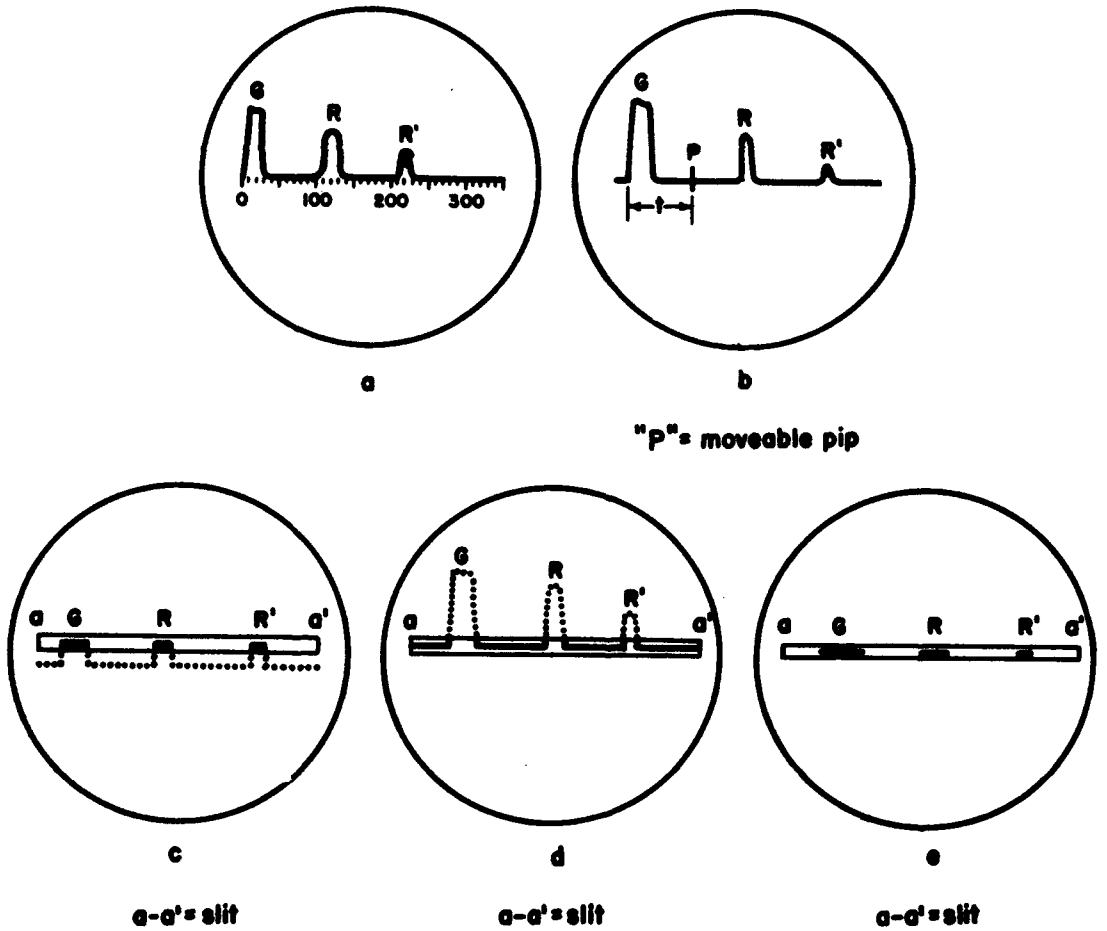


FIGURE 3.1. *Methods of presentation of ionosphere echoes.*

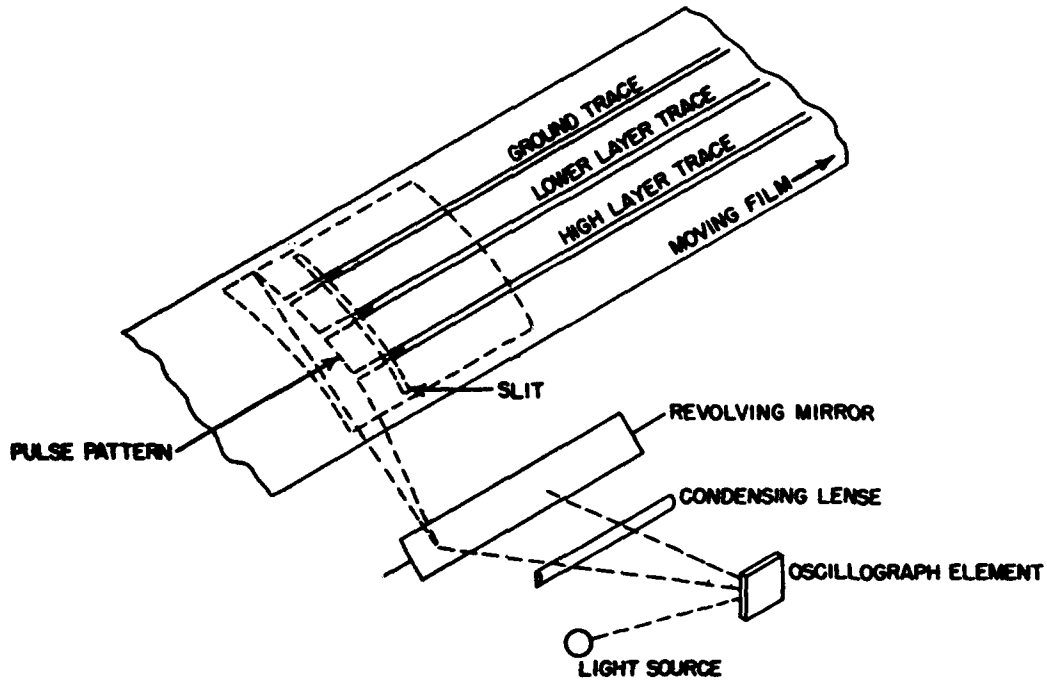


FIGURE 3.2. *Slit method of recording virtual heights.*

The method of presentation of the ionosphere echoes generally employed is the so-called "A" scan, similar to the radar "A" scan. In this method, the time base of the cathode-ray oscilloscope is synchronized with the pulse repetition frequency of the transmitter, and a stationary pulse pattern, of one of the types shown in figure 3.1 is obtained. The retardation is measured from the beginning of the ground pulse "G" to the leading edge of the echo "R", and can be calibrated directly in terms of virtual height in kilometers. Virtual-height markers, similar to radar range markers, may be put directly on the time base, as in figure 3.1,a, or a movable marker, operated by either a phase shifter or a variable delay circuit calibrated in kilometers of virtual height, may be set on each echo in turn and measurements thus made (see fig. 3.1,b).

For manual measurements, the amplitude-modulated A scan, shown in figures 3.1,a, and 3.1, b, is generally used. The equipment needed for this type of measurement is relatively simple; it consists of (1) a pulsed self-excited power-oscillator, delivering perhaps 50- or 100-w peak power and covering the desired frequency range, generally from 1 to 20 Mc, (2) a communications-type receiver, broad-banded to accept the bandwidth of the pulse used, (3) a pulse generator, delivering a pulse about 50 μ sec long to trigger both the transmitter and the cathode-ray oscilloscope circuits, (4) a cathode-ray oscilloscope, with simple triggered sweep and circuit for generating height-marker pips, and (5) broad-band antennas, one for transmitting and one for receiving, such as vertical rhombics or inverted vees having good vertical directivity. The output of the receiver is placed, together with the marker pip voltage, on the vertical sweep plates of the scope; the trigger pulse is used to start the time base sweep of the scope. The transmitter may be placed right beside the receiver, provided the time constants of the receiver are sufficiently small to prevent blocking. In operation the receiver is tuned to a given frequency which is free from interference, the transmitter tuned to the receiver by observing the amplitude of the echo pulses, and the height measurement made and recorded. This procedure is repeated for as many frequencies as desired, generally for a series of frequencies starting at the lowest and going to the highest frequency on which any echo is visible.

For automatic recording, some modification of the basic A sweep is necessary. As in figure 3.1,c, the peaks of the pulses can be clipped, so that all echoes appear in a slit a-a' as a mark, and then a strip of photographic paper moved slowly vertically past the slit, so that there is a continuous record made of the retardation time or virtual heights appearing on the scope, as shown in figure 3.2. Alternatively, the slit may be placed on the base line (fig. 3.1, d) so that the echo appears as a white trace on the recording paper, all the rest of the paper being blackened. The scope may also be intensity-modulated rather than amplitude-modulated (A2 scan), the resultant echo-dots or "blips" appearing on the slit (fig. 3.1,e).

If an automatic height record is made with the receiver and transmitter tuned to one frequency, the resultant record is called a "fixed frequency" or "$h't$" record. If, however, the receiver and transmitter tuning are continuously changed over

Fig. a. Fig. b.

FIGURE 3.3. *Typical automatic ionosphere records.*
(a) $h't$ and (b) $h'f$ traces.

a band of frequencies while the record is being made, we obtain a record of virtual height against frequency; this is known as a "sweep frequency," "multifrequency," or "$h'f$" record. Figure 3.3,a, shows a typical $h't$ record and figure 3.3,b, an $h'f$ record, made in the above manner.

Modern ionosphere recording equipments are fully automatic, in that the tuning and tracking of the receiver, transmitter, and recording camera are all done continuously and automatically with clocks and relays for making sweeps at predetermined intervals.

Another type of presentation on the cathode-ray scope, the "sweep" scan, is being used on modern ionosphere recorders. This is like the A2 scan in principle, except that the time base sweep is vertical; the echoes appear as bright spots on the vertical sweep. As the transmitter and tracked receiver are tuned over the frequency range (1 to 20 Mc, say) each succeeding vertical sweep is displaced slightly in a horizontal direction on the scope, from the preceding ones. The $h'f$ curve is thus built up on the screen of the scope; by use of a fast sweep (1 to 20 Mc in 15 sec, say) and a long persistence screen on the scope, the $h'f$ curve appears as a whole, resembling a radar "B" type scan. Height markers appear as horizontal lines, and frequency markers as vertical lines. Figure 3.6 indicates the appearance of such a record.

Typical of modern ionosphere equipment is the CRPL model C automatic ionosphere recorder. Both A2 and sweep scans are provided, with provision for direct photography of the sweep scan on 16-mm film and moving-film photographs of the A2 scan on 35-mm film. The latter is for regular operations, the former is for studying the ionosphere during storms, eclipses or other unusual phenomena. The frequency range is 1 to 20 Mc, in one band, using a superheterodyne principle in both transmitter and receiver, with a common variable frequency oscillator to ensure frequency tracking. The peak power output of the transmitter is about 10 kw (older types equipment ran about ½ kw). The frequency sweep is rapid, requiring but 15 sec to cover the range from 1 to 20 Mc. The $h'f$ record has a logarithmic frequency scale, so that a logarithmic muf factor slider can be used.

The velocity of light times the delay time measured by this method gives the time a pulse traveling with the velocity of light would require to travel to the virtual height of the layer and back. Since the group velocity (u) decreases as the wave penetrates into regions of higher and higher ionization density, the wave travels over an appreciable part of its path with less than the velocity of light. At vertical incidence it will reach a point finally when it actually decreases to zero, reverses direction, and, when it finally emerges from the ionosphere layer, it has attained its initial velocity again. The virtual height (h') measured by the group retardation or pulse method is thus always greater than the actual height (h), because of the retardation encountered in the ionized region. The pulse method of measuring virtual heights is the one used in all modern ionosphere equipment.

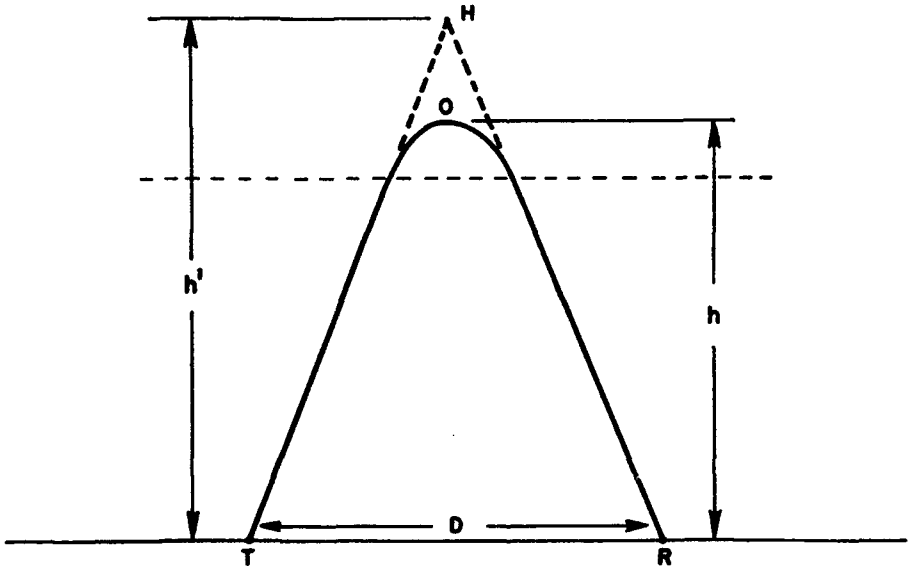


FIGURE 3.4. *Virtual path difference between sky and ground waves.*

b. Frequency Change Method.

In obtaining the virtual heights of the ionosphere layers by the frequency change method developed by Appleton, it is necessary to know both the amount of frequency variation of the transmitter and the maxima and minima observed in the receiver caused by the change of frequency. The transmitter is usually separated from the receiver by several kilometers, so that the skywave field intensities approach those of the ground wave, thus making the interference phenomena more readily discernible.

The virtual path difference between the sky- and ground-wave paths is shown in figure 3.4. It is $\overline{THR}-D$, where D is the distance between the transmitter and the receiver. Then the difference N between the number of wavelengths in the two paths is $N=(\overline{THR}-D)/\lambda=(f/c)(\overline{THR}-D)$. The value of N can be deduced from the interference phenomena of the minima and maxima observed, from which the equivalent path $\overline{THR}$ can be obtained, and by triangulation, the virtual height (h') is derived.

c. Oblique-Angle Method

This method was also developed by Appleton, and is very similar to the frequency change method.

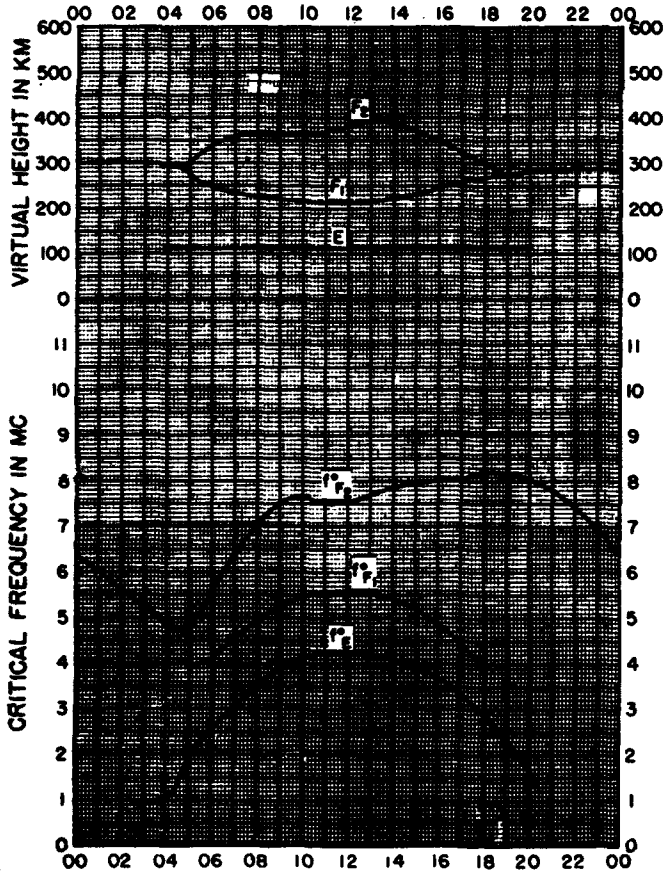


FIGURE 3.5. *Typical dirurnal curves.*

In this method the fading effect on the vertical antenna, and the fading effect on a loop antenna set at right angles to the plane of the wave propagation, are used to derive the angle at which the downcoming skywave meets the surface of the earth. When this angle is found it is a simple triangulation problem to obtain the height of the layer as the distance between the transmitter and receiver is known.

It has been shown that the group retardation, frequency change, and oblique angle methods, all measure the same quantity, i. e., the equivalent or virtual height (h') of the ionosphere layer.

3.3. Ionosphere Records

The technic generally employed at present for obtaining ionosphere characteristics involves taking automatic pulse retardation records at frequent intervals over the 24-hr period of the day and scaling from such recordings the pertinent characteristics, which are then plotted as diurnal curves (fig. 3.5).

Ionosphere recorders have been developed that are completely automatic in that they are switched on at preset times, obtain a multifrequency sweep on photographic film, and switch off until the next sweep is to be taken. A typical multifrequency sweep record is shown in figure 3.6.

There are definite seasonal and diurnal variations in the ionosphere characteristics at a given station. Normal conditions during day and night for different times of the year, may be visualized by examining typical records for (a) a summer night, (b) a summer day, (c) a winter night, and (d) a winter day, as shown in figures 3.6 through 3.9.

3.4. Measurement of Sky-Wave Field Intensity

a. Characteristics of Sky Waves Affecting Field-Intensity Measurement

The analysis of ionosphere heights yields information on delay times of radio waves, the conditions which determine whether reflection will or will not take place, and to some extent, the appearance of downcoming waves reflected from the ionosphere. It is also necessary to know, for propagation studies, the field intensities of the downcoming sky waves.

The technics for measuring field intensities of radio waves have been in general well developed. In a downcoming sky-wave, however, we are not dealing with a steady wave of constant amplitude and phase, but one which fades rapidly and greatly, whose polarization is constantly changing, which is composed of not one but many component waves, which is affected by reflection at the ground near the receiver, and which is subject to the variations in height and energy absorption in the ionosphere, and to focusing by the ionosphere.

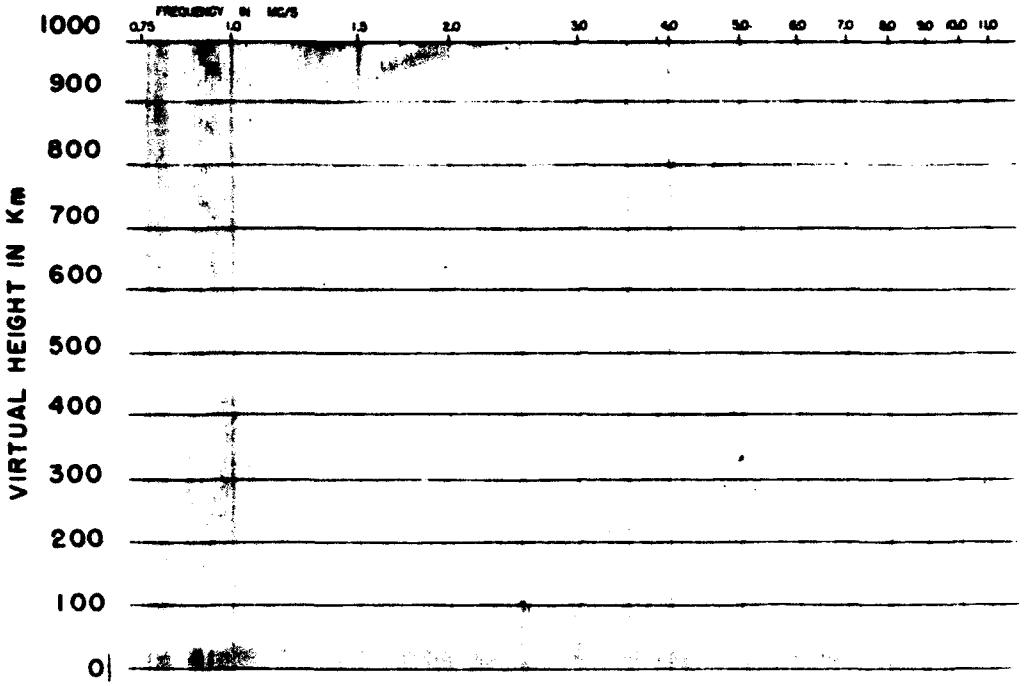


FIGURE 3.6. *Normal summer night.*

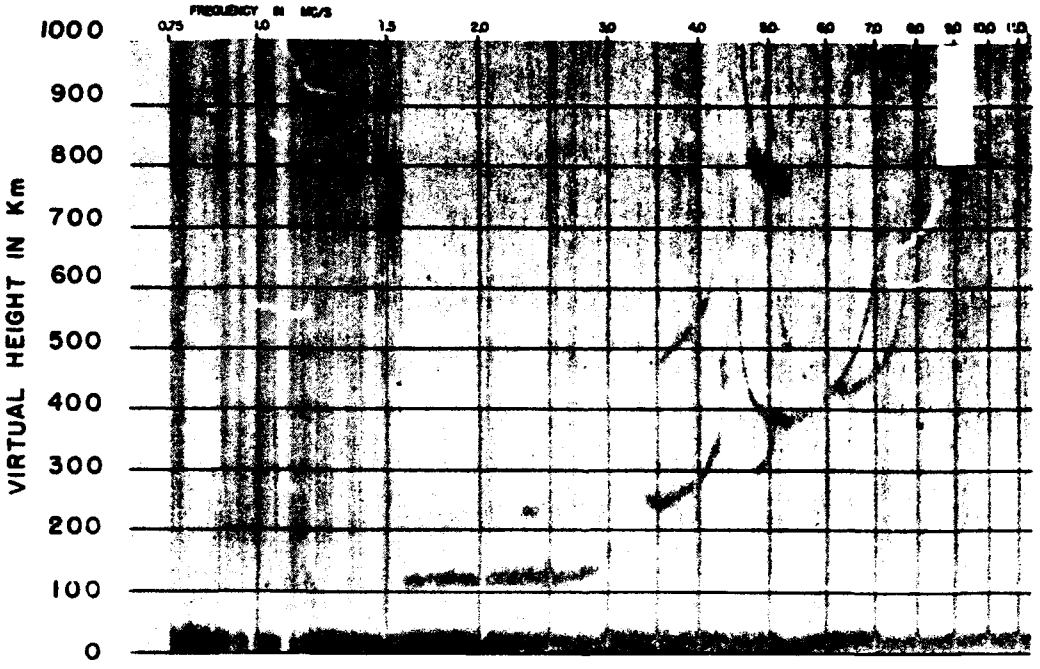


FIGURE 3.7. *Normal summer day.*

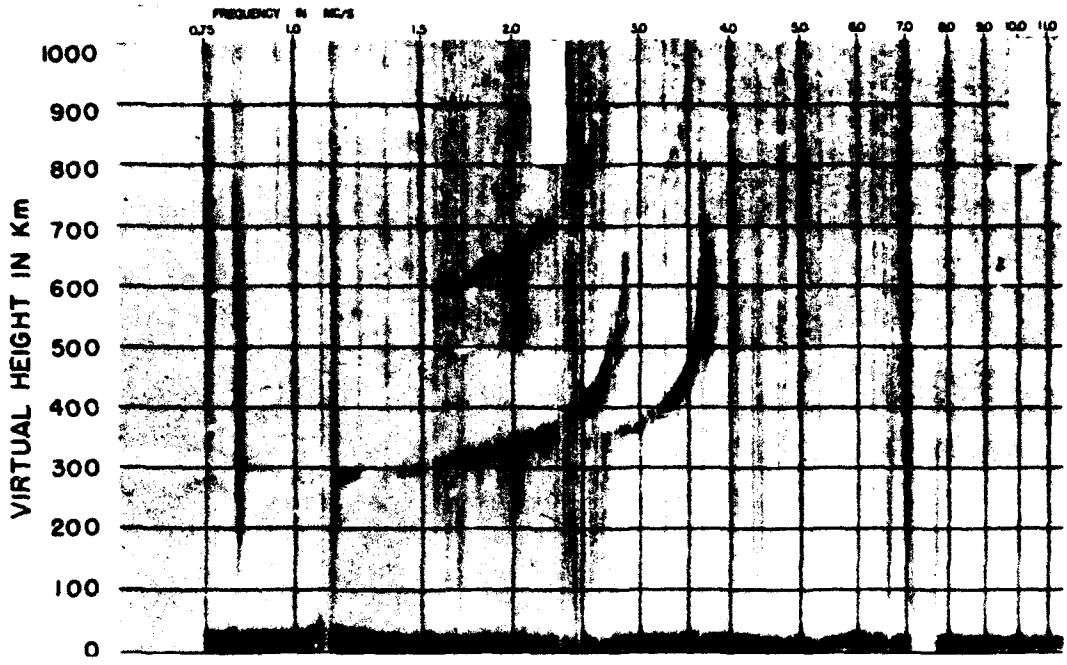


FIGURE 3.8. *Normal winter night.*

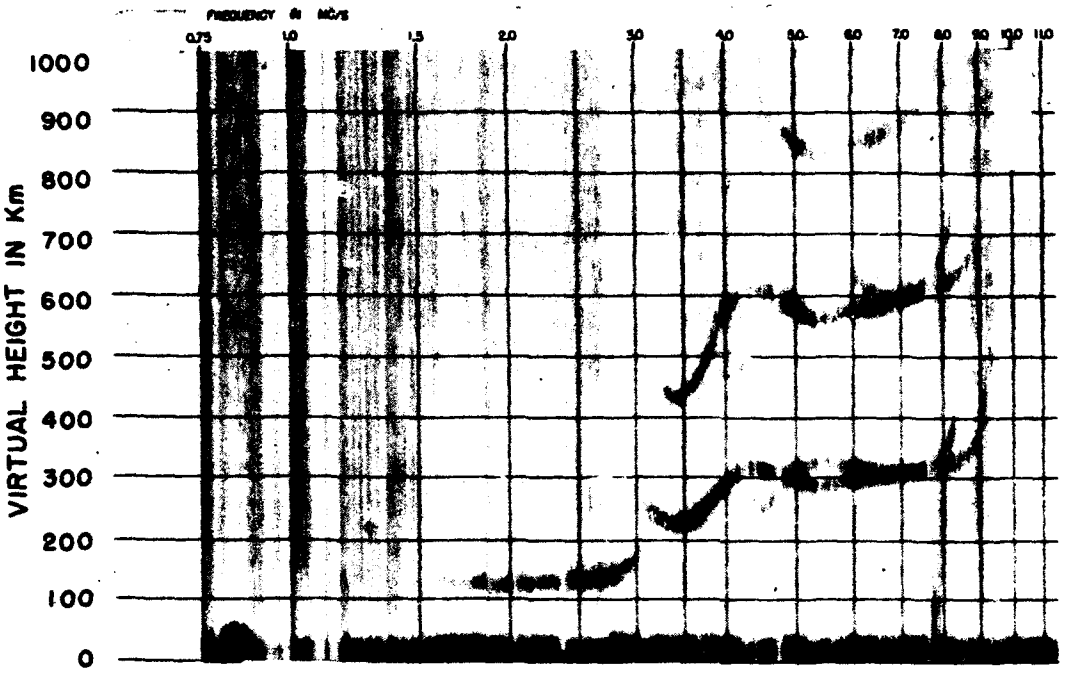


FIGURE 3.9. *Normal winter day.*

In order to interpret field-intensity measurements in terms of the intensity of the downcoming sky wave, it is necessary to determine the pickup factor of the antenna, which is dependent upon a number of parameters. Some considerations which need to be taken into account in deducing the intensity of the downcoming sky wave in ionosphere transmission from the actual voltages induced in the receiving antenna are mentioned in the following paragraph, quoted from the IRE Standards on Radio Wave Propagation: Measuring Methods, 1942:

> In analyzing or reporting measured data, it is important to take into account the influence of the receiving site and surroundings (including topography), electrical properties of the ground, the proximity of disturbing structures, etc., and the orientation and height above ground of the receiving antenna. In view of the effect of ground reflection, the question arises as to whether the measured values of the resultant field should be reported or an attempt should be made to measure (or compute) the value of the incident field. In the case of a well-defined incident angle and when the electrical properties of the ground are accurately known, such a measurement would indicate the actual variation of the field at points equidistant from the transmitter without regard to the nature of the ground at various receiving points. In general, dealing with the incident field will eliminate confusion in the study of laws of attenuation and in the measurement of the directional characteristics of antennas. However, the measurement of the incident field is always an indirect measurement involving a knowledge of the angle or angles of arrival and the electrical properties of the ground, and its proper determination thus introduces some difficulty.

The effective length of the antenna is defined as the length which, multiplied by the field intensity E in volts per meter, gives the open-circuit voltage induced in the antenna when the electric field vector E is parallel to the antenna and the direction of propagation of the received wave is perpendicular to the antenna. For example, the effective length of a half-wave doublet antenna would be the value calculated from a field intensity E_i, polarized as shown in figure 3.10.

This is the maximum value of voltage that would be induced in this antenna from a given strength of field; for fields of this *same* intensity but a different direction or polarization, a voltage pick-up factor is used in place of the effective length. This is particularly important in sky-wave reception, as the pick-up factor varies considerably with the incident angle of the sky wave, whereas the effective length is a function only of the dimensions of the antenna. It is often convenient to express this also in terms of the cross-sectional area over which the antenna can extract energy from the radio wave.

The received field is usually a combination of the direct field due to the incident sky wave, together with that due to the wave reflected from the ground. The resultant electric vector at the antenna is therefore dependent upon variations of the ground-reflection coefficient as well as upon the instantaneous changes in both the amplitude and direction of the incident sky wave. Because of these variations, it is generally best to record the median or rms values of the incident field rather than the instantaneous values. The median sky-wave field intensity is that value which is exceeded 50 percent of the time. Because of roughness of the ionosphere the polarization as well as the intensity of the wave after reflection is random. The rapid variations in field intensity of the reflected sky wave, (that incident upon the receiving antenna) arise from the combination of a large number of electric vectors of random polarization and phase which over a period of time combine to give a distribution of intensity like that shown in figure 3.11. This is called the "Rayleigh distribution" of the instantaneous sky-wave field intensity, and gives the percentage of time that the ratio of instantaneous value to median value of the sky-wave intensity exceeds given values. This distribution is based on a constant value of sky wave incident on the ionosphere; in practical cases, large corrections must also be made for both diurnal and long-period changes in absorption of the wave in the ionosphere.

The curves given in figure 3.12 show the effect

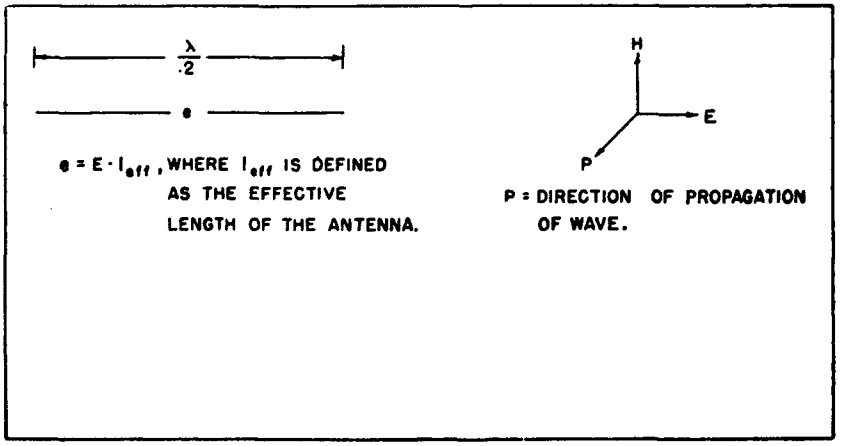


FIGURE 3.10. *Effective length of the antenna in relation to direction of propagation of wave.*

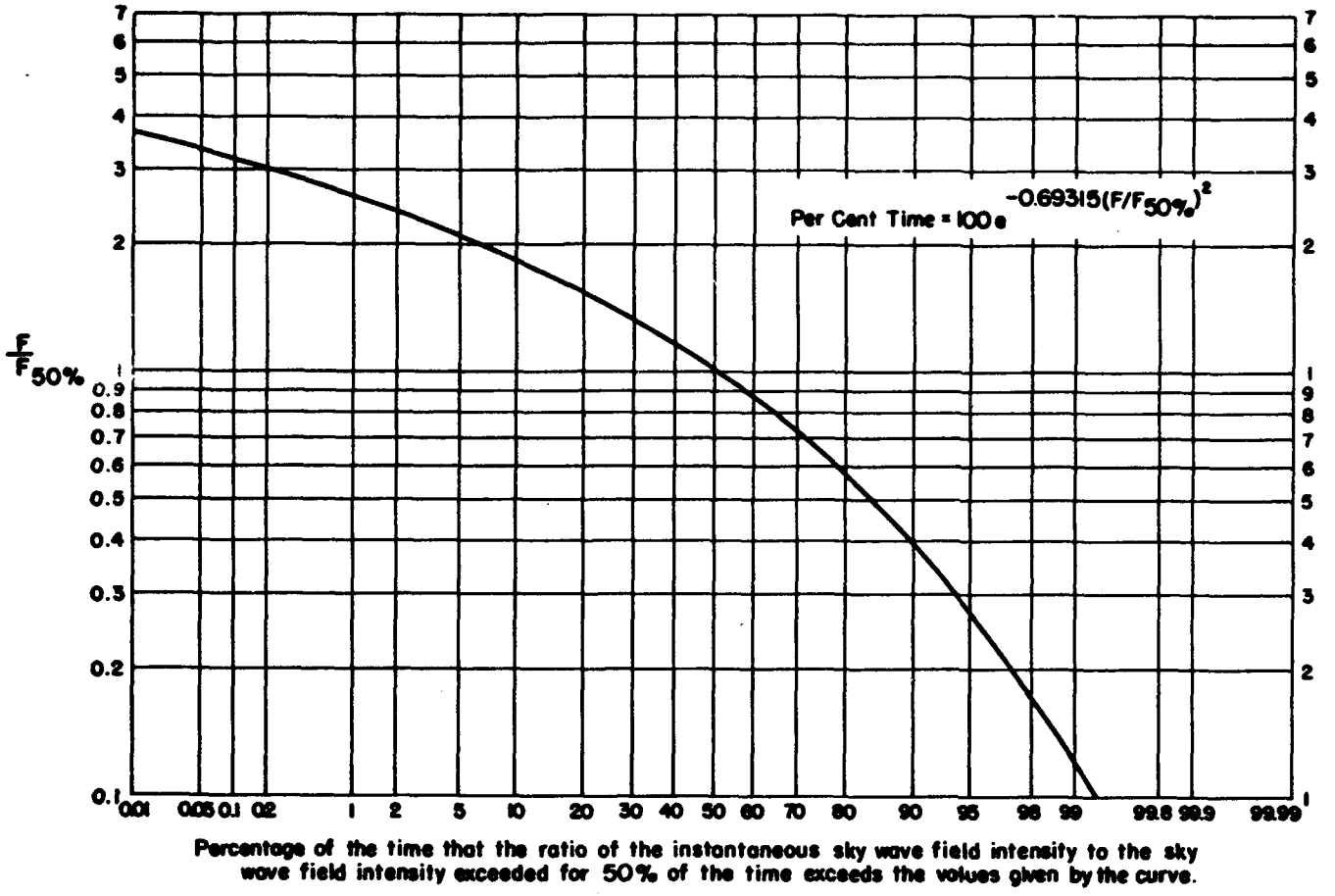


FIGURE 3.11. *Rayleigh distribution of the instantaneous sky-wave field intensity.*

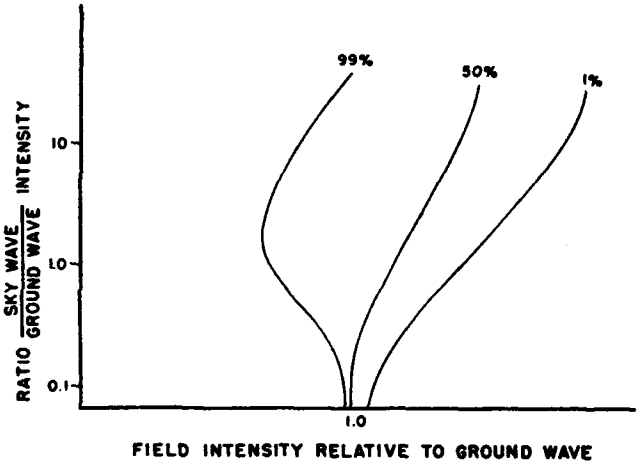


FIGURE 3.12. *Received intensity relative to ground wave.*

of the combination of sky wave and ground wave at the receiving antenna, for various ratios of incident sky-wave to ground-wave field intensity. It may be shown that when the ratio of sky-wave intensity to ground-wave intensity exceeds 10, the resultant field-intensity distribution is essentially the Rayleigh distribution given in figure 3.11 above. As the ratio of sky-wave to ground-wave intensity decreases to 0.1, however, the values of received field intensities converge to narrower limits and approach the ground-wave value in the limit when the sky wave is negligible. Thus the effect of simultaneous sky-wave and ground-wave reception is to increase the peak values of received fields, so that considerable fluctuations occur, and also to increase the rms values. This is shown in figure 3.13, which is a continuous automatic recording of the intensity of received signals from the Allentown, Pa., radio range, on 319 kc, as received at Sterling, Va. The received intensity is practically uniform throughout the day; during the night period, however, when sky-wave reflection occurs at this frequency, fading results and the field intensity fluctuates as shown.

b. Automatic Field-Intensity Recording

A single observation or short series of observations of the field intensity of a sky wave is completely inadequate to yield information of practical use in radio-propagation calculations. It has thus been necessary to develop equipment for recording field intensities continuously and automatically over long intervals of time.

The basic field-intensity recorder is a recording radio-frequency field-intensity meter, consisting generally of a calibrated antenna, a radio receiver and a recording output meter. Many combina-

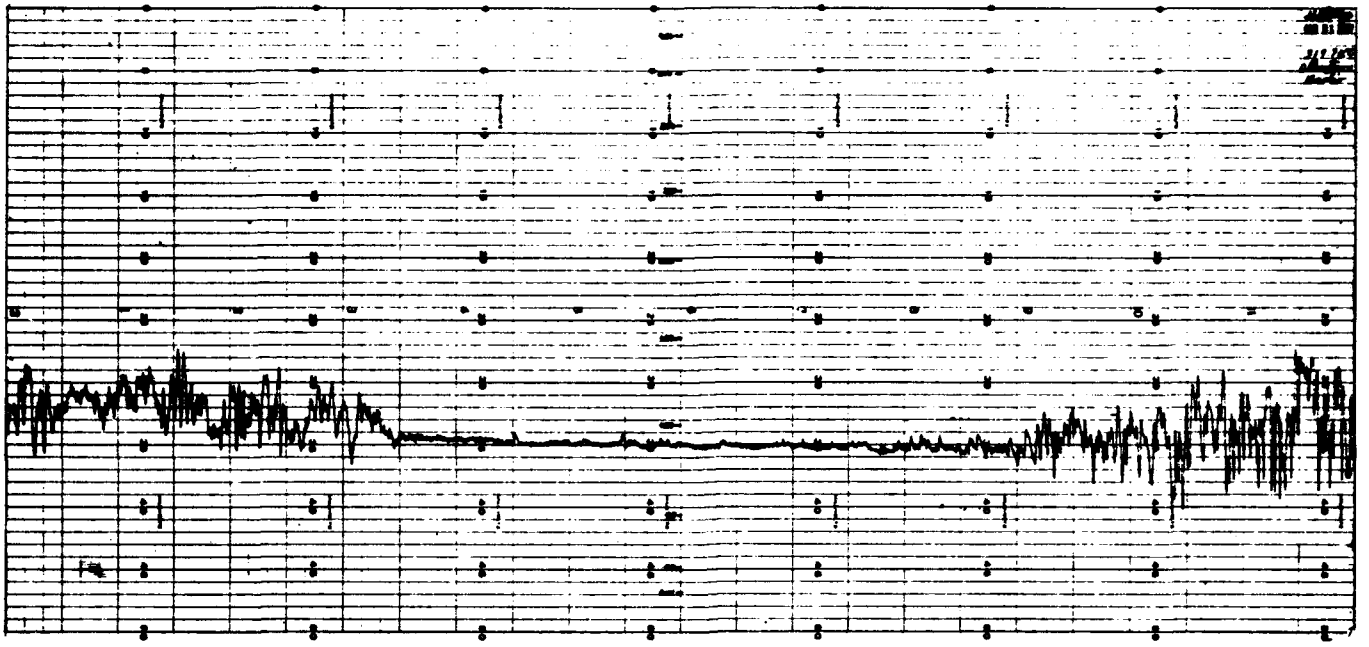


FIGURE 3.13. *Effect of simultaneous ground-wave and sky-wave reception.*

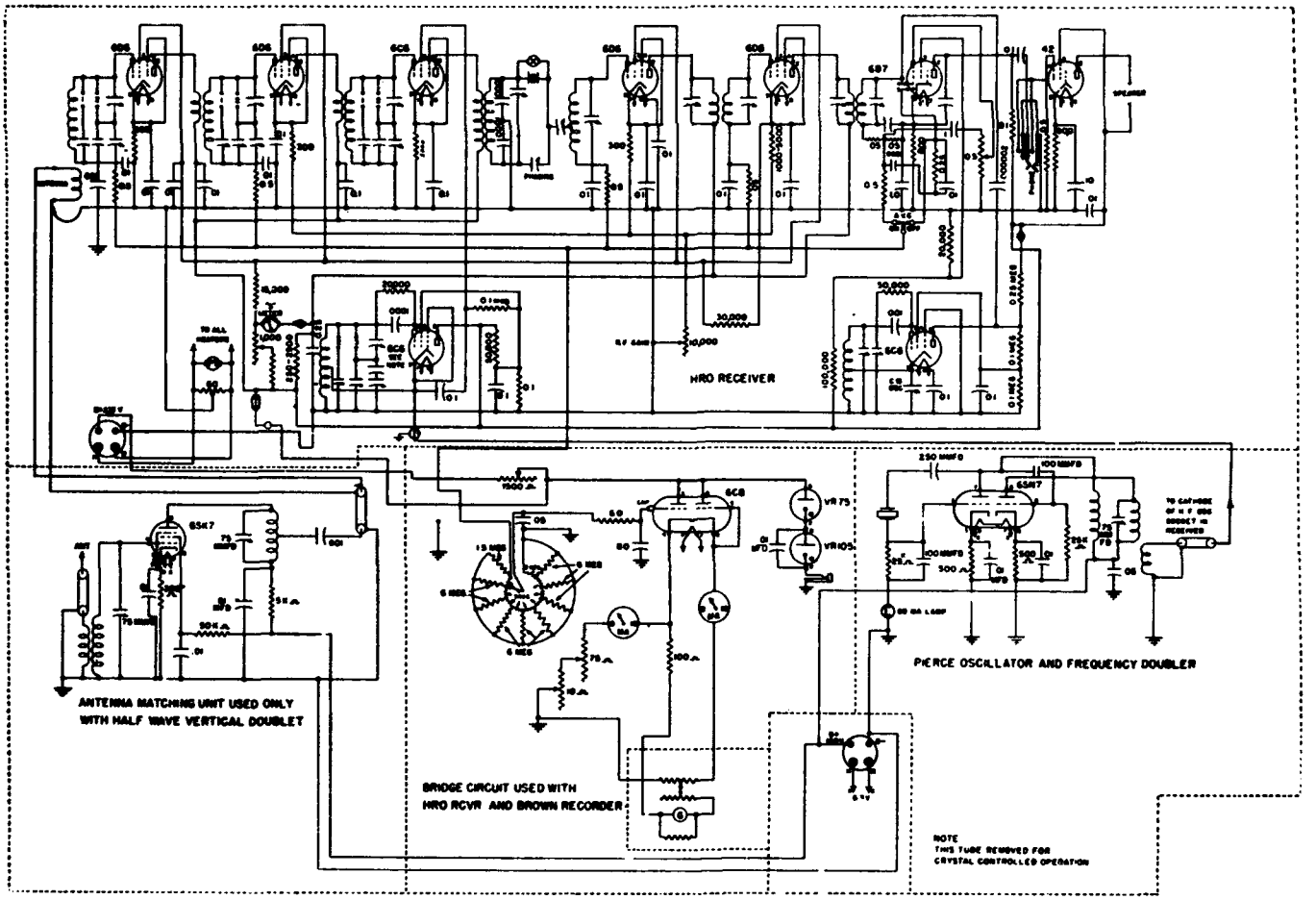


FIGURE 3.14. *Automatic field-intensity recording unit.*

tions of these components have been devised, of which the following, in use at the National Bureau of Standards, is typical.

A communications-type receiver, fed by a calibrated antenna, is used, and the avc voltage is supplied to an attenuator and vacuum-tube bridge circuit, as shown in figure 3.14. Two resistance arms of this bridge and the balance galvanometer are contained in a commercial galvanometer-type recording potentiometer which operates to adjust the slide wire for balance. The voltage across the slide wire is about 1.2 v for this galvanometer-type recorder.

For this type circuit it was found that $R_p \cong \log E$, where R_p is the plate resistance of the tube in the bridge circuit whose grid is supplied from the avc voltage, and E is the voltage input to the receiver.

By adjustment of the attenuator, the recorder can be set to operate on very weak signals or strong local signals, and the scale can be expanded where desired. Where average values of intensity are of interest, a circuit with a long time constant (48 sec or longer) is used to average out the fluctuations and allow easier scaling of the records. In the galvanometer-type recorder, the galvanometer itself is also damped in addition to the use of a suitable time constant.

Figure 3.14 shows the complete circuit of the galvanometer-type field-intensity recorder. Two voltage regulator tubes are used in the bridge-plate supply to obtain constancy of plate voltage under variations in line conditions. It has been found most desirable and practical to use crystal control of the receiver oscillator to prevent changes in the frequency of the receiver with consequent changes in calibration of the recorder, to include a-c line-voltage controls, and to install the equipment in a building in which the temperature is thermostatically controlled.

c. Printing-Type Recording Potentiometer

A printing-type recording potentiometer which operates on a fixed time cycle at 15-sec intervals is also being used to obtain field-intensity recordings. This unit, which was originally designed for use in recording d-c thermocouple voltages, has two separate input terminals, which may be connected in parallel for field intensity recordings. A standard cell is incorporated in the instrument, which gives compensation for bridge current, and the instrument itself records d-c voltages from 0 to 5 mv with an accuracy of 0.5 percent, in spite of line voltage variations from 100 to 130 and frequency changes of from 50 to 70 cycles.

The d-c potential applied to the bridge is changed to alternating current through a vibrator unit, amplified and applied to one winding of a two-phase driving motor. The other winding is supplied with current at line frequency. When current flows in both windings the motor operates and drives the recording arm toward the position of balance.

To provide a long time constant, the unit is connected through a suitable attenuator to the output of the radio receiver. By suitably proportioning the resistances and capacitances usable time constants of the order of 60 to 90 sec may be obtained and give considerable smoothing of the actual impressed voltages.

d. Calibration of Equipment

Calibration of the field-intensity recording equipment may be divided into two phases, (1) calibration of the recorder scale in terms of the microvolts applied by a signal generator to the receiver input, and (2) over-all calibration of the receiving antenna and recorder to indicate absolute values of received fields.

In the case of vertical-incidence field-intensity measurements the angle of arrival remains essentially constant, the pattern of the receiving antenna (including the effect of ground reflection) can be calculated. For oblique-incidence work, some assumptions have to be made regarding the relation of the downcoming sky-wave field to the field at the receiving antenna, and the effective length of the receiving antenna must be calculated or measured. The recorder then measures the field intensity at and in the direction of the receiving antenna, and further deductions as to the character of the sky-wave field involve considerations of polarization and records of sky-wave propagation.

One method of calibrating antennas for oblique-angle field intensity recordings consists of comparing records made by the recorder with records made by a standard field-intensity recording meter. Such a meter may consist essentially of a field-intensity recorder located at ground elevation over a ground mat extending several hundred feet radially from the recorder, and using either (1) a short vertical antenna over ground mat, or (2) a loop antenna. This installation is used as a secondary standard, by calibrating directly the vertical antenna and loop, using a target transmitter at ground level and a loop-type field-intensity measuring set.

3.5. Measurement of Ionospheric Absorption

Two methods are in use for measuring the absorption or reflection coefficient of the ionosphere at vertical incidence, (1) the pulse amplitude method, and (2) the CW recording method.

The basic equipment used in the pulse method is similar to that used for manual ionosphere height measurement, except that a calibrated

attenuator is provided in the receiver, and a reference amplitude mark on the screen of the oscilloscope. The principle of operation is to measure the relative amplitude of each arriving sky-wave echo in terms of the attenuation, α, necessary to bring its amplitude up to the reference mark on the scope. Suppose, at a given frequency, the field intensity of the first sky-wave echo, in the absence of ionospheric absorption, is enough to produce a deflection K/h on the scope, where h is the height of reflection and K is the over-all equipment constant. Let A be the value of attenuation (taken as a fraction less than unity) necessary to bring this unabsorbed echo to the reference mark on the scope. If, at the time when absorption measurements are desired, the reflection coefficient of the ionosphere is ρ, and that of the ground ρ_0, the amplitude of the first reflection will be $K\rho/h$, that of the second $K\rho^2\rho_0/2h$, that of the nth, $K\rho^n\rho_0^{n-1}/nh$. Let A_1, A_2, A_n, etc., be the attenuations necessary to bring the various echoes to the reference mark. Then, from each successive pair of echoes, we can determine ρ by the relation

$$\rho=\frac{A_{(n-1)}}{A_n}\frac{n}{(n-1)\rho_0}.$$

since

$$A_1\frac{K\rho}{h}=A_2\frac{K\rho^2\rho_0}{2h}=\cdots=\frac{A_nK\rho^n\rho_0^{n-1}}{nh}.$$

It is to be noted that the ground-reflection coefficient ρ_0 is generally nearly a constant with time at any frequency; it cannot be determined by using the attenuator readings, but will need to be determined by other means.
In particular for the first and second echoes:

$$\rho=\frac{2}{\rho_0}\frac{A_1}{A_2},$$

and the attenuation corresponding to the unabsorbed echo is

$$A=\rho A_1=\frac{2A_1^2}{\rho_0A_2}.$$

For the unabsorbed echo, usually taken as a night echo, it is possible to determine an attenuator setting A_0, which will bring the first unabsorbed echo to the same reference mark previously used. Then

$$\frac{A_0K}{h}=\frac{A_1K\rho, \text{ etc.}}{h}$$

Thus we can calibrate the equipment in terms of A_0 for each frequency; in making the measurement, then, we can deal, if we wish, only with the first reflection, and obtain, therefore, for each frequency:

$$\rho=\frac{A_0}{A_1}$$

Unfortunately it is not very easy to determine any of the values of A_0 or A_n, because all sky-wave pulse echoes fade severely, continuously and randomly. In order that measurements of ρ should have practical value, the A's should be recorded continuously for a period long enough to insure obtaining a median value, free from abnormalities or focusing effects. Thus far there is in use no practicable form of continuous A-recorder for pulses.

In practice, two systems of observing are used. In one, the observer adjusts the pulse amplitude a dozen or so times at perfectly regular intervals, and averages the results, for each mode of reflection, thereby obtaining an approximation to a random sampling of the fading pattern. In the second, the observer obtains median values of A_0 or A_n, by estimation, for a larger number of frequencies, plots them, and then fits a curve to the plot. The pulse technique is more involved than the CW technique described below, but by its use it is possible to separate out, identify, and analyze separate modes. Since the ordinary and extraordinary waves are oppositely polarized, it is possible to obtain one and exclude the other by the use of crossed dipoles in a properly phased circuit.

The CW recording method of measuring ionospheric absorption has the advantage of continuous recording, thereby obtaining reliable median values, but it has the disadvantage of not being able to separate the modes of reflection. This method involves the use of continuous automatic field-intensity recorders, operating from half-wave horizontal dipole antennas, a quarter-wave high, and broadside to the CW transmitters, which are located a number of miles away in order to minimize the ground wave, and which also operate on half-wave horizontal antennas a quarter-wave high. The maximum radiation and the greatest pick-up factors of these antennas are for vertical sky-wave propagation. The field intensity received at any given time is that of a combination of different modes and ordinary and extraordinary waves, all fading randomly with respect to each other. Approximately, the resultant median field intensity is the square root of the sum of the squares of the median values of all of its components. From a knowledge of the differential absorption of ordinary and extraordinary waves, and using the root-sum-square rule, the field intensity of the ordinary wave, first echo, can be deduced from the records.

It is possible to interpret the measured median field intensity in terms of the reflection coefficient ρ_1 for the ordinary wave. It is necessary to correct for the differences in h' for various frequen-

cies and times of day, using the regular $h'f$ sweeps for this purpose.

Figures 3.15 to 3.18 show sample vertical-incidence field-intensity recordings using emissions from the Beltsville, Md., radio transmitting station of the National Bureau of Standards, as received at the Bureau's Sterling, Va., radio propagation laboratory.

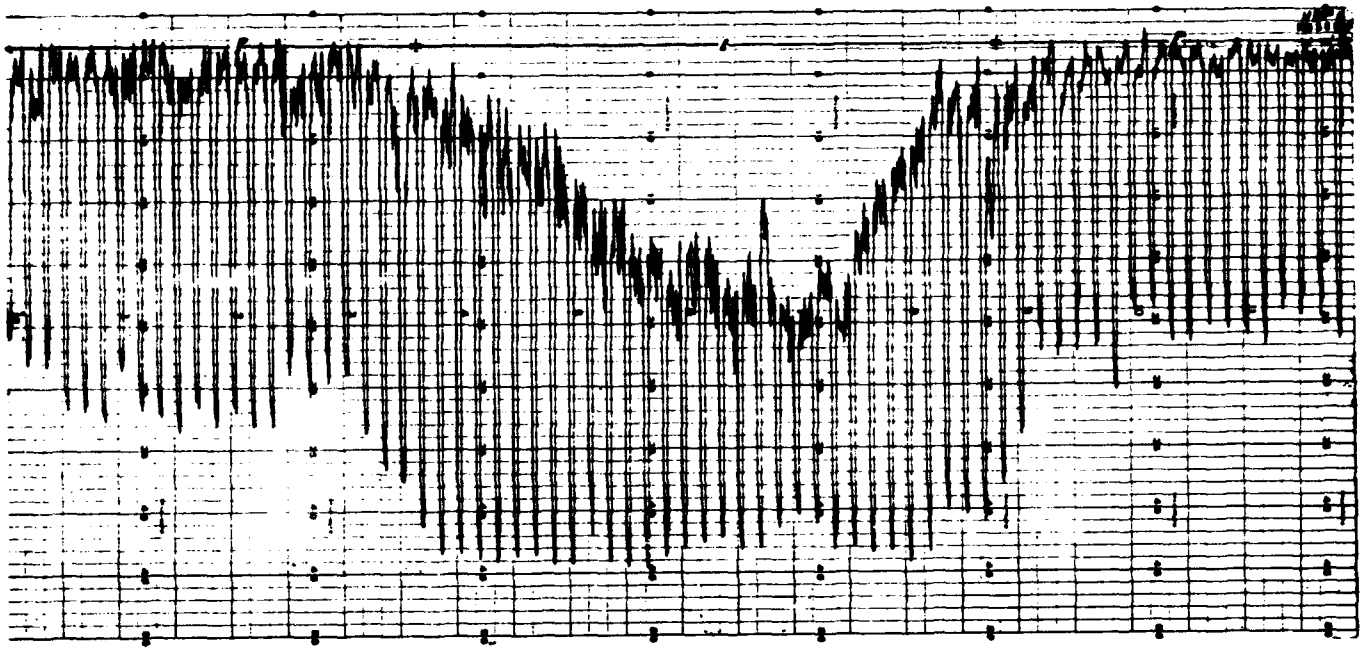


FIGURE 3.15. *Typical vertical-incidence field-intensity recording on 2061.2 kc.*

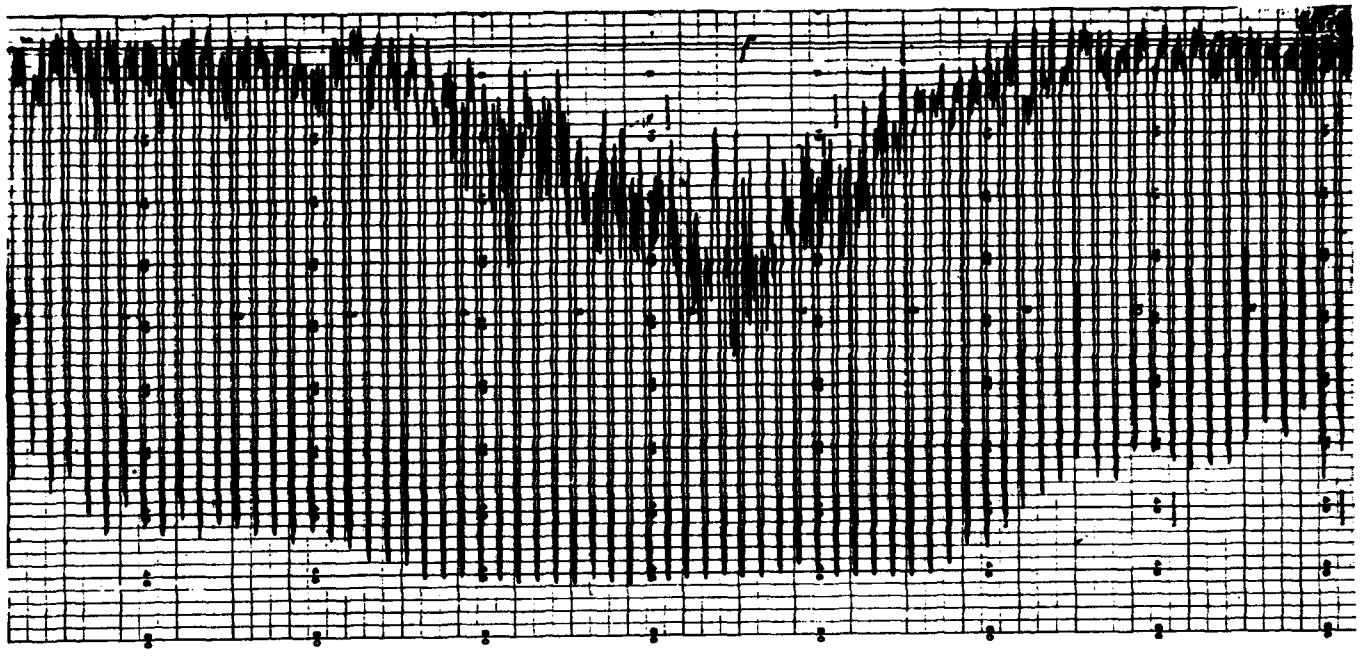


FIGURE 3.16. *Typical vertical-incidence field-intensity recording on 4272.4 kc.*

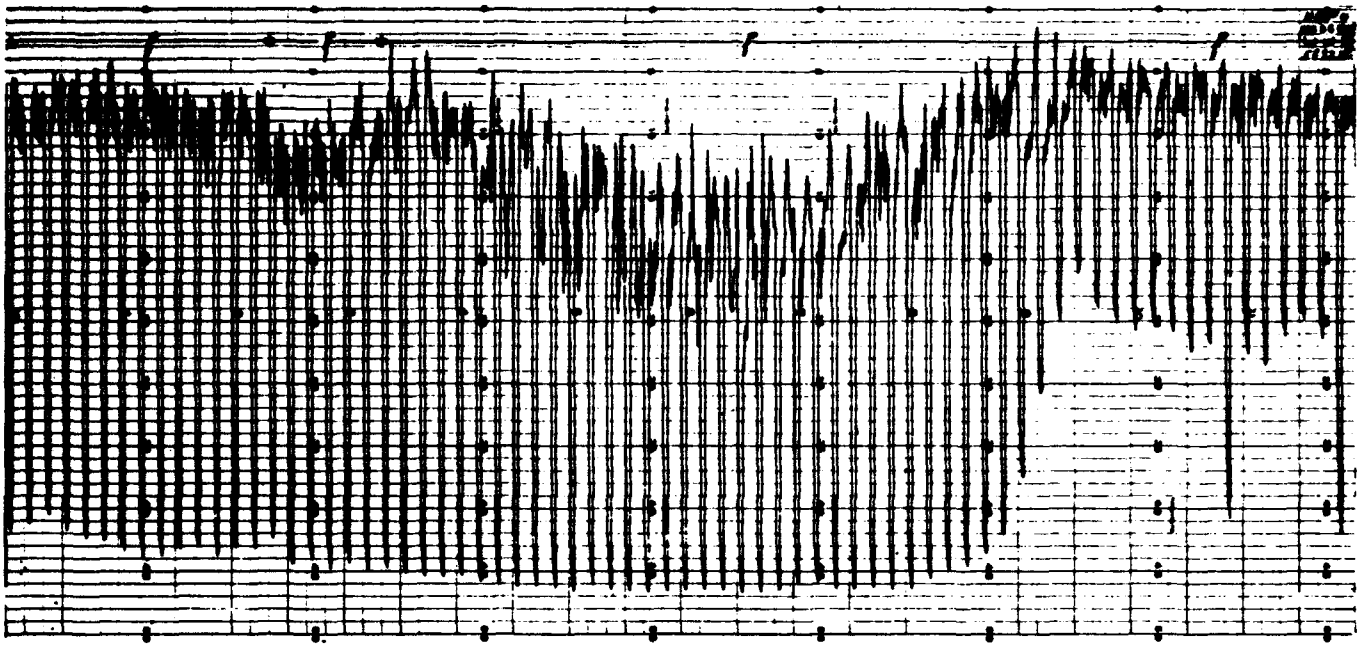


FIGURE 3.17. *Typical vertical-incidence field-intensity recording on 5892.0 kc..*

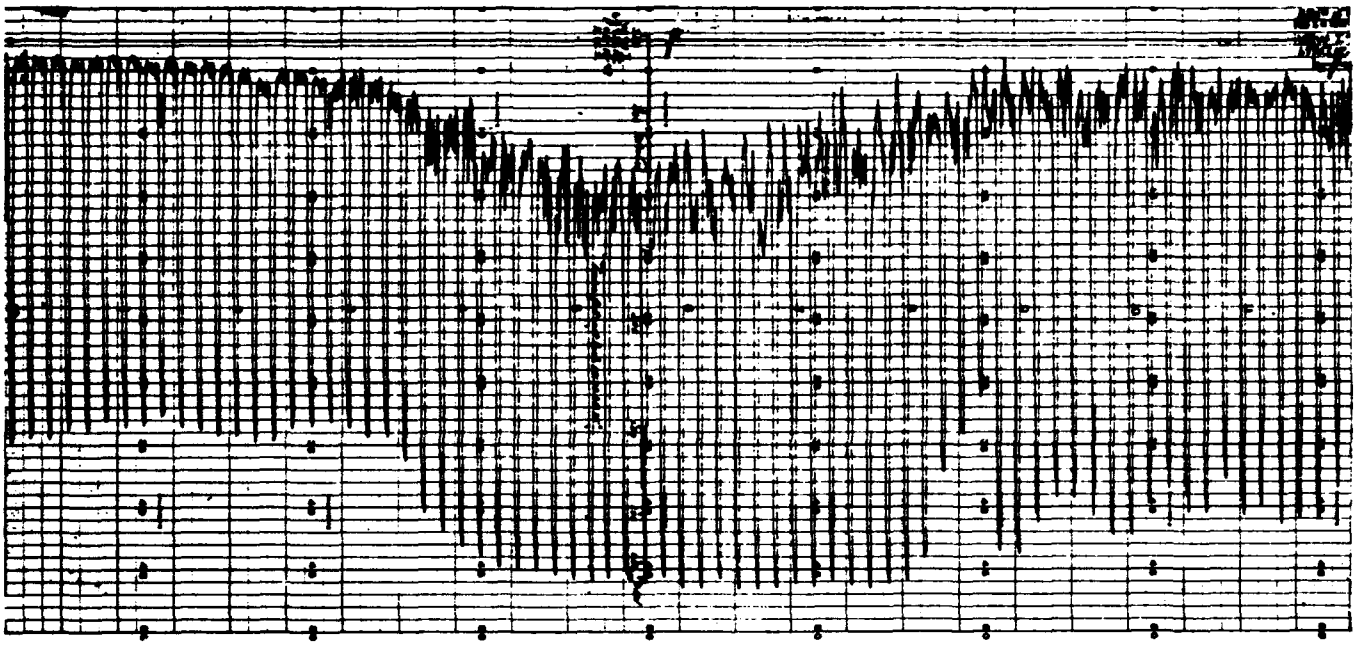


FIGURE 3.18. *Typical vertical-incidence field-intensity recording on 6995.6 kc.*

3.6. Measurement of Sky-Wave Angles of Arrival

In designing an antenna for the reception of radio sky waves, it is desirable to know the angle above the horizontal at which the principal amount of energy arrives. This angle is referred to as the vertical angle of arrival. In designing a transmitting antenna it is equally desirable to know the most favorable angle for the particular path and frequency. Although the vertical angle of arrival can be calculated approximately from a knowledge of ionospheric conditions along the path these conditions are frequently too complex for easy solution of the problem. It is therefore desirable to make direct measurements wherever possible

There are three principal methods of measuring the vertical angle of arrival: (1) the phase measurement method, (2) use of a "musa" or multiunit steerable antenna system, and (3) the use of two antennas having different radiation patterns in the vertical plane.

Because of its relative simplicity the phase measurement method is the one most often used. This method is based on the fact that when a radio wave from a given direction arrives at two receiving antennas separated by a known distance the phase angle ϕ between the induced voltages is

$$\phi=\frac{2\pi d}{\lambda}\cos\theta\cos\psi, \qquad (3.1)$$

where

d=spacing in meters between the two parallel antennas
λ=wavelengths in meters of the radio wave
θ=vertical angle, measured from the horizontal, of the downcoming wave
ψ=bearing angle, measured from a line joining the centers of the two antennas.

In practice the antennas are so oriented in the horizontal plane that $\psi=0$ for the condition of propagation along the great circle. Although propagation is normally by the great-circle path, with minor deviations, it is known that large deviations are sometimes encountered. In order, therefore, to eliminate this possibility of error from the results it is desirable to take bearings with a good direction finder at the same time the vertical angle is being measured. It is to be noted that the accuracy of the phase-measurement method is subject to certain limitations. If measurements are being made on a single downcoming wave front, the accuracy will depend largely on how well the phase difference between the two antennas can be determined. If multipath transmission is taking place, nothing better than a weighted mean figure for the angle of arrival can be expected. In such cases the principal purpose of the observation is however, achieved, i. e., the determination of the conditions under which antennas operate, since any antenna system of dimensions comparable to those used in the tests would work with such a resultant wave. From 3.1 it can be seen that the resolution for small values of θ is poor and that for an antenna spacing greater than half a wavelength the equation yields a multiple solution for θ. These ambiguities can be resolved by using several antenna spacings and noting the value for θ, which repeats in all cases. Having determined approximately the angle or range of angles involved for a particular path, an antenna spacing can be chosen that gives a maximum resolution.

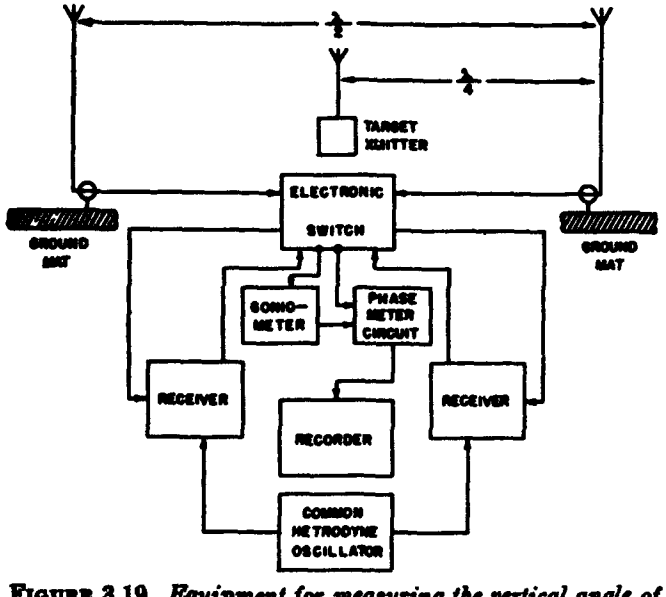


FIGURE 3.19. *Equipment for measuring the vertical angle of arrival by the phase-measurement method.*

Figure 3.19 is a simplified block diagram showing the arrangement of the principal components of an equipment for measuring the vertical angle of arrival by the phase measurement method.

The essential features of such a system are a pair of parallel antennas spaced a known distance along the great-circle path to the transmitter, a means of amplifying the radio-frequency voltages picked up by the antennas so that they can be observed on a cathode-ray oscilloscope or other suitable instrument. In practice, a pair of identical radio receivers, provided with a common heterodyne oscillator, is used. The difference frequency is then amplified and applied to a cathode-ray oscilloscope or a phase-meter circuit as indicated in figure 3.19. If the output voltages from the receivers are applied to the opposite plates of a cathode-ray oscilloscope the phase angle can be determined from an analysis of the elliptical pattern appearing on the screen. A recording phase meter calibrated in terms of vertical angle of arrival can be incorporated in the system to provide a continuous record of the vertical angle.

The musa system for measuring the vertical angle consists of an antenna array having a very sharp lobe which can be steered up and down in the vertical plane. The vertical angle is determined by the position of the lobe at which the maximum voltage is received. An advantage of this system is that, if the lobe is sufficiently narrow, the various angles at which appreciable amounts of energy are arriving can be determined. It can also be made fully automatic and provide a printed record. This system has the disadvantage of requiring an antenna array of formidable dimensions for operation on frequencies used in ionospheric propagation.

The third system mentioned makes use of two antennas having different radiation patterns in the vertical plane. The difference between the voltages induced in the antennas by the downcoming wave will then be a function of the vertical angle of arrival. The antennas may be either a pair of identical antennas spaced at different heights above the ground or one can be a horizontal antenna and the other a vertical antenna.

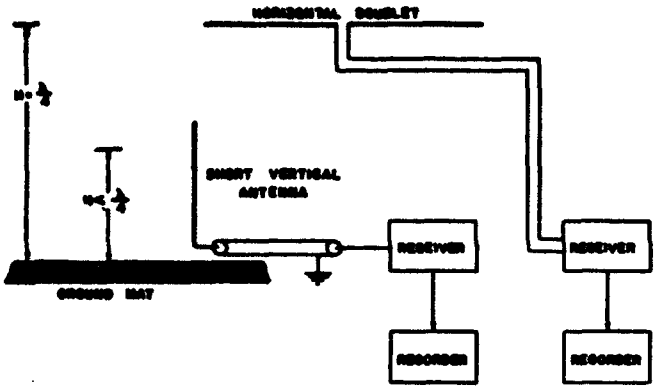


FIGURE 3.20. *Equipment for measuring the vertical angle of arrival by use of two antennas having different radiation patterns in the vertical plane.*

In either case the relative gain as a function of the vertical angle must be known. Like the phase measurement method, this system, under conditions of multipath transmission, yields only a weighted mean value for the angle of arrival. If the approximate range of vertical angles to be encountered can be determined, a combination of patterns can be chosen which provides a maximum resolution. Figure 3.20 shows the arrangement of equipment for making measurements according to this system.

3.7. Measurement of Ionospheric Disturbances

A method of detecting the occurrence and estimating the severity of ionosphere storms that has proven quite useful has been developed during the past few years by use of measurements on radio direction-finding and associated ionosphere equipments. This has been done by regularly observing the bearings and "bearing swings" of a selected group of radio transmitting stations, whose transmission paths pass through or tangent to the north auroral zone.

For this purpose a number of European transmitting stations operating at frequencies between 9 to 21 Mc are regularly monitored with radio direction-finding set, consisting primarily of a four-unit antenna array, a goniometer, a receiver, and a visual bearing indicator. The direction finder is operated in the usual manner; the important characteristic of the direction finder is that it gives an instanteneous bearing indication, so that rapid bearing shifts and swings can be noted.

Bearing deviations, fading characteristics, and received field intensities are measured and an hourly record is plotted for each of the monitored stations which shows changes of bearing, or a period of unusual bearing fluctuations. Fading characteristics are divided into one of two groups, IQW's and RQW's. The former is the normal irregular quick weak fades, usually signifying normal conditions. In the latter group (regular quick weak fades) the signals are subdivided into either light RQW's or heavy RQW's; a great number of RQW's sometimes signifying the irregular conditions preceding an ionospheric storm.

The received field intensities are noted and compared to values on normal days. The bearings are further classified into normal, no observed bearing (N. O. B.) or not able to receive (N.A.R.)

Figure 3.21 shows values plotted for a typical station on a normal day (a) and a disturbed period (b).

Used in conjunction with the bearing data are observations of the fluctuation of the horizontal component of the earth's magnetic field. Readings are taken on a magnetograph giving relative values of field strength. The readings are taken over the most active three-hour period in the preceding 24-hr period and classified into a disturbance factor similar to the geomagnetic *K* activity figures.

By the use of the above techniques, it is often possible to recognize ionosphere storms in their initial phases and to issue warnings of impendi disturbance on various transmission paths.

3.8 The Measurement of Radio Noise

The present state of development of vacuum tubes and amplifier circuit technics is such that, subject to some limitations in bandwidth, the gain of an amplifier can be raised to a very high value. The practical limits to the gain are set by the intensity of electrical interference that may reach the input of an amplifier from the same source from which the amplifier receives its signal input power, and by the intensity of radio noise

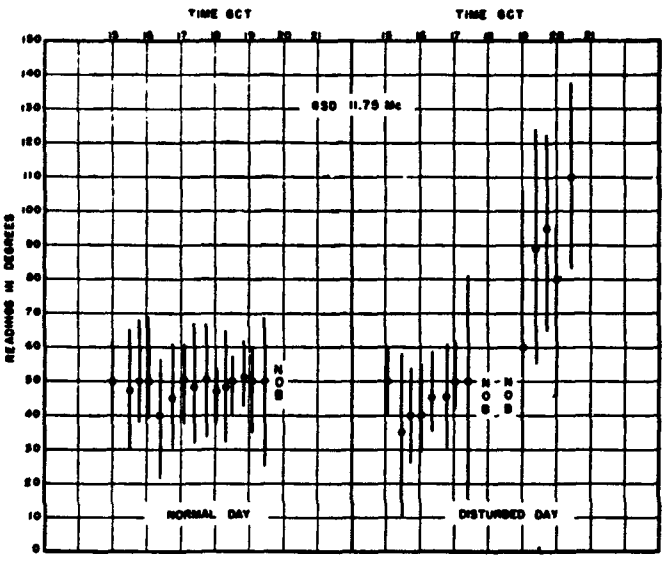


FIGURE 3.21. *Comparison of bearings on normal and disturbed days.*

935333 O - 51 - 3

which is generated within the amplifier. The external radio interference lends itself readily to a classification into man-made interference and natural radio noise. It is not within the scope of this section to discuss the measurement of man-made interference. The natural radio noise, in turn, may be classified into several categories according to origin. In the frequency spectrum below approximately 25 Mc the radio noise is predominantly terrestrial in origin and is known as atmospheric. Above that frequency the atmospheric noise is low in intensity and it is possible to detect and measure radio noise coming, apparently, from interstellar regions. This noise is known as cosmic radio noise. At still higher frequencies, of the order of several hundred megacycles, it is possible, by the use of sensitive and highly directive radio receivers to detect and measure the intensities of radio noise generated by the sun. This noise is known as solar radio noise. The internal noise of an amplifier may, in turn, be classified into circuit and tube noise. The circuit noise is caused by the molecular agitation of electrons in a conductor and is known as thermal or "Johnson" noise after its discoverer. The tube noise is produced in several ways. The most important of these is that due to random fluctuation in emission of electrons by the cathode. This noise is known as "shot effect" ("SCHROTEFFEKT", or "Schottky Effect," so named after its discoverer).

a. Atmospheric Radio Noise

A vast amount of effort has been expended on attempts to measure atmospheric radio noise; the work has been done by many observers using different methods, however, and the results have been difficult to correlate. There are many problems involved in the measurement of noise intensities. Selection of a suitable antenna is one phase of the problem. Vertical-rod antennas possess the advantage of uniform gain in all directions in azimuth, as well as structural advantages, but they do not have the same gain for different vertical angles of arrival. This is particularly true if the antennas are longer than a half wavelength. They also have the limitation of being sensitive only to the vertically polarized waves. Loop antennas possess some mechanical and electrical advantages, but at the same time have a disadvantage in that they have distinct directive patterns.

The task of actual measurement of noise is another phase of the problem. Noise is, by its nature, a random phenomenon possessing no given frequency; it lends itself, however, to a Fourier-integral analysis by which it can be resolved into a disturbance having a general spectral distribution of energy, with varying intensities at various frequencies. The integration of this intensity

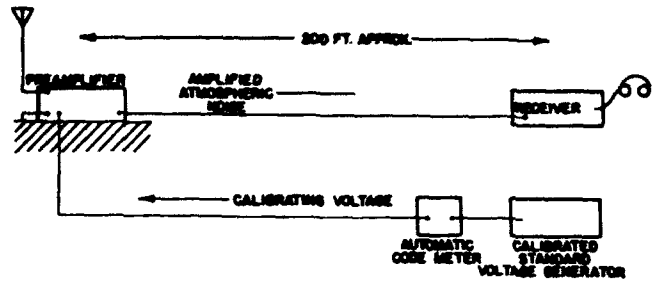


FIGURE 3.22. *Atmospheric-noise measuring equipment.*

over a given frequency band, say the bandwidth of a receiver, gives the total power content within that frequency band. These considerations apply strictly only to the fluctuation type of noise; they apply only approximately to impulse-type noise. It is then necessary to decide, in measuring the noise output of a receiver, whether to measure the average voltage, the peak voltage, or the rms voltage. The conventional sinusoidal relationship connecting these quantities does not hold for the noise. Numerous tests have been made to determine which of these parameters might present the best correlation with the interference effect which the noise has on radio communication, but there is no general agreement at present. Either the peak or the average noise voltages can be measured by the use of a diode rectifier by proper charging and discharging time constants.

There is a system, now in wide use geographically, for measuring required field intensities for a particular service in the presence of atmospheric noise. A block diagram of this system in improved form is shown in figure 3.22. It consists of an untuned 25-ft rod antenna connected to a receiver. A standard voltage generator, keyed to give a Morse code signal, is also connected to the input of the receiver. The output of the voltage generator is attenuated until the signal is barely intelligible. The intensity of the field which would produce this signal is then noted. This is a strictly subjective test, in that the equipment depends on the judgment of the operator, for proper evaluation of noise. This equipment has been in use for approximately three years at a number of stations scattered over the world and has been used to collect a substantial amount of data.

b. Cosmic and Solar Radio Noise

The methods of measurement of cosmic and solar radio noise are substantially the same as those of atmospheric radio noise, except in the selection of an antenna. A vertical rod antenna, because of its "blind spot" in the zenith, does not lend itself very well to the measurement of cosmic radio noise that comes from a general overhead direction; a horizontal dipole, on the other hand, is good, particularly when it is spaced one-quarter wave-length above ground.

Most of the measurements of cosmic radio noise published thus far have employed directive antennas. Only a moderate degree of directivity can be obtained with any reasonable size antennas at these frequencies (a few degrees' beamwidth at a few hundred megacycles). A more precise study of this phenomenon is necessarily dependent on the development of more directive systems.

One of the problems that arises in connection with measurement of cosmic noise is the calibration of equipment for sensitivity and for noise inherent in the receiving system. One possible way of doing it is by substitution of a resistor equal in its resistance value to the radiation resistance of the dipole. As will be discussed in greater detail later on, a resistor is a source of noise power proportional to its resistance value and to its temperature. By substituting a resistor at ambient temperature one calibrating point will readily be available. If means are provided for varying the temperature of the resistor, without changing its resistance value, a number of noise-calibrating voltages can be obtained. Another method which is being used in practice is to replace, for calibration purposes, the dipole by a resistor of equivalent value of resistance and to send known noise current through it. This noise current is readily obtainable from a temperature-limited diode and the method will be explained in greater detail later on in this section. By the use of this method a number of calibration points may be secured for the equipment. This method is applicable to both single dipole antennas and to parabolic reflector antennas which are used for detailed exploration of the sky.

c. Solar Radio Noise

The methods of measuring solar radio noise are identical in principle to those for measuring cosmic noise. Although parabolic reflectors offer the most promising means for obtaining the directivities required for solar measurement, a number of observers have used successfully dipole arrays similar to those used on low frequency radar equipment and even yagi antenna systems. The problem of calibrating the equipment is the same as that for calibrating cosmic noise equipment.

d. Thermal Noise

In the 1920's it was discovered that an electrical resistance maintained at a temperature above absolute zero has generated within it an emf which has equal components at all frequencies. This emf is proportional to the square root of the absolute temperature of the resistance, and to the square root of its resistance value. Theoretical considerations in fact predict such an emf on the basis of molecular motion within the resistor. The expression is known as the Nyquist formula and is as follows:

$$e_T^2 = KTR\Delta F,$$

e_T = the noise voltage in volts
K = Boltzmann constant 1.38×10^{-23} joules per degree kelvin
T = absolute temperature in degrees Kelvin
R = resistance in ohms
ΔF = width frequency band over which the measurements are made, in cycles per second.

Thus a 100-ohm resistor at room temperature (300° K) will have an open-circuit voltage, due to thermal noise, of approximately 0.1 μv in a circuit having a 6-kc bandwidth.

Thermal noise is extremely important, because it represents a theoretical limit to the sensitivity of receivers. Although a high signal-to-noise ratio may be attained over almost any path by the use of directional antennas, there is no known method of eliminating or reducing the interference of the thermal noise below the theoretical minimum.

e. Shot Noise and Other Sources of Noise

In addition to the thermal noise which is generated in the electrical circuits of a receiver there are several other sources of noise which are present in an amplifier system. For example, noise may enter from outside because of faulty shielding or filtering, or it may be generated by poor contacts within the receiver or by defective batteries. These sources are here ignored. However, the vacuum tubes themselves are at all times a source of noise. Tubes with oxide-coated cathodes have a phenomenon known as the "flicker" effect which results from shifting active spots on the coated cathode. The noise components of this effect are principally in the audio-frequency band. Vacuum tubes may also contribute to the noise by ionization of residual gases in the tube. This noise is a function of imperfections in vacuum exhausting systems and, except for qualitative knowledge that it can be reduced by improving exhausting technics, does not lend itself to precise analysis.

There are two sources of radio noise in a vacuum tube which are more important than the others. One is known as the "shot effect" and the other as "partition noise." Shot noise is radio noise generated in the plate circuit of a vacuum tube because of the corpuscular nature of electricity. The intensity of shot noise, for a temperature-saturated diode, is given by

$$i_N^2 = 2eI\Delta F,$$

where

i_N = noise current in amperes
e = charge of electron in coulombs $(=1.59 \times 10^{-19})$
I = diode space current in amperes
ΔF = the width of the band of frequency over which the noise is measured in cycles per second.

Although this formula is not especially significant for the purpose of computing the internal noise of an amplifier circuit, a temperature-limited diode is extremely useful as a generator of radio noise with an accurately known, uniform frequency spectrum (limited by the band-width of associated circuits), and its performance is given by the formula. Random or fluctuation radio noise, generated either in the receiving system or outside of it, can be readily measured by a substitution method, using such a diode as a standard noise generator.

Shot effect is also present in an amplifier triode or pentode, but these operate under space-charge conditions. The effect of the space charge is to reduce greatly the noise due to shot effect. The mechanism by which it effects this reduction is fairly well understood, but does not lend itself to a simple mathematical analysis. The space charge, in effect, acts as a cushion, reducing random fluctuations in the emission of electrons by turning back to the cathode a greater number of electrons when emission is at peak, and vice versa. Comparative magnitudes of the noise currents of a temperature-limited diode and of a good space-saturated triode may be of the order of 0.0044 μa for the former and 0.0014 μa for the latter, both values being computed for a bandwidth of 6,000 cycles.

Partition noise is a property peculiar to multigrid tubes, where the cathode current is the sum of plate and screen currents and where the output is taken only from the plate.

There is no simple direct way to compute the noise due to this source; it is nevertheless quite important, exceeding, in a space-charged tube, the shot effect in magnitude. It effectively bars the use of a pentode tube in the low-level stages of sensitive amplifiers, in which a low noise figure is the most important criterion.

f. Noise Figures

For a number of years the science of radio engineering has been faced with a problem of devising a system of rating a receiver or an amplifier on its merits from the standpoint of low noise. The problem has been complicated by the fact that in addition to the useful voltage output of a generator (the generator, under operating conditions, would be an antenna and the useful voltage, the desired signal voltage) there is always present in the output of a generator a certain noise voltage. In case of an antenna this voltage would include thermal resistor noise voltage, and atmospheric and cosmic noise voltages; in case of a standard voltage generator this voltage would include only the thermal resistor noise voltage. Because the atmospheric and cosmic noise voltages are subject to fluctuation with time, location and construction and orientation of antenna, they do not offer a satisfactory standard for rating a receiver or an amplifier; however, thermal noise, presenting a readily computed voltage offers a satisfactory standard against which the noise introduced by a receiver or an amplifier can be rated. Based on this principle a system of rating a receiver in terms of "noise figure" has been devised for this purpose. According to this method the noise figure of a receiver is a quotient of the ratio of available useful (signal) power output to the available thermal noise power output of the voltage generator divided by the ratio of available useful (signal) power output of receiver to the total available noise output of the receiver. Because the receiver will contribute noise of its own the second ratio will have a smaller value and the noise figure of a real receiver will always be greater than unity. It can be shown that the ratio of useful voltage to the thermal noise voltage output depends on the load as well as the useful (signal) voltage emf and internal resistance of the voltage generator. For this reason a concept of available power is introduced. The available useful (signal) power of the generator is defined as the maximum which it can deliver. This maximum would be obtained under matched load conditions, when the load resistance would be equal to the internal resistance of the generator. If R_G is the internal resistance of the generator and e_G is its useful (signal) emf, the available useful power is equal to $e_G^2/4R_G$. In a somewhat similar manner the available thermal noise power of the generator is defined as $e_T^2/4R_G$, where e_T is the thermal noise voltage of the generator. As $e_T^2=4KTR_G\ \Delta F$, then the available thermal noise output power of a generator is equal to $KT\Delta F$.

If the bandwidth of the receiver is known it is possible to measure its noise figure by the use of a standard single-frequency voltage generator. However, because the bandwidth of the receiver is not always known very accurately and also because the bandwidth between the half-power gain points is not exactly equivalent to the bandwidth for noise comparison purposes, it has been found very useful to employ the shot effect of a temperature-limited diode for measuring the noise figures of receivers and amplifiers. The noise generator employing such a diode has uniform frequency distribution of noise power just as the thermal noise power of a resistor. Therefore no correction for bandwidth is necessary.

Figure 3.23 illustrates the conventional circuit connections for the use of diode noise generator in measuring the noise figure of a receiver. Resistor R_G represents and must equal the internal resistance of the voltage generator (antenna, in case of a receiver) from which the receiver or amplifier normally receives its input. The procedure for measuring the noise figure is as follows: with noise diode current equal to zero, the output of the receiver E_0, which will be amplified thermal

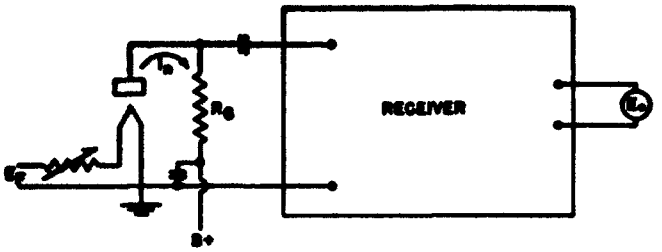


FIGURE 3.23. *Circuit connections for measuring noise figure of receiver.*

noise input to the receiver plus internal circuit and tube noise is measured and recorded. The filament power of the noise diode is then adjusted until the diode space current I is such that the output power of the receiver is doubled. The additional power output is the amplified noise diode output which, in this test, is the useful power output. Under these conditions the useful power output is equal to the noise power output and their ratio is equal to unity.

The available useful power of the diode noise generator is equal to $e_N^2/4R_G$, where e_N is the open-circuit shot-noise voltage developed across R_G owing to shot noise current generated in the diode, i_N.

$$e_N^2 = i_N^2 R_G^2$$
$$= 2eIR_G^2\Delta F.$$

Available useful power input (shot noise) is

$$\frac{2eIR_G^2\Delta F}{4R_G} = 0.5eIR_G\Delta F.$$

The ratio of available useful (shot-noise) power input to available thermal noise power input is

$$\frac{0.5eIR_G\Delta F}{KT\Delta F} = \frac{0.5eIR_G}{KT}.$$

As the noise figure, $\overline{NF}$, is the quotient of the above quantity divided by the ratio of available useful (shot-noise) power output of the receiver to the available noise power output of the receiver, and, as the latter ratio was made unity during measurement, the noise figure is equal to

$$\overline{NF} = \frac{0.5eIR_G}{KT}.$$

Substituting the values previously given for e and K and 288° K for T,

$$\overline{NF} = 20IR_G,$$

where I is in amperes and R_G is in ohms. In a special, but not uncommon case where R_G is equal to 50 ohms, the noise figure is given directly by the current I in milliamperes.

3.9 References

L. W. Austin, Some quantitative experiments in long distance radiotelegraphy, Bul. BS **7**, 315 (1911) S159.

W. H. Eccles, On measurement of signal strength, Proc. Inst. Radio Engrs. **7**, 267 (1919).

R. Bown, C. R. Englund, and H. T. Friis, Radio transmission measurements, Proc. Inst. Radio Engrs. **11**, 115 (1923).

J. Hollingsworth, A resume of modern methods of signal measurement, Wireless World & Radio Rev. **14**, 485 (1924).

H. H. Beverage and H. O. Peterson, Radio transmission measurements on long wave lengths (RCA), Proc. Inst. Radio Engrs. **11**, 661 (1923).

L. W. Austin, A method of measuring radio field intensities and atmospheric disturbances, Proc. Inst. Radio Engrs. **12**, 521 (1924).

H. J. Round, T. L. Eckersley, K. Tremellen, and F. C. Lunnon, A report on measurements made on signal strength at great distances during 1922 and 1923 by an expedition sent to Australia, J. Inst. Elec. Eng. (London) **63**, 933 (1925).

J. A. Ratcliffe, and M. A. F. Barnett, On the attenuation of wireless signals in short distance overland transmission, Proc. Phil. Soc. (Cambridge) **23**, 288 (1926).

H. T. Friis and E. Bruce, A radio field-strength measuring system for frequencies up to 40 Mc, Proc. Inst. Radio Engrs. **14**, 507 (1926).

B. Saltmarsh, A radio signal-intensity recorder, Exp. Wireless (London) **4**, 743 (1927).

E B. Judson, An automatic recorder for measuring strength of radio signals and atmospheric disturbances, Proc. Inst. Radio Engrs. **16**, 666 (1928).

J. Hollingsworth, A new universal long wave radio intensity measuring set, J. Sci. Instr. (London) **5**, 1 (January 1928).

T. L. Eckersley, Field-strength measurements of short wave transmissions, Marconi Rev. **30**, 1 (1931).

P. A. de Mars, G. W. Kenrick, and G. W. Pickard, Use of automatic recording equipment in radio transmission research, Proc. Inst. Radio Engrs. **19**, 16 (1931).

L. V. Berkner, Some studies of radio transmission over long paths made on the Byrd Antarctic Expedition, BS J. Research **8**, 265 (1932) RP412.

W. W. Mutch, A note on an automatic field strength and static recorder, Proc. Inst. Radio Engrs. **20**, 1914 (1932).

A. L. Green and H. B. Wood, A field intensity set, J. Inst. Engrs. (Australia) p. 1–8; (1933).

G. C. De Contouly, Portable long-wave testing apparatus, Bell labs. Record **11**, 178 (1933).

K. A. Norton and S. E. Reymer, A continuous recorder of radio field intensities, BS J. Research **11**, 373 (1933) RP597.

E. B. Judson, LF radio receiving measurements at Bureau of Standards in 1931 and 1932, Proc. Inst. Radio Engrs. **21**, 1354 (1933).

C. M. Jansky, Jr., Automatic recording of field strength, Electronics **7**, 148 (1934).

G. W. Kenrick and G. W. Pickard, Phase interference phenomena in low frequency radio transmission, Proc. Inst. Radio Engrs. **22**, 344 (1934).

H. H. Beverage and G. W. Kenrick, LF transmission over transatlantic paths, Proc. Inst. Radio Engrs. **24**, 472 (1936).

J. J. Vormer, Arrangement for simultaneously registrating the field intensities of three transmitters, Wireless Engr. **14**, 113 (1937).

W. W. Smith, A highly sensitive field-strength meter, Radio No. **22**, No. 247, 37 (1940).

Standards on radio wave propagation measuring methods, Supplement to Proc. Inst. Radio Engrs. **30**, No. 7, part III; (1942).

H. W. Kline, New field intensity recorder, Electronics **14**, 50 (1942).

W. H. Eccles, On measurement of signal strength, Proc. Inst. Radio Engrs. 7, 267 (1919).
R. Bown, C. R. Englund, and H. T. Friis, Radio transmission measurements, Proc. Inst. Radio Engrs. 11, 115 (1923).
H. H. Beverage and H. O. Petersen, Radio transmission measurements on long wave lengths (RCA), Proc. Inst. Radio Engrs. 11, 661 (1923).
T. R. Gilliland, Note on a multifrequency automatic recorder of ionosphere heights, BS J. Research 11, 561 (1933) RP608; Proc. Inst. Radio Engrs. 22, 236 (1934).
T. R. Gilliland, Continuous measurements of the virtual heights of the ionosphere, BS J. Research 11, 141 (1933). RP582; Proc. Inst. Radio Engrs. 21, 1463 (1933).
P. A. De Mars, T. R. Gilliland, and G. W. Kenrick, Kennelly-Heaviside layer studies, Proc. Inst. Radio Engrs. 19, 106 (1931).
M. A. Tuve and O. Dahl, A transmitter modulating device for the study of the Kennelly-Heaviside layer by the echo method, Proc. Inst. Radio Engrs. 16, 794 (1928).
G. Breit and M. A. Tuve, A test of the existence of the conducting layer, Phys. Rev. 28, 554 (1926).
L. R. Hafsted and M. A. Tuve, Further studies of the Kennelly-Heaviside layer by the echo method, Proc. Inst. Radio Engrs. 17, 1513 (1929).
E. L. C. White and J. A. Ratcliffe, An automatic recording method for wireless investigations of the ionosphere, Proc. Phys. Soc. 45 (1933).
O. O. Pulley, A self-synchronized system for ionospheric investigation by the pulse method, Proc. Phys. Soc. 46 (1934).
R. K. Potter, High frequency atmospheric noise, Proc. Inst. Radio Engrs. 19, 1731 (1931).
R. K. Potter, An estimate of the frequency distribution of atmospheric noise, Proc. Inst. Radio Engrs. 20, 1512 (1932).
E. V. Appleton, and F. W. Chapman, On the nature of atmospherics, Proc. Roy. Soc. 158 (1937).
L. W. Austin, Long distance radio receiving measurements and atmospheric disturbance at the Bureau of Standards in 1925, Proc. Inst. Radio Engrs. 14, 663 (1926).
R. Bureau, Diurnal variations of atmospherics at Paris 1928–31. Influence of source and propagation, Compt. rend. 194, 1368 (1932).
R. Bureau, Goniometric researches on atmospherics, Compt. rend. 194, 2073 (1932).
E. V. Appleton and J. A. Ratcliffe, Some simultaneous observations on downcoming wireless waves, Proc. Roy. Soc. (London) 128, 133 (July 1, 1930).
G. Breit, M. A. Tuve, and O. Dahl, Effective height of the Kennelly-Heaviside layer in Dec. 1927 and Jan. 1928, Proc. Inst. Radio Engrs. 16, 1236 (1928).
E. V. Appleton, On some measurements of the equivalent height of the atmospheric ionized layer, Proc. Roy. Soc. (London) [A] 126, 542 (March 3, 1930).
E. V. Appleton and A. L. Green, On some short wave equivalent-height measurements of the ionized regions of the upper atmosphere, Proc. Roy. Soc. (London) [A] 128, 159 (July 1, 1930).
T. Parkinson and T. R. Gilliland, A radio method for synchronizing recording apparatus, BS J. Research 6, 195 (1931) RP269; Proc. Inst. Radio Engrs. 19, 335 (1931).
T. R. Gilliland and G. W. Kenrick, Preliminary note on an automatic recorder giving a continuous height record of the Kennelly-Heaviside layer, BS J. Research 7, 783 (1931) RP373; Proc. Inst. Radio Engrs. 20, 540 (1932).
S. S. Kirby, L. V. Berkner, T. R. Gilliland, and K. A. Norton, Radio observations of the Bureau of Standards during the solar eclipse of Aug. 31, 1932, BS J. Research 11, 829 (1933) RP629; Proc. Inst. Radio Engrs. 22, 247 (1934).
R. Bureau, Direction of the summer sources of atmospherics, Compt. rend. 198, 1057 (1934).
R. Bureau and M. Douguetr, Atmospherics in the southern regions, Compt. rend. 209, 117 (1939).
J. C. Jaeger, Atmospherics and noise level, CSIR (Australia) (1943).
K. G. Jansky, Directional studies of atmospherics at high frequencies, Proc. Inst. Radio Engrs. 20, 1920 (1932).
K. G. Jansky, Electrical disturbances apparently of extra-terrestrial origin, Proc. Inst. Radio Engrs. 21, 1387 (1933).
K. G. Jansky, A note on the source of inter-stellar interference, Proc. Inst. Radio Engrs. 23, 1158 (1935).
K. G. Jansky, Minimum noise levels obtained on short wave radio receiving systems, Proc. Inst. Radio Engrs. 25, 1517 (1937).
K. G. Jansky, An experimental investigation of the characteristics of certain types of noise, Proc. Inst. Radio Engrs. 27, 763 (1939).
G. Reber, Cosmic static, Astrophys. J. 91, 621 (1940).
G. Reber, Cosmic static, Proc. Inst. Radio Engrs. 30, 367 (1942).
G. Reber, Cosmic static, Proc. Inst. Radio Engrs. 28, 68 (1940).
G. Reber, Cosmic static, Astrophys. J. 100, 279 (1944).
D. F. Martyn, Temperature radiation from the quiet sun in the radio spectrum, Nature (London), 158, 632 (1946).
G. C. Southworth, Microwave radiation from the sun, J. Franklin Inst. 239, 285 (1945).
J. B. Johnson, Thermal agitation of electricity in conductors, Phys. Rev. 32, 97 (1928).
H. Nyquist, The thermal agitation of electric charge in conductors, Phys. Rev. 32, 110 (1928).
C. V. Aggers, D. E. Foster, and C. S. Young, Instruments and methods of measuring radio noise. Elec. Eng. 59, 178 (1940).
H. T. Friis, Noise figures of radio receivers, Proc. Inst. Radio Engrs. 32, 419 (1944).
K. A. Norton and E. W. Allen, FM and television signal propagation, FM & Television 6, 44 (1946).

CHAPTER 4

STRUCTURE OF THE IONOSPHERE

1. Description of the Atmosphere

a. General Nature of the Earth's Atmosphere

The general properties of the atmosphere at the earth's surface are well known. The atmosphere is composed of molecules of nitrogen, about 80 percent, and oxygen, about 20 percent with a minor admixture of other gases including helium, there being, at the earth's surface, about 2.6×10^{25} molecules per cubic meter in all; these molecules are in a state of thermal agitation traveling with random velocities of about 400 m/sec and colliding with each other at the rate of about 4×10^{9} collisions per molecule per second. As we go to greater heights, the pressure decreases and with it the molecular density and collisional frequency.

The relationship between density and height is given by

$$\rho=\rho_0\epsilon^{-\frac{h}{H}}, \qquad (4.1)$$

ρ_0 being the density at ground level, ϵ, the base of the system of natural logarithms, h, the height, and H, the "scale-height", equal approximately to 7.7 km close to the earth's surface. Thus for an increase in height of 7.7 km, the density decreases about 63 percent to 1/2.718 of what it was at the lower level. The scale height is given by $H=kT/mg$ (k is the gas constant; T the absolute temperature, m the molecular mass of the gas; and g the gravitational constant). The greater the value for scale height the more slowly the density changes with height. Factors of importance in increasing the scale height are increases in temperature and decreases in the molecular weight. Thus if the temperature were 273° C (=546° absolute) instead of 0° C (=273° absolute), the scale height would be increased to 15.4 km. Similarly, if all the diatomic molecules of oxygen and nitrogen were changed to monatomic molecules, the scale height would be increased to 15.4 km. With these basic facts in mind the observational facts of the atmosphere may be examined.

b. Variation of Temperature and Density With Height

As we proceed upward from the earth's surface the temperature and density of the atmosphere decrease. The decrease in temperature is approximately such that if a mass of air is moved upward it cools by expansion to the temperature which prevails at the height to which it is raised (adiabatic expansion). Owing to this condition, localized heating of air masses will cause them to move upward so that complete mixing of the air is possible. This condition extends to a height of about 10 km. The region through which it prevails is known as the troposphere. Its upper limit is called the tropopause.

Above the tropopause the temperature is about −55° C (218° absolute) and remains essentially the same at all heights up to about 30 km above the earth's surface. As a mass of air cools when expanded, any upward motion of air masses in the stratosphere is impossible. Under these conditions the various gases of the atmosphere would be in diffusion equilibrium, that is, each would be distributed according to its own partial pressure and scale height. The formula for the atmospheric composition would then be

$$\rho_a=\rho_{0,1}\epsilon^{-h/H_1}+\rho_{0,2}\epsilon^{-h/H_2}+\rho_{0,3}\epsilon^{-h/H_3}\ \ldots, \qquad (4.2)$$

where the additional subscripts a, 1, 2, 3 . . . refer to the molecules of all gases, and the molecules of the individual gases present, respectively. As a molecule of hydrogen weighs one-sixteenth as much as a molecule of oxygen, the scale height for hydrogen will be 16 times as great as for oxygen, or about 100 km. This means that at a height of 100 km the number of molecules of hydrogen per milliliter will be 1/2.73 of what it was at ground level, and the number of oxygen molecules would be 1/10,000,000 of the ground-level value. Thus gases which are negligible at the ground might be the dominant constituents of the atmosphere at great heights.

This was the conception of the upper atmosphere which persisted until recent years. Observations of meteors, refraction of sound waves, sounding balloon measurements, and of late, measurements made by rocket-borne instruments indicate that the temperature above 30 km increases with height, reaching a temperature around 80° C (353° absolute) at about 60 km, after which it decreases again to stratospheric or substratospheric temperatures at slightly above 80 km. This condition—a decrease of temperature with altitude—is favorable to the presence of ascending air currents, so that vigorous mixing of the constituent gases is likely. It is probable that at this height the composition of the atmosphere is nearly the same as at sea level. Above 80 km the temperature is thought to rise again, but any statements about temperature and composition in this region and above are largely conjectural. This

region above the stratosphere is called the ionosphere. From such information as is available regarding temperature, density, and composition of the upper atmosphere, the velocity, mean free path, collisional frequency, and other parameters may be calculated. As all of the variables are not definitely established, the calculated values are different and dependent on the assumptions which are made. Various hypothetical distributions of the several quantities are shown in table 4.1, together with the assumptions on which they are based. As new evidence becomes available from future observation, an estimate can be made as to which assumption most closely approximates the physical facts. The peculiar structure of the upper atmosphere is largely the result of the various physical and chemical actions of ultraviolet light from the sun. The most important of these is the ability of ultraviolet radiation (1) to break down the oxygen molecule into two unattached atoms and (2) to separate an electron from an atom whether it is attached to another atom or not. The process of ion production—the separation of an electron from a molecule or atom—is the one which most closely concerns the propagation of radio waves, although the other facts are important too.

TABLE 4.1.—*Properties of the atmosphere for tentative standard temperatures* [1]

Altitude, h	Absolute temperature, T	Pressure, p	Specific weight, $W=g\rho$	Speed of sound, a	Mean free path of molecules, λ	ν	N (per m³)
(a) FOR BOTH DAY AND NIGHT							
m	°K	kg/m²	kg/m³	m/sec	m		
0	288.0	10332.3		340.2	0.00007×10^{-3}	5832×10^{4}	2710×10^{23}
20,000	218.0	568.4	8851×10^{-4}	296.0	.00109	325.9	149.1
30,000	218.0	120.1	1864	296.0	.00476	74.62	31.49
40,000	276.7	29.35	358.0	333.5	.0246	16.27	7.696
50,000	350.0	9.970	95.80	375.1	.0914	4.925	2.615
60,000	350.0	3.820	36.59	375.1	.238	1.891	1.002
70,000	288.9	1.335	15.46	340.7	.560	0.73	0.350
(b) FOR DAY ONLY							
80,000	240.0	0.3675	5102×10^{-6}	310.6	1.68×10^{-3}	2219×10^{2}	9639×10^{17}
90,000	265.5	.1029	1165	347.1	6.64	616.8	2698
100,000	302.0	.03734	338.4	391.5	20.7	221.3	979.4
110,000	338.5	.01589	128.1	414.5	54.4	89.15	416.8
120,000	375.0	.007398	53.67	436.3	129	39.57	194.0
(c) FOR NIGHT ONLY							
80,000	240.0	0.3675	5102×10^{-6}	310.6	1.68×10^{-3}	2219×10^{2}	9639×10^{17}
90,000	265.5	.09640	1206	326.7	7.08	553.7	2528
100,000	302.0	.02992	328.0	348.4	25.9	161.4	784.5
110,000	338.5	.01082	98.63	384.3	79.9	57.24	284.0
120,000	375.0	.004800	34.82	436.3	100	25.65	125.9

[1] Data in the above table are from "Tentative tables of the upper atmosphere, National Advisory Committee for Aeronautics Technical Note No. 1200 (January 1947)." Columns 1 to 6 are as given in the table, and columns 7 and 8 were computed from constants appearing elsewhere in the note. Mean free paths for electrons and for ions may be calculated from those given for molecules by the approximate formulas $\lambda_e=4\sqrt{2}\,\lambda_m$, $\lambda_i=\lambda_m/5$. ν_e and ν_i may be similarly calculated but require assumptions as to extent of thermal equilibrium attained. Assuming equilibrium, they are given approximately by $\nu_e=33\nu_m$, $\nu_i=5\nu_m$.

4.2. Formation of Ionized Layers

a. Ionizing Agents

Ions may be produced in a gas by electromagnetic radiation, including ultraviolet light, X-rays, gamma rays, and cosmic rays. They may also be produced by the impact of high-speed particles on the molecules or atoms of a gas. Both causes of ionization appear to be present in the upper atmosphere, but by far the predominant ionizing agent is the ultraviolet light from the sun.

b. Recombination and Attachment

When ions (and among ions are included free electrons) are formed they tend to unite with ions of the opposite sign to form neutral molecules or atoms. The rate of change of the ion density is given by the equation

$$dN/dt=q-aN^2, \tag{4.3}$$

in which N is the number of ions of one sign (or ion pairs); q, the rate of production, and a the coefficient of recombination. For equilibrium conditions, $q=aN^2$. The value of a is dependent on conditions in the gas. There is in addition another process that is important for electrons. As long as an electron is free it can contribute as much to the propagation of radio waves as 30,000 ordinary ions, but if an electron becomes attached to a neutral molecule it is no more effective than the heavier positive ion which was formed when the electron was split from the original molecule. Attachment of electrons to molecules takes place very rapidly at high pressures, where the molecular density is great. In the troposphere attachment takes place so rapidly that the life of a free electron is only a few millionths of a second; attachment probably remains an important process in the lower portion of the ionosphere but becomes negligible at great heights.

c. Theory of Layer Formation

The theory of the production of ions in the atmosphere by the absorption of monochromatic ultraviolet radiation from the sun has been developed in detail by Chapman. If dS is the radiation absorbed in passing through a thin layer of the atmosphere, then

$$dS=A\rho S(\sec\chi)dh$$

or

$$dS/S=(A\rho_0\sec\chi)\epsilon^{-h/H}dh, \tag{4.4}$$

in which A is the absorption coefficient, χ the sun's zenith angle, ρ the density of the atmosphere at the height h, and S the intensity of the radiation at that height. As S has its maximum value,

S_∞, outside of the atmosphere, where $h=\infty$, we may solve eq 4.4 and obtain

$$S=S_\infty \epsilon^{-(A\rho_0 H \sec \chi)\epsilon^{-h/H}}. \quad (4.5)$$

The number of pairs of ions produced in each unit of volume of air is equal to the derivative with respect to h of eq 4.5 multiplied by cos χ and a constant, β. Designating by h_0 the level where the rate of ion production is a maximum when $\chi=0$ and by I_0, the rate of ion production at level h_0 when $\chi=0$, we have

$$I=I_0\epsilon^{[(h_0+H-h)/H]-[\sec\chi]\epsilon^{(h_0-h)/H}} \quad (4.6)$$

h_0 and I_0 may be determined by the relationships

$$\epsilon^{h_0/H}=A\rho_0 H$$

and

$$I_0=\beta S_\infty/H\epsilon. \quad (4.7)$$

Equation 4.6 may be written

$$\log_\epsilon I/I_0=(h_0-h)/H+1-(\sec\chi)\epsilon^{(h_0-h)/H}. \quad (4.6a)$$

I is a maximum where $(h_0-h)/H=\log_\epsilon(\sec\chi)$, and where

$$I_{max}=I_0\cos\chi. \quad (4.8)$$

Equation 4.6 is plotted in figure 4.1 for various values of χ. This shows that the rate of ion production varies with height, and has a maximum from which the value falls off in both directions, i. e., it has the characteristics of a layer. The height of the maximum varies with χ and the lowest value is determined by the absorption coefficient for the radiation. The intensity of the ion-production is determined by the coefficient β and the intensity of the radiation. Changing the intensity of the radiation does not alter the height at which maximum ion-production occurs. Thus if there are several spectral components in the ultraviolet light for which the absorption coefficients are markedly different, several distinct layers of ionization will result, the most strongly absorbed radiation producing the uppermost layer and the least strongly absorbed radiation producing the lowermost layer. Likewise, ionized layers would result if the gases of the atmosphere were distributed according to their partial pressures as given in eq 4.2, assuming that the gases of lowest molecular weight absorb most strongly.

As I in eq 4.6 is the rate of ion-production, we can identify it with q in eq 4.3 and write

$$dN/dt=I_0\epsilon^{[(h_0+H-h)/H-[\sec\chi]\epsilon^{(h_0-h)/H}]}-aN^2. \quad (4.9)$$

For the steady state N is equal to the square root of the first term on the right-hand side divided by the recombination coefficient a. That is, $N=\sqrt{I_0/a}\epsilon^{[(h_0-h)/2H]+[1/2][1-(\sec\chi)\epsilon^{(h_0-h)/H}]}$. It is assumed in the Chapman theory that a is the same throughout a given layer. The maximum ion density, corresponding to I_{max}, is then

$$N_{max}=\sqrt{(I_0\cos\chi)/a}.$$

The maximum ion density of a Chapman layer is thus proportional to the square root of the cosine of the sun's zenith angle. The behavior of the E-layer ionization follows this law reasonably well.

It is of interest to see how the energy absorbed from a radio wave passing through such a region and reflected by a higher one varies with χ. In chapter 2, p. 13, the absorption coefficient k is

$$k=\frac{\nu Ne^2}{2c\eta\kappa_0 m\omega^2}$$

where ν is the collisional frequency (roughly proportional to atmospheric density) and N is the electron density, corresponding to the ion density N derived above.

The total absorption index α, for a vertically incident wave, is the integral, $\int k ds$, over the path, If we consider only the nondeviative absorption α_m.

$$\alpha_m=\int k ds=2\int_0^{h_0}\frac{\nu Ne^2}{2c\eta\kappa_0 m\omega^2}dh=\frac{e^2}{c\eta\kappa_0 m\omega^2}\int_0^{h_0}\nu N dh$$

If the absorption is nondeviative, the quantity νN will be negligible at heights much less than h_0 due to the low value of N, and also at very great heights due to the low value of ν, so that we are interested in the integral $\int_{-\infty}^{+\infty}\nu N dh$. Writing $\nu=\nu_0\epsilon^{(h_0-h)/H}$, where ν_0 is the collisional frequency

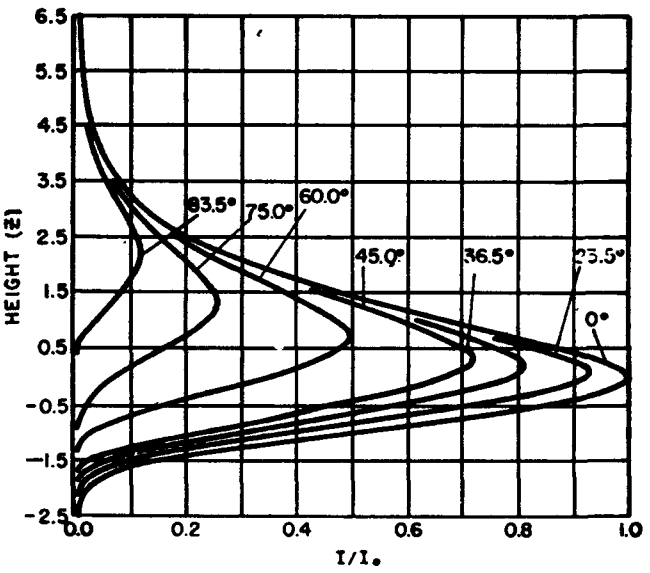


FIGURE 4.1. *Rate of ion production at various heights* z $(=(h-h_0)/H)$ *for different values of* χ.

at height h_0, and

$$N=\sqrt{I_0/a}\epsilon^{(h_0-h)/2H}\epsilon^{(1/2)(1-\sec\chi)\epsilon^{(h_0-h)/H}}$$

$$\int_{-\infty}^{+\infty}\nu N dh=(2\nu_0 H\sqrt{I_0/a})\epsilon^{(1/2)(1+2h_0/H)}\int_0^{+\infty}x^2\epsilon^{-bx^2}dx$$

$$=2H\nu_0\sqrt{I_0\pi\epsilon/a}\cos^{3/2}\chi,$$

in which $x=\epsilon^{-h/2H}$ and $b=(1/2)(\sec\chi)\epsilon^{h_0/H}$. Thus the nondeviative absorption for a wave penetrating a layer of the Chapman type is found to be proportional to the 3/2 power of the cosine of the sun's zenith angle.

If we consider that the electrons are removed by the process of attachment of electrons to molecules, a different type of ion distribution results. The attachment of electrons to molecules is proportional to the number of molecules present, which is proportional to the density of the atmosphere. Therefore, we must substitute $Na'\rho_{h_0}\epsilon^{(h_0-h)/H}$ for aN^2 in eq 4.9, a' being the attachment coefficient and ρ_{h_0} the density at $h=h_0$. The equation becomes

$$dN/dt=I_0\epsilon^{[(h_0+H-h)/H]-(\sec\chi)\epsilon^{(h_0-h)/H}}-$$

$$Na'\rho_{h_0}\epsilon^{(h_0-h)/H} \qquad (4.10)$$

and in the steady state when $dN/dt=0$

$$N=\frac{I_0}{a\rho_{h_0}}\epsilon^{1-(\sec\chi)\epsilon^{(h_0-h)/H}} \qquad (4.11)$$

This type of ion bank has no maximum for the steady state except at very great heights where a maximum is attained by all of the molecules becoming ionized. A plot of ion density versus height for this type of ion bank is shown in figure 4.2. It may be noted that the lower part of the ion bank closely resembles the distribution deduced by Chapman. Since the steady state is reached more rapidly at lesser heights there is a tendency for a process of this type to lead to layer formation owing to fluctuations in the rate of ion-production as the zenith angle of the sun changes. The process will also lead to the formation of distinctly separated banks of ions if the incoming ultraviolet radiation is selectively absorbed or if the distribution of the atmospheric gases is in accord with their molecular weights. A similar condition obtains if, in the Chapman theory, the value of a is dependent on molecular density.

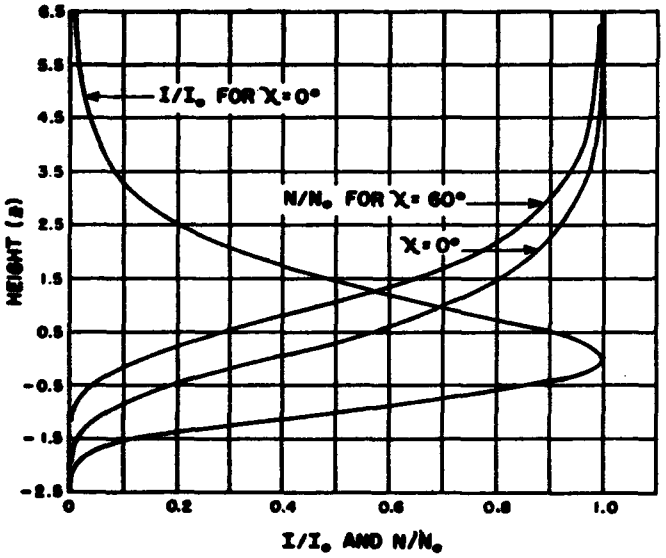


FIGURE 4.2. *Distribution of ion density with height z (=(h−h₀)/H), assuming ions are removed by attachment and that a steady state exists.*

It is entirely likely that both types of processes apply in the formation of ions in the atmosphere, the process cited by Chapman predominating in the highest levels and the alternate process in the lower levels. These processes are capable of accounting for most of the observed effects of the ionosphere.

4.3. Observed Distribution of Ion Densities With Height

Let us examine the facts of the ionosphere as supplied by radio measurements. Since there is a dependence of ionization on the sun's zenith angle, the ionization might be expected to exist only during the daylight hours. However, because of the fact that recombination of electrons with ions or their attachment to neutral molecules takes place at a limited rate, some of the ions persist through the night hours. It is well, therefore, to consider daytime and nighttime conditions separately.

Figure 4.3 shows a typical distribution of ion densities at midday at a tropical station as deduced from automatic multifrequency ionospheric records obtained by the method described in chapter 3. Figure 4.4 shows the distribution at midday during the summer at a midlatitude station. There are striking similarities in both figures. The presence of three distinct banks of ions is clear, one at a height of around 300 km, one at a height of 200 km and one at a height of 100 km. These are referred to as the $F2$, $F1$, and E layers, respectively. Ion densities below about 10^{10} equivalent electrons per cubic meter are not shown because the measuring equipment used is not sensitive enough to obtain reflections of radio waves at frequencies corresponding to those ion densities during daylight hours, although the frequency range of the equipment is adequate. This indicates the existence below the E region, of a fourth region of ionization called the D region. This D region exists practically only in the daytime, and so strongly absorbs radio waves of frequencies around 1 Mc (corresponding to an equivalent electron density of $1.24\times10^{10}/m^3$ for

reflection), that reflections at that frequency are extremely weak, as compared with reflections at the same frequency at night.

Comparison of figure 4.4 with a corresponding representation given in figure 4.5 for the same station, for the same time of day but during the winter, shows striking differences. As might be expected, the ion densities of the *E* layer and the *F*1 layer are lower during winter when, due to obliquity, the ionizing effect of the sun is less. The same is true of the *D* region, as may be inferred from the fact that lower ion densities may be "seen" through it. But the *F*2-layer ion density does not exhibit this characteristic; in fact, at the midlatitude station selected (the CRPL station near Washington, D. C.) the *F*2-layer ion density is greater in winter than in summer.

Similar sets of curves are shown in figures 4.6, 4.7, and 4.8 for night hours at the same times and stations. In these it is seen that the *F*1-layer shrinks to indiscernible proportions at night, while the *E* layer is still present but with greatly reduced ion density. Appreciable ionization persists in the *F* region, throughout the night. It may be "seen" down to the lower limit of the frequency range of the measuring equipment indicating that the ions in the *D* region have been largely removed.

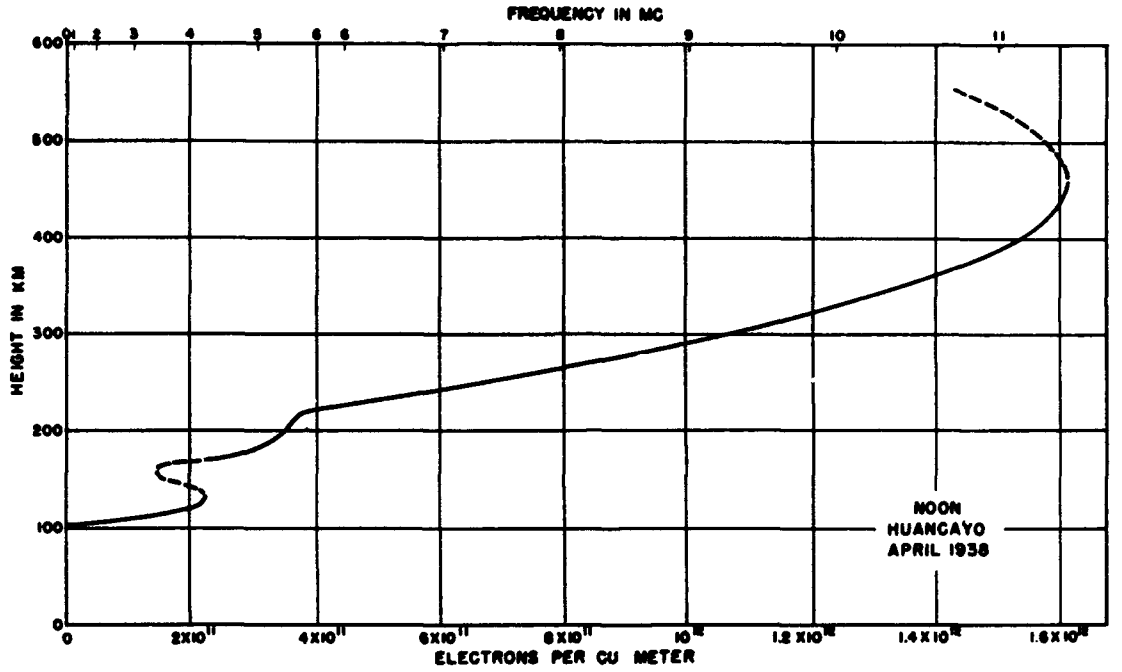


FIGURE 4.3. *Distribution of ion density with height for quiet conditions.*

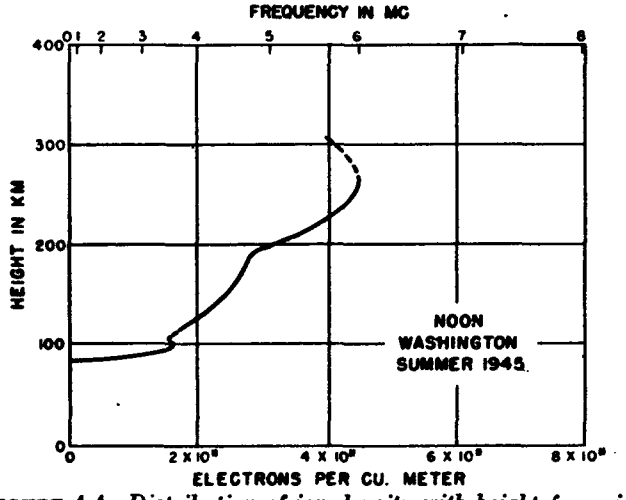


FIGURE 4.4. *Distribution of ion density with height for quiet conditions.*

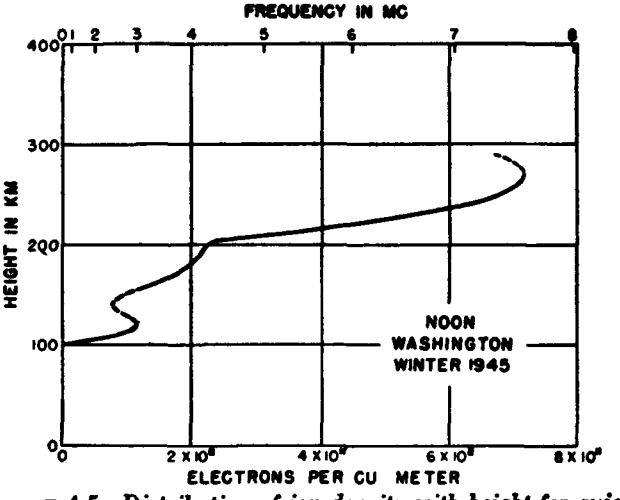


FIGURE 4.5. *Distribution of ion density with height for quiet conditions.*

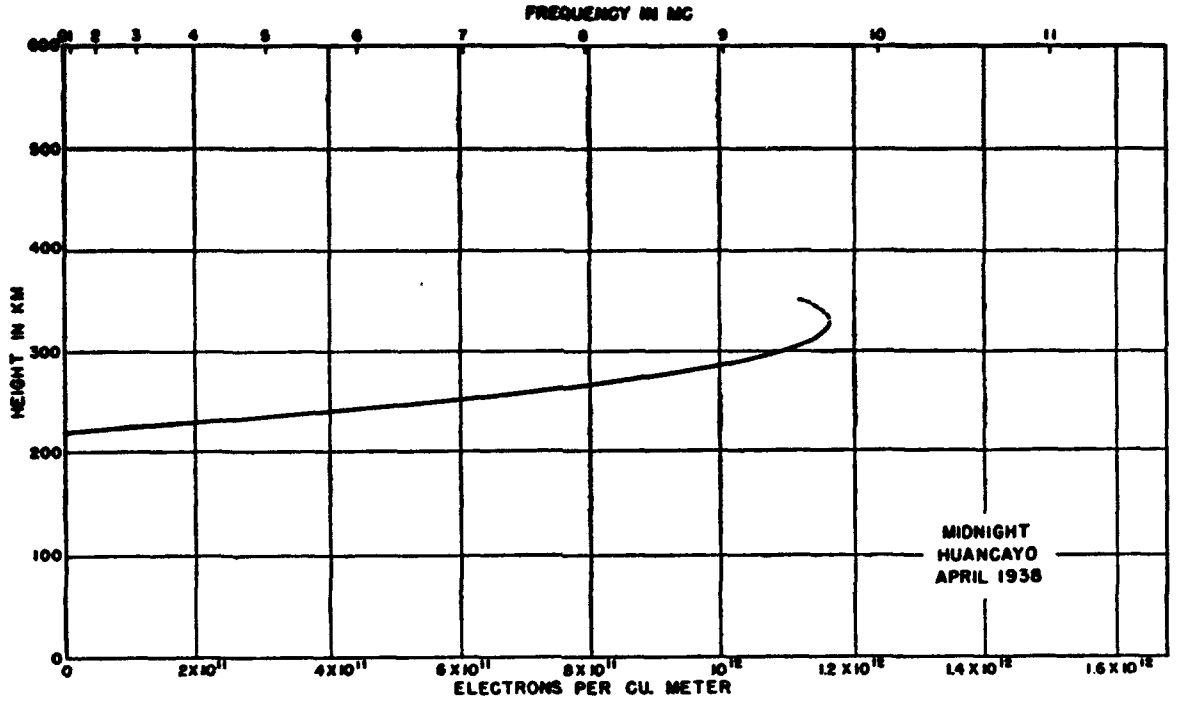


FIGURE 4.6. *Distribution of ion density with height for quiet conditions.*

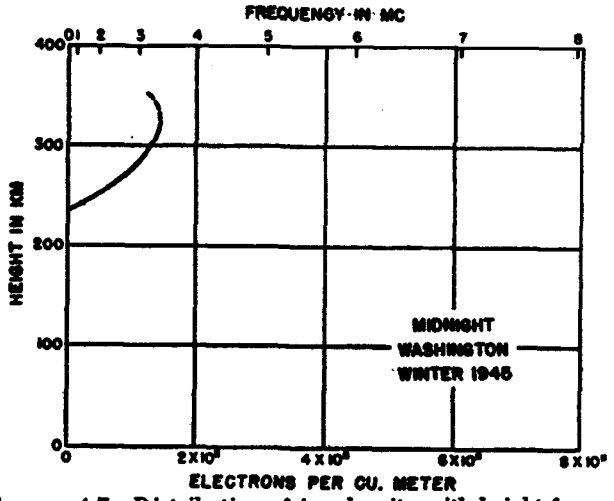


FIGURE 4.7. *Distribution of ion density with height for quiet conditions.*

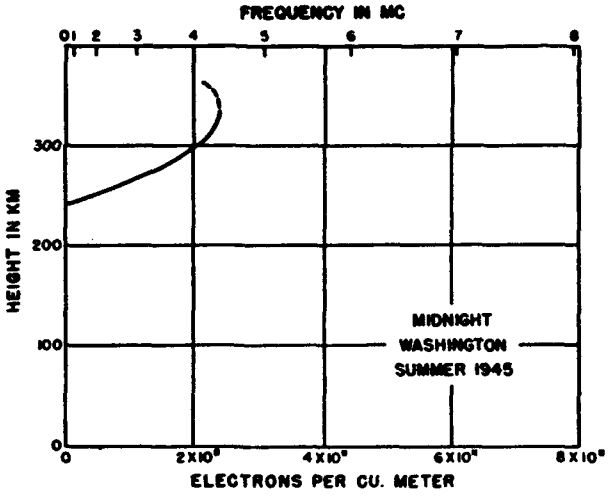


FIGURE 4.8. *Distribution of ion density with height for quiet conditions.*

The curves in these six figures are plotted for true heights so that the ion densities may be compared directly with the molecular densities given in table 4.1, on the basis of the various assumptions. It is important to realize that the ion densities are computed from directly observed quantities and although certain assumptions are made in the calculations the values of ion densities given are far more reliable than the molecular densities. It is assumed, of course, that the ions effective for reflection of the radio waves are free electrons. The radio observations verify this assumption, although they do not preclude the possibility that molecular or atomic ions may be several hundred times as numerous as the electrons. It was assumed in making these calculations that the Lorentz polarization correction should not be applied. If the correction were applied, the values of ion density would be increased by exactly 50 percent.

There is sufficient evidence presented in these six figures to place confidence in the theory of production of ions in the atmosphere as outlined earlier in this chapter, at least insofar as the *D*, *E*, and *F*1 regions are concerned. Additional evidence from the many ionospheric observing stations scattered over the surface of the globe further supports the theory. The *F*2 region

appears to depart decidedly from the simple theory and needs separate consideration.

4.4. Geographical Distribution of Ionized Layers

a. *F*1, *E*, and *D* Regions

The general features brought out by figures 4.3 to 4.8 are supported by observations from the worldwide network of ionospheric stations. These are represented in plots shown in figures 4.9, 4.10, 4.11, and 4.12 of the worldwide distribution of critical frequencies in megacycles (which when squared and multiplied by 1.24×10^{10} give the densities of electrons per cubic meter) for the *E* region and for the *F*1 region at the equinox and the solstices for 1945. In constructing the figures it was assumed that seasonal antisymmetry applied with respect to the geographic equator, i. e., that observations during June in a northern latitude correspond to observations during December in the same southern latitude. The abscissas in the charts are local mean times, so that the data apply for all longitudes for the times shown on the charts. Because of the antisymmetry with respect to the equator, the charts shown for the December solstice have only to be reversed, south for north, to represent the June solstice.

According to the theory of ion production the maximum ion density for any latitude should occur shortly after noon for both the *E* and *F* layers, and the value of the maximum should be greatest at that latitude where the sun passes most nearly overhead. Values of maximum ion density for any latitude at the solstices should be nearly the same as the value at the equinox for a latitude which differs from it by the average difference in declination of the sun. (Since the seasons are formed here by grouping the months January, February, November, December for the December solstice, and May, June, July, August for the June solstice and March, April, September, and October for the equinox, the mean declination of the sun at the solstices is $\pm19°$, not $\pm23°$).

The charted values of critical frequency (or ion density) agree very well with the theoretical values. The tendency for the gradient of critical frequency with time to be less steep during the evening hours than during the morning hours, is also in accord with theory. The different shapes of the plots for the solstices as compared with that for the equinox results largely from distortions introduced by the projection used, although there are some real differences which result from the fact that the apparent course of the sun at the solstices is not a great circle.

The charts of minimum virtual heights of the *E* and *F*1 layers given in figures 4.13 and 4.14 also support the theory. For greater zenith angles of the sun the ions are formed at lower levels in the atmosphere, as is indicated by the theoretical curves in figure 4.1. This general feature is verified by the charts of observed heights. As predicted by the theory, the lowest virtual heights appear near the subsolar point.

These charts show the minimum virtual heights from which radio waves are actually returned; therefore they do not correspond exactly to the true heights shown in figures 4.3 to 4.8. In the case of the *E* layer, radio waves of low frequencies, which would give evidence of the minimum height of the layer, are absorbed by the *D* region. The minimum heights of the *F*1 layer are similarly screened by the *E* layer; in addition waves which are reflected back to earth by the *F*1 layer must pass through the *E* layer where they suffer retardation as explained in chapter 2. Therefore, the virtual heights are all greater than the true heights.

b. The *F*2 Layer

This close agreement with the theory outlined above does not apply for the *F*2 layer. The assumption of simple antisymmetry relative to the equator does not hold for the *F*2 layer; neither does the assumption that the *F*2-layer ionization is the same along a given parallel of latitude for the same local time. To represent the *F*2 layer four charts are necessary. Figures 4.15 and 4.16 show the distribution with latitude and longitude of critical frequencies for the equinox and the December solstice when the sun is over the 69° W meridian, and figures 4.17 and 4.18 show the same when the sun is over the 111° E meridian. In constructing these charts the assumption has been made that the *F*2 layer at a given point when the sun is over the 69° W meridian is the same as it would be in the reverse latitude, with a change of 180° in longitude and when the sun is over the 111° E meridian. That is, the map for the June solstice corresponding to figure 4.16 would be obtained by inverting the contour lines on figure 4.18, south for north and vice versa. This constitutes an over-simplification of the actual conditions, but it serves to illustrate the nature of the problem involved. Distribution of virtual heights of the *F*2 layer are shown in figures 4.19 and 4.20.

Clearly, a distribution of ionization like this cannot be explained as the result of an ionizing process which is a simple function of the sun's zenith angle, although, because of the distinct dependence on solar illumination, there can be no doubt that the main, if not the only, ionizing agent is solar ultraviolet light.

An important clue to the explanation of the phenomenon is the distinct difference between the distribution of ionization when the sun is over the 69° W meridian and when it is over the 111° E meridian. The northern pole of the earth's uni-

form magnetization is in longitude 69° W, latitude 78.5° N, and the southern pole is in longitude 111° E, latitude 78.5° S. The division of figures 4.15 to 4.18 is based on whether the sun is over the meridian containing one pole or the other; the difference between the two conditions is distinct. It is evident that the earth's magnetic field has a strong influence on the distribution of $F2$-layer ionization. This is known as the geomagnetic effect on the $F2$ layer. Another feature of the figures which indicates a geomagnetic effect is the lower ionization which predominates along the geomagnetic equator, indicated in the figures by a dashed line. A smooth approach to a maximum of ionization in equatorial regions would be expected according to our simple theory of undisturbed effects of ultraviolet light.

No complete theory has been worked out to explain the anomalous characteristics of the $F2$ layer. It is likely that the effects will find explanation in the simple ultraviolet light theory coupled with the effects of the earth's magnetic field and of electric forces produced by changes in the magnetic field on migration of ions, in accordance with the equations set forth in chapter 2.

These anomalous characteristics of the $F2$ layer have a profound bearing on the phenomena of radio wave propagation over paths in different parts of the earth. It is not possible to consider the characteristics for long propagation paths as simple functions of time of day and of latitude; separate considerations must be given to the parts of the earth over which the propagation takes place. The details of these considerations will be presented in chapter 6.

c. Effects in the Auroral Zone

In figure 4.21 are indicated the zones of maximum auroral frequency. In these regions ions are produced in the atmosphere by corpuscular emissions from the sun in addition to the effects of ultraviolet light. The effects of these corpuscles extend all the way down to the E region. Their usual effect is to produce very intense ionization in the lower levels where absorption takes place so that propagation over paths involving reflections in the auroral zones is seriously hampered. Detailed consideration of these matters properly belongs in later chapters.

4.5. References

S. Chapman, The absorption and dissociative or ionizing effect of monochromatic radiation in an atmosphere on a rotating earth, Proc. Phys. Soc. (London) **43,** 26 (1931).

B. Haurwitz, Physical state of the upper atmosphere, J. Roy. Astron. Soc. Canada **30,** 315 (Oct. 1936), 349 (Nov. 1936), 397 (Dec. 1936).

B. Haurwitz, Physical state of the upper atmosphere, J. Roy. Astron. Soc. Canada **31,** 19 (Jan. 1937), 76 (Feb. 1937).

N. E. Bradbury, Ionization negative-ion formation, and recombination in the ionosphere, Terr. Mag. & Atmos. Elec. **43,** 55 (March 1938).

O. R. Wulf and L. S. Deming, Production of ionospheric E and F regions and lower altitude ionization causing radio fadeouts, Terr. Mag. & Atmos. Elec. **43,** 283 (Sept. 1938).

J. A. Fleming, Terrestrial magnetism and electricity (McGraw-Hill Book Co., New York, N. Y., 1939).

E. O. Hulburt, E region of ionosphere, Phys. Rev. **55,** 629 (1939).

S. Chapman and J. Bartels, Geomagnetism, 2 volumes (Clarendon Press, Oxford, 1940).

F. L. Mohler, Recombination and electron attachment in the F layer of the ionosphere, J. Research NBS **25,** 507 (1940) RP1342.

H. S. W. Massey, Recombination and attachment processes in the ionosphere, Bul. Int. Union Geod. & Geophys., No. 11, 521 (1940).

S. Chapman, The sun and the ionosphere, J. Inst. Elec. Engrs. (London) **88,** part I, 400 (Nov. 1941).

O. E. H. Rydbeck. A theoretical survey of the possibilities of determining the distribution of the free electrons in upper atmosphere, Trans. Chalmers Univ. Tech., Gothenburg, Sweden, No. 3, **74** (1942).

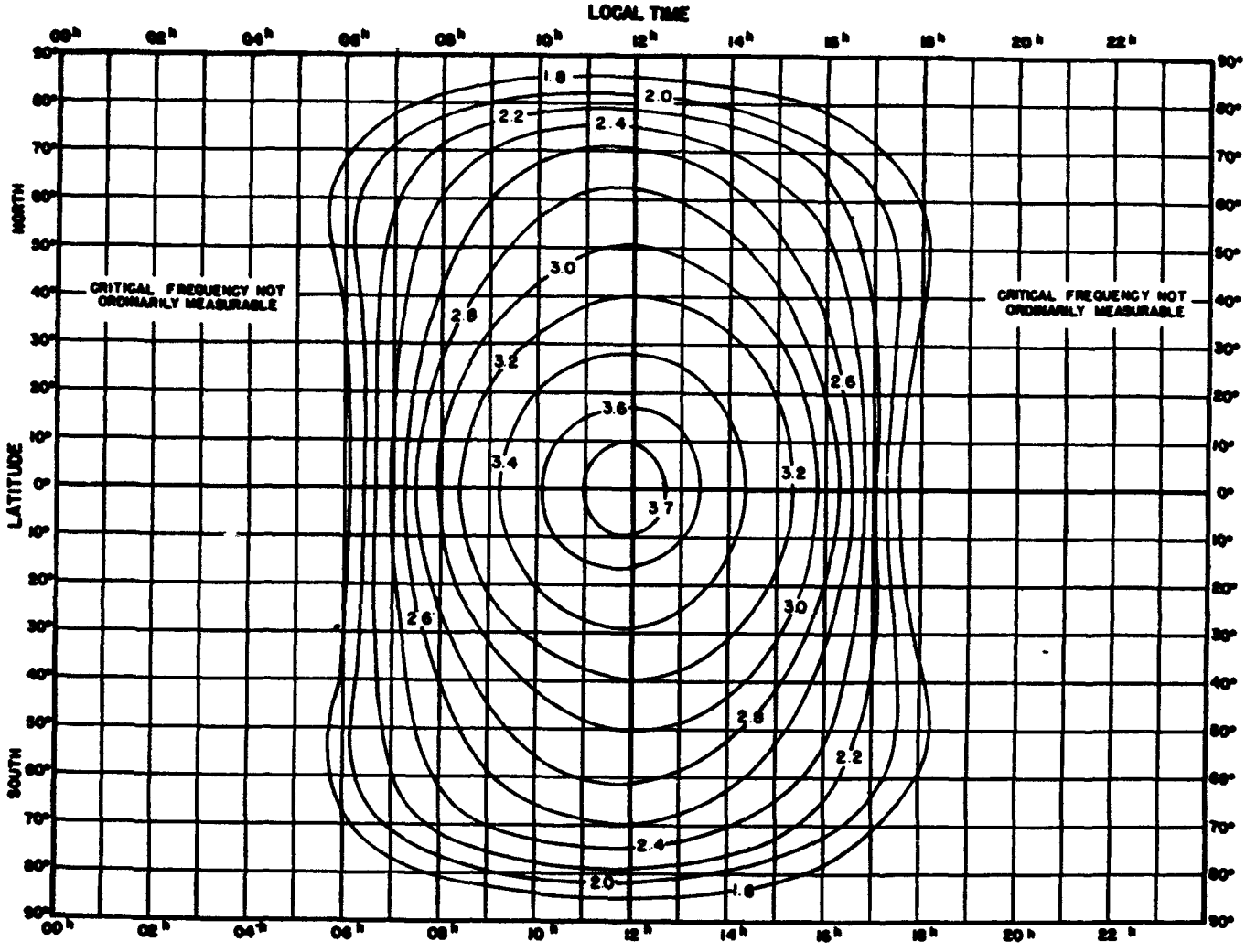


FIGURE 4.9. *World-wide distribution of critical frequency of E layer (f_0E) in megacycles for equinox, 1945.*

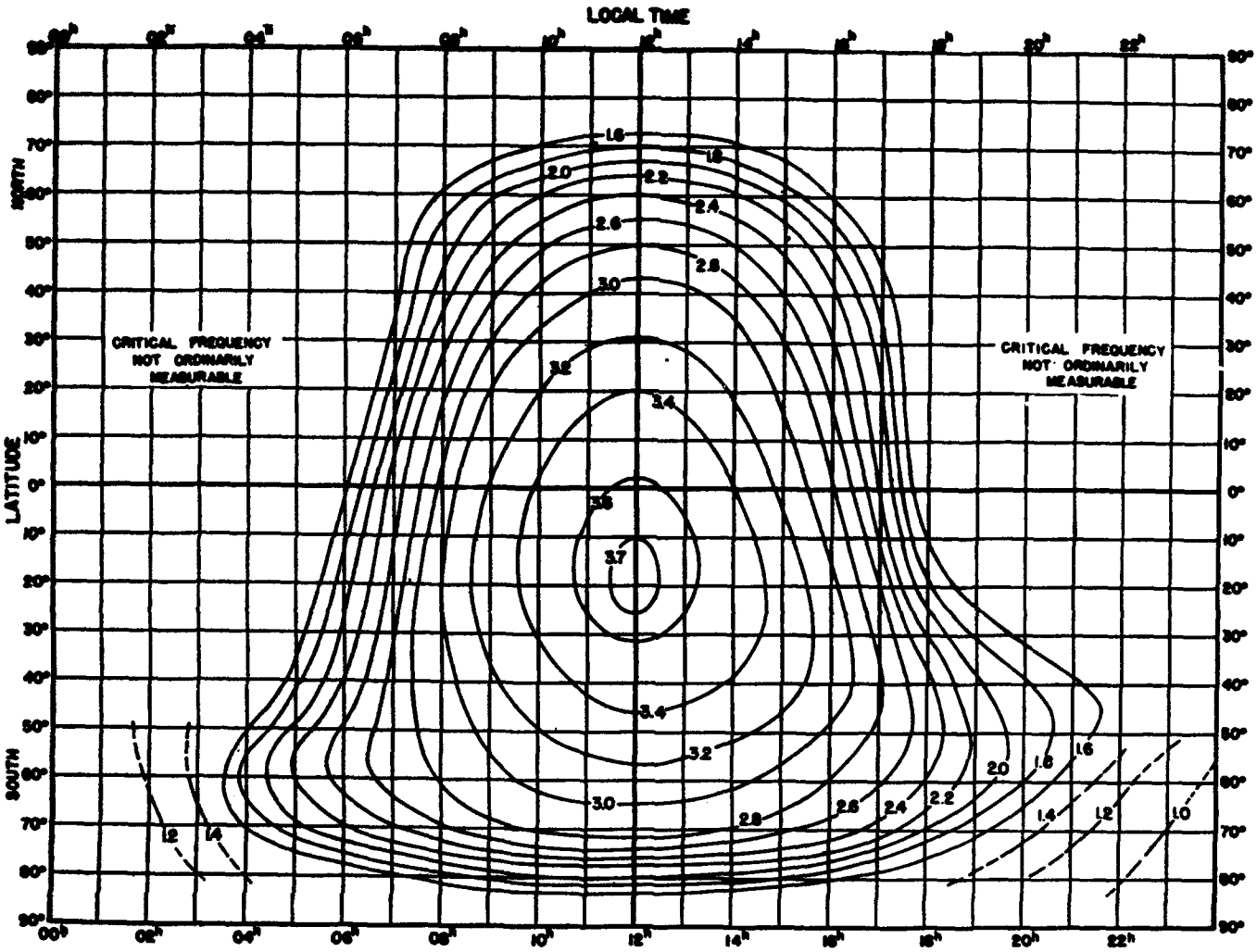


FIGURE 4.10. *World-wide distribution of critical frequency of E layer (f_0E) in megacycles for December solstice, 1945.*

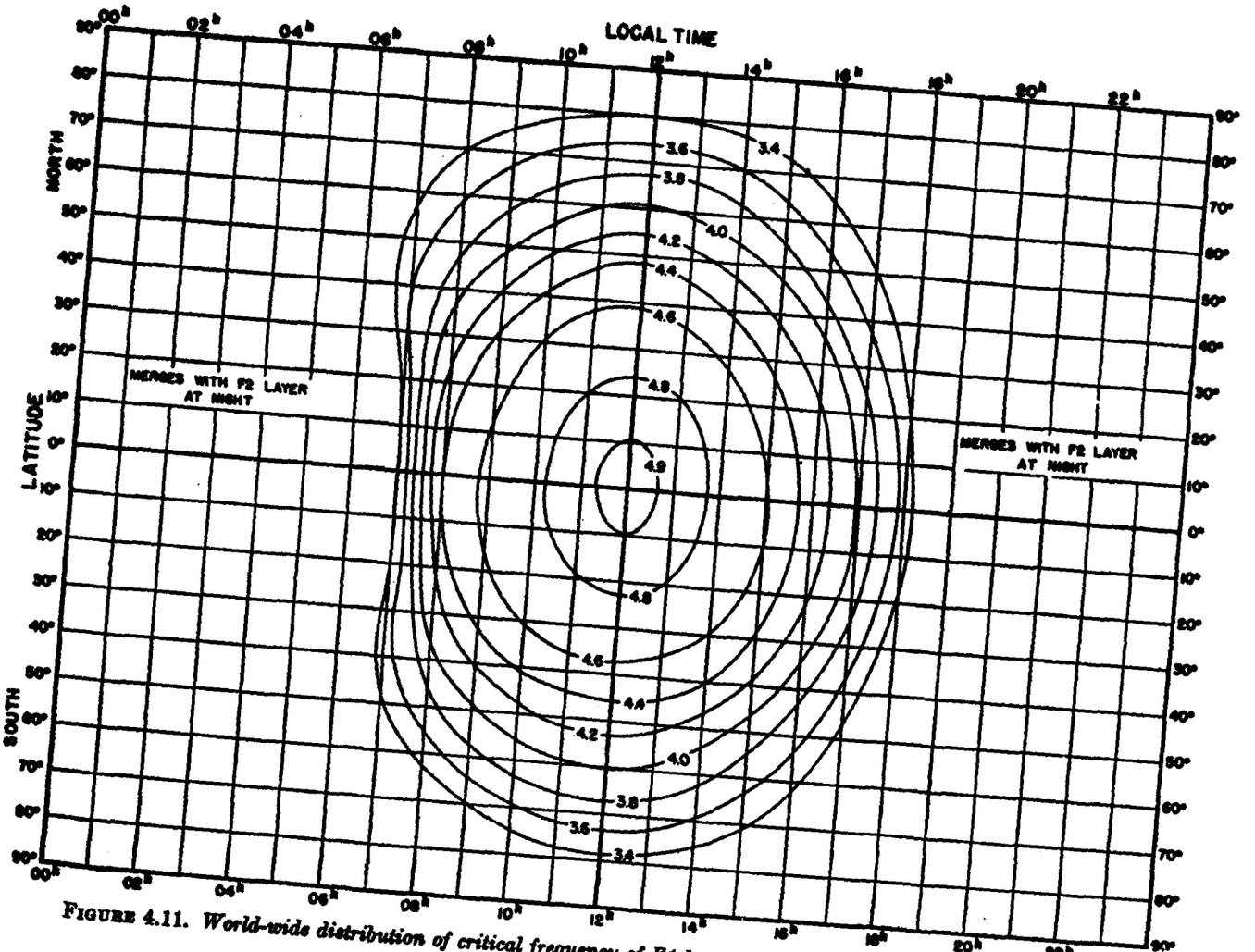


FIGURE 4.11. *World-wide distribution of critical frequency of F1 layer (f·F1) in megacycles for equinox, 1945.*

935333 O - 51 - 4

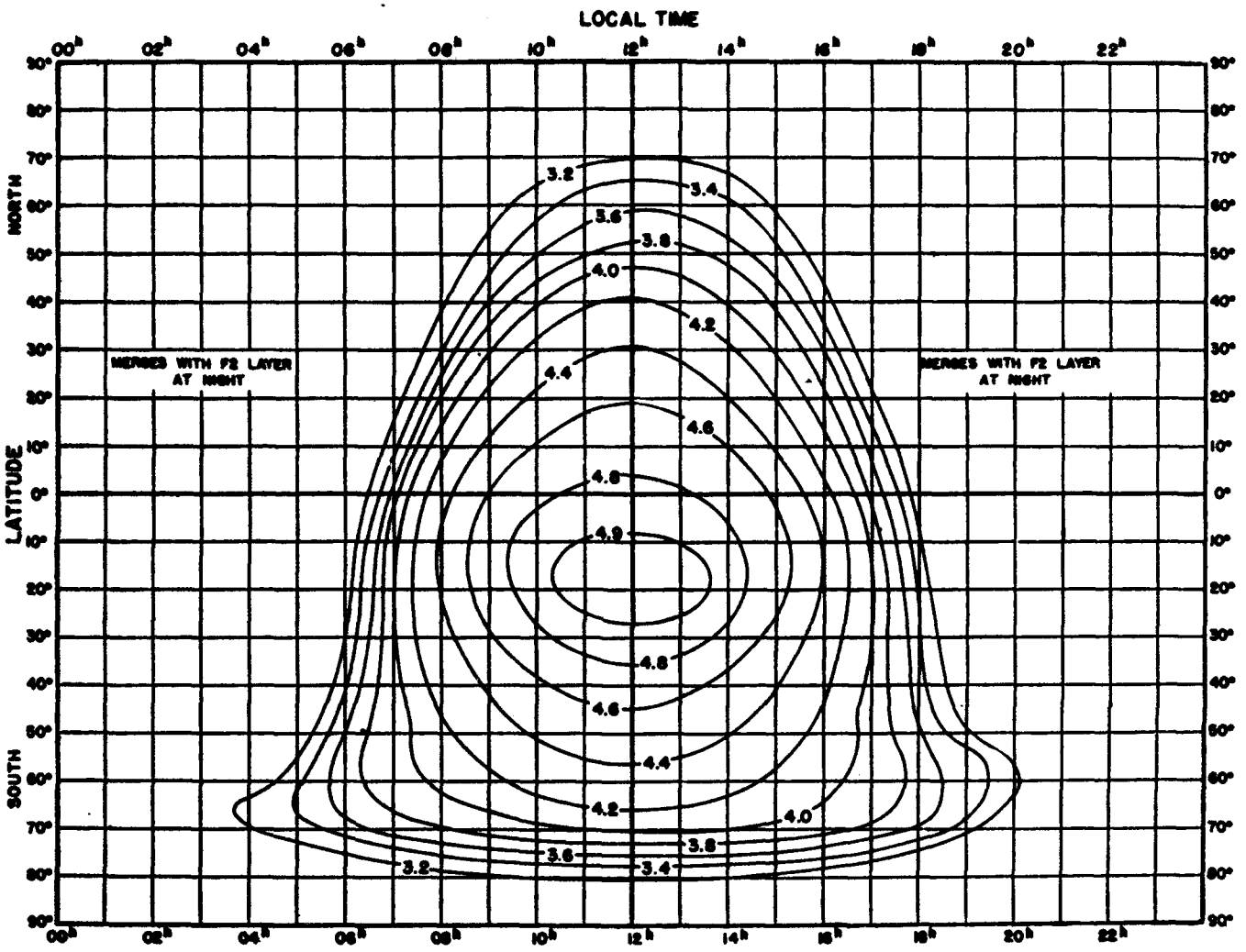


FIGURE 4.12. *World-wide distribution of critical frequency of F1 layer (f_0F1) in megacycles for December solstice, 1945.*

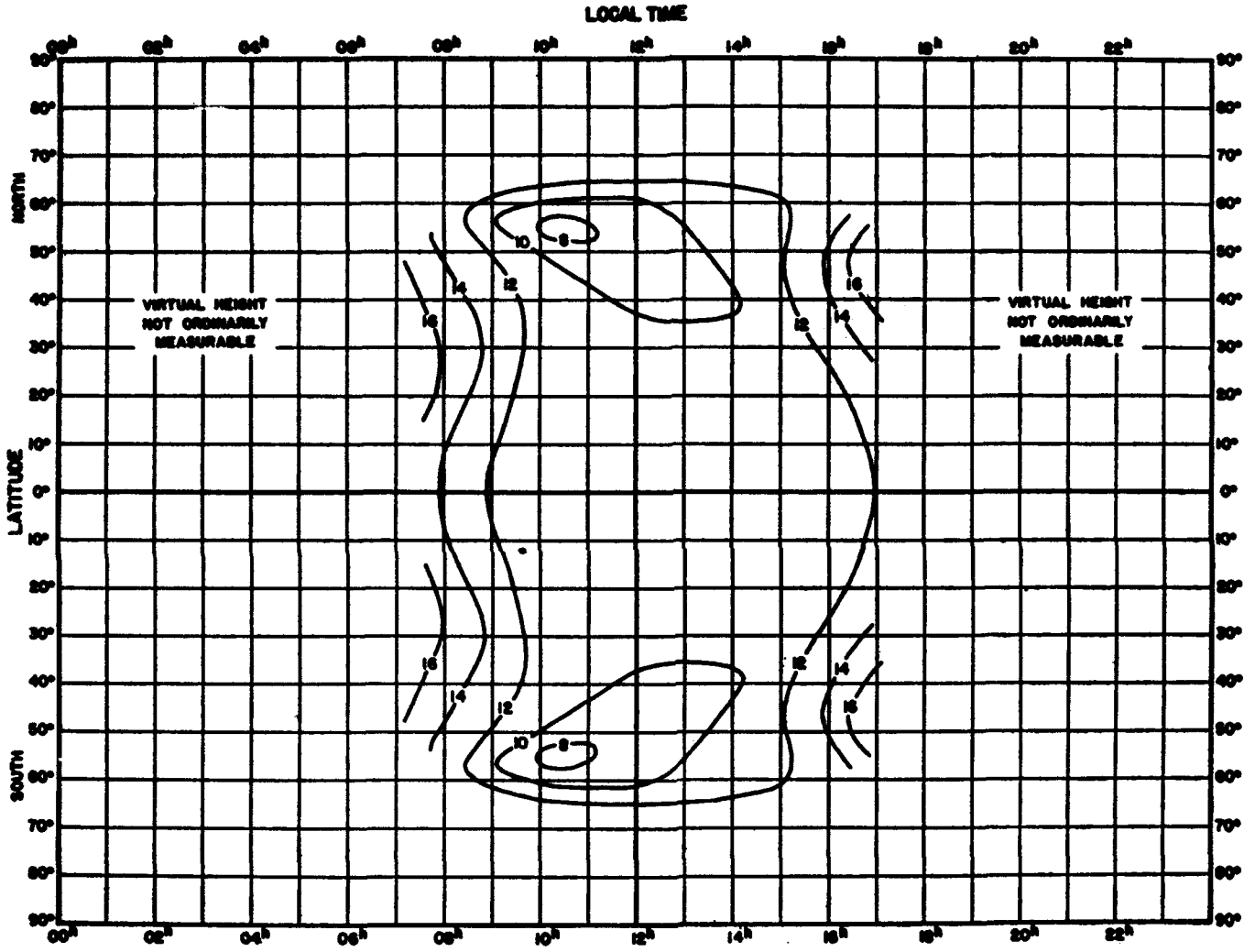


FIGURE 4.13. *World-wide distribution of minimum virtual height of E layer (h'E) in tens of kilometers for equinox, 1945.*

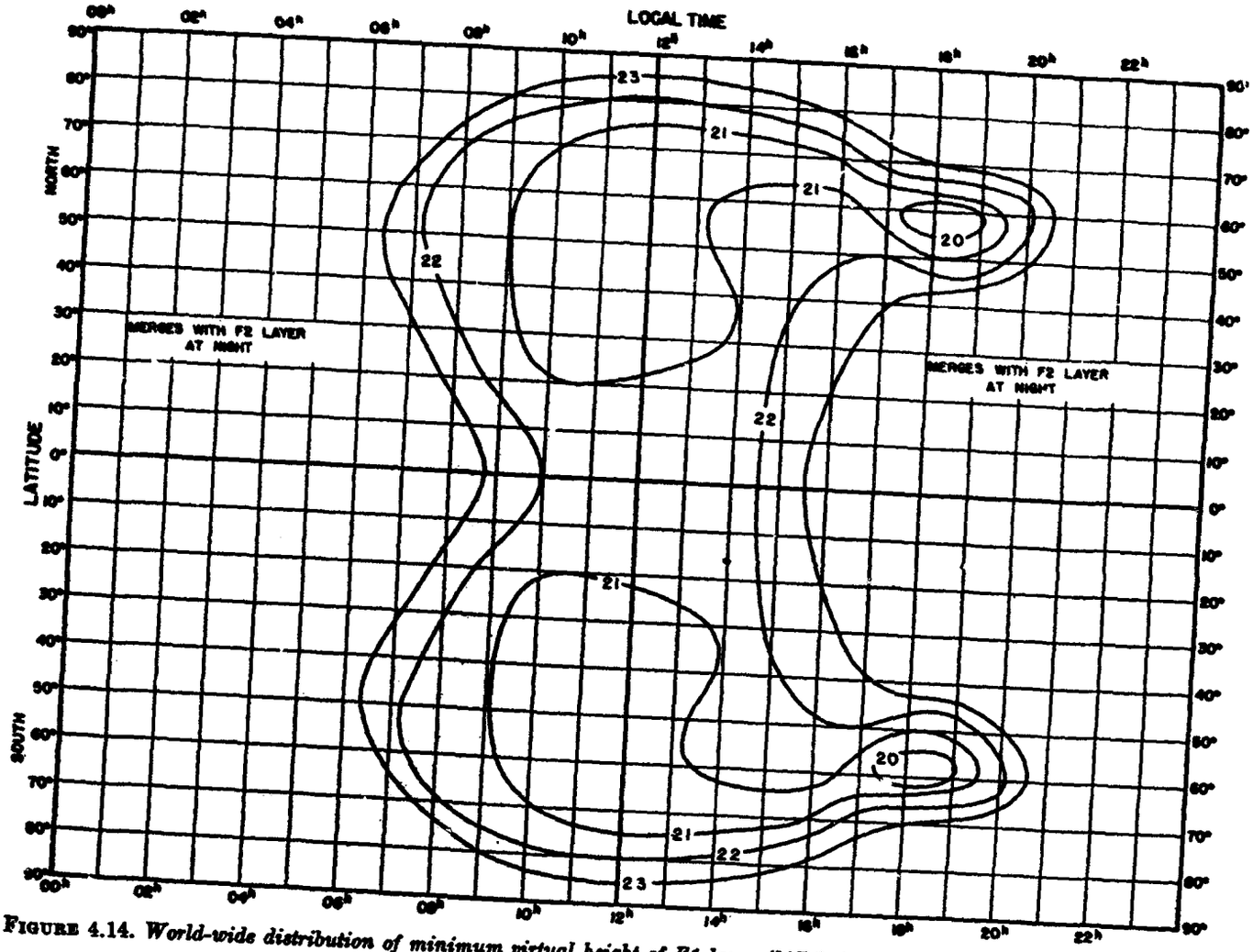


FIGURE 4.14. *World-wide distribution of minimum virtual height of F1 layer (h'F1) in tens of kilometers for equinox, 1945*

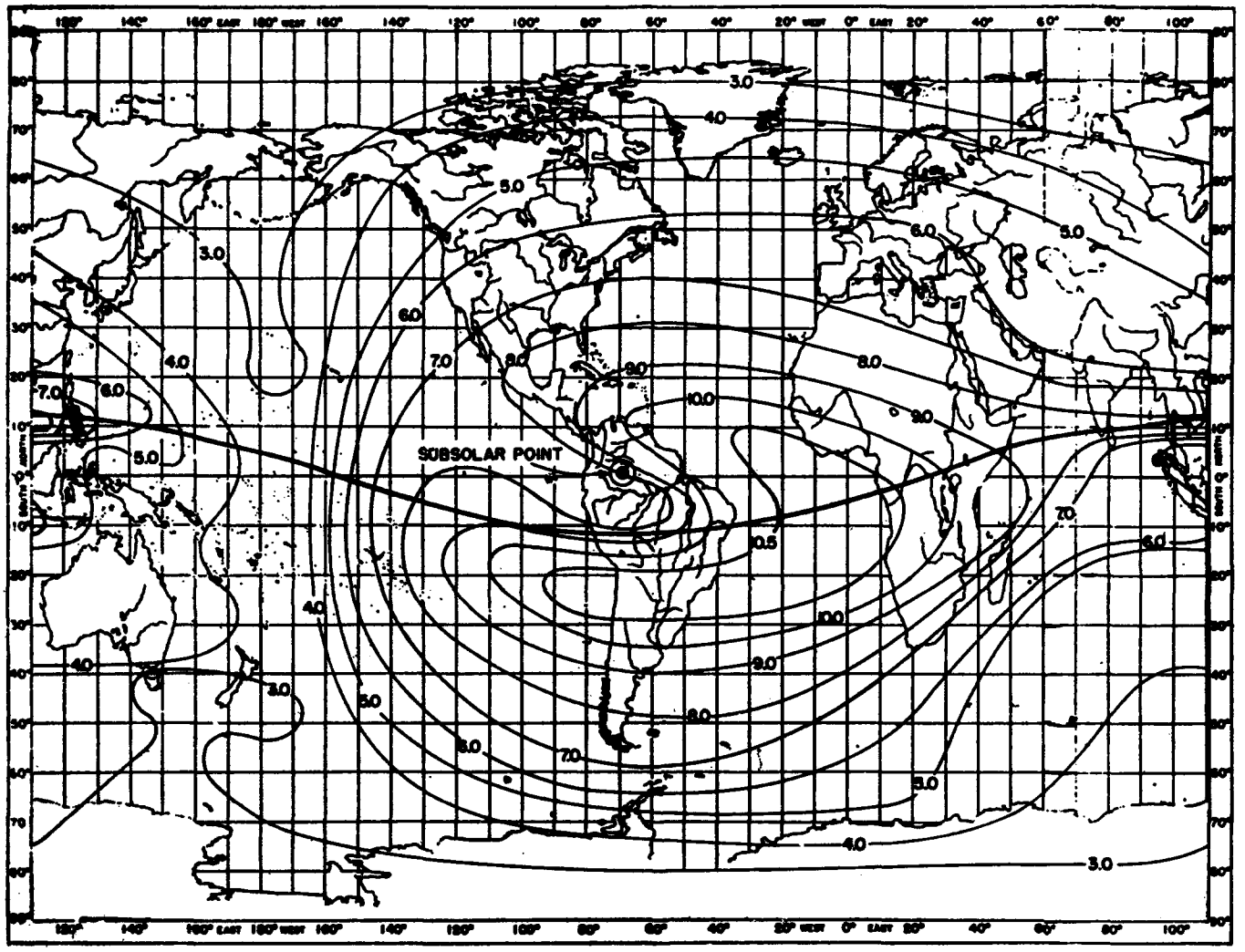


FIGURE 4.15. *World-wide distribution of critical frequency of F2 layer (f°F2) in megacycles for equinox, 1945, when it is noon at 69° W.*

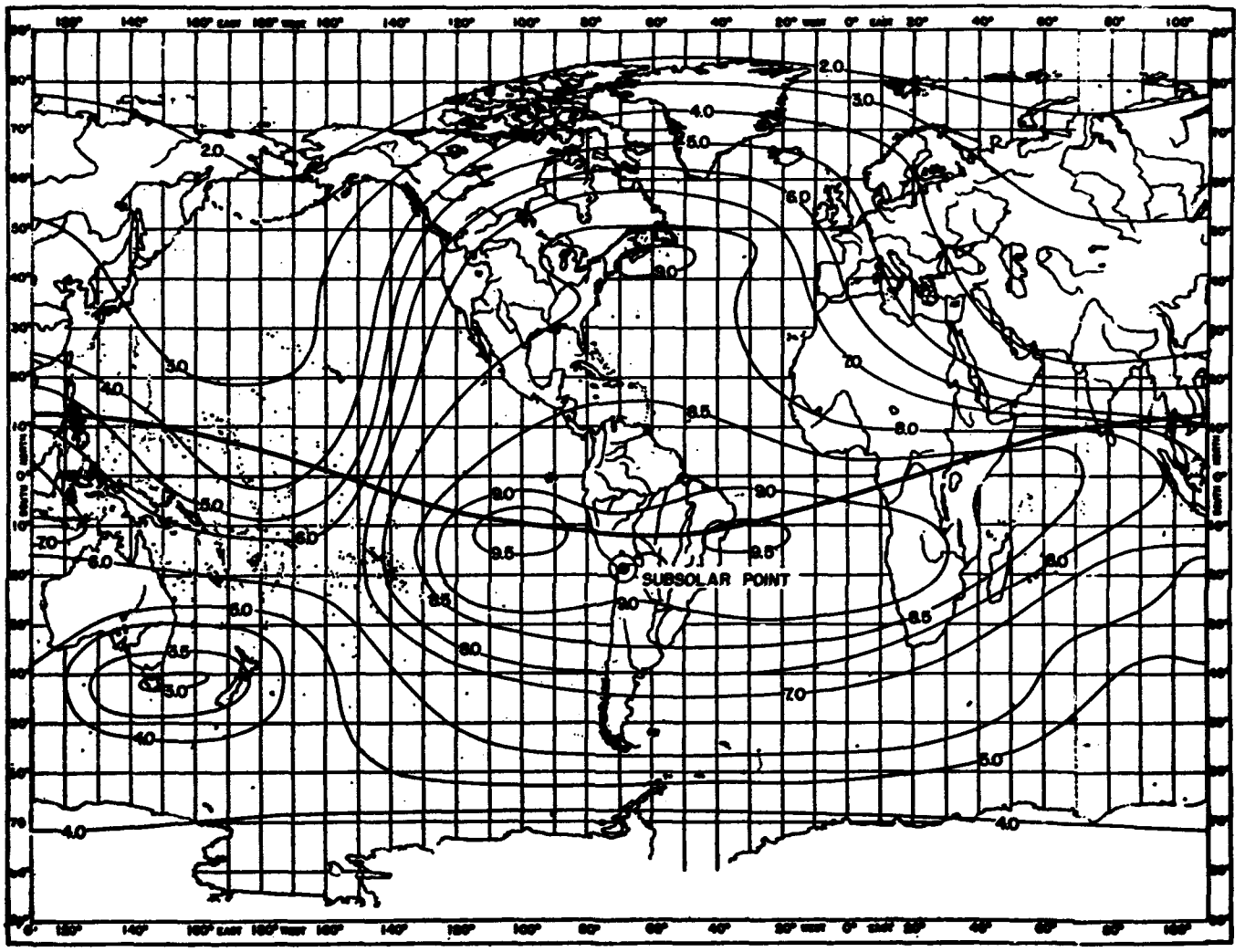


FIGURE 4.16. *World-wide distribution of critical frequency of F2 layer (f_0F2) in megacycles for December solstice, 1945, when when it is noon at 69° W.*

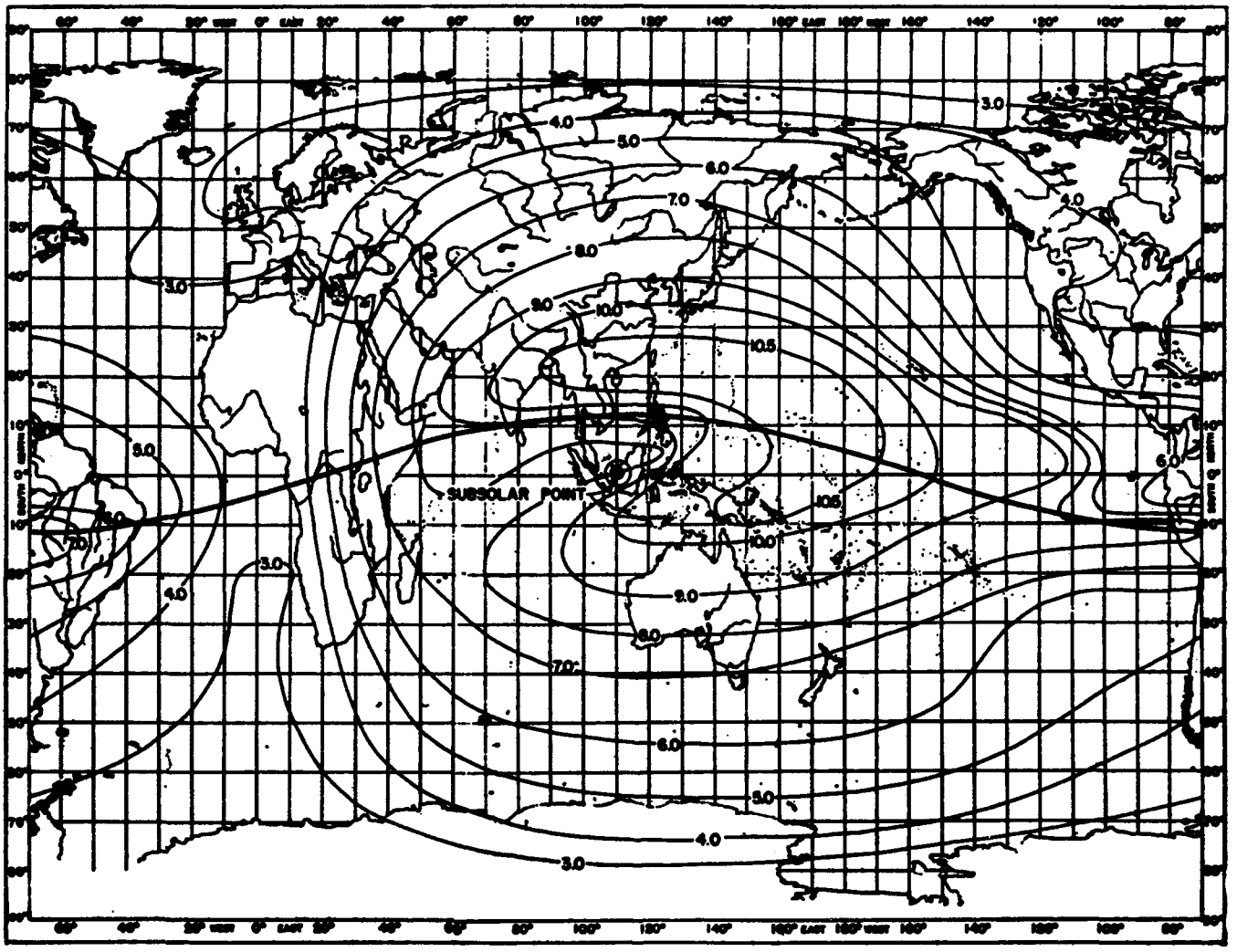


FIGURE 4.17. *World-wide distribution of critical frequency of F2 layer ($f^{\circ}F2$) in megacycles for equinox, 1945, when it is noon at 111° E.*

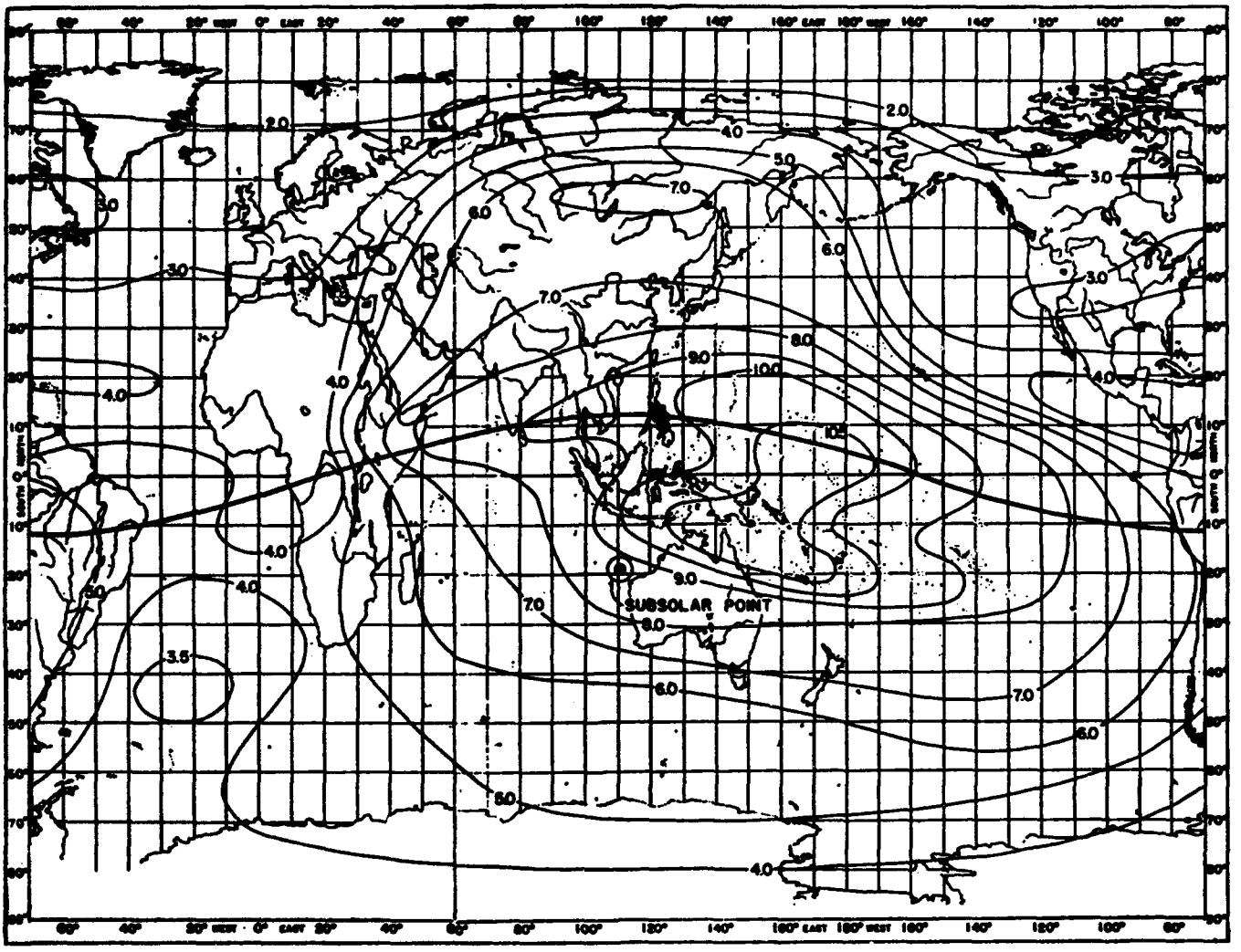


FIGURE 4.18. *World-wide distribution of critical frequency of F2 layer (f°F2) in megacycles for December solstice, 1945, when it is noon at 111° E.*

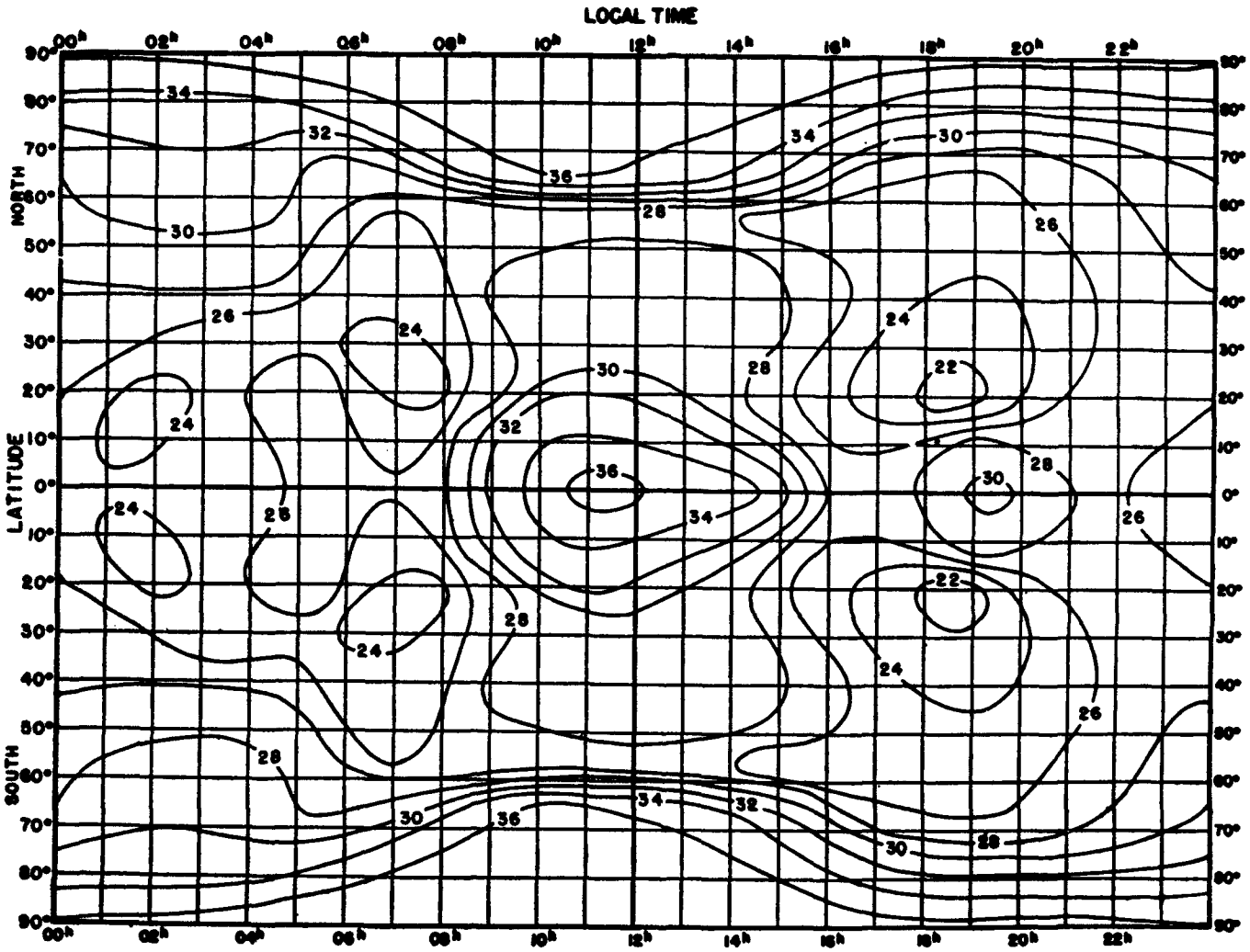


FIGURE 4.19. *World-wide distribution of minimum virtual height of F2 layer (h'F2) in tens of kilometers for equino[illegible]45.*

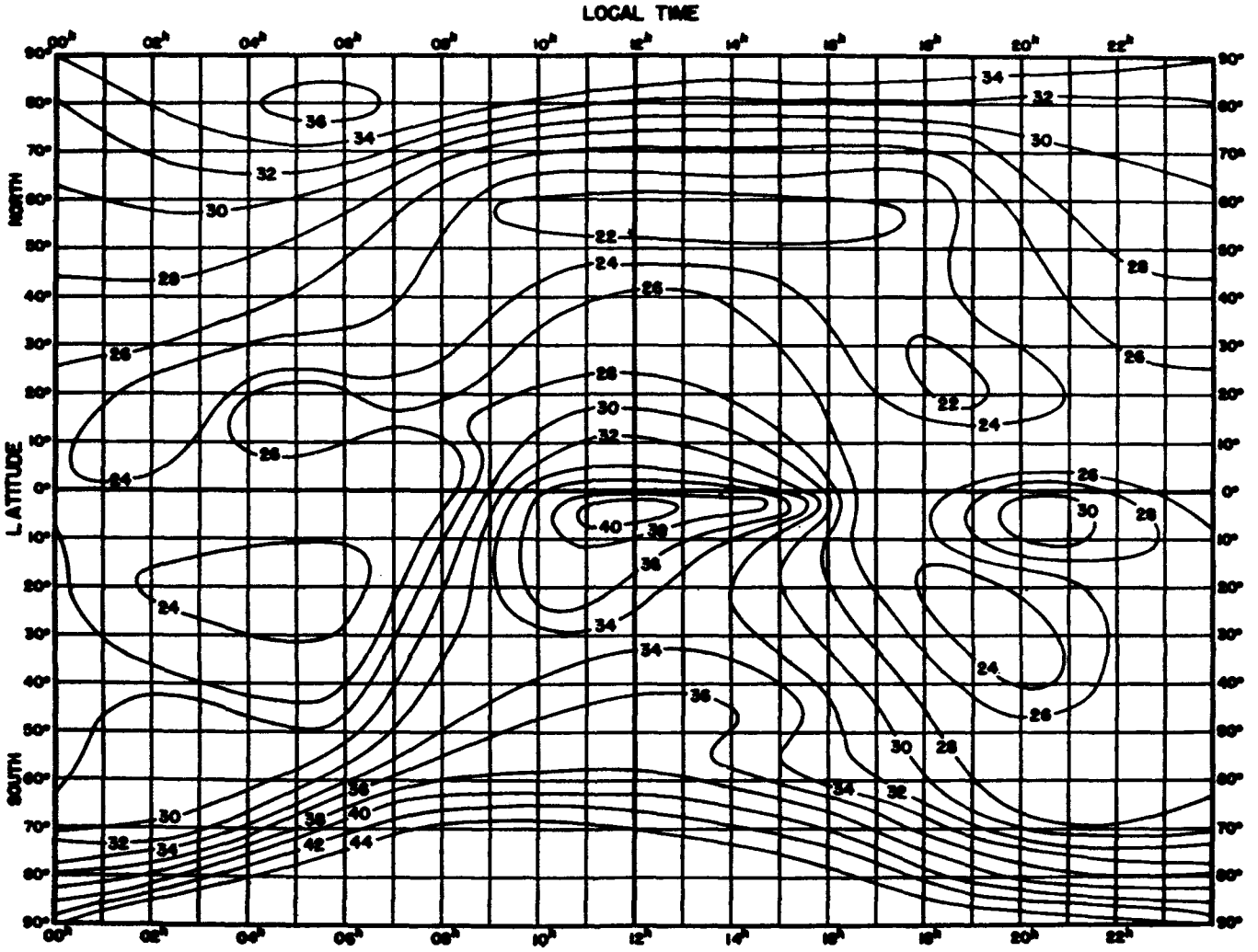


FIGURE 4.20. *World-wide distribution of minimum virtual height of F2 layer (h'F2) in tens of kilometers for December solstice, 1945.*

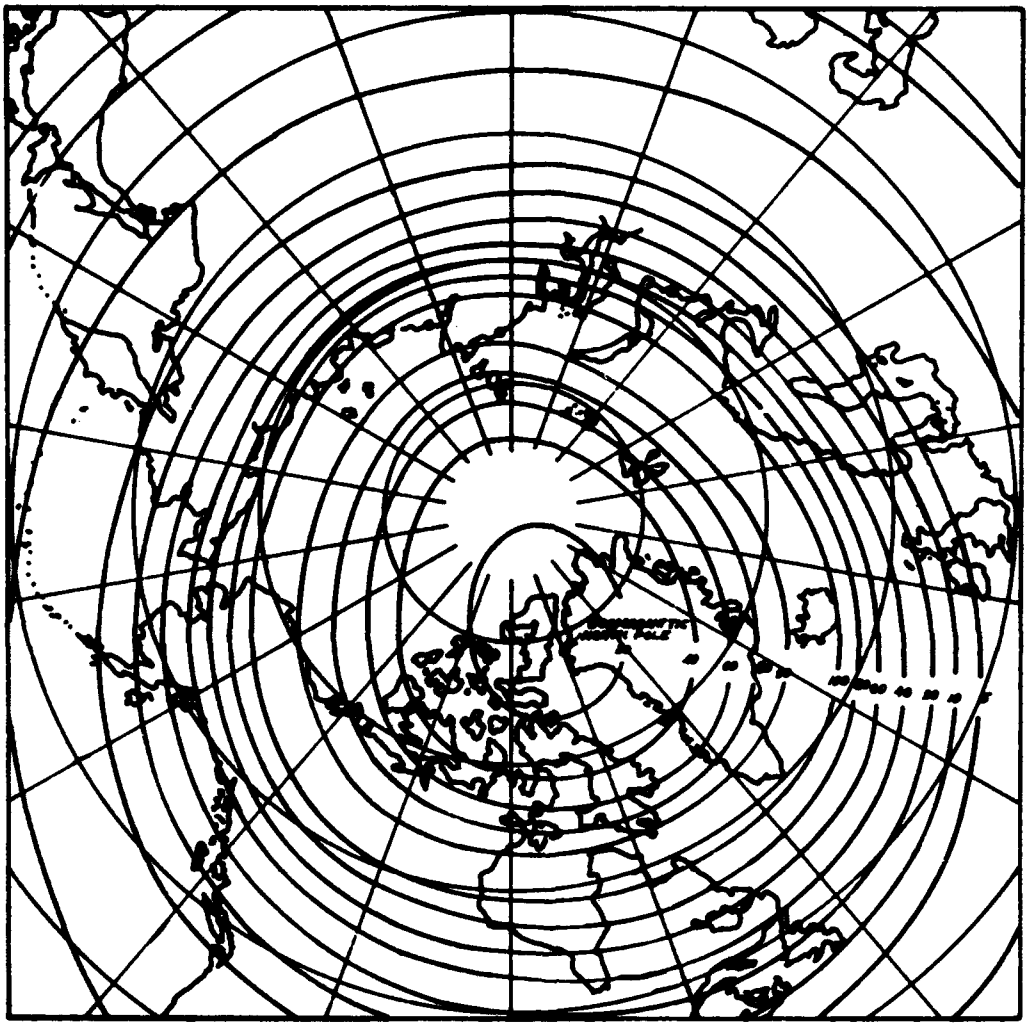


FIGURE 4.21. *Estimated percentage-frequency of days with occurrence of aurora, clear, dark, nights, high latitudes, northern hemisphere.*

CHAPTER 5

VARIATIONS OF THE IONOSPHERE

5.1. Diurnal Variations

a. E Layer

The most regular variations of ion density in the ionosphere are the diurnal variations. The characteristics of the diurnal variation of ion density in any layer (remembering that ion density is proportional to the square of the critical frequency) may be determined from examination of the charts in figures 4.9 to 4.12 and 4.15 to 4.18 in chapter 4. The distribution of critical frequencies for the E and $F1$ layers may be considered as fixed in the atmosphere with respect to the sun so that the earth turns under them. For this reason the E or $F1$ critical frequency at any place on the earth for any time of day may be ascertained by entering the charts pertaining to the appropriate season for the E or $F1$ layer with the latitude of the station and following across to the vertical line for a given time.

As the earth turns under these layers of ionization the critical frequency of the E layer rises from practically zero just before dawn to a maximum shortly after noon and then falls in a regular manner to practically zero shortly after sunset. Some vestige of the E layer appears to continue through the night. The diurnal variation is much the same at all latitudes, the height of the midday maximum depending upon the altitude to which the sun rises at noon.

There is distinct evidence of the production of ions in the E layer both before and after the sun is visible at the earth's surface. This does not seem to be ascribable to the fact that "sunrise" occurs earlier and "sunset" later at the altitude of the E layer, for direct rays of the sun, arriving tangentially, must pass through a great thickness of atmosphere to reach the portion of the atmosphere where the presunrise and post-sunset ionization occurs. It may be possible that scattering of sunlight from higher layers is responsible for this effect. This is one of the details which has not been cleared up in the theory; otherwise the behavior of the regular E layer is in accord with what one would expect from the theoretical considerations outlined in the preceding chapter.

b. D Region

Data regarding the D region are, in a sense, much more limited than those regarding the E layer. This is because reflections at vertical incidence from this layer are rarely obtained, due to excessive absorption at this altitude, where the collisional frequency is so great. Our information regarding the D region is derived almost entirely from observations of absorption of radio waves which are reflected by the other layers and which must pass through the D region on their round trip between the upper layers and the earth. All information about the D region indicates that its regular diurnal variation is substantially similar to that of the E layer in all particulars.

c. F1 Layer

The $F1$ layer too exhibits most of the characteristics of the E layer, but its characteristics are complicated by the fact that the $F1$ layer is formed at nearly the same height as the $F2$ layer and at times merges with it to such an extent that the two are scarcely distinguishable in the vertical-incidence records of ionospheric reflections. Throughout the night hours ionization persists in the upper part of the ionosphere, commonly called the F region. But around sunrise at most stations the nighttime layer splits into two separate layers, the lower of which is the $F1$ layer and the upper is the $F2$ layer. The critical frequency of the $F1$ layer increases to a maximum shortly after noon and then decreases, merging with the $F2$ layer around sunset. It is purely a matter of definition to state whether the ionization which persists throughout the night is the $F1$ or the $F2$ layer. For this reason no values of $F1$ critical frequencies are shown in figures 4.11 and 4.12 of chapter 4 for the night hours, it having been arbitrarily decided, for reasons which will appear later, to consider the persistent ionization in the F region during the night hours as a continuation of the $F2$ layer.

d. F2 Layer

Although the diurnal variation of the $F2$ layer exhibits a sufficient number of the characteristics of the E and $F1$ layers to assure that the $F2$ layer too is produced mainly by the effects of ultraviolet light from the sun, there are a number of conspicuous irregularities which cannot be explained by the simple theory. At most stations the maximum ion density occurs well after noon and declines rather gradually throughout the night. The rise in ion density in the morning is much more rapid for the $F2$ layer than for either the E or $F1$ layers. Some of these facts can be fairly well explained by the low rate of recombination of the electrons with positive ions in the $F2$ layer, due to its existence at great heights where collisions are few and

attachment of electrons to molecules is probably a process which may be ignored.

The vagaries of the $F2$ layer which do not fit into a simple theory include its lack of symmetry with respect to longitude. This is shown in figures 4.15 to 4.18 of the preceding chapter in which it was necessary to show different configurations of the $F2$-layer critical frequencies according to whether the sun was over the 69° W or the 111° E meridian. This indicates that the diurnal variation of the $F2$ critical frequency will be different at two stations in different longitudes although their latitudes may be the same. In contrast to our models for the E and $F1$ layers, in which the pattern of the critical frequency contours is regarded as fixed with respect to the sun while the earth turns under them, the pattern of the contours of the $F2$ layer changes as the earth rotates, the configuration shown in figure 4.15 gradually changing into that shown in figure 4.17 and vice versa as the earth rotates, while the contours in figure 4.16 are similarly worked into those of figure 4.18. Another striking characteristic of the $F2$ layer is that critical frequencies around noon are not highest in equatorial regions. The highest values occur around 20° from the geomagnetic equator, with a trough along the geomagnetic equator itself.

The geomagnetic equator appears to bear a special significance in the configuration of the contour lines; also, an adequate representation of the world-wide distribution of the $F2$ layer requires separate presentation according to whether the sun is over the 69° W or the 111° E meridian, which meridians contain respectively the northern and the southern geomagnetic poles of the earth. These facts indicate that the earth's magnetic field has an effect on the distribution of the ion density of the $F2$ layer, an effect which must be taken into account in predicting frequencies to be used for radio communication.

5.2. Seasonal Variations

Seasonal variations of the D, E, and $F1$ layers conform closely to what is to be expected from the theoretical considerations. As the apparent position of the sun moves with the changes in season from one hemisphere to the other the maximum of ionization in those layers shifts correspondingly. The lag in meteorological conditions behind the astronomical seasons appears to be, at the most, only weakly reflected in the ionosphere. There are a few irregularities in the observed data which have not yet been marshalled into a complete picture, but these are of minor importance.

The $F2$ layer exhibits marked departures from this idealized conception. In most localities the $F2$ ion density is greatest in winter and least in summer, quite the reverse of what would be expected from the simple theory. The distinct separation of the $F1$ and the $F2$ layers is also less conspicuous in winter than in summer, but the result of simple addition of the ions of these two layers is not great enough to account for the high ion densities observed in the $F2$ layer during winter. It has been hypothesized that the lower ion density in summer is due to accumulated heating and expansion of the layer which causes the ions to be dispersed over a greater height in summer, but it is difficult to understand why increased radiation, which is assumed to produce higher temperatures, does not also produce higher ionization as well. The close connection which exists between the distribution and changes in ion density of the $F2$ layer and the earth's magnetic field suggests that many of the phenomena may find explanation in the effect of the magnetic field and the electromotive forces produced by changes in the magnetic field on migration of the ions in accord with the relationships set forth in chapter 2.

5.3. Solar Phenomena

a. Solar Radiation

As has been pointed out, the main ionizing agent of the various layers of the ionosphere appears to be ultraviolet radiation from the sun. Our knowledge of the solar spectrum in the ultraviolet must be obtained by inference, because absorption of ultraviolet light by the upper atmosphere prevents its direct observation at the earth's surface, and even observations by rockets are subject to severe limitations.

The intensity of the spectrum in the ultraviolet may be calculated from the intensity of visible light of the sun by assuming that the sun radiates as a black body. These calculations lead to an intensity of radiation which seems too low to account for the intensity of the ionization which is observed. Furthermore, since various nuclear and atomic processes are taking place in the sun and in its atmosphere, there is reason to suspect that departures from black-body radiation in the high frequency end of the spectrum may be considerable. Many of our inferences regarding solar radiation in this part of the spectrum are drawn from observations of the ionosphere.

b. Sunspots

Among the most conspicuous of solar phenomena are the sunspots. Their exact nature is not known, but they appear to be vortices in the gases which comprise the visible surface of the sun. Although the sunspots themselves appear dark, the physical processes they produce (or which produce them) involve radiations of very high energy. Around the sunspots are areas which emit bright radiation from ionized calcium atoms, from hydrogen atoms, helium atoms, and numerous other

gases of the solar atmosphere. It is reasonable to assume that these visible emissions are associated with corresponding emissions in the ultraviolet; if so, the ultraviolet radiation coming from the sun must be far more variable than the visible radiation from the sun's whole disk, which radiates nearly as a black body, for the visible bright-line radiation is highly variable.

Associated with sunspots are strong magnetic fields, running up to several tenths of a weber per square meter, extending over areas of billions of square kilometers. In certain types of groups of sunspots, known as multipolar groups, the magnetic fields are extremely complicated and subject to rapid changes. Such groups contain many distinct spots and are designated as active groups. They are frequently large enough to be visible to the naked eye.

In addition to the electromagnetic radiations emanating from regions around spots, occasional outbursts of corpuscular radiation occur which reach the earth and cause striking ionospheric effects. There is evidence that these radiations sometimes attain cosmic ray energies. It has been suggested that the changing magnetic fields of active sunspot groups, acting on the same principle as betatrons, are responsible for these radiations. These sunspot regions are also sources of radiations in the radio spectrum which are known as solar radio noise.

The sunspots seem to be imbedded in the general body of the sun and are carried with it in its rotation, which has an average period of about 27 days. The life of an individual group may amount to several of these rotation periods, although its active stage may last through only one or two rotations. Thus sunspots exhibit a quasi-periodic variation which persists for several solar rotations as spot groups are turned toward or away from the earth.

c. The Sunspot Cycle

The most important variation of sunspots is the 11-year sunspot cycle. Shortly after this cycle was discovered in 1851 a method was devised for measuring the intensity of sunspot activity by use of the so-called Wolf "sunspot number". A number is formed for each day by multiplying the number of distinct visible sunspot groups by 10 and adding thereto the number of individual spots observable in the groups. A sunspot number is observed for each day and monthly and annual means formed therefrom. In the following discussions all mention of sunspot numbers refers to this Wolf sunspot number. The 11-year sunspot cycle is clearly revealed by this series of numbers, which extends back as a homogeneous series for nearly 100 years. For months at a time the visible surface of the sun may be devoid of spots so that the sunspot number is zero. This frequently occurs at times of sunspot minima. At other times even the mean annual number has been known to rise as high as 140, with daily values running into the hundreds. These conditions occur at the maximum of the sunspot cycle (see fig. 5.1).

The time from minimum to minimum of the sunspot cycle is variable, but it averages around 11 years. The height of the maximum and the depth of the minimum vary from cycle to cycle. Usually the time from minimum to maximum is less than the time from maximum to minimum. A tendency for high maxima to occur early in the cycle has been pointed out by numerous investigators.

This 11 year sunspot cycle is reflected in the variation of several geophysical phenomena involving the electric and magnetic condition of the earth. Records of terrestrial magnetism covering the last 100 years show that magnetic storms are more numerous at times of sunspot maxima than at times of sunspot minima (see fig. 5.2). Auroral activity, which is closely related to magnetic activity, also reflects the sunspot cycle. The diurnal variations in terrestrial magnetism, which involve other processes than those responsible for magnetic storms, vary in intensity with the sunspot cycle, increasing about 50 percent in intensity from minimum to maximum. As the propagation of radio waves at the frequencies discussed in this volume depends upon the electric state of the upper atmosphere, it is to be expected that effects of the sunspot cycle will appear in propagation phenomena.

5.4. Effects of the Sunspot Cycle on the Ionosphere

One important ionospheric effect associated with increases in solar activity is an increase in ion density of all layers, resulting in higher critical frequencies for the E, $F1$, and $F2$ layers and higher absorption in the D region. This permits the use of higher frequencies for communication over long distances at times of sunspot maxima than would be usable at times of sunspot minima. The increased absorption in the D region, which has the greatest effect on the lower frequencies, requires that higher frequencies be used, but the over-all effect is an improvement in propagation conditions during sunspot maxima as the critical frequencies are enhanced more than the absorption limits.

The changes in noon values of the critical frequencies of the E, $F1$, and $F2$ layers with sunspot numbers are shown in figures 5.3, 5.4, and 5.5. These data are based upon the records of the CRPL station near Washington, D. C., and the Carnegie Institution stations near Huancayo. Peru, and near Watheroo, Western Australia, These stations have the longest series of observa-

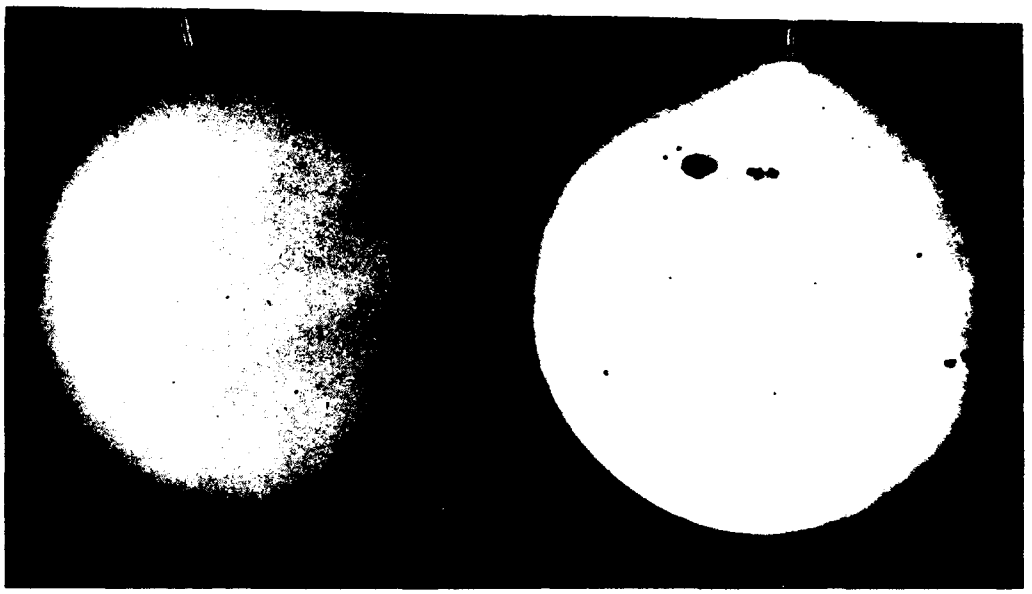


FIGURE 5.1. *Photographs of sun on December 14, 1943, during a period of low sunspot activity, and on February 4, 1946, during a period of high sunspot activity.*

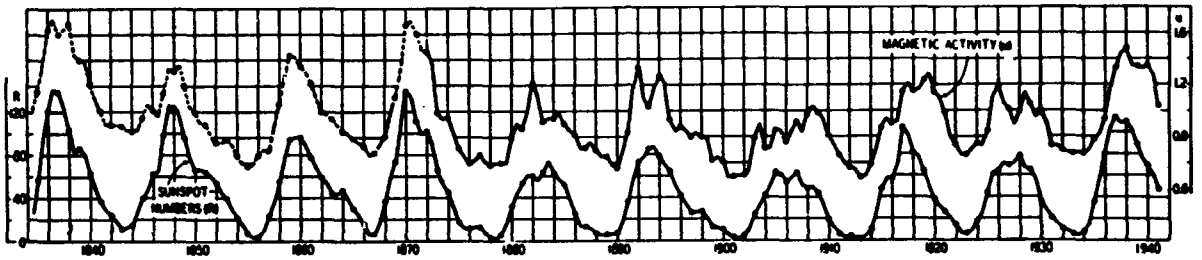


FIGURE 5.2. *Geomagnetic activity (U) and relative sunspot numbers (R), annual means, 1835–1941.*

tions available at the present time, covering more than one complete sunspot cycle. They also represent wide geographic areas, Washington being in the northern hemisphere, and Watheroo in the southern, with Huancayo near the equator. Observations from all these stations indicate the same characteristics for the various layers. The greatest variation with sunspot numbers occurs in the $F2$ layer, whereas the $F1$ and E layers exhibit nearly the same variation. Variations in the D region seem to parallel closely the variations in the $F1$ and E layers, but the observational data for that layer are not so complete and definite.

It is of interest to consider the implications of these variations. As may be seen from the figures, the relationship between critical frequencies and sunspot numbers is substantially linear for all three layers. From eq 2.26 and eq 4.3, there is a definite relationship between critical frequency, f_0, maximum ion density, N, and intensity of the rate of ion production, q, given by $q \propto f_0^4$. This follows from the relations

$$q = aN^2 \text{ and } f_0 = \sqrt{Ne^2/4\pi^2\kappa_0 m}$$

if we assume that electrons are removed from the ionosphere by recombination. In the $F2$ region recombination is probably the most important process. As f_0 for this layer increases by a factor of approximately two from sunspot number 0 to sunspot number 100, the ionizing radiation of the sun responsible for the formation of that layer must increase sixteen-fold for the same variation in sunspot numbers. The variation of visible solar radiation from sunspot minimum to sunspot maximum is almost imperceptible.

5.5. Short-Term Fluctuations

Superposed on these long-term trends of the ionosphere are short-term fluctuations, many of which appear to be related to concurrent fluctuations in the sunspot numbers. As the number of visible sunspots changes from day to day because of the solar rotation or the formation of new spots or disappearance of old ones on the visible part of the sun, absorption by the D region also changes (see fig. 5.6). Similar changes are observed in the E-layer critical frequency. These changes in D-region absorption and E-layer critical frequencies do not exhibit a one-to-one correspondence with the daily sunspot numbers, presumably because

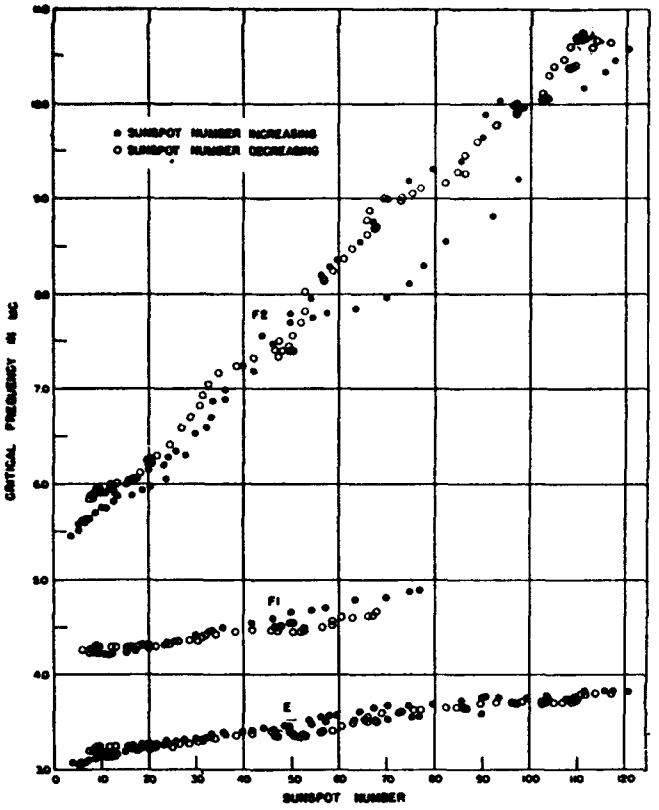


FIGURE 5.3. *Variation of 12-month running-average f°E, °F1, f°F2, at noon for Washington, D. C., with 12-month running-average sunspot number.*

the number of sunspots is not an exact measure of solar ultraviolet radiation. These variations exhibit wide geographic range; they are not effects which are observed at one station and not observed at others.

Fluctuations in $F2$ critical frequencies from day to day are greater than the fluctuations of the E, $F1$, and D layers. Including the effects of magnetic and ionospheric storms, to be discussed later in this chapter, the variability of $F2$ critical frequencies is such that, on the average, in one occasion out of 10 the critical frequency at a given station differs from the running average value by more than ± 15 percent. This corresponds to a range of ± 30 percent in ion density or, assuming the entire effect arises from changes in the rate of ion production, a range of ± 61 percent in that quantity. However, the $F2$-layer fluctuations are not in general world-wide in character. For illustration, the correlation between daily fluctuations in values of critical frequency at noon for stations in approximately the same longitude but in different latitudes decreases rapidly as the separation between stations increases, so that for separations of 1,000 km the correlation coefficient is only 0.5, whereas for separations of 3,000 km it is negligible (see fig. 5.7). If one station lies within the equatorial trough and the other lies outside of it, the correlation becomes negative, indicating that high values of critical frequency at one station are associated with low values at the other (see fig. 5.8). This is additional evidence that the $F2$ region is affected by the earth's magnetic field; it shows that the major day-to-day fluctuations of the $F2$ layer are not due to changes in ultraviolet radiation, as that would require that the world-wide fluctuations should be in unison.

Because of this variability of the $F2$ layer, precise predictions of its critical frequencies cannot be made for individual days, although seasonal and long-term trends, and geographic distribution may be accurately outlined in advance. It is necessary in selecting frequencies for long-distance communication to allow for these fluctuations.

5.6. Sporadic-E Layer

In addition to these rather regular characteristics of the ionosphere there are a number of anomalous effects which have important bearing on propagation phenomena. One of the most important of these is the occurrence of a sporadic-E

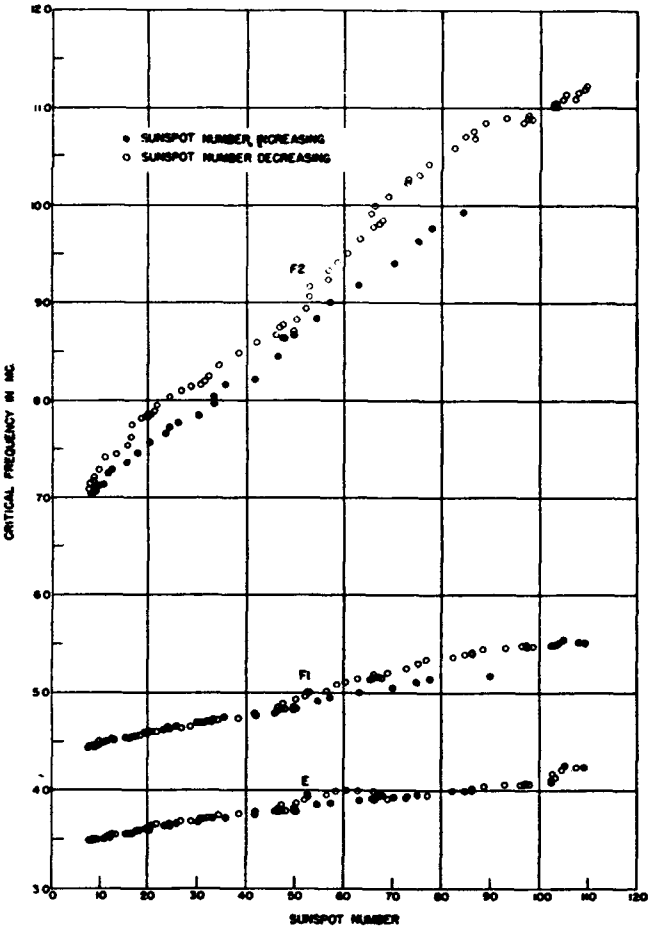


FIGURE 5.4. *Variation of 12-month running-average f°E, f°F1, f°F2, at noon for Huancayo, Peru, with 12-month running-average sunspot number.*

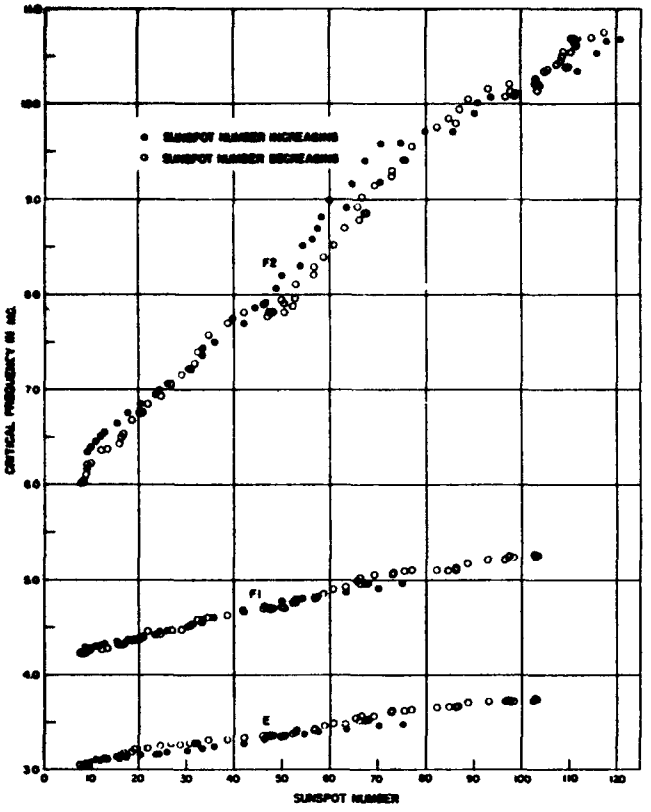


FIGURE 5.5. *Variation of 12-month running-average f°E, f°F1, f°F2, at noon for Watheroo, W. Australia, with 12-month running-average sunspot number*

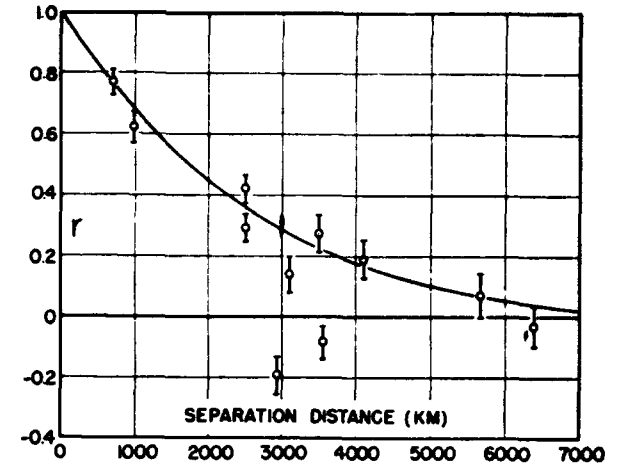


FIGURE 5.7. *Correlation coefficient for pairs of stations versus separation distance.*

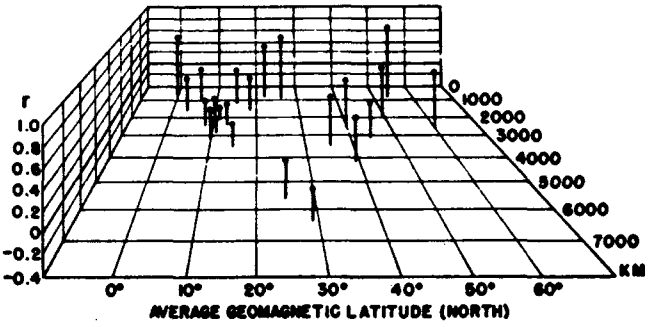


FIGURE 5.8. *Correlation coefficient between daily noon values of f°F2 at pairs of stations as a function of geomagnetic latitude and station separation.*

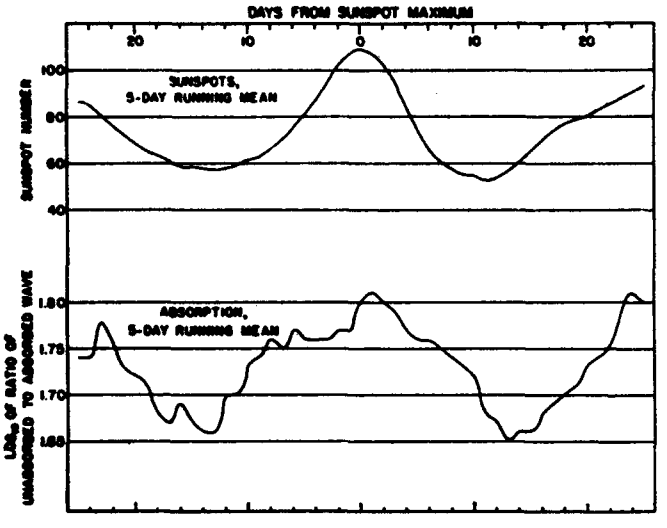


FIGURE 5.6. *Effect of sunspot variations on short-time variations in radio absorption at vertical incidence on 2,061 kc/sec. (superposition of 13 epochs selected for sunspot maximum 1945–46).*

layer, usually designated as *Es*, which cannot be accounted for by the processes which give rise to the normal *E*-layer ionization.

Sporadic-*E* layer phenomena are manifested on multifrequency ionospheric records by the appearance of reflections from the *E* region at frequencies above the normal *E*-layer critical frequency. Usually these reflections come from a slightly greater height than the normal *E* reflections, since the normal critical frequency can be measured sometimes while *Es* prevails. Sometimes *Es* consists of an extremely efficient radiating surface, which is capable of reflecting so much of the energy radiated from the transmitter, even at frequencies of 10 to 15 Mc, that reflections from the other layers of the ionosphere are completely blanketed out. At other times *Es* may be so "thin" that although vertical-incidence reflections from it occur well above the normal *E* critical frequency, reflections from the higher layers may still be "seen" through it. *Es* reflecting low frequencies is observed more often than that reflecting high frequencies. The data obtained so far fit fairly well the hypothesis that logarithm of percentage of time of occurrence of *Es* capable of reflecting any frequency is inversely

935333 O - 51 - 5

proportional to the frequency. Records of reflections from *Es* do not exhibit a cuspate form at their upper limits such as are associated with critical frequencies of the regular ionospheric layers.

This sporadic-*E* layer may occur during the day or during the night. The exact diurnal, seasonal, and long-term distributions of its occurrence are difficult to ascertain because of absorption during daylight hours by the *D* region and variable sensitivity of the measuring equipment at different frequencies. Its occurrence is frequent enough that from 25 to 50 percent of the time long-distance propagation at frequencies up to 15 Mc in middle latitudes is rendered possible by *Es*. Occurrence of *Es* is not simultaneous at all stations.

Tropical stations, in general, exhibit less *Es* than stations in higher latitudes. At tropical stations during daylight hours, however, reflections from the *E* layer, at frequencies above the regular critical frequency of that layer, are of common occurrence due to sharp boundary reflections, but reflections from the upper layers are regularly "seen" through them. In the auroral regions reflections from the *E* region which seem to be associated with the appearance of auroras are frequently obtained.

The nature and cause of the sporadic-*E* layer are not yet understood. There is no evidence that the sporadic appearance of abnormal reflections from the *E* layer in middle latitudes is the same phenomenon as the prevalent daytime sharp boundary reflections above the normal *E* critical frequency in the tropics, or the *E* reflections concurrent with auroral manifestations in high latitudes. It is fairly clear that all abnormal reflections from the *E* region cannot be due to the same cause. In certain cases it has been shown that meteors produce ionization in the *E* region which is capable of reflecting radio waves, even at very high frequencies, in fact, some investigators believe that meteors are the predominant cause of sporadic-*E* layer formation.

5.7. Solar Flare Disturbances

At times communication on high frequencies by sky-wave propagation over the daylight hemi-

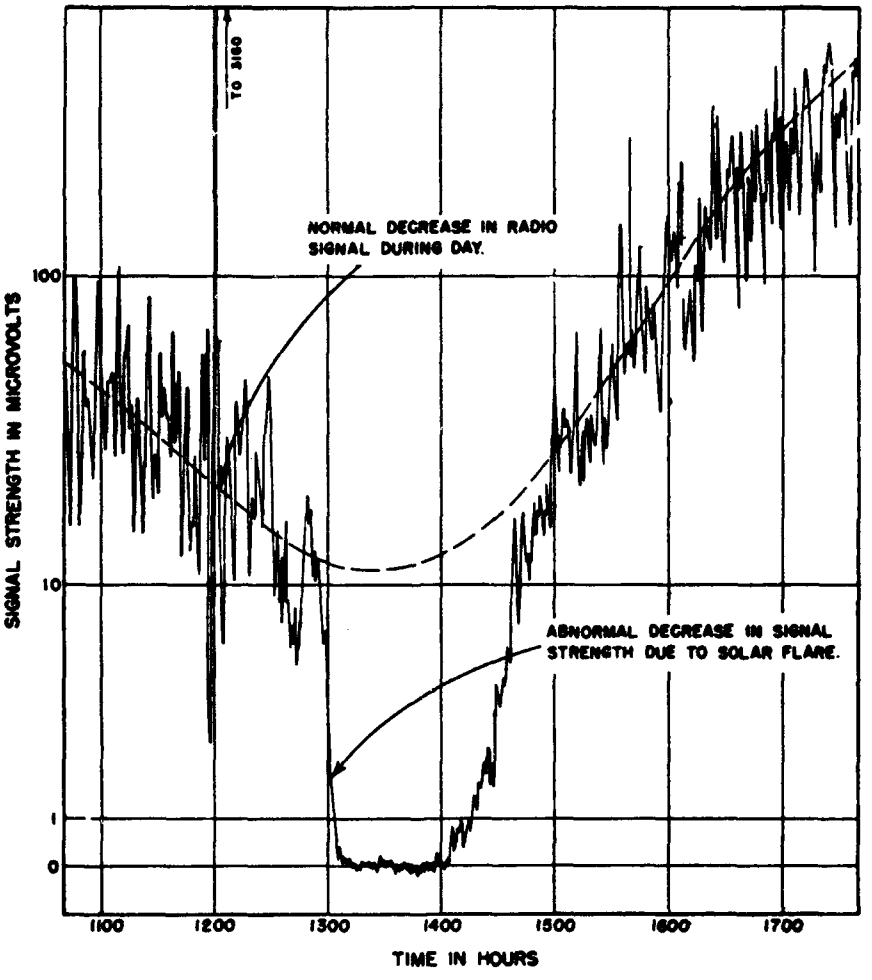


FIGURE 5.9. *Effect of solar flare on radio signals received from W8XAL, Mason, Ohio, as measured at Sterling, Va.*

sphere of the earth is rendered impossible by abnormally great absorption. This condition may prevail for a length of time from a few minutes to several hours. Onset of this condition of absorption is usually very sudden; recovery is more gradual (see fig. 5.9). As revealed by the multifrequency records this anomalous absorption appears as a complete blanking out of reflections from all layers of the ionosphere; no reflections are returned from the absorbing region either. As reflections from all layers are affected, it is clear that the absorption must occur in a region between the earth and these layers, which means that this abnormal absorption is probably a *D*-region phenomenon. When the absorbing region disappears by recombination of the ions or electrons, reflections from the upper layers appear again and indicate that no change has occurred in them either with respect to virtual height or ion density, except such as is due to normal variation. Occasionally the absorbing layer is sufficiently transparent for the upper layers to be "seen" through it. At such times it is observed that they are unaffected by the agency producing the absorption.

Examination of the sun at the times of occurrence of these effects revealed that in all cases where suitable solar data were available the appearance of this anomalous absorption was simultaneous with the appearance of an outburst of bright light in the solar chromosphere. These bright chromospheric eruptions are usually seen best in the alpha line of hydrogen, but may be seen in the light from other chromospheric elements also. In clearly defined cases the onset of absorption is coincident with the onset of the eruption and its duration is the same as that of the eruption, indicating that the disturbing agency is propagated from the sun to the earth with the velocity of light. These solar flare disturbances, often referred to as sudden ionosphere disturbances (SID), are the only ionospheric phenomena discovered so far which have been directly associated with individual events observed on the sun; relationships between ion densities and sunspots, both long-term and short-term trends, are statistical effects which are revealed by average characteristics of both phenomena, not by individual events.

5.8. Ionospheric Storms

a. Magnetic Storms

The most striking of anomalous ionospheric phenomena are ionospheric storms. They are in no way associated with the meteorological storms which occur in the lower part of the troposphere. They derive their name from the similarly erratic character of both phenomena. They are one aspect of the severe electromagnetic disturbances to which the earth is frequently subjected and which have been known for a long time as magnetic storms. Because of our extensive knowledge of magnetic storms derived from observations of the earth's magnetic phenomena, extending back over more than a century, the general characteristics of magnetic storms will be discussed first and then their associated ionospheric characteristics will be brought out.

Magnetic storms are disturbances of the earth's magnetic field lasting for a length of time from several hours to several days during which the magnetic field fluctuates over much wider limits than it ordinarily does. These fluctuations are not only marked by their magnitude, but also by their irregularity. During magnetic storms the field may change by several percent in middle latitudes and the rate of change may be as great as one percent per minute, as contrasted with quiet conditions when the field changes are only a few tenths of 1 percent in several hours (see fig. 5.10). In spite of the irregularity of the field changes, a number of systematic characteristics have been isolated. The field changes associated with a typical magnetic storm may be described, but a particular magnetic storm may differ from the typical storm in many ways.

Many magnetic storms begin with a sudden commencement which is simultaneous all over the earth to within a few seconds. During the next few hours the horizontal component of the earth's magnetism remains above normal. This is called the initial or positive phase of the storm. Then the horizontal force drops rapidly to well below normal values. This is called the main, or negative phase. Recovery continues for the next few days in what is known as the post-perturbation period. The most violent and erratic changes occur during the transition from the positive to the negative phase. The field changes are such as might be due roughly to a ring current encircling the earth at a distance of several thousand kilometers, flowing first in one direction (positive phase) and then in the other (negative phase).

In the auroral zones the changes are most erratic. These auroral zones are belts about 20° in radius, encircling the geomagnetic north and south poles, which are in 78.5° N. latitude, 69° W. longitude, and 78.5° S. latitude, 111° E. longitude. Here the effects are such as would be caused by intense east-west currents flowing along the zones associated with disperse return currents in lower latitudes. There is conclusive evidence that the currents flowing along the auroral zone are confined to the ionosphere, their most probable heights being about 100 km, which places the currents in the *E* region.

The auroral zones are so-called because aurora are most frequently seen there. In the zones aurora may be seen on every clear night, but as one moves from the auroral zones either toward the geomagnetic poles or toward the geomagnetic

equator, aurora are seen less frequently. During intense magnetic storms the auroral zones shift to lower latitudes and with them shift the characteristic magnetic disturbances. During unusually stormy conditions aurora have been seen even at Samoa.

b. Relation to Sunspots

A close relationship, in the statistical sense, between magnetic storms and sunspots has been established. During and slightly after the maximum of the sunspot cycle magnetic storms are most frequent and tend to be most violent (see fig. 5.2). A tendency of magnetic storms to recur at 27-day intervals, that is, the rotation period of the sun, is also established, but like recurrences of sunspots, a series of 27-day recurrences of magnetic storms does not continue indefinitely. It has also been established that magnetic storms occur, in about 50 percent of the cases, within a day to two of the time when an active sunspot group passes the central meridian of the sun. These principles, the 27-day recurrence, frequency of magnetic storms at and after the maximum of the sunspot cycle, and association of magnetic storms with central meridian passage of sunspots furnish a basis for prediction of magnetic storms. During the later part of a sunspot cycle the 27-day recurrence tendency is pronounced, and then affords the most reliable single means for predicting magnetic storms. During the early part of the cycle the recurrence tendency is much less marked so that

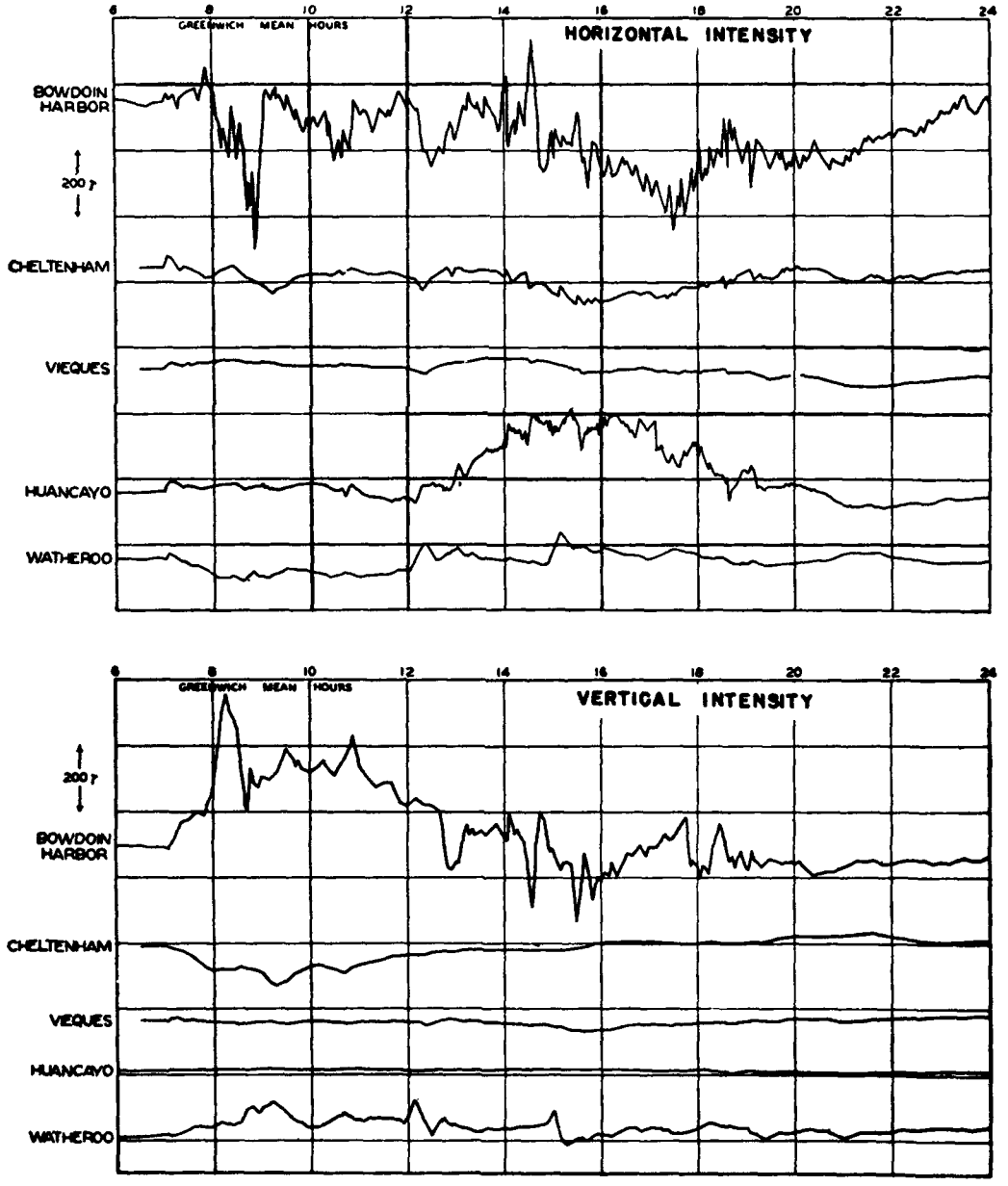


FIGURE 5.10. *Magnetic storm of March 14, 1922, as recorded at different stations.*

considerable reliance must be placed on predictions from visible solar phenomena.

c. Causes

A better understanding of the vagaries of these electromagnetic disturbances may be derived from examination of the theory of their cause. The most satisfactory theory of these phenomena attributes them to the effects of charged corpuscular radiation from the sun. These corpuscles are shot out in individual blasts from active regions on the sun which are usually associated with active sunspot groups. The bursts of corpuscles do not always leave the sun perpendicular to its surface so that whether or not they engulf the earth is largely a matter of chance. As they approach the earth they are deflected by the magnetic field in accordance with the principles developed in chapter 2; some come down the lines of force in polar regions, producing the aurora, while others encircle the earth to produce the ring current. Reaching the denser parts of the upper atmosphere, the polar corpuscles produce ionization at all levels down to about 80 km.

d. Depression of Critical Frequencies

The disturbances in the ionosphere that are associated with these electromagnetic storms seriously impair communications over long distances. The most conspicuous effects are lowering of the critical frequencies and increase in absorption. Thus the band of usable frequencies, those low enough not to penetrate the ionosphere but great enough not to be excessively absorbed, is considerably narrowed or, during the more intense storms, wiped out altogether. Communication during these disturbances, when possible, is considerably impaired by various types of fading.

During the early part of a storm, corresponding to the positive phase of the magnetic disturbance, ionospheric conditions are usually not greatly affected. For this reason it is frequently possible to warn radio operators several hours in advance when disturbances are likely to occur, although the intensity of the disturbance cannot be foretold. As the disturbance develops, the ion density in the F region drops, with attendant lowering of the critical frequencies, the main effect being associated with the $F2$ layer. During daylight hours depression of the ion density is not as marked as during night hours because then the continual flux of ultraviolet light from the sun serves to maintain the ionization against the adverse effect of the storm.

e. Absorption

When violent changes in the magnetic field occur, even during the early stages of the storm, the ionosphere is most likely to be affected. The appearance of auroral displays (visible, of course, only during the dark hours) in lower latitudes produces high absorption due to the creation of ionization at low levels where the ions engage in frequent collisions with the neutral molecules of the air. This is the absorption effect which is usually observed during ionospheric storms. Shifting of the auroral activity to lower latitudes during intense magnetic storms explains in part the impairment of radio communication during such times. Communication over paths where the reflection takes place in very low latitudes is less likely to be affected. For this reason communication between North America and Europe which, foll[illegible]t-circle path, is reflected by the iono[illegible]igh latitudes, is frequently disrupted by electromagnetic storms. Communication is frequently maintained at such times by relaying the traffic through low latitude stations, which are less subject to disturbance.

The E layer is not appreciably affected by the storms except when the intensity of disturbance is very great. However, absorption at low levels during storms often impairs propagation conditions to such an extent that communication by modes involving E-layer reflections is not satisfactory.

f. Changes in Layer Heights and Anomalous Reflections

In addition to changes in ionic density of the layers during ionospheric disturbances great changes in layer heights occur, the $F2$ layer often rising to heights well above the normal ones. Abnormal stratification of the region also occurs so that overlapping reflections are returned from the layer and the otherwise clearly defined critical frequencies do not appear.

In the auroral zones, ionization capable of reflecting radio waves is often produced in the E region during disturbances although the region at the time may not be illuminated by ultraviolet light. It is presumed that such ionization is due to corpuscular radiation. Similar effects have also been noted at the geomagnetic equator, but there is reason to doubt that the two effects are due to the same mechanism.

The extensive knowledge of electromagnetic storms, gained from the lengthy series of observations of the earth's magnetism, may be applied directly to the problems of radio propagation. Thus, had the sunspot-cycle relationship and 27-day recurrence of magnetic storms not been established, it is doubtful if we would have confidence in the reality of the corresponding ionospheric effect, owing to the limited interval of time over which we have comprehensive ionospheric observations.

5.9. References

J. H. Dellinger, Sudden ionospheric disturbances, Terr. Mag. & Atmos. Elec. **42**, 49 (March 1937).

N. E. Bradbury, A. G. McNish, J. H. Dellinger, and R. S. Richardson, Astrophysical problems of the ionosphere, J. Applied Phys. **8**, 709 (1937).

N. Smith, T. R. Gilliland, and S. S. Kirby, Trends of characteristics of ionosphere for half a sunspot cycle, J. Research NBS **21**, 835 (1938) RP1159.

L. V. Berkner, H. W. Wells, and S. L. Seaton, Ionospheric effects associated with magnetic disturbances, Terr. Mag. & Atmos. Elec. **44**, 283 (Sept. 1939).

J. A. Fleming, Terrestrial magnetism and electricity (McGraw-Hill Book Co., New York, N. Y., 1939).

L. V. Berkner, and S. L. Seaton, Ionospheric changes associated with the magnetic storm of March 24, 1940, Terr. Mag. & Atmos. Elec. **45**, 393 (Dec. 1940).

S. Chapman, and J. Bartels, Geomagnetism, 2 volumes (Clarendon Press, Oxford, 1940).

CHAPTER 6

MAXIMUM USABLE FREQUENCIES

6.1. Introduction

Early experimenters in high-frequency radio transmission noted that for any fixed distance of transmission there was an upper limit of frequency which could just be transmitted. This upper-limit frequency was in general greater for greater distances up to 4,000 km, was greater in the daytime than at night, and was greater on a winter day than on a summer day. The existence of this upper-limit frequency depends on the fact that the ionization in the ionosphere will reflect only waves of frequencies less than a certain critical value; this value is called the "maximum usable frequency" (abbreviated muf). At frequencies above the maximum usable frequency for a given distance the wave is said to "skip."

The maximum usable frequency for high-frequency transmission via the ionosphere over any given path at any instant is an important quantity. If the operating frequency is above the maximum usable frequency the wave skips, since, as will be seen below, it then will not be reflected from the ionosphere. As the operating frequency is decreased below the maximum usable frequency in the daytime, it becomes increasingly attenuated, since as pointed out in chapter 7, in the high-frequency range the lower the frequency the greater the ionospheric absorption. Hence it is usually desirable for transmission to occur on a frequency as near to the maximum usable frequency over the path at any instant as is feasible under conditions prevailing in practice. For this reason a technic of prediction whereby the proper operating frequency may be chosen with regard to the maximum usable frequency is highly desirable.

There exists a direct relationship between the maximum usable frequency and the condition of the ionosphere. Knowledge of this relationship makes it possible to predict median values of maximum usable frequency for propagation over any path at any time of day in any future month. It is the purpose of this chapter to describe this relationship, the method of predicting basic world charts of maximum usable frequency, and the method of problem solution by using world charts.

6.2. Obtaining Maximum Usable Frequencies From Vertical-Incidence Sweeps

a. Transmission-Curve Technics

In an ionized layer there exists a "maximum usable frequency" for reflection at oblique incidence which corresponds to the critical frequency at vertical incidence. Whereas the critical frequency is, for vertical-incidence reflection, the frequency below which waves will be reflected by a given layer of the ionosphere and above which they will penetrate that layer, the maximum usable frequency for a given great-circle distance along the earth is the highest frequency of waves that will be reflected from a given layer of the ionosphere and return to the earth at that distance from the transmitter.

In the ionosphere the refractive index η' for a radio wave of frequency f' has a value less than unity. If we neglect the effect of the earth's magnetic field, η' as a function of the transmitted frequency is given by

$$\eta' = \sqrt{1 - \frac{f_N^2}{f'^2}}, \tag{6.1}$$

where

$$f_N^2 = \frac{Ne^2}{4\pi^2 k_0 m}$$

f' = wave frequency or operating frequency (i. e., the frequency of the oblique-incidence wave)
N = equivalent electron density per cubic meter
e = electronic charge in coulombs
m = electronic mass in kilograms
k_0 = dielectric constant of free space $(1/36\pi)10^{-9}$.

Considering a plane ionosphere and plane earth, as illustrated in figure 6.1, with a wave frequency f' incident on the ionosphere at an angle ϕ_0, it is evident that the wave will become horizontal at a true height at which the refractive index for the given wave is $\eta' = \sin \phi_0$.

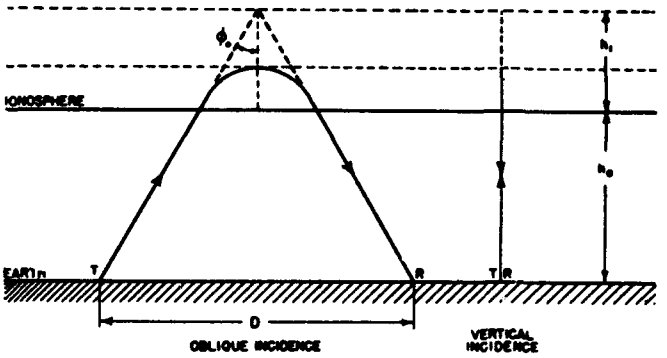


FIGURE 6.1. *Illustration of equivalence theorem for a flat earth and flat ionosphere.*

Now at that same height there will be a frequency f lower than f', which will just be reflected at vertical incidence. As for this case the angle of incidence $\phi_0=0°$, the refractive index η for the equivalent vertical-incidence frequency f at this height is equal to 0 or in eq 6.1

$$\sqrt{1-\frac{f_N^2}{f^2}}=0$$

or

$$f_N=f.$$

Then substituting in eq 6.1 and solving for η'

$$\eta'=\sqrt{1-\frac{f^2}{f'^2}}=\sin\phi_0,$$

from which

$$f'=f\sec\phi_0. \tag{6.2}$$

This equation, the well-known "secant law", states that the frequency which will be reflected from a plane ionosphere at a given height at oblique incidence is equal to the secant of the angle of incidence times the frequency which will be reflected at vertical incidence at the same true height.

Sweep records ($h'f$ records) which are obtained by pulse technic, using multifrequency equipment, give the virtual heights at which vertical-incidence frequencies throughout a range are reflected. Although true heights cannot be obtained directly from these records by a simple method, there fortunately exists an important relationship that makes the records of virtual height versus frequency directly usable.

This important relationship is the equivalence theorem. For a flat earth and a flat ionosphere, without consideration of the earth's magnetic field, this relationship states that the virtual height of reflection of a wave propagated at oblique incidence over a path, i.e., the height of the equivalent triangular path, is equal to the virtual height of reflection for the wave of the equivalent vertical-incidence frequency propagated vertically through the ionosphere. The relationship is illustrated in figure 6.1, where h_0+h_1 is both the height of the equivalent triangular path and the virtual height of reflection of the vertical-incidence wave. The equivalent vertical-incidence frequency, for a wave propagated obliquely through the ionosphere, is the frequency of the wave reflected, at vertical incidence, at the same actual height as is the oblique-incidence wave. The equivalent vertical-incidence frequency and the oblique-incidence wave frequency are related by the secant law.

The importance of this relationship is that, at least for the case of the flat ionosphere, it is possible to apply the secant law directly to the $h'f$ records. This follows as a result of the equivalence theorem, since a frequency f picked off the ionosphere sweeps at any virtual height h' is the equivalent vertical-incidence frequency corresponding to an actual transmitted frequency f' propagated at oblique incidence, which enters the ionosphere at such an angle ϕ_0 that the height of the equivalent triangular path is equal to the above-mentioned virtual height h'

Then for any frequency f scaled off the record, still assuming a flat earth and a flat ionosphere, eq 6.2, $f'=f\sec\phi_0$, still holds.

Considering any distance of propagation D, the angle of incidence ϕ_0 is, by geometry, a simple function of the height of the equivalent triangular path. For the simple case of a flat earth, for example, shown in figure 6.1

$$\phi_0=\tan^{-1}\frac{D}{2h}. \tag{6.3}$$

Taking account of the earth's curvature, ϕ_0 is given by

$$\phi_0=\tan^{-1}\frac{\sin\frac{1}{2}\theta}{1+\frac{h}{R}-\cos\frac{1}{2}\theta}, \tag{6.4}$$

where R is the radius of the earth, and θ is the angle at the center of the earth subtended by the transmission path.

Assuming a flat earth and a flat ionosphere, then, if we take the $h'f$ curve for the $F2$-layer ordinary wave, such as the one illustrated in figure 3.8 (chapter 3), and scale a frequency f at a low virtual height far down on the trace and then obtain sec ϕ_0 for that height from eq 6.3, we can substitute in eq 6.2 and obtain a value of f'.

As we scale higher values of frequency off the graph and repeat the process the increased values of f tend to make f' larger. However, as the frequency f increases, the virtual height of reflection h' becomes larger, and so ϕ_0 and sec ϕ_0 become smaller, which tends to bring the value of f' down. At first the frequency increase is the predominant factor, so that f' increases. Eventually, however, as the frequency f approaches its critical value the virtual height rises sharply, and sec ϕ_0 decreases so rapidly that f' diminishes. It is evident that somewhere between there is a maximum value of f', which is termed the maximum usable frequency for the path.

Note that in this development the ordinary-wave trace on the record was scaled. The detailed development involving the earth's magnetic field requires this; however, the consideration of the extraordinary wave results in but slight difference in the muf, except at short distances.

From the above considerations it can be said in general that in order to calculate the muf for any given transmission path and state of the ionosphere, it is necessary to satisfy two conditions—(1) As the virtual height of reflection varies with

frequency at vertical incidence, the height of reflection of an oblique-incidence transmission will vary both with frequency and distance, and thus the frequency and the angle of incidence of the waves on the layer must be such that they will be reflected at the proper height to reach the receiver when they are reflected at the given angle. In other words, the frequency $f' = f\sec\phi_0$ and the height h' must satisfy simultaneously the $h'f$ curve and the geometrical relationship between D, h', and ϕ_0, given by eq 6.3, noting that by the equivalence theorem $h' = h$. (2) As the maximum usable frequency is to be found, it is necessary to determine the maximum value of f' that will satisfy the first condition.

The maximum usable frequency may be calculated for any given $h'f$ curve, as follows: A coordinate scale, coinciding with that on which the original $h'f$ curves are plotted, is used to plot a family of curves of h' against f, where f is obtained from eq 6.2 and 6.3 for values of sec ϕ_0 corresponding to different values of h' for a fixed distance D. These graphs, known as "transmission curves," are made parametric in f', the wave frequency. If a transparent overlay is made to scale with this family of transmission curves drawn on it, and if this overlay is placed over the $h'f$ graph, the intersections of the curves with the $h'f$ graph will give, for the fixed distance, values of the wave frequency reflected at oblique incidence from the virtual heights of the intersection, and also the values of f corresponding to the reflection heights.

Figure 6.2 shows such a family of transmission curves superimposed on an ionosphere sweep. It may be seen that most of the curves intersect the F2-layer portion of the $h'f$ graphs in two places, corresponding to a "low wave" and a "high wave" at the indicated wave frequency of the parametric curve. In this case the low wave corresponding to the intersection of the 14-Mc curve at a will be attenuated by the deviative absorption of the $F1$ layer (chapter 2, sec. 2.6), and the high wave corresponding to the intersection of that curve at a' will be attenuated by the deviative absorption of the $F2$ layer. At a higher wave frequency, 18 Mc, the low wave corresponding to the intersection of the curve at b should be strong, other things being considered, whereas the high wave corresponding to b' should still be somewhat attenuated.

As the frequency f' of the parametric curves increases, the two points of contact come closer together, culminating at a tangent contact at one point c, giving a maximum usable frequency for this particular ionosphere sweep equal to 20 Mc.

At this point, where the low and high waves merge ("focusing effect"), the received wave should be strongest. This partially accounts for the observations of radio operators that a station is received best just before skipping.

At higher values of f' (higher operating frequencies) there is no value of f sec ϕ_0 that will become equal to f', and hence no height from which the wave will be returned to earth. Hence it is said to skip at the distance concerned. The distance for which a given frequency is the maximum

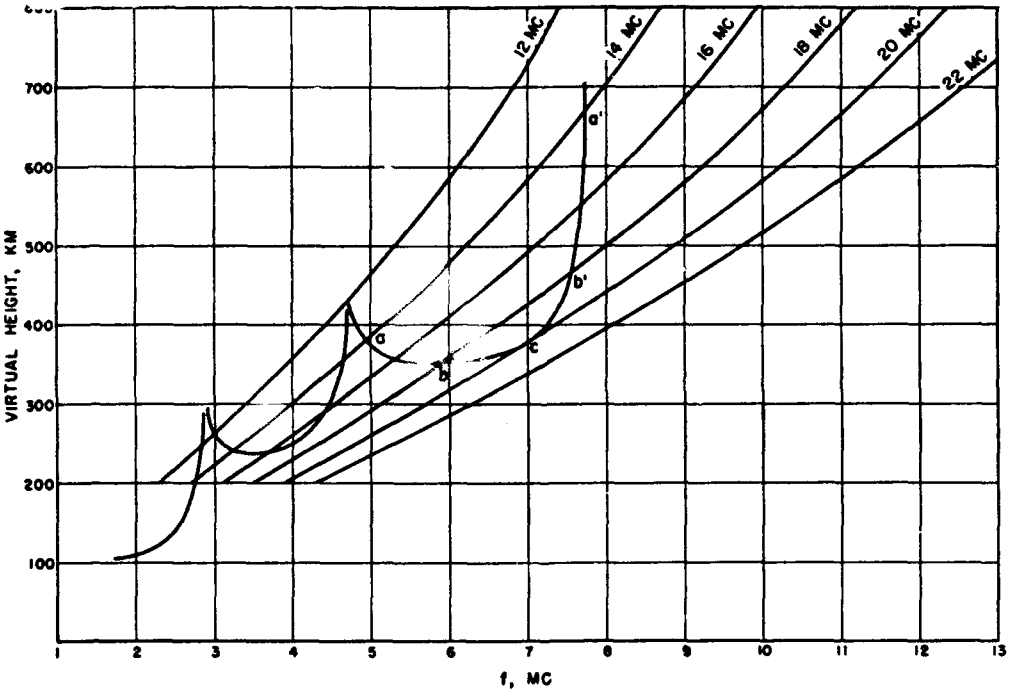


FIGURE 6.2. *Family of transmission curves parametric in frequency for a fixed distance of 2,000 km superimposed on an h'f graph.*

A flat earth and flat ionosphere assumed.

usable frequency is called the "skip distance" for that frequency.

Figure 6.3 illustrates the paths through the ionosphere traversed by the waves of figure 6.2 as frequency is increased. The low and the high paths of the 14-Mc wave being reflected from the true heights corresponding to the virtual heights of reflection of points a and a' are indicated as dashed lines. The paths for the points b and b' at a higher frequency, 18 Mc, are indicated as a pair of dot-dash lines and the case for the maximum usable frequency, 20 Mc, reflected at c is indicated as a single solid line. A still higher frequency, 22 Mc, penetrating the ionosphere and thus skipping is illustrated as a dotted line.

The effect of varying the distance on a fixed frequency is made evident by drawing a family of curves of h' against f, for a fixed wave-frequency parametric in distance in a similar manner to the way in which the curves parametric in wave frequency for a fixed distance were drawn. Figure 6.4 shows such a family for 12 Mc superimposed on an ionosphere sweep, and here again it may be seen that most of the curves intersect the $h'f$ graph in two places, corresponding to a low wave and a high wave at the indicated distance of the parametric curve. Considering the curve for 1,600 km, the low wave, corresponding to its intersection with the $h'f$ graph at a, will be attenuated by the deviative absorption of the $F1$ layer, and the high wave, corresponding to intersection of that curve at a', will be attenuated by the deviative absorption of the $F2$ layer. At a shorter distance, 1,200 km, the low wave, corresponding to the intersection of the curve at b, should be strong, other things being considered, whereas the high wave, corresponding to b', should still be somewhat attenuated.

As the value of distance of the parametric curves decreases, the two points of contact come closer together, culminating at a tangent contact at one point, c, giving a skip distance of 1,060 km for the particular wave frequency for the particular ionosphere sweep.

At this point the two waves focus, and the received wave should be the strongest; at still shorter distances there is no value of $f \sec \phi_0$ that will become equal to f', so that skip will occur.

Figure 6.5 is analogous to figure 6.3, except that the frequency is held constant and the distance varied. The distances are chosen so that the reflection point is constant. The low and the high paths of the wave being reflected over a path 1,600 km long from the true heights corresponding to the virtual heights of reflection a and a' are indicated as dashed lines. The paths for the points b and b' at a shorter distance, 1,200 km, are indicated as a pair of dot-dash lines, and the case for the maximum usable frequency reflected at c

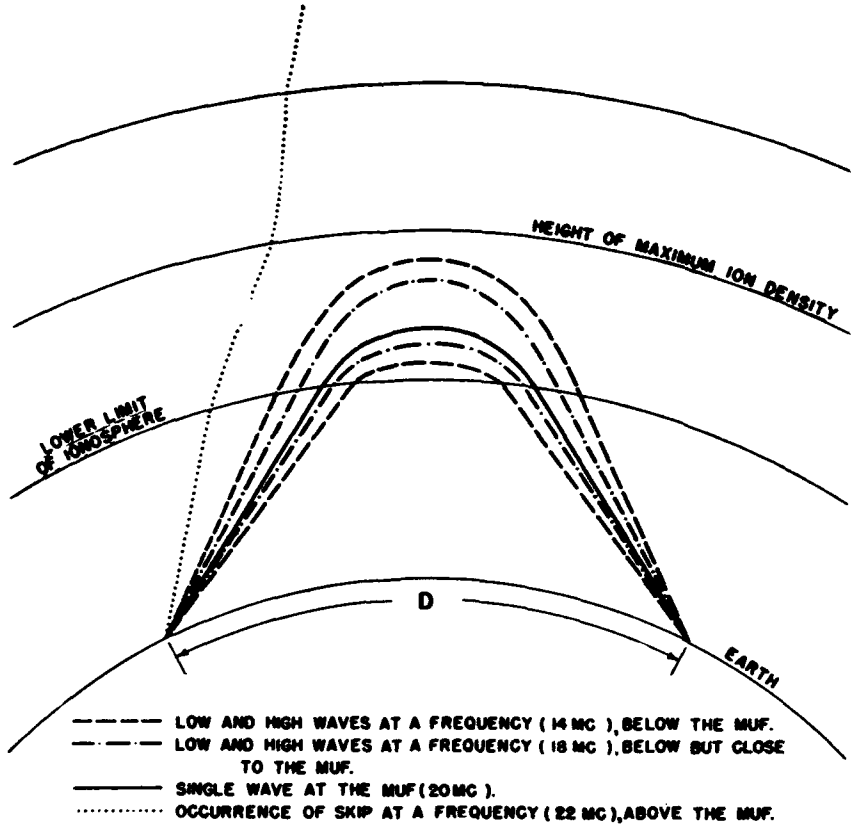


FIGURE 6.3. *Fixed distance—varying frequency (curved earth, curved ionosphere assumed).*

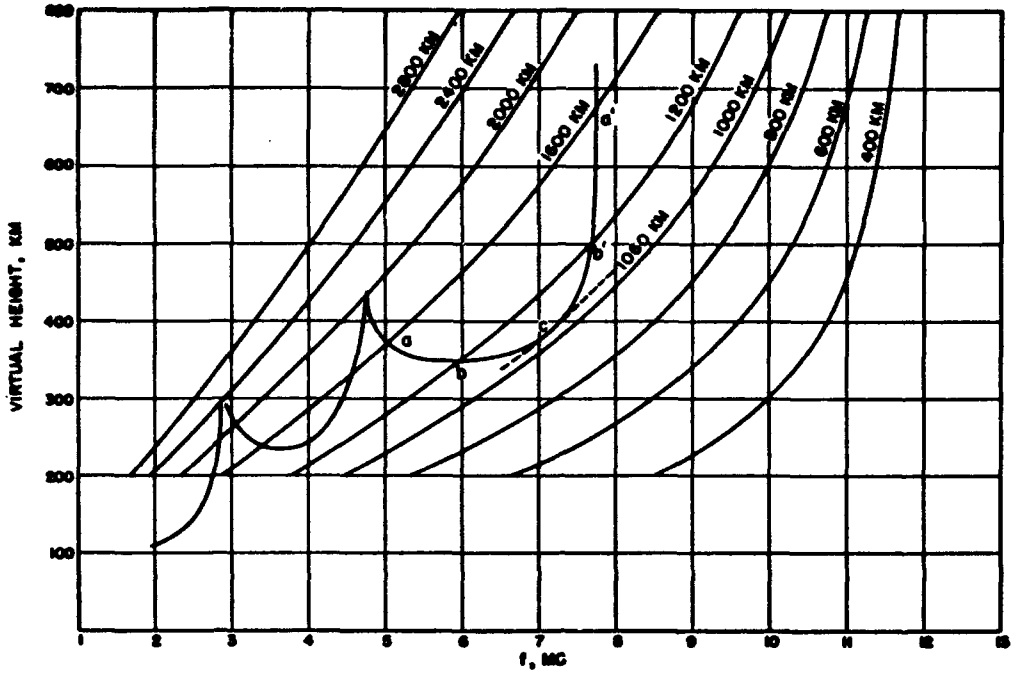


FIGURE 6.4. *Family of transmission curves parametric in distance for a fixed frequency of 12.0 Mc superimposed on an h'f graph.*
A flat earth and flat ionosphere assumed.

for a distance of 1,060 km is indicated as a single solid line.

Summarizing the behavior of a sky wave as indicated by the above discussions and figures 6.2 through 6.5, assuming a constant set of ionospheric conditions, it can be said that at a constant distance of transmission up to the limit of one hop with a variable wave frequency:

(1) Frequencies considerably below the muf will be greatly attenuated by nondeviative ionosphere absorption.

(2) Frequencies somewhat below the muf will be reflected as low and high waves, either or both of which may be greatly attenuated by deviative absorption.

(3) Frequencies near the muf will be reflected as low and high waves, both of fair strength.

(4) Frequencies at the muf will be received in the greatest possible strength as one wave.

(5) Frequencies above the muf will skip and not be heard (except by "scatter transmission").

For a fixed wave frequency, assumed somewhat above the extraordinary-wave critical frequency, being received over a variable distance up to the limit of one hop:

(1) At short distances the wave will skip.

(2) Just within a certain distance called the skip distance the wave will be received strongest and as one wave.

(3) At a greater distance the wave will still be received but as a low and a high wave.

(4) At still greater distances the low and high waves will be received, but either or both will be greatly attenuated by deviative absorption.

(5) At even greater distances the wave will be greatly attenuated by deviative and nondeviative absorption.

One important fact to be borne in mind is that the waves are receivable at distances greater than the skip distance, but that as this distance is appreciably increased, increased attenuation results.

The simple relationship of eq 6.2 is not strictly true for a curved earth and curved ionosphere, nor does the equivalence theorem hold, as the virtual height of reflection at oblique incidence is

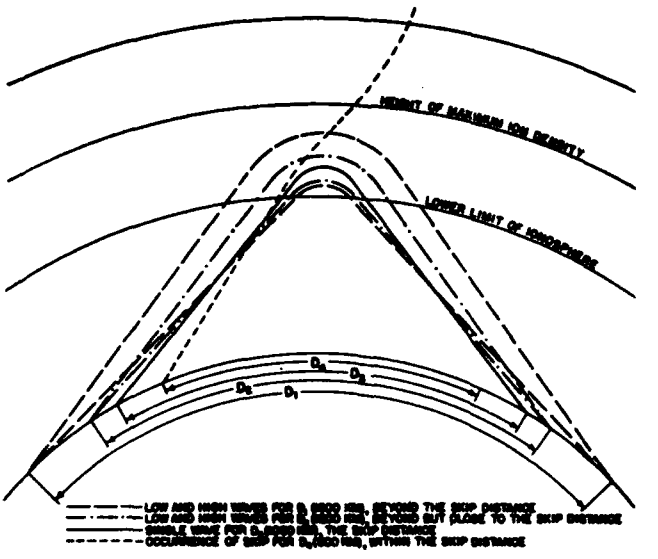


FIGURE 6.5. *Fixed frequency—varying distance (curved earth, curved ionosphere assumed).*

no longer equal to the height of the equivalent triangular path. The relationship needs to be corrected to:

$$f' = kf \sec \phi_0, \qquad (6.5)$$

where k is a correction factor, usually greater than unity; $k \sec \phi_0$ is usually referred to as "sec ϕ_0 (corrected)."

The value of k is a function of the distance and also of the ion distribution and is very complicated to calculate. However, for any given virtual height, it is possible to calculate this factor for the limiting cases of two different types of ionization, one for which the wave path itself is almost triangular, penetrating far into the medium, and the other for the case where the wave is deflected at the lower boundary of the ionosphere. From a curve drawn between the two limiting values by an approximate method it is possible to calculate a fairly good correction factor.

In practice, ionosphere sweeps are scaled for maximum usable frequencies or maximum-usable-frequency factors (described below) for a standard distance and the data used in predicting the maximum usable frequencies for future months. For such purposes it is possible to use a family of transmission curves parametric in wave frequency for the standard distance.

As an alternative to using a family of transmission curves for each desired distance, it is possible to construct a type of transmission curve whereby a single curve suffices for each distance and a family of curves parametric in distance can be drawn on one transparency. This is done by making the frequency scale logarithmic, and requires that sweeps be obtained originally on a standardized logarithmic scale or else be replotted on such a scale.

The factor $k \sec \phi_0$ for a given distance is plotted as abscissa on the same logarithmic scale as that used for the $h'f$ graph but with the numbers running in the opposite direction, the virtual height h' being plotted in the usual way as ordinate. Figure 6.6 shows a plot of a family of such logarithmic transmission curves. The transmission curve transparency laid over the $h'f$ curve with the height scales coinciding will, in general, intersect the $h'f$ curve in two places. Opposite the abscissa 1.0 used as an index on the scale on the transparency will appear the value of the product of f times $k \sec \phi_0$ or f' in eq 6.5. (Because of the logarithmic frequency scale, the same number, except for the decimal point, will appear on the transparency opposite the 1.0 index on the logarithmic $h'f$ chart). The transmission curve can be slid along the abscissa until it is just tangent to one curve on the f' sweep. The product as then read at the index is the value of the maximum usable frequency. The procedure is illustrated in figure 3 of the first reference listed in section 6.7.

For many purposes, the maximum-usable-frequency factor or ratio of the maximum usable frequency to the critical frequency for the same layer is desired. This may be obtained directly from the logarithmic scale on the transparency opposite the critical frequency cusp when the transparency is set to read the maximum usable frequency.

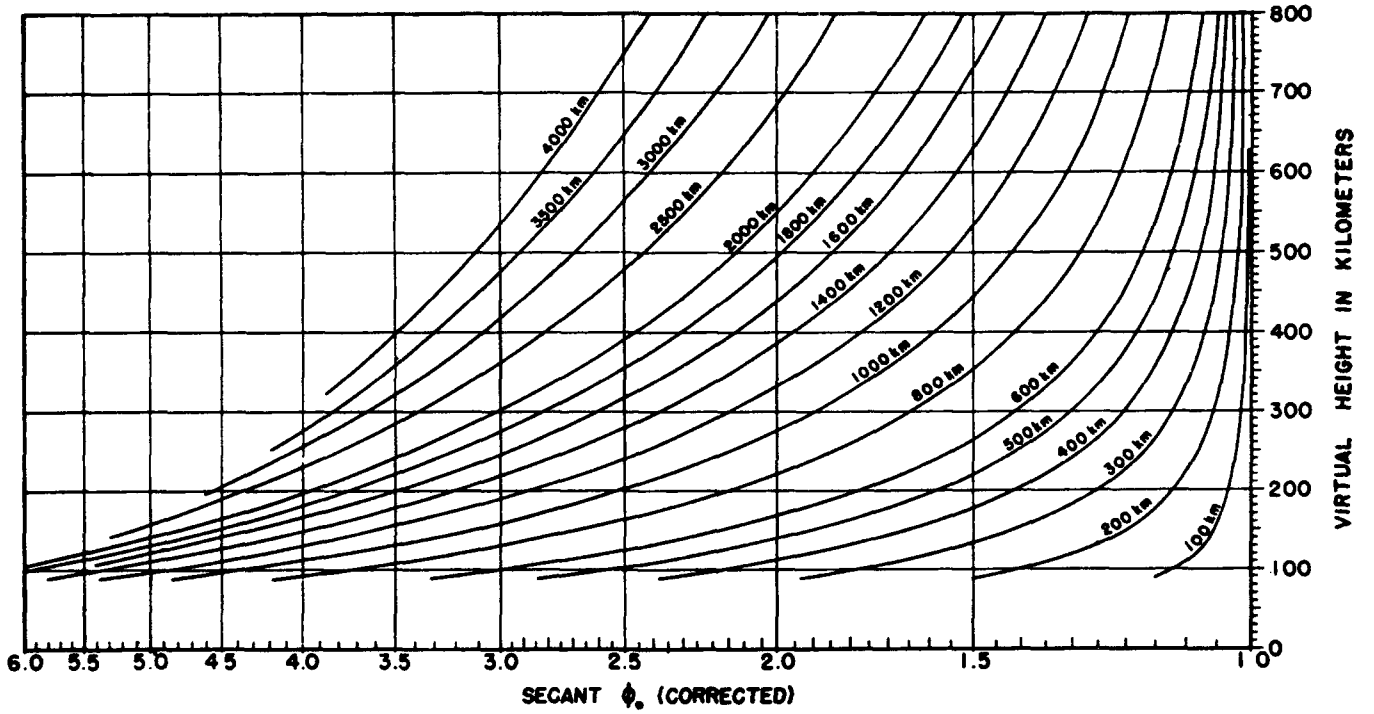


FIGURE 6.6. *Logarithmic transmission curves for a curved earth and curved ionosphere.*

The above procedure gives the reflection conditions and the muf for the ordinary wave only. For short distances, if greatest accuracy is desired, a mere scaling of the muf from the ordinary wave sweep is not sufficient. The relation of the path to the earth's magnetic field must be considered. However, this relationship is very complicated. For the simple cases where the longitudinal component of the earth's magnetic field is equal to zero, the transverse component is equal to zero or the longitudinal and transverse components are equal to each other, it is possible to obtain a reasonably accurate solution for the extraordinary-wave muf by a method indicated in the first reference listed in section 6.7, under VIII, "Effect of the Earth's Magnetic Field."

In the limiting case of vertical-incidence sky-wave transmission, the solution is simple, the maximum usable frequency for zero distance, the zero-muf, then being the extraordinary-wave critical frequency.

For long-distance propagation, the transmission curve applied to the *o*-trace gives approximately the correct answer. At shorter distances something might be gained by use of the method just mentioned if an exact analysis of the maximum usable frequencies indicated by the ionosphere sweep for a particular time and place were desired; usually, however, the *x*-muf is only very slightly greater than the *o*-muf, and the uncertainties introduced by normal ionosphere variations are so great that the additional precision is not justified.

The transmission curves of figure 6.6 are most accurate at *F*2-layer heights, as they were calculated by using the correction factor k of eq 6.5 for typical *F*2-layer ion distributions. At short distances they will be accurate for any layer because, as the distance approaches zero, k approaches a value of 1.0.

b. The Parabolic-Layer Method

Another method, developed in Great Britain, exists for obtaining maximum usable frequencies from vertical-incidence sweeps. This method, the parabolic-layer method, is based on the validity of the equivalence theorem and the secant law just as is the transmission-curve method. The difference is that in the transmission-curve method the secant law is applied to the sweep by means of a standardized graphical overlay, whereas in the parabolic-layer method the ionization of the layer is assumed to have a parabolic distribution, the parameters of the parabola corresponding to the important part of the sweep for transmission are found graphically for the particular sweep being examined and the muf factor determined by reference to a previously calculated set of curves for parabolic layers having the same parameters.

The parabolic-layer method has certain advantages, particularly when manual ionosphere recorders are used, because only a few specific virtual heights need to be measured. However, it is a longer method, and is therefore generally not used as widely as the transmission-curve method.

6.3. Angles of Departure and Arrival; Maximum Usable Frequencies for Distances Greater than One-Hop-F2

a. General

Angles of departure and arrival are of importance chiefly because a knowledge of these angles for the prevalent mode or modes of propagation makes possible the design of the best antennas for transmission and reception of such modes.

b. Single-Hop Transmission

The angle of departure or arrival usually assumed for single-hop transmission is that obtained from the straight geometry of a triangular path over a curved earth with the apex of the triangle placed at the virtual height assumed for the level of reflection. Families of curves parametric in angle of take-off superimposed on the logarithmic transmission curves of figure 6.6 are shown in figure 6.7. Used in conjunction with an $h'f$ graph, they give angles of departure for the low and high waves for any operating frequency for any distance out to the limiting distance for one hop for the layer considered. The procedure is almost the reverse of that used in obtaining maximum usable frequency. If the transparency is slid to where the appropriate index reads the operating frequency, the transmission curve will intersect the sweep at two virtual heights, through each of which will pass the curve for the angle of departure concerned. In practice the low-angle wave is the one usually considered.

It should be noted that this technic is only approximate, inasmuch as the equivalence theorem is assumed to be valid, i. e., the virtual heights on the $h'f$ graph are assumed to be the virtual heights for oblique-incidence reflection.

If it is desired to predict angles of departure for any given path less than 4,000 km in length, approximate calculations can be made by applying the transmission curves to median values of the minimum virtual heights, such as the ones on the world charts of figure 4.19, for the *F*2 layer. For the *E* layer a height of 110 km can be assumed. For the *F*1-layer angles of departure between those for the *E* layer and *F*2 layer can be assumed; if actual values of the minimum virtual heights of the *F1* layer are required, they may be obtained as medians of values observed at the location of an actual ionosphere station, or the world charts

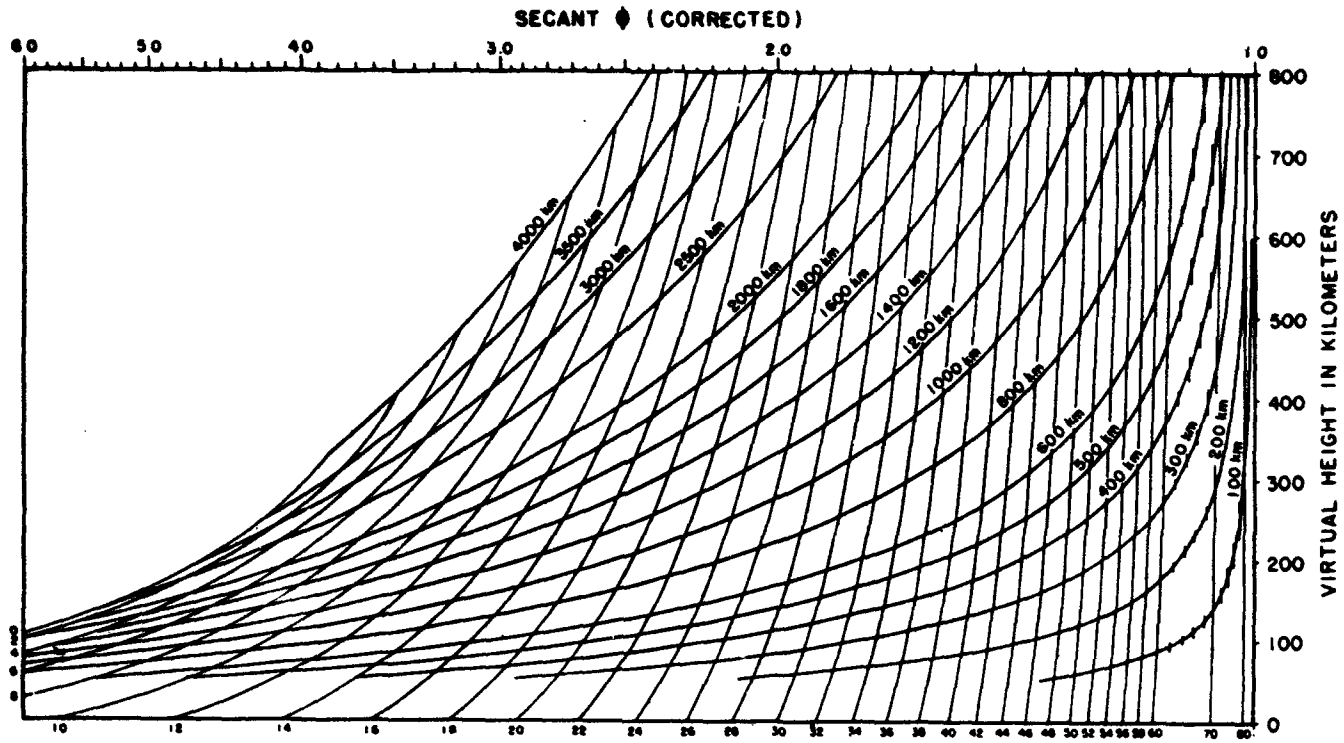


FIGURE 6.7. *Transmission curves showing vertical angles.*

of $F1$-layer minimum virtual heights (fig. 4.14) may be consulted.

For greater accuracy, it would be necessary to predict a hypothetical median ionosphere sweep for the location of the midpoint for the desired month, and proceed in the manner described above. These results would still be subject to the error of assuming that the equivalence theorem holds, and furthermore would be only median values. However, the values of the critical frequencies and heights of the layers fluctuate widely from hour to hour and from day to day, so that for any given path of intermediate length the modes of propagation and the layer heights, and hence the optimum angle of departure, may change widely from hour to hour and from day to day.

With these considerations in view, it would seem that for best practical results an antenna for intermediate-distance sky-wave propagation should be designed for a minimum of low-angle propagation and with most of its radiation upward in a broad enough radiation pattern to cover the range of possible angles necessitated by variations in heights, critical frequencies and modes of propagation, and possibly also distances of transmission when the service is not point-to-point. The whole period of the day when transmission is to be made on a given operating frequency should be considered.

An idea of the possible modes of propagation can be obtained by finding out if the operating frequency will be lower than the maximum usable frequency in the median case by use of the world prediction charts, as described under section 6.6. These charts do not indicate, however, whether or not the E layer shields the $F2$ layer from reflecting the waves at the same time. This phenomenon sometimes occurs and can be determined from the transmission curves and angle of departure curves of figure 6.7. If the muf for E-layer propagation, at such a distance that the angle of take-off is the same as that for $F2$-layer propagation over the desired distance, is greater than the operating frequency, the E layer shields the $F2$ layer. See also section 7.1.

If both E and $F2$ modes can appear simultaneously, the range of angles of departure ought to be fairly wide for general coverage. A range of from 20° to 60° is not improbable in some cases.

c. Transmission Over Distances Greater Than 4,000 Km

It would appear that the limiting distance for single-hop propagation by any layer is that determined by the earth's curvature, at which distance the angle of departure should be zero. This distance depends, of course, on the reflection height; for the $F2$ layer, in various parts of the world and at various frequencies, the limiting distance may range from 3,500 to 5,000 km or greater. However, experience points to 4,000 km

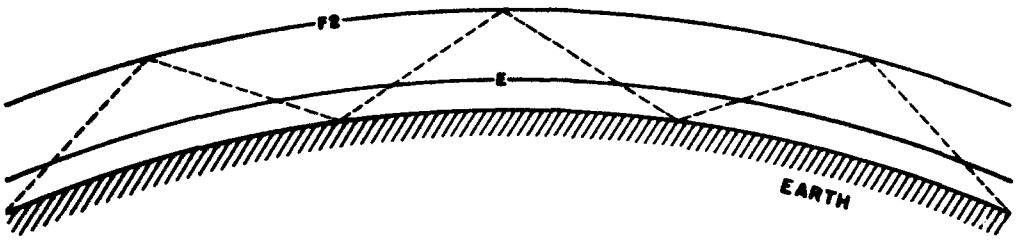


FIGURE 6.8. *Propagation by a number of geometric hops.*

as being about the average limit. Furthermore, this limit has been reached in practice with transmitting and receiving antennas whose heights are too low to radiate appreciable amounts of energy at angles lower than about 5° or so. Detailed consideration of this propagation is beyond the scope of this book; such phenomena as ionospheric focusing, scattering, and low-mode guided-wave propagation are involved.

Beyond the single-hop limit, the simple picture of transmission involves propagation by a number of geometrical hops, as illustrated in figure 6.8. It was at one time believed, because of this picture, that failure of propagation at any one of the ionosphere reflection points would cause failure of propagation altogether. Empirically, however, it has been observed in a number of cases that propagation did not fail until the ionosphere at "control points" 2,000 km from each end of the path failed to support the transmission. This was apparently true, within limits, no matter how long the great-circle path was between the control points and no matter what the condition of the ionosphere was in between.

A detailed discussion of the reasons for using the 2,000-km control points is also beyond the scope of this book; it is related to the low-angle radiation problem discussed above. We may consider, however, the way in which scattered transmission affects the modes of propagation. Considering a single end of a propagation path, any wave which leaves the earth and strikes the *F*2 layer excites scatter sources on its way up through the *E* region and again on its way down after being reflected from the *F*2 layer. Reradiation takes place from both regions, the first being called "short scatter" and the second "long scatter". The short-scatter source provides fairly intermittent energy because of the intermittent nature of the ionization causing the reradiation. However, by the time the wave has gone up to the *F*2 layer obliquely and returned, it is spread over a sufficiently wide area that there are always enough patches of ionization to provide an appreciable amount of reradiation.

Considering a hop of a given length, if propagation just fails for that length of path it will, assuming a uniform ionosphere, persist at a slightly greater distance and continue to "illuminate" the scatter sources mentioned above. The limiting length of path for which this could happen would be 4,000 km and the midpoint of this path would, of course, be 2,000 km from the transmitter. Energy from the long-scatter sources would in each case fall into regions where propagation could occur by some mode, such as by means of a regular layer, *Es*, or other scatter sources. Figure 6.9 shows a possible manner in which long scatter might contribute to propagation over a path after failure of the geometric modes.

Now returning to the idea of propagation over a long path, if we assume propagation by a number of geometric hops and then let the muf at, say, the first reflection point go below the operating frequency, it is evident that waves leaving the transmitting antenna at a lower angle of departure than that which was necessary to sustain the geometric modes will be propagated and excite

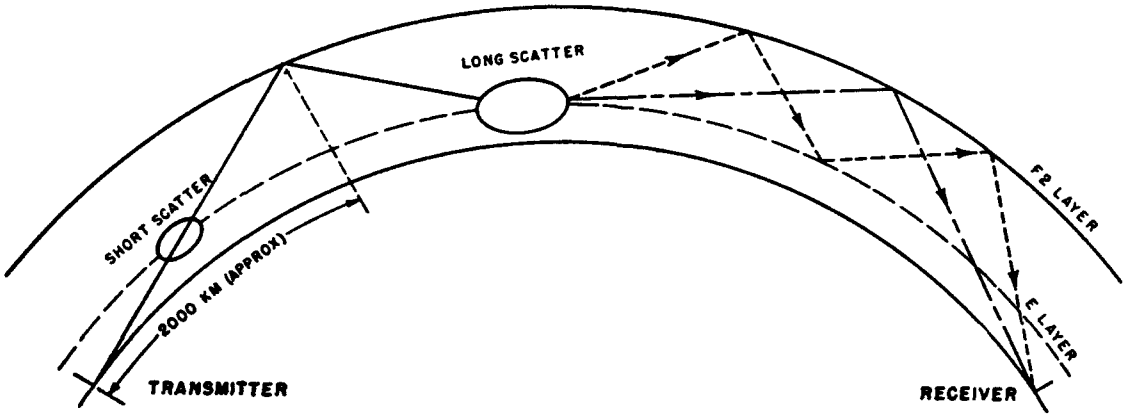


FIGURE 6.9. *Some possible modes of propagation over great distances involving long scatter.*

scatter sources, the energy, though attenuated, still reaching the receiver by any or all of the devious modes mentioned above. The limiting case for which energy can take off and still be propagated becomes, under these assumptions, the condition where regular $F2$-layer propagation is just possible at a point 2,000 km from the transmitting end.

By the same reasoning it can be shown that at the receiving end the limiting case for which waves over the path can be propagated to the receiver is that in which regular $F2$-layer propagation is just possible at a point 2,000 km from the receiving end.

In the practical solution of problems of finding the maximum usable frequency by $F2$ layer for paths over 4,000 km in length two "control points" 2,000 km from each end of the great-circle path are chosen and the muf for the path at any hour taken as the lower of the two determined for each control point separately. The technic is described in section 6.6, e.

When it is desired to include the effects of the E and $F1$ layers and sporadic-E ionization an additional pair of control points for use with E-layer data are placed 1,000 km from each end. There is, in this case, less theoretical justification than in the case of the $F2$-layer control points. The procedures for these cases are given in section 6.6.

The problem of long-path propagation seems to involve two considerations: (1) that of obtaining the greatest received intensity, during the hours when the geometric modes prevail, using a minimum of power, and (2) that of obtaining the greatest number of hours of propagation by any possible mode with the greatest amount of power it is reasonable to use. In each case much can be gained from beamed antennas. In the first case the proper angle of departure may be that which will excite the appropriate geometric modes. It should be pointed out, however, that the geometric modes are not clearly defined in the case of long-distance propagation; they are split by scatter and by ionosphere roughness, and the predominant energy may not be carried by the mode that would seem by simple theory to be the best one. There is some reason to believe that on paths greater than 4,000 km long the 4,000-km mode may be the dominant mode, even though a geometric construction does not show this as a possible mode. In the second case the lowest possible angles of departure would be the best. In designing antennas for low angles of departure the effect of ground reflection in producing nulls and maxima in the antenna pattern is very important. In order to bring the lowest maximum close to the ground the center of the antennas should in practice be mounted several wavelengths above the ground.

The question of the distinction between angles of departure for greatest intensity by geometric modes and angles of departure for the greatest number of hours of propagation by any possible mode needs to be examined in more detail. Most users have enough power available so that they are interested only in the greatest number of hours of frequency usage. However, although most long-distance installations are lacking in radiation at the lower angles it is believed that too sharp a beam in the vertical direction is often used, especially where rhombic antennas are used. As a general rule, it is believed that a large amount of power should be radiated at the lowest possible angles but also a reasonable amount should be radiated at angles high enough to excite such geometric modes as would fit the path in the fewest possible number of hops and, say, that number of hops plus one additional.

Frequently, when transmitters are beamed off the great-circle path, they can be received during hours when propagation should have failed even according to calculations made by the conventional control-point method. Under such conditions, transmissions from northern Europe have often been observed in the United States; a high-frequency broadcasting station in England beamed toward Africa, for example, illuminates a long-scatter source far to the south. Such stations can be heard long after failure of propagation over the great-circle path, and direction finders trained on these stations give at these times badly swinging bearings which point on the average to a source far to the south of England. The phenomenon has been recognized for a long time; in fact, on certain trans-Atlantic circuits the antennas are beamed for a path to the south of the great-circle path.

6.4. Current Methods in Use at CRPL for the Prediction of F2-Layer Characteristics

The average variations of the critical frequencies and muf factors of the various ionosphere layers are sufficiently well known to permit long-range predictions to be made for average conditions on ionospherically quiet days (days without ionosphere disturbances). One of the basic principles employed in all ionosphere predictions is the relation of ionospheric characteristics to the sunspot cycle. The actual prediction consists essentially of first predicting the solar activity and then deducing, from the mass of data available, the corresponding trends of seasonal, diurnal, and geographical variations of the ionosphere characteristics. No attempt is made at present in these long-range predictions, to estimate the detailed day-to-day variations in the ionosphere, these being rather localized and depending on conditions of solar radiation and terrestrial effects on particu-

lar days. Also, no account is taken of ionospheric disturbances, either of the ionosphere-storm type or of the sudden ionosphere disturbance type. These abnormalities constitute a different forecasting problem.

The basic ionospheric data used in the predictions are the E-, $F1$-, and $F2$-layer ordinary-wave critical frequencies (f^oE, f^oF1, and f^oF2, respectively) and the $F2$–M3000 factor (the muf factor for a transmission distance of 3,000 km, consisting of the ratio of the 3,000-km $F2$-layer muf divided by the f^oF2).

a. Basic Relations

All predictions are based on the well-established relationship between critical frequency and sunspot number. As figures 6.10 and 6.11 show, the assumption of a linear correlation is satisfactory for all practical purposes.

For each station, a separate graph is made for each hour of each month of the year. The monthly median critical frequency is plotted as ordinate against sunspot number as abscissa for all years for which data are available. Two points are plotted for each value of the critical frequency, one point using the 12-month running average of the monthly average Zurich sunspot numbers centered on the month in question, and the other using the monthly average Zurich sunspot number for that month. A straight line is drawn through the points, as illustrated in figures 6.10 and 6.11 by the line segments marking the zero and 100 sunspot number intercepts. These figures present two examples of this type of graph for Washington, D. C.

In general, points for the 12-month running average sunspot number seem to deviate less from a straight line than points for the monthly average sunspot number. Consequently, in drawing the line, greater weight is given to the points for the 12-month running average sunspot numbers.

The slope of the line varies with the hour of the day and with the month of the year for each station, and also varies considerably between different stations for the same hour and month.

b. Prediction Nomograms

The values of f^oF2 for sunspot number zero and 100 are read off the graphs, and diurnal curves of these values are plotted for each month for each station. These curves are then compared with previously observed monthly diurnal curves of f^oF2 for that station for that month, as a check on their shape. Minor irregularities in the diurnal curves are smoothed at this point. Figure 6.12 presents a sample of the f^oF2 zero and 100 sunspot-number diurnal curves.

The smoothed values of the zero and 100 sunspot-number curves for each hour are then used to establish points on a nomogram of the type shown in figure 6.13.

A set of these nomograms is prepared for each month for each station. The nomograms are then used in the preparation of all predictions and revisions of predictions of $F2$-layer critical frequencies and muf.

c. Preparation of F2-Layer Predictions

The first step in prediction is the selection of the predicted sunspot number, which is the expected 12-month running average Zurich number for the month of prediction. This can be done by a subjective extrapolation of the observed trend of sunspot activity. It may also be done by a more objective system using least squares, giving prediction coefficients. These are multiplied by the deviations of current sunspot numbers from the mean value for that time of the sunspot cycle and this predicted deviation added to the mean value of the sunspot number for the month in question. The point of reference is from the last observed sunspot minimum. Reliable estimates can be made for a year in advance, but more advanced estimates are subject to considerable uncertainty.

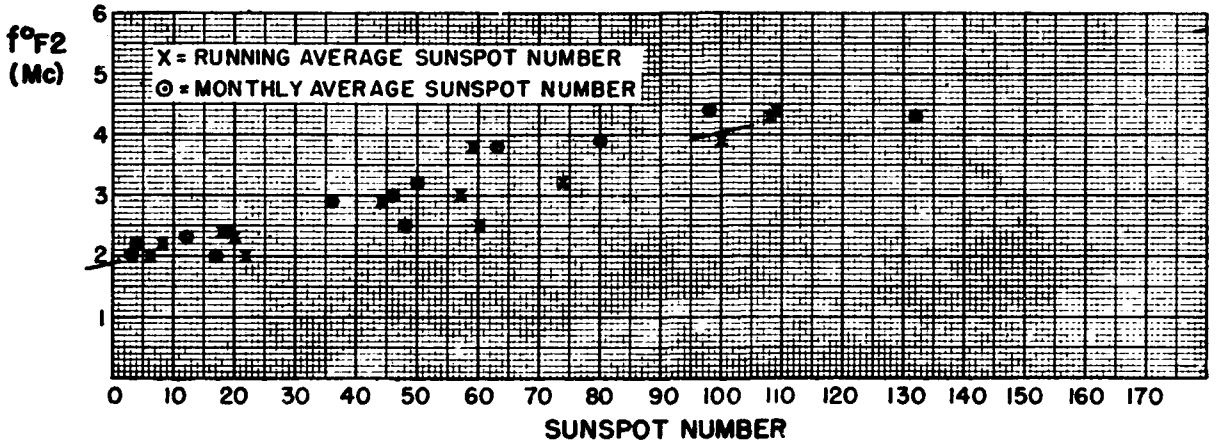


FIGURE 6.10. *Points plotted for January, 0000 local time, for years 1934–1946, at Washington, D. C.*

935333 O - 51 - 6

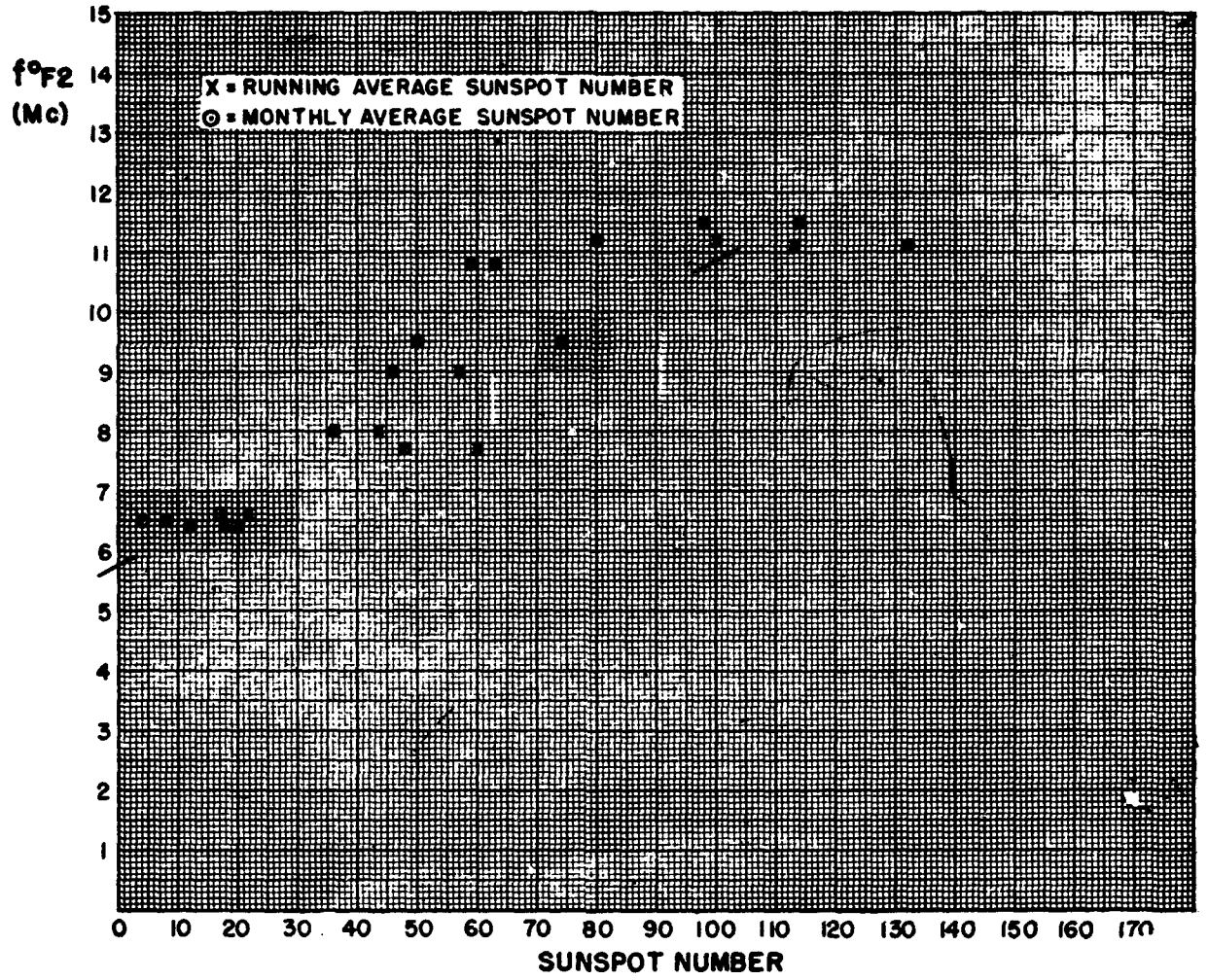


FIGURE 6.11. *Points plotted for January, 1200 local time, for years 1935–1946, at Washington, D. C.*

By using the predicted sunspot number, corresponding values of $f^{o}F2$ are read off the prediction nomograms. Diurnal curves of the predicted $f^{o}F2$ values are then drawn, smoothed where necessary, and compared with previously observed data from that station for that month as a check on their shape.

For a given hour of day, latitude variation curves are prepared, the values being read from the smoothed predicted diurnal curves. To assist in drawing these, similar latitude variation curves are drawn using observed data for that month for the previous year for comparison with the predictions. Separate curves are drawn for each of the three zones, *E*, *I*, and *W*, into which the world is divided because of the longitude effect.

As no data are available for many large areas of the world, predictions for these areas using the estimated sunspot number for the month of prediction, are prepared from the nomograms for areas in the opposite hemisphere, if data for those regions are available, but using the reversed season. These are plotted on the latitude variation curve sheet with their latitudes and zones reversed (i. e., north latitude E-zone stations are plotted as south latitude W-zone stations, etc.). The "reversed" points help determine the shape of the curve where no direct data are available, but the level of the curve is established by the directly predicted points.

Figures 6.14 and 6.15 present samples of predicted latitude variation curves.

As there is some longitude effect within each zone, as discussed in section 4.4, the latitude variation curves cannot be drawn to pass through each direct predicted point without introducing complex and improbable curve shapes. Also, predicted data from some stations are relatively unreliable for some or all hours because of observing difficulties or because of insufficient observed data to establish good trends. Consequently, the latitude variation curve must be drawn and smoothed by an experienced person who knows the data well and who can take these factors into account.

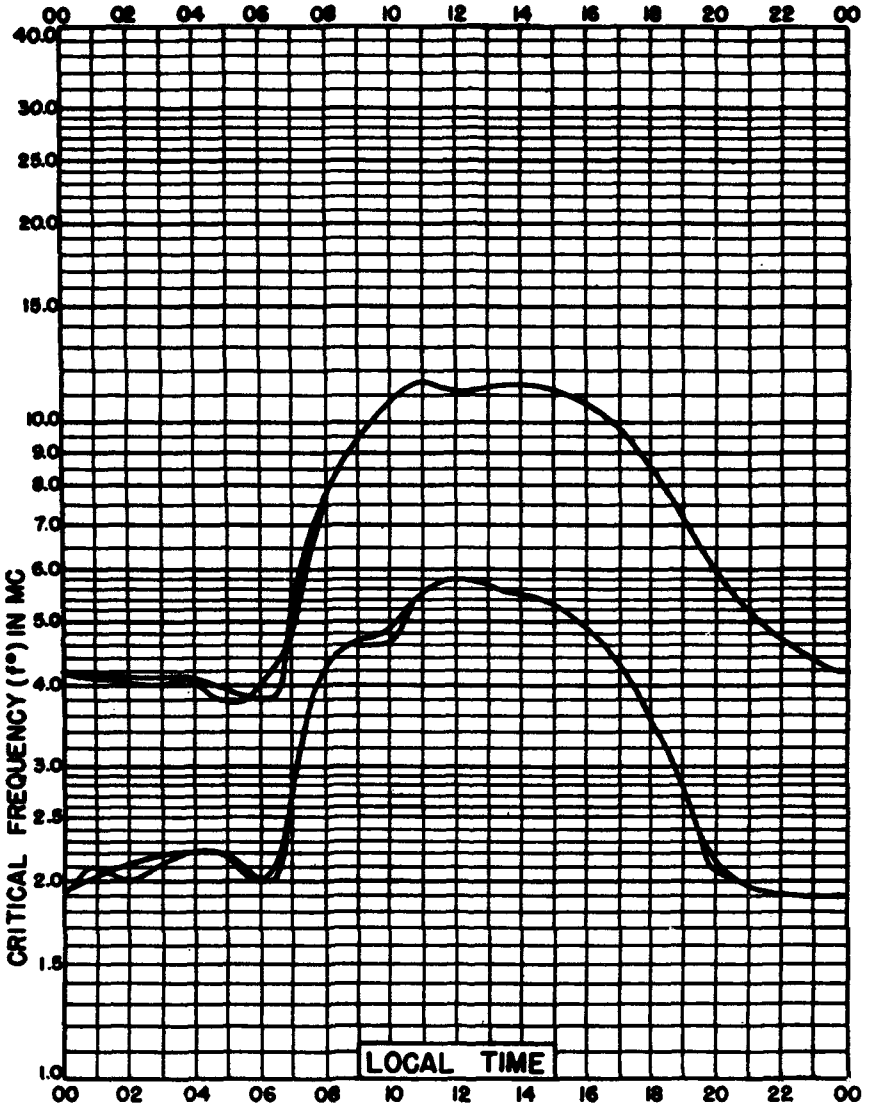


FIGURE 6.12. *Washington; January, 0 and 100 intercepts.*

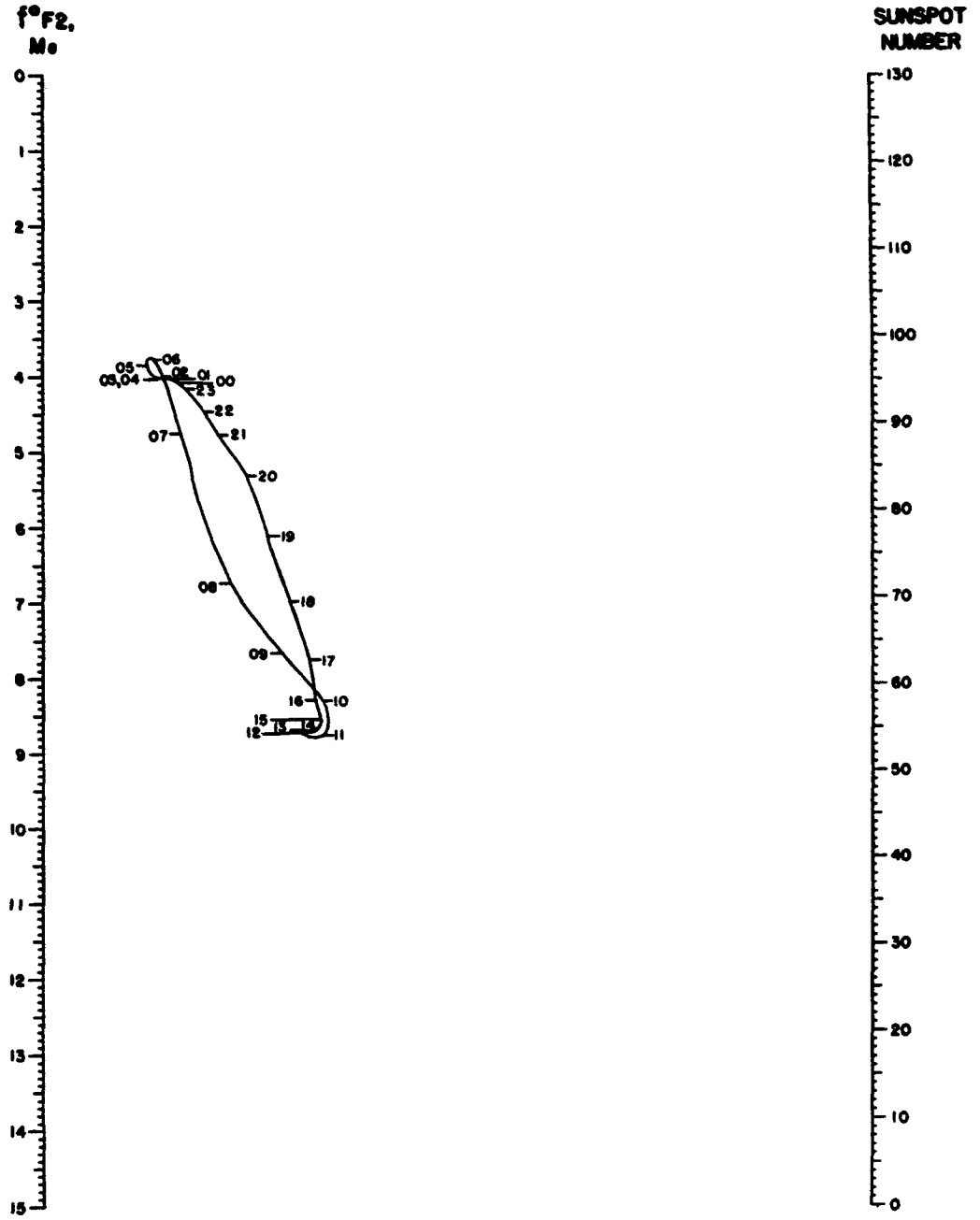


FIGURE 6.13. *Prediction nomogram for Washington, D. C. January.*

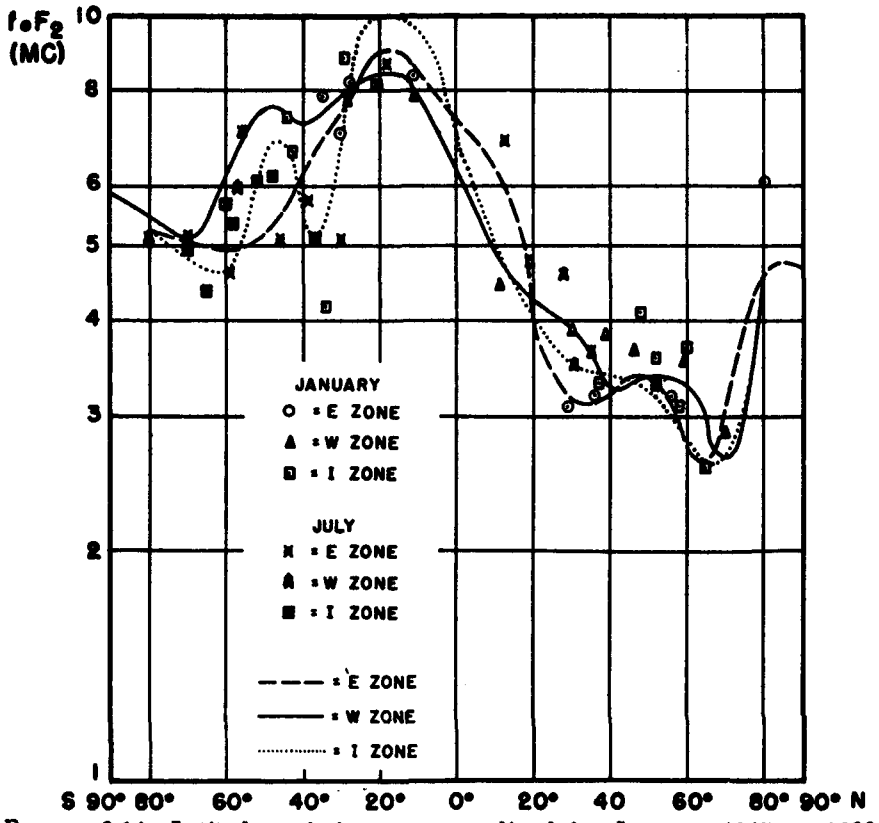


FIGURE 6.14. *Latitude variation curves predicted for January 1947 at 0000 local time.*

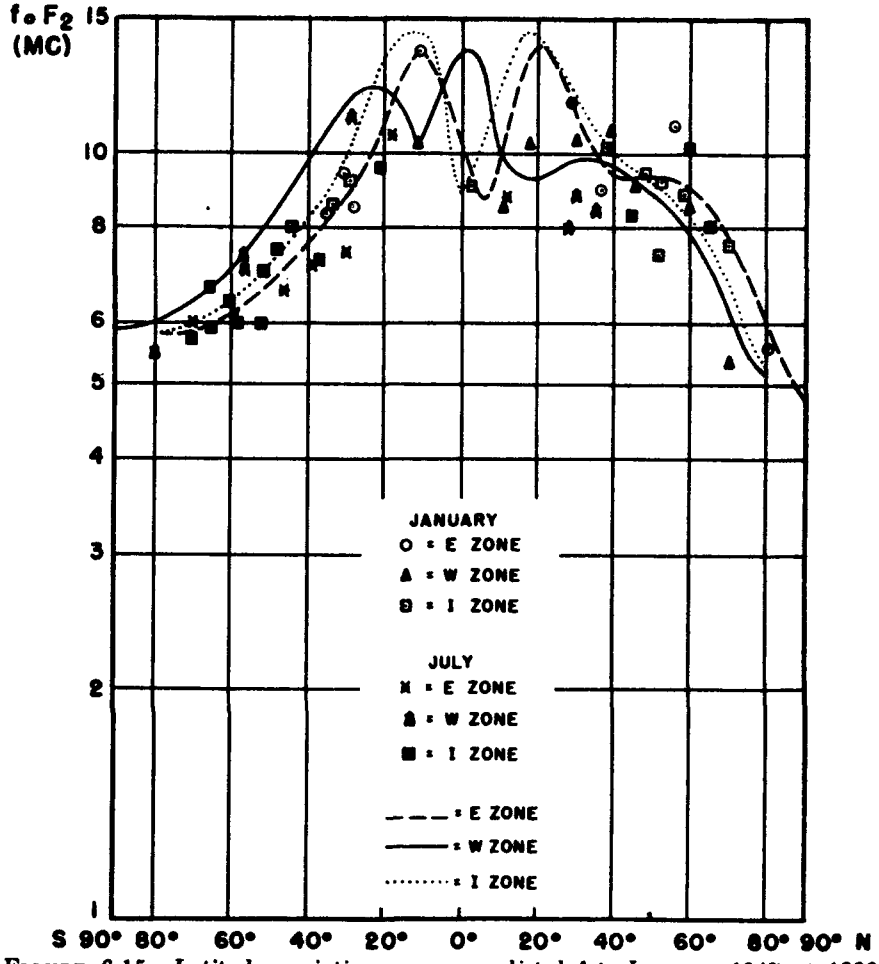


FIGURE 6.15. *Latitude variation curves predicted for January 1947 at 1200 local time.*

The predicted zero-muf (extraordinary-wave) and 4000-km-muf charts described in section 6.5, e, and published in CRPL–D series publications are prepared by reading the predicted $f°F2$ from the latitude variation curves and adding an appropriate amount (section 2.3) or multiplying by a suitable factor to convert to zero-muf or 4000-km-muf, respectively. Contours are drawn and smoothed where necessary. Predicted diurnal curves for a number of stations are read from the predicted zero-muf contour charts and compared with previous data for those stations as a final check, alterations in the chart being made where necessary to eliminate inconsistent values.

The critical frequencies of the E layer, like those of the $F2$ layer, vary approximately linearly with the sunspot number. The variations with season, location and local time are generally smaller and more regular than those of the other regular ionosphere layers.

The diurnal variation of the critical frequency expressed as percentage of the noon value for the E layer is nearly independent of sunspot number; therefore, it is only necessary to estimate the critical frequency for one time of day, usually noon. Latitude variation curves, 0 and 100 sunspot number values, and prediction nomograms, of the type described for the $F2$-layer critical frequencies, are used to obtain the predicted values of noon $f°E$ for every 10° of latitude (fig. 6.16).

The prediction of the values at other hours is made by multiplying the predicted noon $f°E$ by the average ratio of the critical frequency at other hours to that at noon obtained from all previous data for the same month. As the muf factors are nearly constant for all times and latitudes, the E-2000-muf is obtained by multiplying the predicted critical frequency by a single factor, 4.78. These multiplications are effected by means of a second central scale, B, as shown in figure 6.16.

d. Prediction Services in Other Countries

The problem of predicting critical frequencies is one that engaged the attention of many of the Axis and Allied Nations during the war from 1939 to 1945. Among the other countries actively working in this field were Australia, Canada, England, Japan, and the Soviet Union. Since the war, prediction services have been initiated in France and China.

The Canadian report, listed in section 6.7, follows closely the principles and methods in current use at the CRPL, the chief differences bein in the type of nomograms constructed for practical application.

Typical of a somewhat different approach is the method used in Australia. A mean monthly value is predicted for each characteristic for each station, first by studying the numerical differences that have been observed between successive monthly means, and second, by comparing the variations in ionospheric characteristics with the variations in sunspot number. Adjustments are made to reconcile any differences that may exist and to obtain a final value. Then the mean hourly values which make up the predicted monthly mean, are predicted. A family of curves, composed of all the curves depicting one characteristic during the same month for all years is plotted on one sheet. By extrapolation, a diurnal curve can be built up and drawn in. A separate check is obtained by drawing a series of 24 curves, representing the variations that have occurred at each hour of the day in the given month since observations first began. Mean hourly values are read off the predicted curve, averaged, and compared with the value of the monthly mean already obtained.

6.5. Technic of Predicting Maximum Usable Frequencies for Transmission Over Any Path

a. Scaling Predicted Sweeps

The most accurate way of obtaining maximum usable frequencies for any location is to construct an average ordinary-wave $h'f$ sweep for the desired hour and month for that location and apply the transmission curves to the sweep. If the location is near an ionosphere station of long standing, the availability of empirical information on the characteristics of the ionosphere records from that part of the world is of great advantage.

If only minimum virtual heights, critical frequencies and $F2$–M3000 factors (3,000-km maximum usable frequency factors) are available, it is still possible to construct fairly satisfactory sweeps. A template consisting of a 3,000-km transmission curve to the proper scale cut concavely along the curve can serve as a guide for drawing that part of the sweep to which the 3,000-km transmission curve was originally tangent.

If greatest accuracy is desired for short distances, it is necessary to take account of the earth's magnetic field as described in section 6.2, a.

b. Distance Factors

Practical problems in radio sky-wave propagation require a rapid and easy method of solution, since the problem usually has to be solved for a great many hours of the day. The method of constructing and scaling of sweeps described in section 6.5, a is much too slow and cumbersome even without consideration of the earth's magnetic field. A desirable method would involve the application of a simple factor or nomographic procedure to numerical predictions of standardized characteristics of the ionosphere at the location

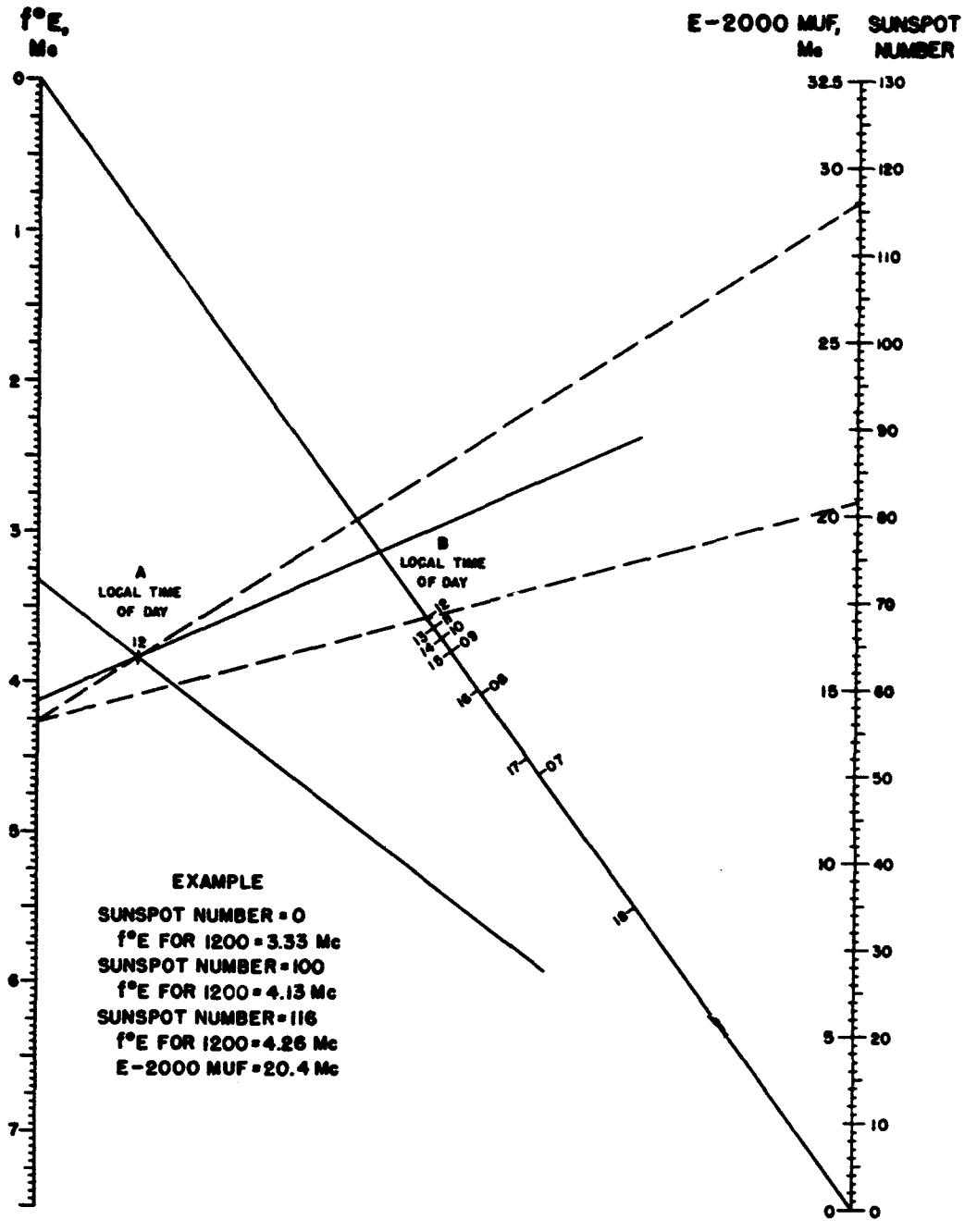


FIGURE 6.16. *Prediction nomogram (f°E) for July, latitude 10° S.*

concerned. Such a method exists involving the use of empirical "distance factors," one type representing the ratio of the muf for a given distance to the muf for zero distance (the extraordinary-wave critical frequency f_x), another type representing the ratio of the muf for a given distance to the 4000-km-muf.

There is no simple exact relationship between the muf at one distance and that at another, since the value of muf depends upon where the transmission curve falls tangent to the sweep, which is a function of the shape of the sweep, i. e., of the vertical distribution of ionization. It is possible to obtain, from some hundreds of scalings of muf for various distances from a large number of sweeps taken in various parts of the world in various seasons, empirical average distance factors relating the muf for all distances out to 4,000 km. In the case of the $F2$ layer, these factors are averages for a rather wide distribution of empirical factors. If one starts with a predicted value of $F2$-zero-muf (i. e., the $F2$-layer extraordinary-wave critical frequency, or f_xF2) the application of these factors to obtain, say, the 1500-km-muf, may yield results vastly different from those obtained if one started with the 4000-km-muf and worked down

to the 1500-km-muf, since the ratio of the zero-distance-muf itself to the 4000-km-muf varies widely. For this reason $F2$-layer distance factors based on f_0F2 are of value only for short distances, and factors based on the 4000-km $F2$-layer muf are of value only near 4,000 km. Such distance factors are seldom used; instead, the grid nomogram described in the next section is employed. In the case of the E layer, the scatter of distance factors is very small, so that either the factors or a nomogram can be used; the nomogram is ordinarily chosen because of its convenience.

Distance factors must not be confused with the muf factors (M-factors) scaled as described under section 6.2, a, and used in the prediction of world charts as mentioned under section 6.4. The muf factor is the ratio of the muf to the ordinary-wave critical frequency whereas one type of distance factor just described is the ratio of the muf to the extraordinary-wave critical frequency. Neither should be confused with sec ϕ_e (corrected), which is the ratio of the muf or any other oblique-incidence frequency to the frequency which would just be reflected at vertical incidence from the same height in the ionosphere.

c. Use of Distance Factor Nomograms

In order to avoid the ambiguities involved in the use of distance factors for $F2$-layer calculations and to provide a rapid method of obtaining the muf for any distance up to 4,000 km, the muf grid nomogram of figure 6.17 was devised. In the muf grid nomogram the ordinates are labeled in distance and are spaced apart proportionally to the empirical factors of section 6.5, b. The oblique lines connect equal values of frequency on the left-hand and on the right-hand scales. The left-hand scale is linear and is labeled "zero muf" (i. e., the $F2$-zero-distance muf, approximately $f_0F2+1/2$ gyrofrequency). The right-hand scale is labeled "4000 muf" (i. e., the $F2$-4000-muf or the $F2$-layer muf for 4,000 km).

The nomogram is used in connection with the $F2$-zero-muf and the $F2$-4000-muf values for the problem to be solved. These values may be scaled from the $F2$-zero-muf and the $F2$-4000-muf charts of section 6.6. To find the muf at any distance, then, the $F2$-zero-muf is located on the left-hand scale and the $F2$-4000-muf is located on the right-hand scale. The two points are joined with a straightedge and the muf for the desired distance read off as the value of the interpolated oblique line passing under the straightedge at the desired distance. An example of the use of this nomogram is given under section 6.6, b.

Any reasonably large reference distance such as 3,000 km may be used instead of 4,000 km, the given reference muf being located on the ordinate for that distance.

Results have been obtained with this nomogram using data for areas remote from those from which the original data were gathered in establishing the factors upon which it is based. Agreement with answers obtained by using the empirical relationships established in these other areas has been very close. Nonetheless, it is recognized that as new data from various areas are analyzed it may become necessary to revise the nomogram using revised factors, or even to construct additional ones for certain parts of the world.

The results of a survey made to determine the accuracy of the muf grid nomogram indicated that its error was nearly always less than 10 percent. The grid nomogram affords a quick method for determining the muf at any distance and is accurate enough for the solution of practical problems. It should not however be used in the study of conditions of the ionosphere on individual days, because individual $h'f$ sweeps may be too different from the median sweep. In these cases the sweep should be scaled directly with a transmission curve for the desired distance as described in section 6.2.

In the case of the E layer the small distribution of individual values of distance factor about the median makes a simple distance factor nomogram feasible. Such a nomogram is shown in figure 6.18, with an example of its use. The nomogram applies a factor for the desired distance to the E–2000-muf which must be obtained, usually from the world charts, and applied to the left-hand scale of the nomogram.

The effect of the $F1$ layer is approximately taken care of in this nomogram, by the assumption that the $F1$ layer controls propagation at distances between 2,000 and 4,000 km at a value of muf which is a few percent above the E-layer 2000 muf.

The same nomogram is, for convenience, used with sporadic–E, although the sporadic–E factors are slightly different.

d. ISIB Method

A method of obtaining maximum usable frequencies for any distance by $F2$ layer was developed by the Inter-Services Ionosphere Bureau in England during World War II. This method is more accurate than the one in which the muf grid nomogram is used but involves an extra operation.

The method is based on a family of muf factor curves derived from a number of $h'f$ sweeps, or from factors calculated for parabolic layers of different dimensions. An assumption was made that no matter what the layer dimensions are, considering any given pair of sweeps, if the value of the muf factor is the same for a given distance, the values of the muf factor for both sweeps will be equal for any other distance. With this assump-

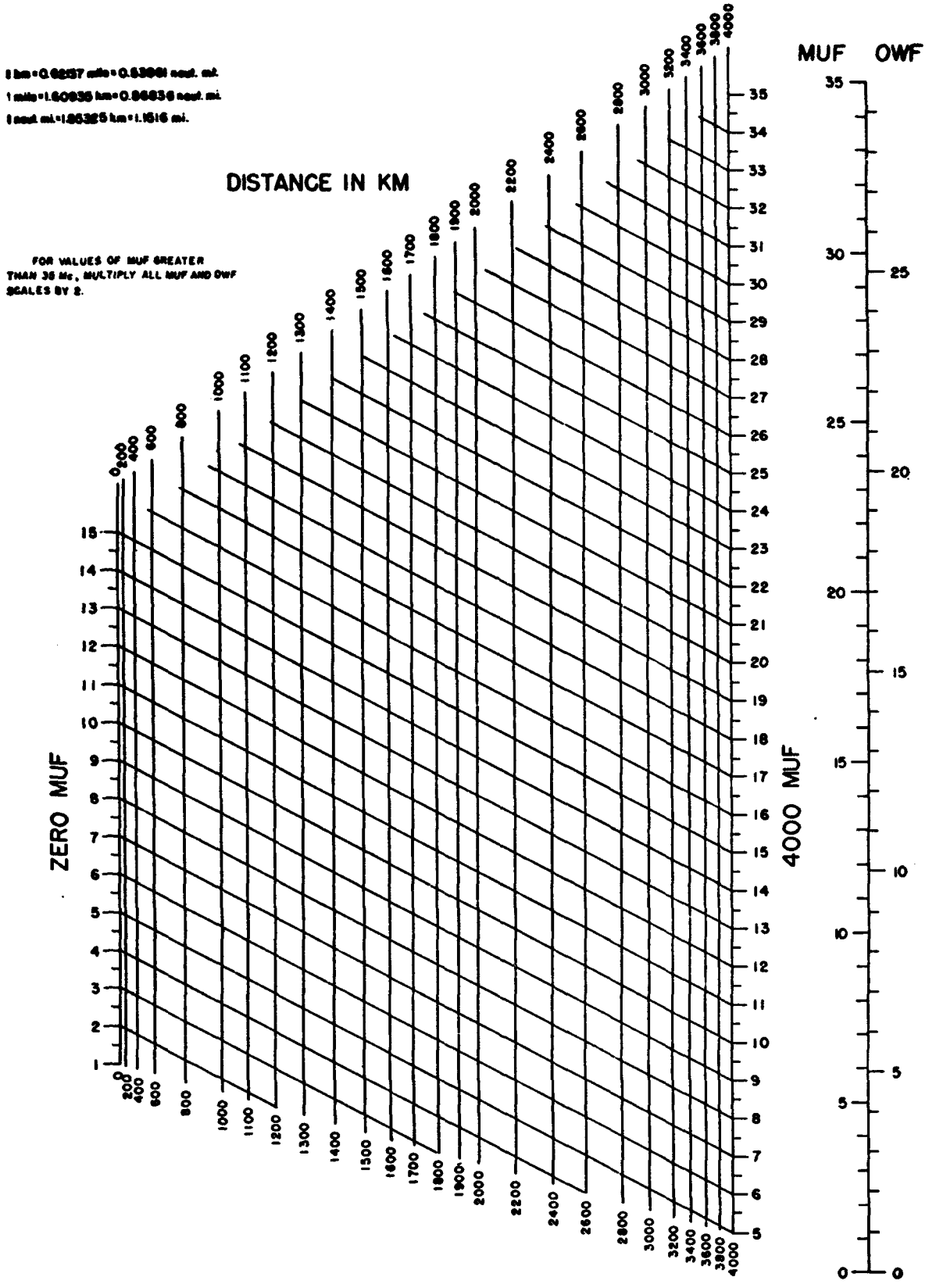


FIGURE 6.17. *Nomogram for transforming F2-zero-muf and F2-4000-muf to equivalent maximum usable frequencies at intermediate transmission distances; conversion scale for obtaining optimum working frequencies.*

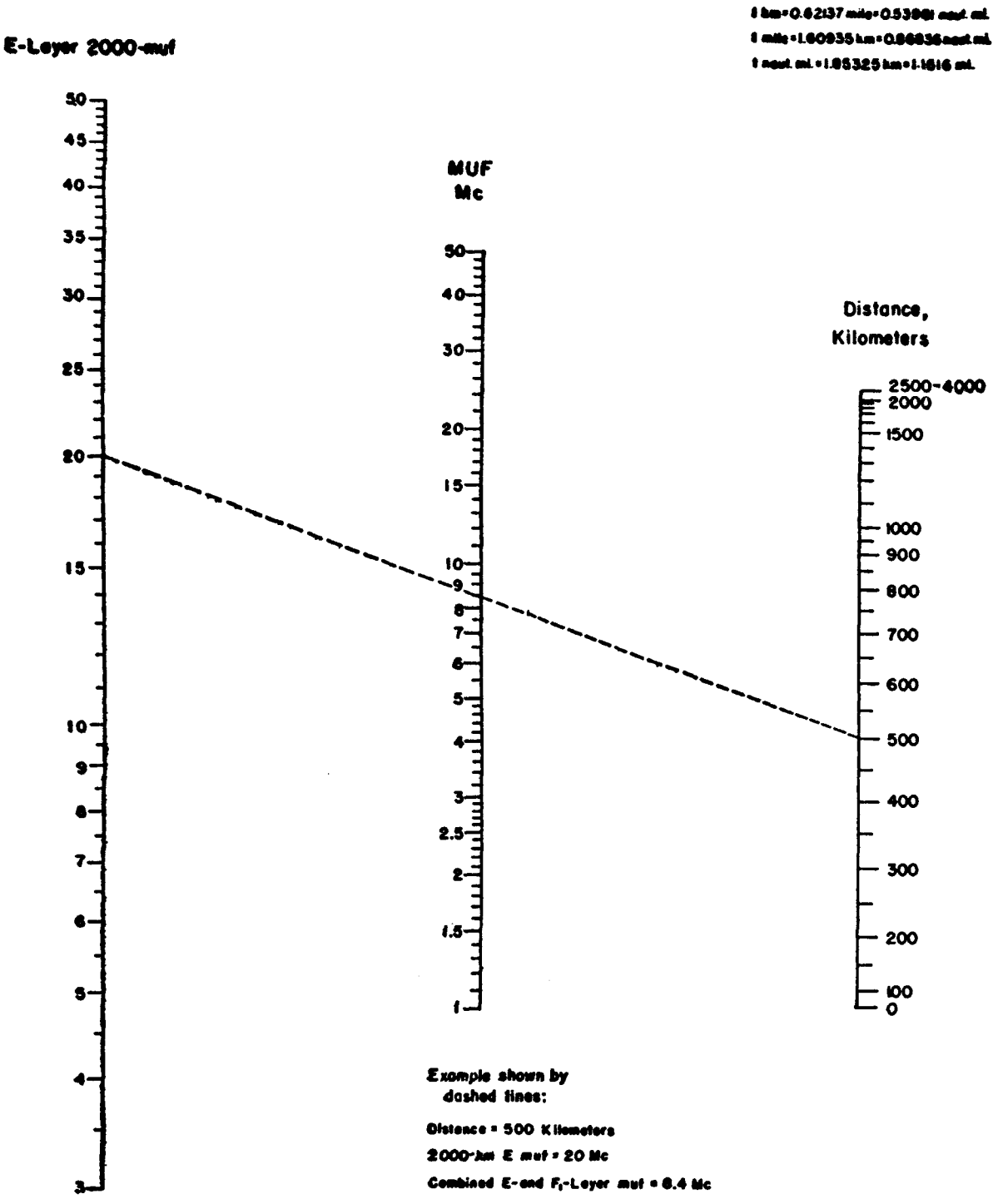


FIGURE 6.18. *Nomogram for transforming E-layer 2000-muf to equivalent maximum usable frequencies and optimum working frequencies due to combined effect of E layer and F1 layer at other transmission distances.*

tion, all possible variations of conditions were expressible as families of curves of muf factors related to the muf factor for a standard distance; the standard-distance factor is thus the parameter, instead of the layer dimensions. The standard distance used was 2,500 km. Actually, the muf factor for any distance can be regarded as a parameter simply by relabeling the curves with the appropriate value of the factor for that distance. Figure 6.19 shows these curves.

In practice the muf factor for the standard distance is found from the $h'f$ curve by any method, such as the transmission-curve method, and the muf factor for the desired distance found by following the appropriate curve down to that distance. Another set of families of curves is provided by which the value of f^oF2 scaled from the $h'f$ graph can be multiplied by the muf factor just obtained and at the same time corrected for the effect of the earth's magnetic field, the result being the desired muf. These families of curves are given for different values of gyrofrequencies and for two sets of conditions of direction of transmission with respect to the earth's magnetic field, one for north-south transmission or the earth's magnetic field horizontal, and one for east-west transmission or the earth's field vertical. Figure 6.20 shows such a family of curves of the latter type, for a gyrofrequency of 1.4 Mc.

In the same study of accuracy as that made for the muf grid nomogram of section 6.5, c, the ISIB method was found to have an error nearly always less than 4 percent; it is thus somewhat more accurate than the method using the grid nomogram.

e. World Contour Charts

The world contour chart of maximum usable frequency is a convenient device for representing worldwide variation of muf with local time. Figures 6.21 to 6.28 are examples of such charts.

It must be emphasized that the world contour chart is a time chart and in no sense a map. It is actually a plot of variation of median monthly values of muf against local time at all latitudes within a longitude zone for which the data are reasonably homogeneous, usually being constructed from individual graphs for each hour showing latitude variation of muf for that hour such as the latitude variation graphs described under section 6.4. At each hour integral values of muf are plotted at the latitude where they fall. These points are joined to form contours of equal muf.

Three longitude zones are chosen: East, West, and Intermediate (E, I, and W), in accordance with the considerations of section 4.4.

These time charts can be used to find the muf at any location on a standard world map such as the one of figure 6.29 at any world time, since they are drawn to the same longitude scale as the standard map and have a 24-hr time scale the same length as 360° on the standard map.

Methods of using world charts for muf predictions over any path are described in section 6.6.

6.6. Calculation of Maximum Usable Frequencies by Use of World Charts

The term "maximum usable frequency" as used in the following discussion refers to monthly median values such as appear on the world prediction charts. Thus communication at the muf calculated by the methods given here should be effective approximately 50 percent of the time during undisturbed periods.

a. General

The calculation of maximum usable frequencies involves knowing the length of the transmission path and the location of certain points on the path, called "control points," explained in section 6.3. A convenient method for obtaining this information consists in making use of figures 6.29 and 6.30.

Figure 6.29 is a map of the world. Figure 6.30 is a chart to the same scale as figure 6.29, on which the solid-line curves crossing the equator at two points 180° apart, represent great circles. The numbered dot-dash lines crossing the great circles indicate distances along them in thousands of kilometers. In using figures 6.29 and 6.30, proceed as follows:

(1) Place a piece of transparent paper over the map, figure 6.29, and draw the equatorial line (zero degrees). Place dots over the locations of the transmitting and receiving stations. Also mark the meridian whose local times are to be used as the times for calculation. Usually the Greenwich meridian is used.

(2) Place this transparency over the chart, figure 6.30, and, keeping the equatorial line of the transparency always on the equatorial line of figure 6.30, slide the transparency horizontally until the terminal points marked on it fall either on the same great circle or the same proportional distance between adjacent great-circle curves. Draw in the great-circle path which passes through the terminal points.

(3) For paths shorter than or equal to 4,000 km, locate the midpoint of the path, keeping the transparency in position on figure 6.30 and using as a distance scale the points at which the numbered lines in figure 6.30 cross the path as drawn on the transparency.

(4) For paths longer than 4,000 km, designating the ends as the A-end and B-end, respectively, locate on the path and mark with a dot the following "control points", scaling the distances as in (3) above:

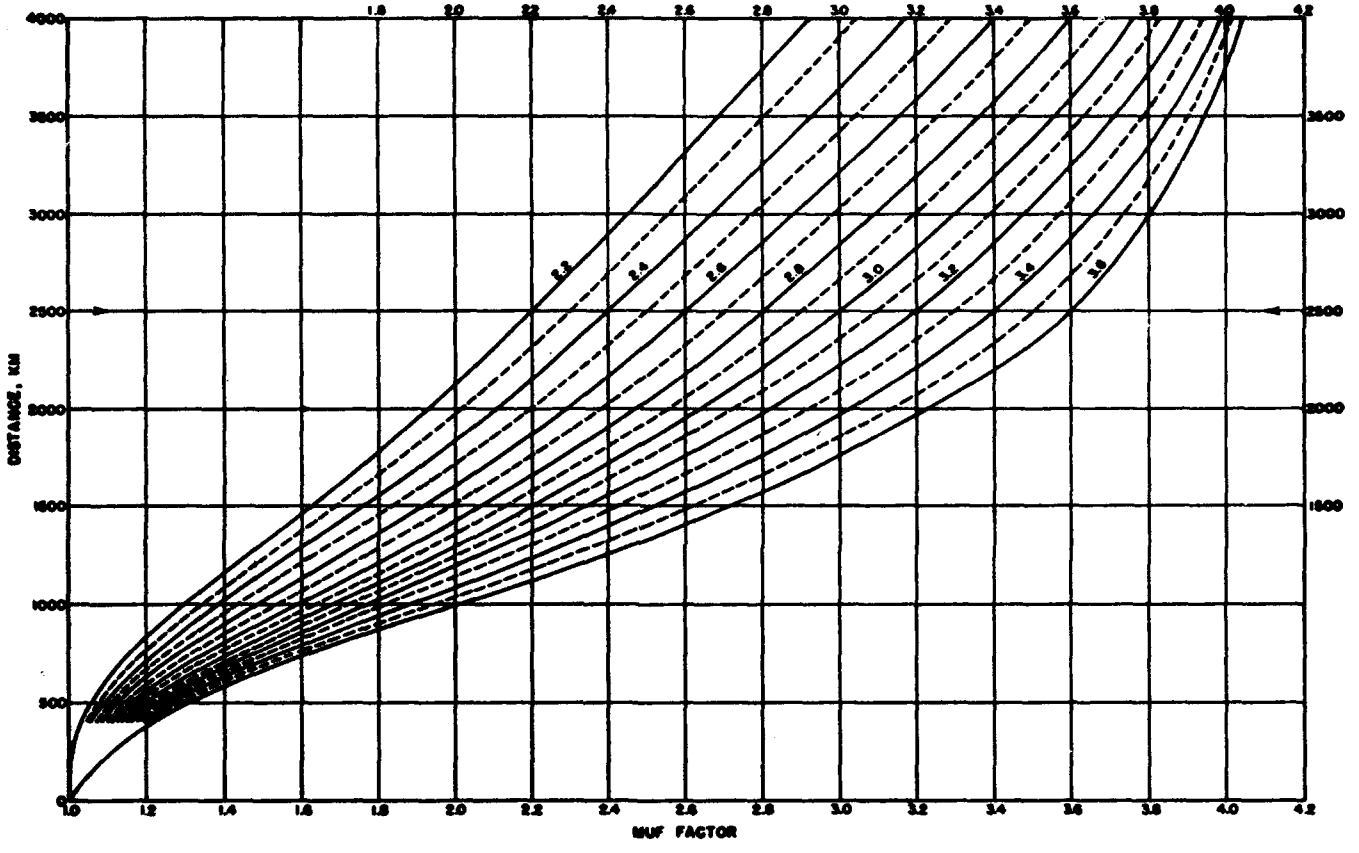


FIGURE 6.19. *ISIB muf factor chart (as of May 20, 1945).*

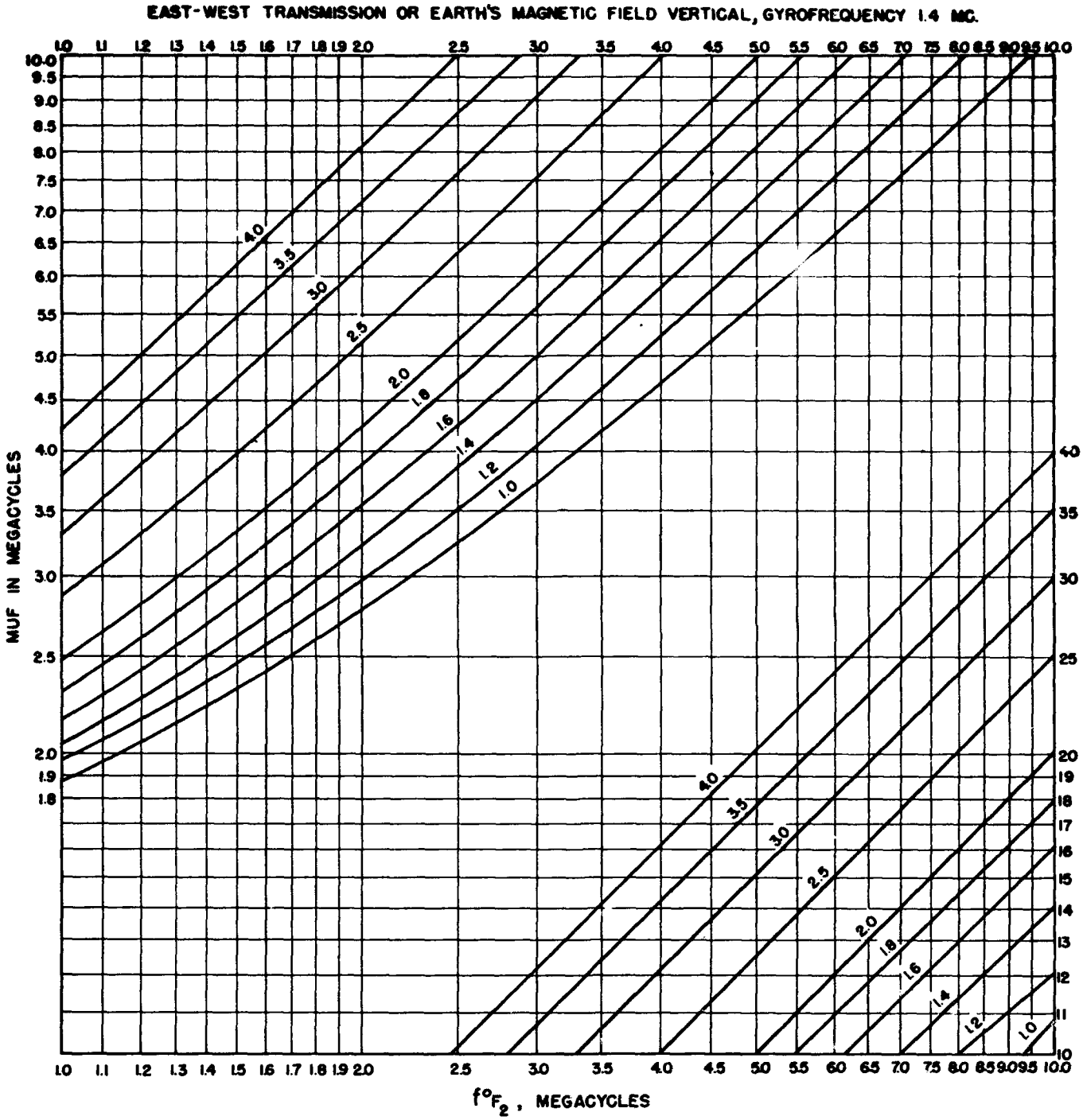


FIGURE 6.20. *Sample ISIB chart for obtaining muf from f°F2 and muf factors.*

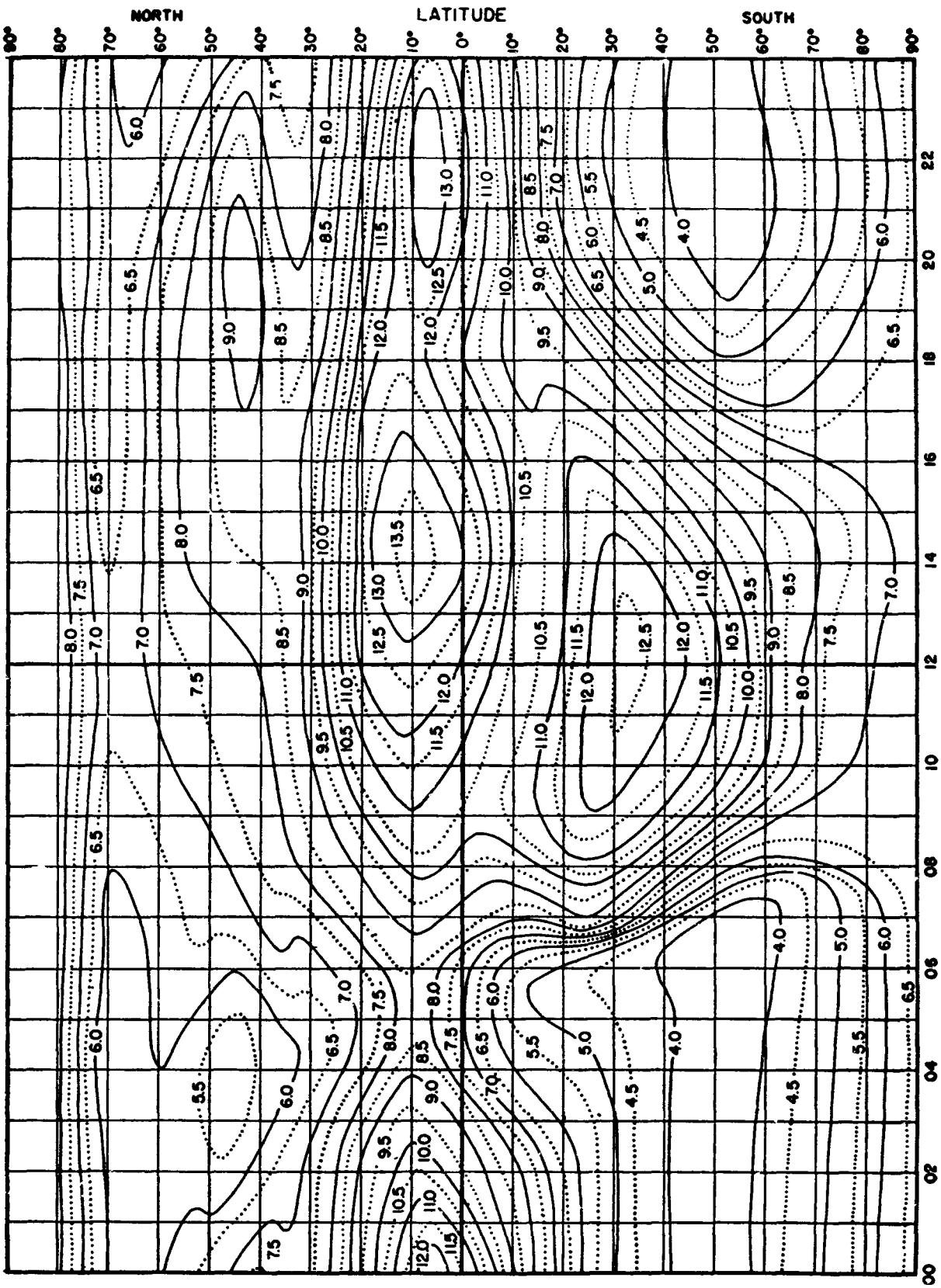


FIGURE 6.21. *F2 zero-muf, in megacycles, W zone, predicted for June 1947.*

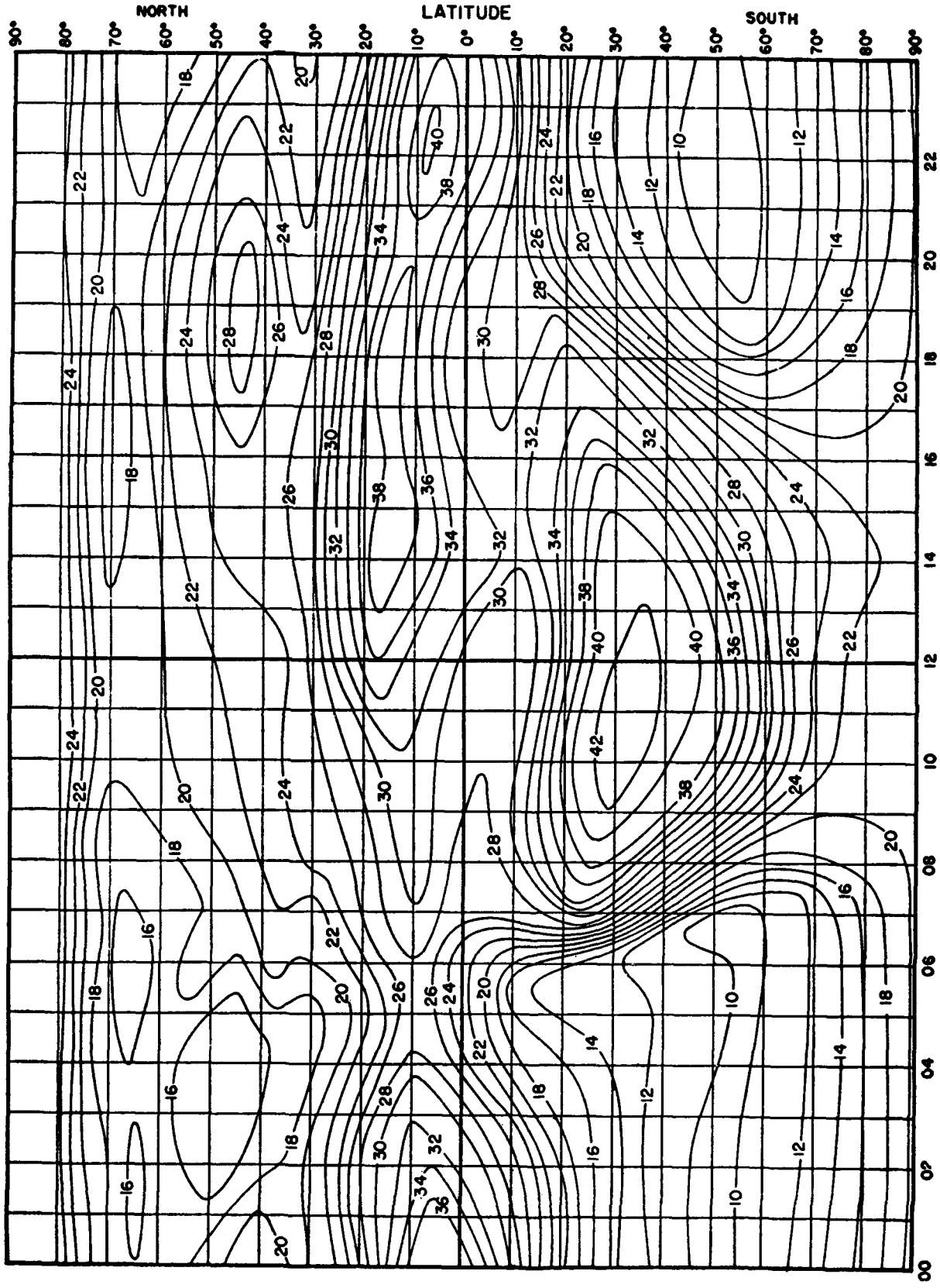


FIGURE 6.22. *Fs 4000-muf, in megacycles, W zone, predicted for June 1947.*

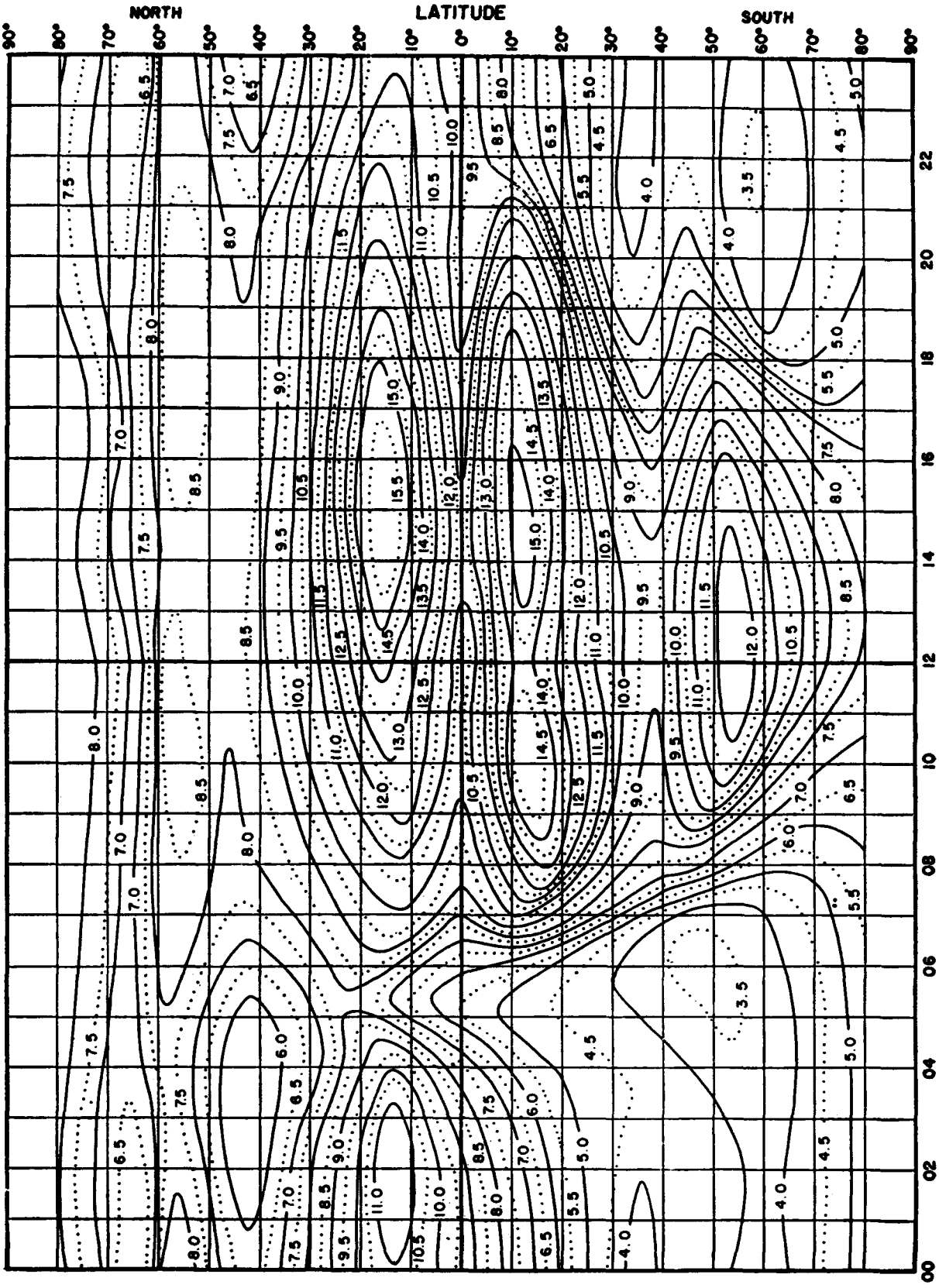


FIGURE 6.23. *F2 zero-muf, in megacycles, I zone, predicted for June 1947.*

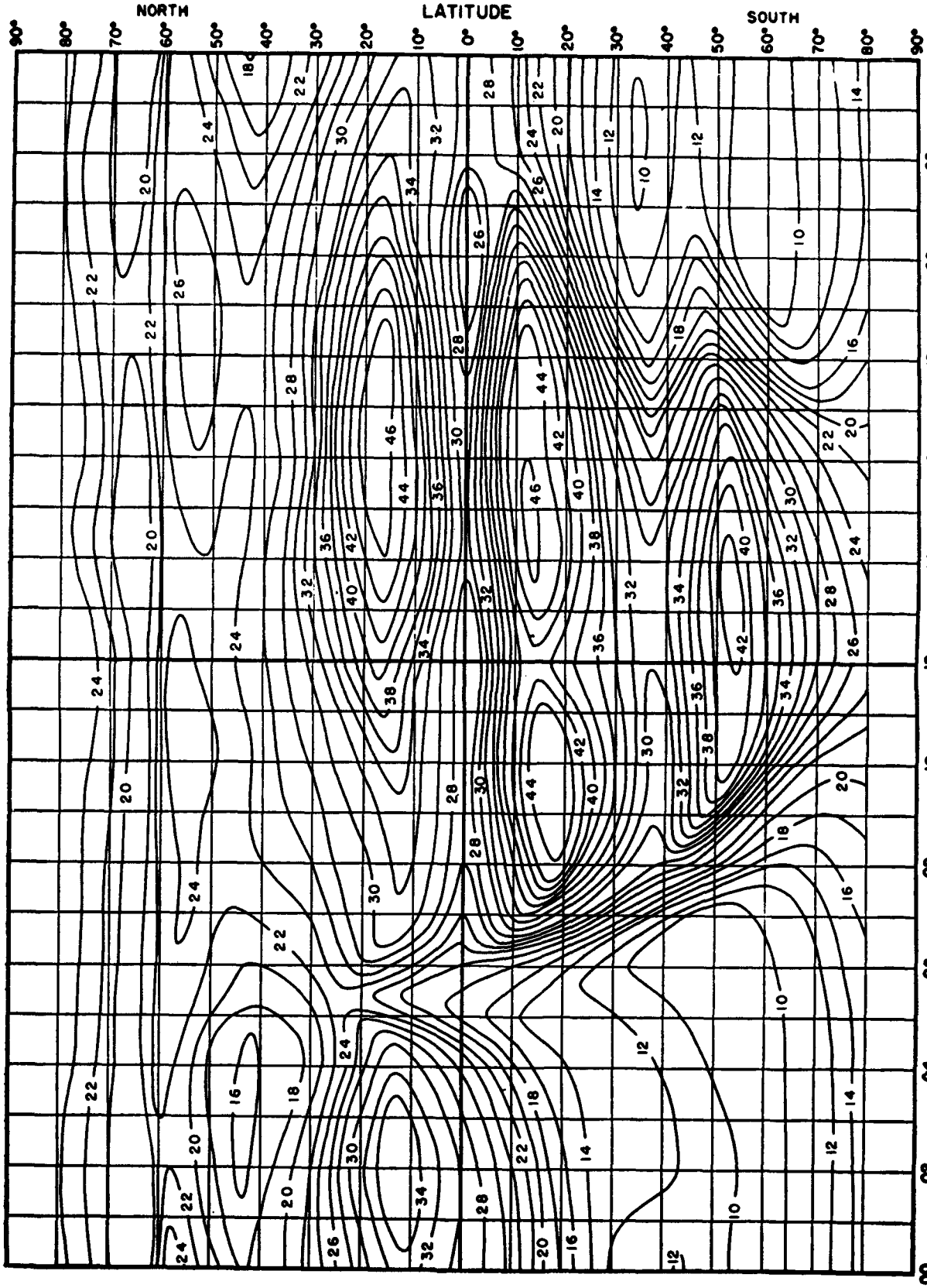


FIGURE 6.24. *Fs 4000-muf, in megacycles, I zone, predicted for June 1947.*

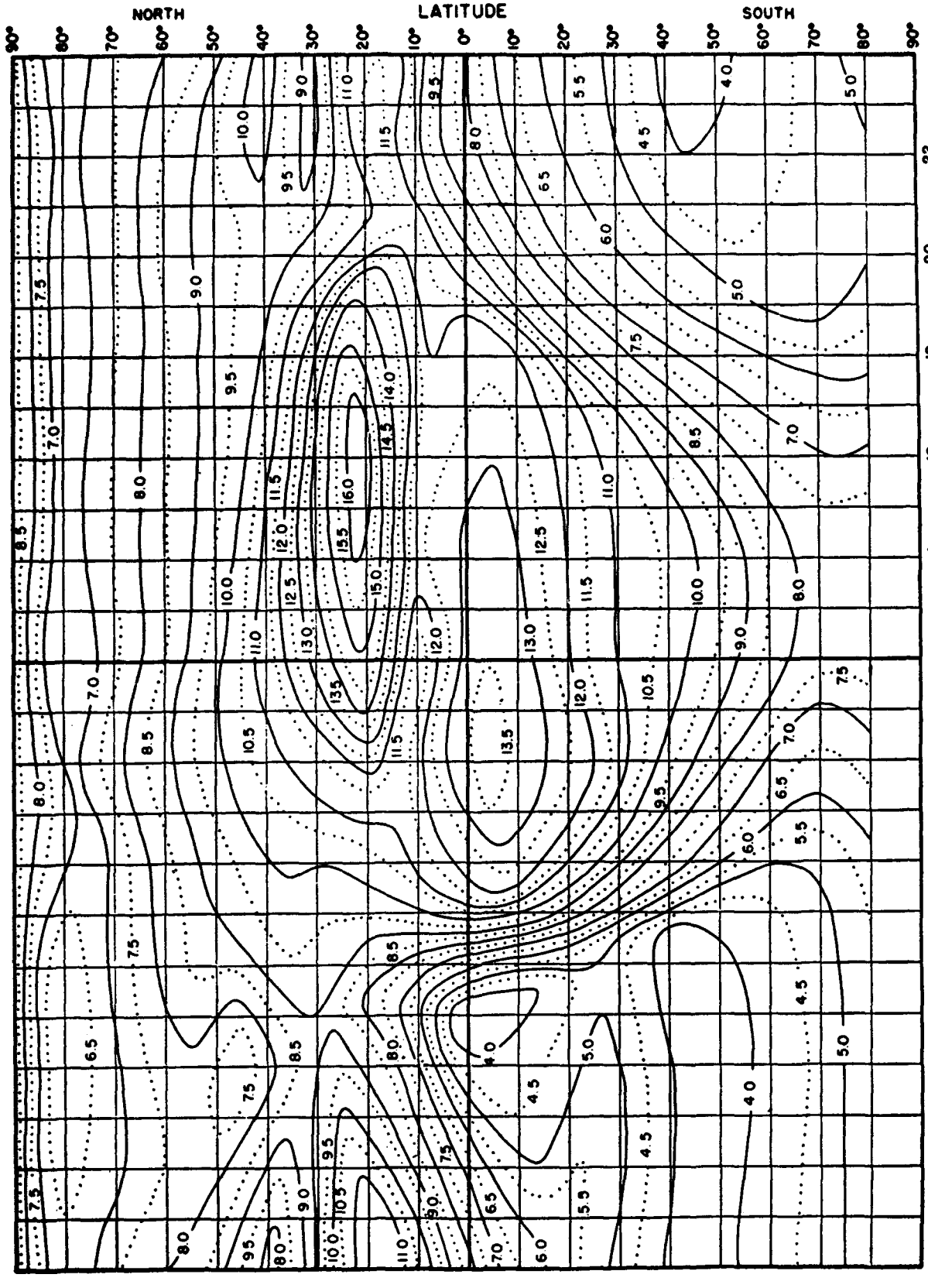


FIGURE 6.25. *F2 zero-muf, in megacycles, E zone, predicted for June 1947.*

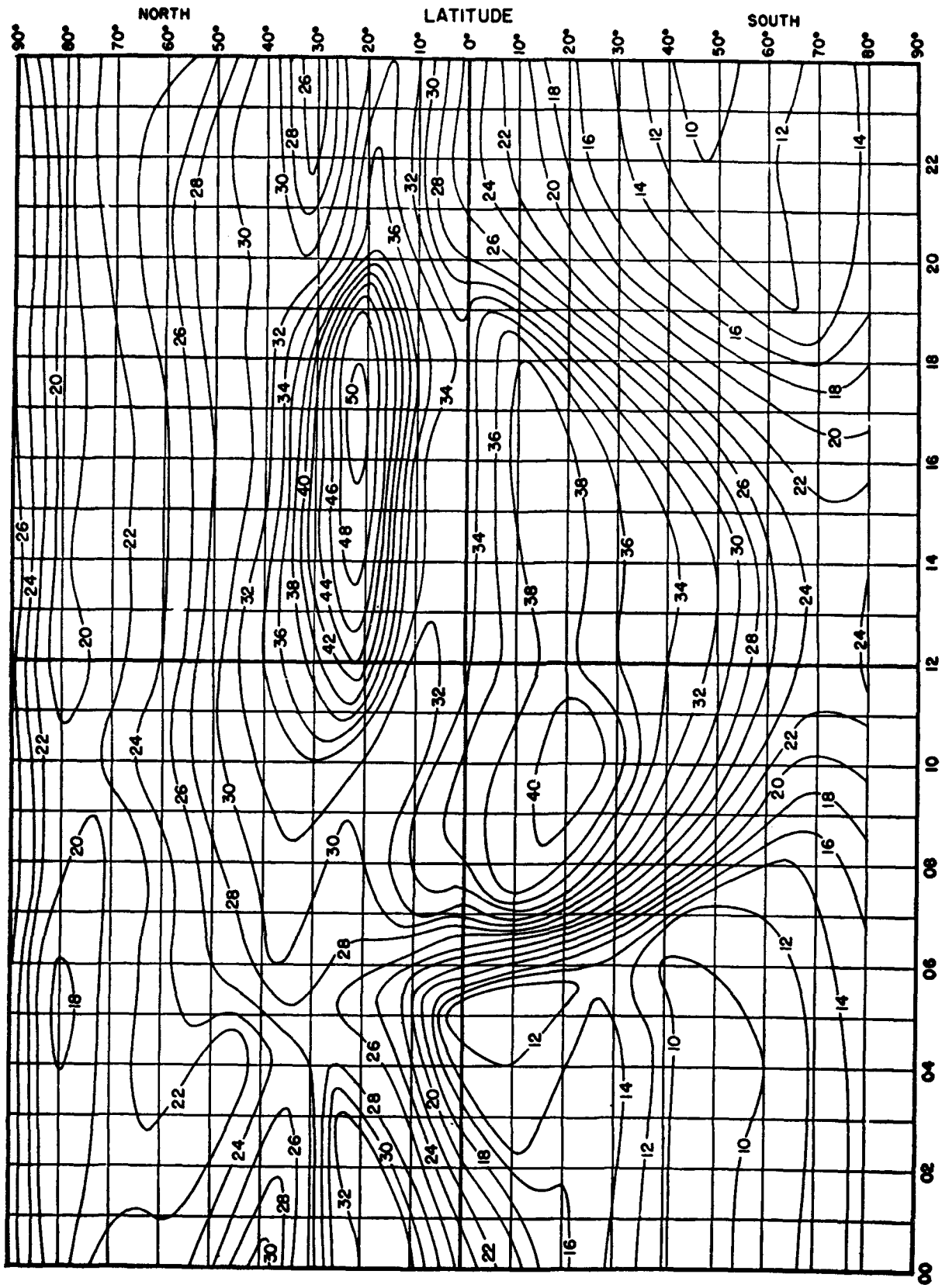


FIGURE 6.26. *F2 4000-muf, in megacycles, E zone, predicted for June 1947*

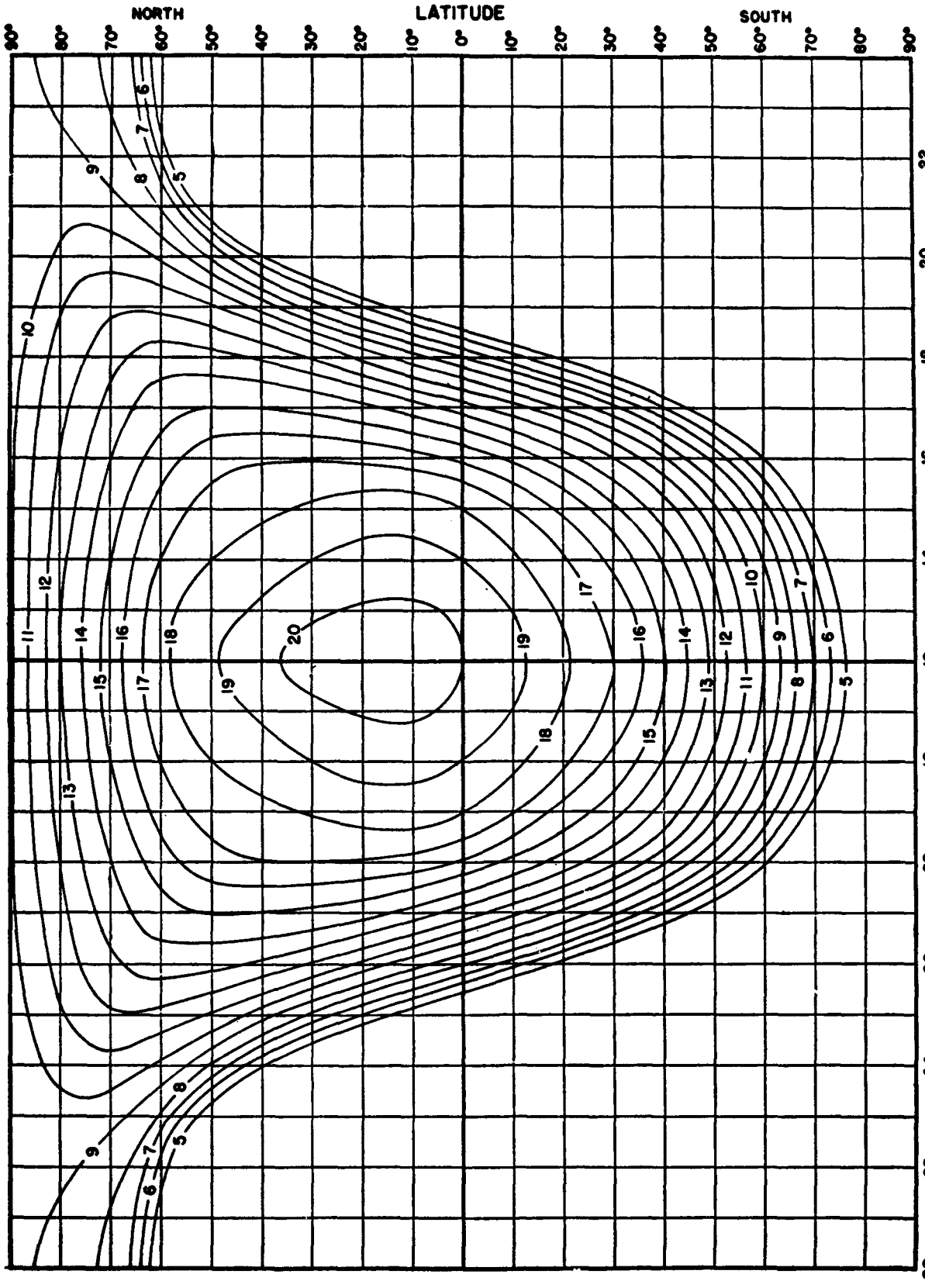


FIGURE 6.27. *E-layer, 2,000 muf, in megacycles, predicted for June 1947.*

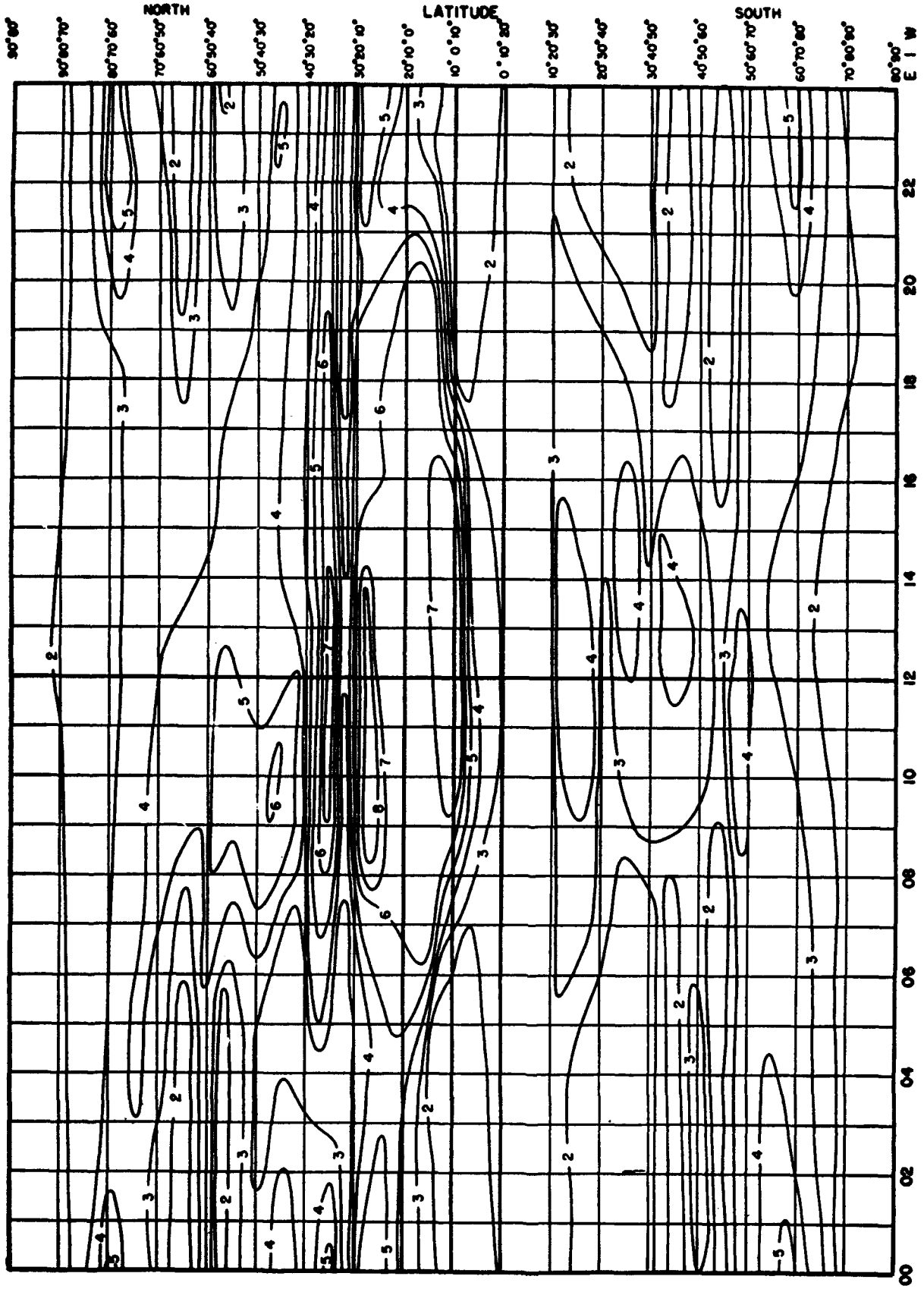


FIGURE 6.28. *Median fEs, in megacycles, predicted for June 1947.*

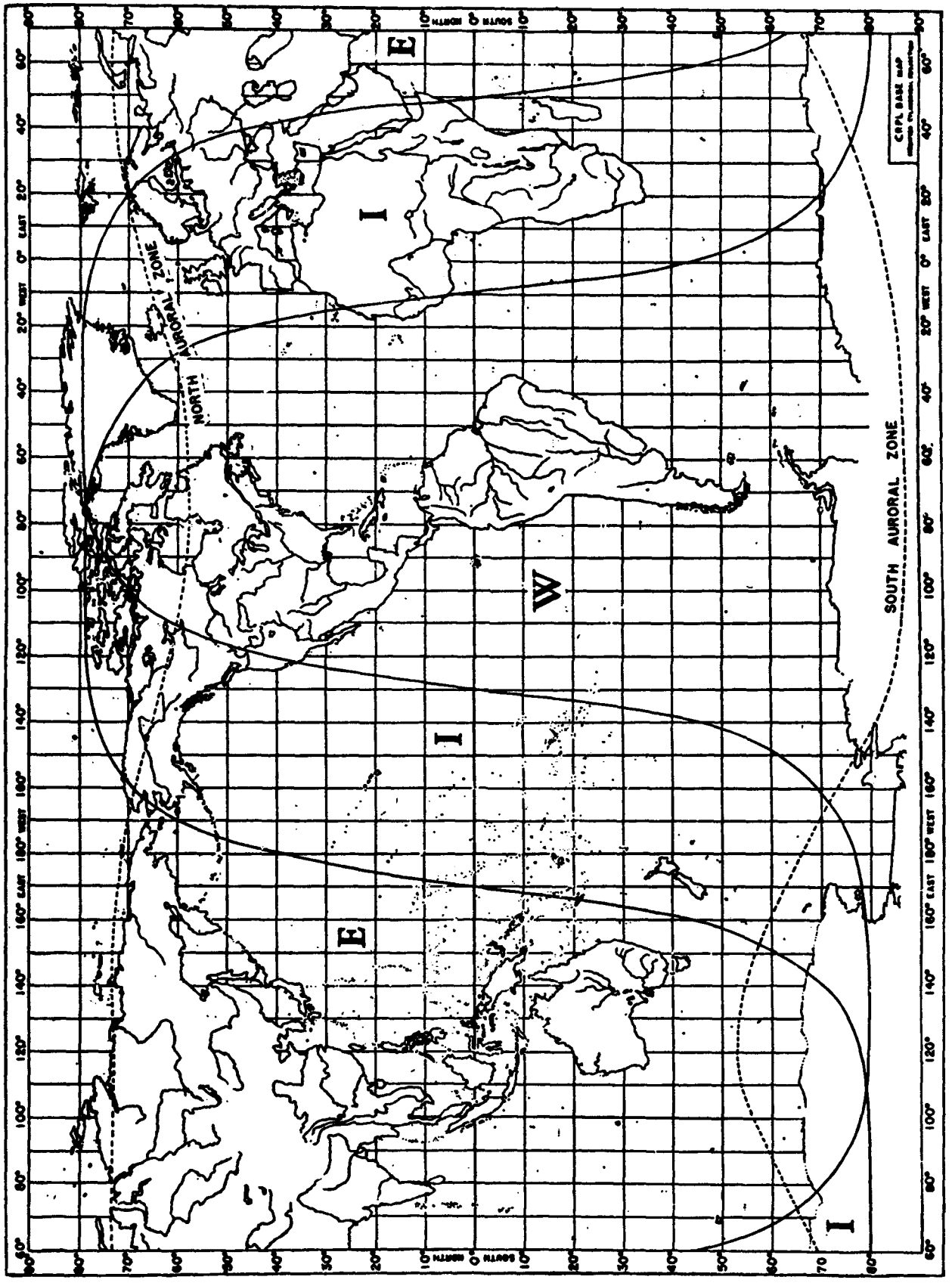


FIGURE 6.29 *World map showing zones covered by predicted charts, and auroral zones.*

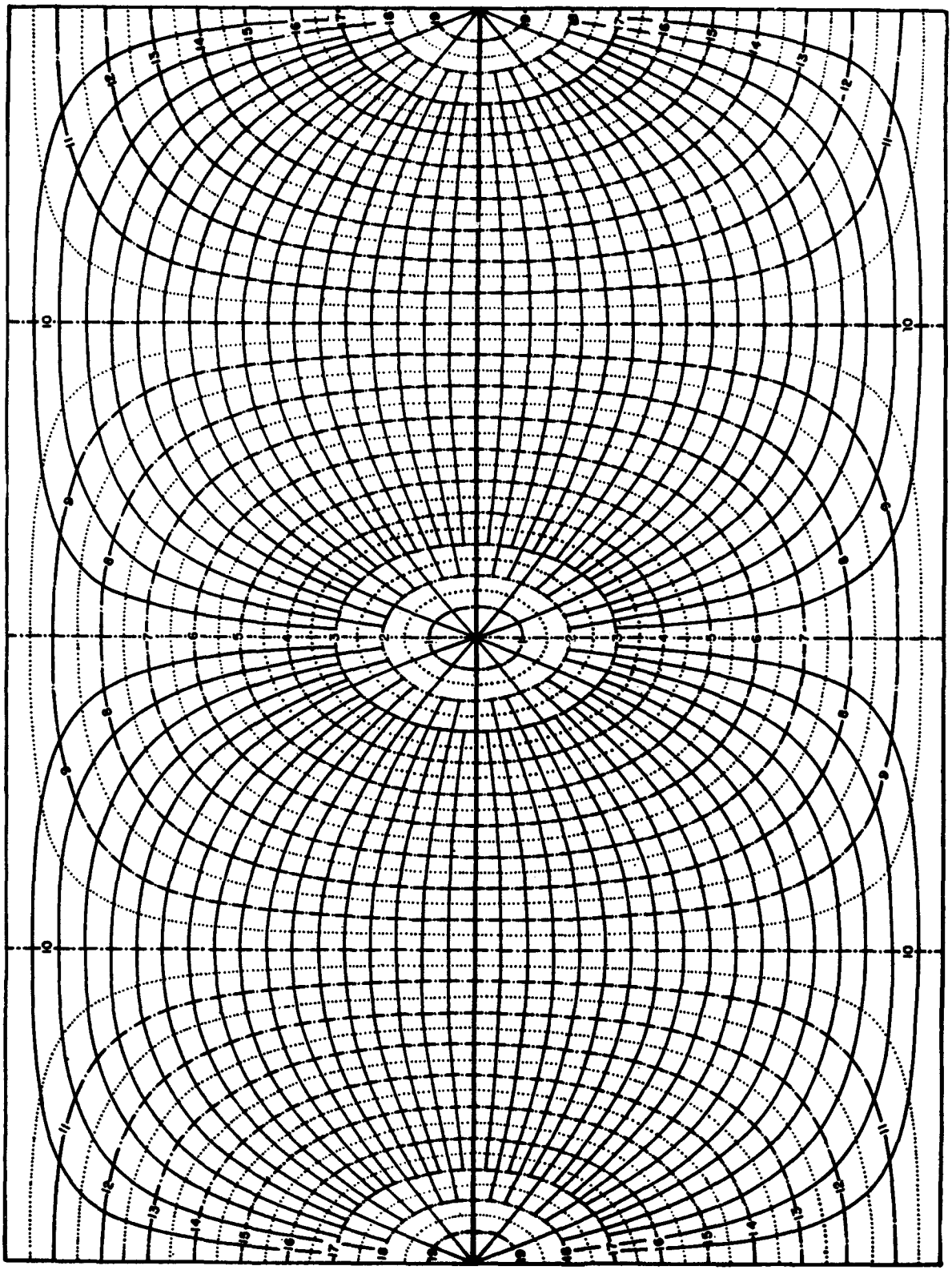


FIGURE 6.30. *Great circle chart centered on equator. Solid lines represent great circles. Numbered dot-dash lines indicate distances in thousands of kilometers.*

For $F2$ layer, points A and B, 2,000 km from each end.

For E layer, points A' and B', 1,000 km from each end.

b. Calculation of Muf for Transmission by the F2 Layer Alone, for Distances of 4,000 Km or Less (Single-Hop Transmission)

(1) Locate the midpoint of the transmission path by using the method of a above. Lay the great-circle path transparency back on the world map of figure 6.29, and, with the ends of the path in their proper location, determine in which geographical zone, E, I, or W, the midpoint falls.

(2) Place the great-circle transparency over the F2-zero-muf chart (the world chart for the calculation of muf for transmission by F2 layer at zero distance) for the proper zone of the midpoint of the path, and, keeping the equatorial line of the transparency over the equatorial line of the chart, slide the transparency horizontally until the Greenwich meridian coincides with either 00 or 24 (not labeled) on the time scale. Note that all points on the great-circle path are in their proper local time relationship to Greenwich because 24 hr on the time scale of a muf chart is drawn to the same scale as 360° of longitude on the world map.

(3) Read the value of $F2$-zero-muf for the midpoint of the path.

(4) Repeat steps (2) and (3) for the $F2$-4000-muf chart for the proper zone.

(5) Place a straightedge between the values of $F2$-zero-muf and $F2$-4000-muf at the left- and right-hand sides, respectively, of the grid nomogram, figure 6.17, which was described under section 6.5 above, and read the value of the muf for the actual path length at the intersection point of the straightedge with the appropriate vertical-distance line.

Example: $F2$-zero-muf=6.8 Mc. $F2$-4000-muf =23.0 Mc. For a distance of 3,000 km the $F2$ muf is 20.8 Mc.

(6) Repeat steps (2), (3), (4), (5), for 01, 02, etc., on the time scale. At times it will be necessary to make the Greenwich meridian of the transparency coincide with an imagined 25, 26, etc., on the time scale. A convenient aid is to place marks at 1-hr intervals on the equatorial line of the transparency.

(7) As an example, the following values (in Mc) are given for transmission by the $F2$ layer between Washington, D. C. (39.0° N. 77.5° W.). and Miami, Fla. (25.7° N. 80.5° W.) for average conditions during the month of June 1947 (figs. 6.21 and 6.22).

GCT	00	02	04	06	08	10	12	14	16	18	20	22
F2-muf for path	14.7	13.4	12.8	12.1	11.4	10.5	12.7	14.8	15.4	15.7	16.3	15.8

c. Calculation of Muf for Transmission by the Regular Layers (F2, E-F1) for Distances of 4,000 Km or Less

(1) Calculate the values of muf for transmission by the $F2$ layer, as in b above.

(2) Place the great-circle transparency over the E-layer 2000-muf chart, figure 6.27, and sliding the transparency in accordance with the instructions of b above, read off the hourly values of muf for the midpoint of the path.

(3) These values are for transmission by the E layer for a distance of 2,000 km. To convert them to values of muf for the actual transmission distance and to provide for the inclusion of the effects of the $F1$ layer, the nomogram of figure 6.18 is used, described in section 6.5. For each hour place a straightedge between the value of the E-layer 2000-muf on the left-hand scale of the nomogram, figure 6.18, and the value of the path length on the right-hand scale, and read the E-$F1$-muf for that path length, off the central scale. (Example on nomogram).

(4) It is assumed that the higher of the two values obtained, one for transmission by the $F2$ layer, the other for transmission by the E-$F1$ combined layer, is the muf for the path at any given hour.

(5) As an example, the following values (in Mc) are given for transmission by the regular layers between Washington, D. C. and Miami, Fla., for average conditions during the month of June 1947 (figs. 6.21, 6.22, and 6.27).

GCT	00	02	04	06	08	10	12	14	16	18	20	22
F2-muf	14.7	13.4	12.8	12.1	11.4	10.5	12.7	14.8	15.4	15.7	16.3	15.8
E-*F1* muf	8.9	----	----	----	----	6.2	13.1	16.4	18.0	18.4	17.0	14.3
Muf for path (regular layers)	14.7	13.4	12.8	12.1	11.4	10.5	13.1	16.4	18.0	18.4	17.0	15.8

d. Calculation of Muf for Distances of 4,000 Km or Less; Provision for the Effects of Es

(1) Calculate the values of muf for transmission by the $F2$, E-$F1$ layers, as in a, b, c above.

(2) Place the great-circle path transparency over the median fEs chart, figure 6.28, being careful to use the latitude scale for the zone containing the control point. Read off the values of fEs for the midpoint of the path for the desired hours. Multiply each value by 5, thus obtaining the Es-2000-muf.

(3) By use of the nomogram, figure 6.18, convert the Es-2000-muf into the Es-muf for the proper distance.

(4) It is assumed that the highest of the three values obtained: one for transmission by the $F2$ layer, one for transmission by the combined E-$F1$ layer, and the third for transmission by sporadic E, is the muf for the path at any given hour.

(5) As an example, the following values (in

Mc) are given for transmission between Washington and Miami for June 1947, (figs. 6.21, 6.22, 6.27, and 6.28):

GCT	00	02	04	06	08	10	12	14	16	18	20	22
*F*2-muf	14.7	13.4	12.8	12.1	11.4	10.5	12.7	14.8	15.4	15.7	16.3	15.8
E-*F*1-muf	8.9					6.2	13.1	16.4	18.0	18.4	17.0	14.3
Es-muf	18.4	14.7	14.7	14.7	14.3	14.3	18.1	25.1	25.1	21.8	19.8	19.3
Muf for path	18.4	14.7	14.7	14.7	14.3	14.3	18.1	25.1	25.1	21.8	19.8	19.3

e. Calculation of Muf for Transmission by the F2 Layer Alone, for Distances Greater than 4,000 Km (Multihop Transmission)

(1) Locate two control points, *A* and *B*, 2,000 km from each end of the path, on the great-circle transparency for the transmission path, constructed in accordance with the instructions given above.

(2) Determine, by placing the transparency on the world map of figure 6.29, in which geographical zone, E, I, or W, each control point falls.

(3) Place the great-circle transparency over the *F*2–4000–muf chart for the zone of control point *A* and, keeping the equatorial line of the transparency over the equatorial line of the chart, slide the transparency horizontally until the Greenwich meridian coincides with either 00 or 24 (not labeled) on the time scale.

(4) Read the value of *F*2–4000–muf for control point *A*.

(5) Repeat steps 3 and 4 for 01, 02, etc., on the time scale.

(6) Repeat steps 3, 4, 5 for point *B*, making sure that the chart for the proper zone is used.

(7) The lower of the two values found at any given hour for control points *A* and *B*, is the muf for transmission over the path by the *F*2 layer, at that hour.

(8) As an example, the following values (in Mc) are given for transmission by the *F*2 layer between Washington, D. C. (39.0° N. 77.5° W.) and Trieste (45.7° N. 13.8° E.) for average conditions during the month of June 1947, between the hours of 0800 and 1400 GCT (figs. 6.22 and 6.24).

GCT	Control point *A* (W-zone) *F*2-4000-muf	Control point *B* (I-zone) *F*2-4000-muf	*F*2-muf for path
0800	15.6	23.6	15.6
1000	18.6	24.2	18.6
1200	20.0	24.2	20.0
1400	21.5	23.0	21.5

f. Calculation of Muf for Transmission by the Regular Layers (F2, E-F1), for Distances Greater than 4,000 Km

(1) Locate on the transparency two control points for *E*-layer transmission, *A′* and *B′*, 1,000 km from each end of the great-circle path, the point *A′* being on the same end of the path as point *A* of e above.

(2) Place the great-circle transparency over the *E*-layer 2000-muf chart, figure 6.27, and, sliding the transparency in accordance with the instructions of e above, read off the hourly values of muf for control point *A′*. Compare results with the values for the corresponding hours obtained in e above, for control point *A*. The higher of each pair of values is taken as the muf for transmission by the regular layers (*F*2, *E*–*F*1) at the *A*-end of the transmission path.

(3) Repeat for control point B′ and compare results with those previously obtained for control point *B*. The higher of each pair of values thus obtained is taken as the muf for transmission by the regular layers (*F*2,*E*-*F*1) at the *B*-end of the transmission path.

(4) The muf for the transmission path at each hour is found by comparing the muf for the *A*-end with the muf for the *B*-end, and taking the lower of the two values.

(5) As an example, the following values (in Mc) are given for transmission by the regular layers between Washington and Trieste for average conditions during the month of June 1947, between the hours of 0800 and 1400 GCT (figs. 6.22, 6.24, and 6.27).

GCT	Control point *A* (*W*-zone) *F*2-4000-muf	Control point *A′* *E*-2000-muf	Muf *A*-end	Control point *B* (*I*-zone) *F*2-4000-muf	Control point *B′* *E*-2000-muf	Muf *B*-end	Muf for path
0800	15.6		15.6	23.6	16.7	23.6	15.6
1000	18.6	11.5	18.6	24.2	18.1	24.2	18.6
1200	20.0	15.8	20.0	24.2	18.9	24.2	20.0
1400	21.5	17.9	21.5	23.0	18.0	23.0	21.5

g. Calculation of Muf for Distances Greater Than 4,000 Km; Provision for the Effects of Es

(1) Determine, by placing the transparency on the world map of figure 6.29, in which geographical zone, E, I, or W, each control point *A′* and *B′* falls.

(2) Place the great-circle transparency over the median *fEs* chart, figure 6.28, using the latitude scale for the zone containing the control point *A′*. Read off the values of *fEs* for control point *A′*, for the desired hours. Multiply each value by 5, thus obtaining the *Es*-2000-muf at the *A*-end of the path.

(3) It is assumed that the highest of the three values obtained at the *A*-end: one, for transmission by the *F*2 layer at control point *A*, as outlined under e above, one for transmission by the combined *E*-*F*1 layer at control point *A′*, as outlined under f above, and the third for transmission by sporadic *E* at control point *A′*, as outlined in (1) and (2) immediately above, is the muf for the *A*-end of the path at any given hour.

(4) In a similar manner, obtain the muf for the *B*-end of the path.

(5) The muf for the transmission path is found by comparing the muf for the *A*-end with the muf for the *B*-end, and taking the lower of these two values at any given hour.

(6) As an example, the following values (in Mc) are given for transmission between Washington and Trieste for June 1947 (figs. 6.22, 6.24, 6.27, and 6.28).

GCT	Control point *A* (W-zone) *F*2-4000-muf	Control point *A'* (W-zone)			Muf *A*-end
		E-2000-muf	*fEs*	*Es*-2000-muf	
0800	15.6				15.6
1000	18.6	11.5			18.6
1200	20.0	15.8	2.9	14.5	20.0
1400	21.5	17.9	4.6	23.0	23.0

	Control point *B* (I-zone) *F*2-4000-muf	Control point *B'* (I-zone)			Muf	
		E-2000-muf	*fEs*	*Es*-2000-muf	*B*-end	For path
0800	23.6	16.7	5.0	25.0	25.0	15.6
1000	24.2	18.1	5.0	25.0	25.0	18.6
1200	24.2	18.9	4.9	24.5	24.5	20.0
1400	23.0	18.0	4.2	21.0	23.0	23.0

h. Relative Importance of the Layers in Affording Transmission

For short paths, transmission is frequently controlled by the *E* layer, or by sporadic *E*, especially during periods of relatively low solar activity. For long paths, the *F*2 layer is almost always controlling. Thus for transmission distances over 4,000 km in length, it is usually necessary only to calculate the *F*2-4000-muf at control points 2,000 km from each end of the path, and to use the lower of each pair of values for a given hour.

i. Reliability of Layers in Affording Transmission; Optimum Working Frequency (Owf)

In the following discussion by owf is meant that frequency effective for communication 90 percent of the time during undisturbed periods.

The most reliable layer is the *E* layer; in fact, the day-to-day variation in critical frequency from the median is so small that the optimum working frequency for transmission by the combined *E*-*F*1 layer is taken as equal to the maximum usable frequency for the *E*-*F*1 layer.

The optimum working frequency for transmission by the *F*2 layer is taken as 85 percent of the *F*2-muf, in order to allow for day-to-day variations in critical frequency and virtual height from the median values of these characteristics. Thus, cases may arise, over a given path, in which, although the *F*2-muf is higher than the *E*-*F*1-muf, the *E*-*F*1-owf is higher than the *F*2-owf.

Because of the variable nature of *Es*, and the relative uncertainty with which it is known, it is often desirable to operate near the optimum working frequency for the regular layers (*F*2, *E*-*F*1) only, without the inclusion of *Es*.

At present the *Es*-2000-owf is found by subtracting 4 Mc from the *Es*-2000-muf. In determining the *Es*-owf for a path 4,000 km long or less, it is necessary to make use of the nomogram of figure 6.18, in order to convert the *Es*-2000-owf into the *Es*-owf for the proper distance.

In finding the over-all owf for a path falling into any of the groups considered in previous paragraphs of this section, calculate the muf for each layer for each control point as previously described. Then in the case of paths 4,000 km or less in length, choose the over-all owf as the highest of the owf by *F*2, by *E*-*F*1, and by *Es*, exactly as was done for the muf. For paths over 4,000 km in length, choose the owf for each path end as the highest of the owf by *F*2, by *E*-*F*1, and by *Es*. Then choose the over-all owf as the lower of the owf for each end.

During periods of ionospheric disturbance (section 5.8), the critical frequencies are lower than usual. Consequently, it may be necessary, at such times, to lower the operating frequency in order to insure communication. When, however, the frequency is lowered to the extent that the energy of the wave is completely absorbed (section 2.6) communication is impossible. Since the "ionosphere-storm" type of disturbance is most severe in polar regions, with lessening severity with decreasing latitude, relaying through points within about 35° of the equator is often indicated.

These measures do not suffice in the case of so-called "sudden ionosphere disturbances", which are characterized by simultaneous fadeouts in the entire high-frequency spectrum, on paths in the daylight side of the world. In contrast to the "ionosphere-storm" type, such disturbances are unpredictable, but fortunately rarely last more than two hours. Recovery is usually first at the higher frequencies.

6.7 References

Newbern Smith, The relation of radio sky-wave transmission to ionosphere measurements, Proc. Inst. Radio Engrs. **27**, 332 (1939).

E. V. Appleton and W. J. G. Beynon, The application of ionospheric data to radio communication problems. Part I (May 6, 1940), and part II (June 12, 1942), Radio Department, National Physical Laboratory, Teddington, England.

Note—Part I appears in Proc. Phys. Soc. **52**, 518 (1940); part 2 was originally a classified Radio Research Board Paper R. R. B. /C. 52.

T. L. Eckersley, Studies in radio transmission, J. Inst. Elec. Engrs. **71**, 405 (1932).

T. L. Eckersley, Analysis of the effect of scattering in radio transmission, J. Inst. Elec. Engrs. **86**, 548 (1940).
T. L. Eckersley and F. T. Farmer, Short period fluctuations in the characteristics of wireless echoes from the ionosphere, Proc. Roy. Soc. (London) **12**, 146 (1941).
C. B. Feldman, Deviation of short radio waves from the London-New York great circle path, Proc. Inst. Radio Engrs. **27**, 635 (1939).
M. L. Phillips, The ionosphere as a measure of solar activity, Report Interservice Radio Propagation Laboratory-R26 (November 15, 1945).
Methods used by Interservice Radio Propagation Laboratory for the prediction of ionospheric characteristics and maximum usable frequencies, Report Interservice Radio Propagation Laboratory-R4 (December 31, 1943).
M. L. Phillips, A nomographic method for both prediction and observation correlation of ionosphere characteristics, Report Interservice Radio Propagation Laboratory-R11 (May 15, 1945).
Predicted limits for F2-layer radio transmission throughout the solar cycle, Report Interservice Radio Propagation Laboratory-R15 (July 12, 1945).
Report on Japanese research on radio wave propagation, GHQ, USAFP, OCSigO, Tokyo **1**, 22 (May 1946).
Commonwealth of Australia, Australian Radio Propagation Committee of the Radio Research Board: Ionospheric Bulletin, p. 4–6 (January 1944).
ARPC–A series: Radio propagation bulletin (monthly).
ARPC–C series: Maximum usable frequencies and optimum working frequencies (yearly).
Canadian Radio Wave Propagation Committee Report JCC/WP/1P/No. 11, Radio frequency prediction for Canada 1946–55.
England: Department of Scientific and Industrial Research, National Physical Laboratory, Teddington: Predictions of radio wave propagation conditions (monthly).
France: Service des Etudes, Direction Centrale des Constructions et Armes Navales: Prévision de la Propagation Radioélectrique (monthly).
Soviet Union: Scientific Research Institute of Terrestrial Magnetism: "Temporary instructions for the determination of maximum usable frequencies", Moscow 1945.
China: Radio Wave Research Laboratories, Central Broadcasting Administration: Short Wave Radio Frequency Predictions for China in Summer Season of 1947, Chungking April 1947.
A. G. McNish and J. V. Lincoln, Prediction of Annual Sunspot Numbers, Report Central Radio Propagation Laboratory–CRPL–1–1 (May 15, 1947).

CHAPTER 7

IONOSPHERIC ABSORPTION AND SKY-WAVE FIELD INTENSITY

7.1. Modes of Propagation

a. General

Sky-wave propagation between two points on the earth's surface takes place by combinations of reflections from the ionosphere layers and the earth's surface. A particular combination of reflections constitutes a mode of propagation.

The various ionosphere layers, which act as reflectors for radio waves, were described in chapter 4. Examples of simple modes of propagation are the 1-hop-$F2$ and 1-hop-E modes, which involve a single reflection from the $F2$ and E layers, respectively, at the midpoint of the path. The 2-hop-$F2$ mode consists of two single $F2$ hops laid end to end. Similarly, the 2-hop-E mode is composed of two single E hops laid end to end. Various other modes are possible, involving the $F1$ layer as well as the E and $F2$ layers, and involving reflections from more than one layer in a single mode. However, for the purposes of this chapter, modes involving the $F1$ layer and modes involving more than one layer at a time may be ignored.

The 1-hop-$F2$, 2-hop-$F2$, and 1-hop-E modes are illustrated in figure 7.1. It is evident that each mode involves a particular radiation angle, Δ (vertical angle of departure), which depends upon the height of the reflecting layer, the length of the transmission path, and the number of hops, and that for a particular mode the angles at which the wave path penetrates layers beneath the reflecting layer are fixed. It is also evident that the vertical angle of arrival at the receiving station is equal to the radiation angle for the same mode, unless the reflecting layer is tilted.

As the path length increases, the radiation angle and angle of arrival decrease. The limiting path length for a particular mode is the distance for which this angle is zero. The limiting distances for the 1-hop-E and 1-hop-$F2$ modes are approximately 2,400 and 4,000 km, respectively. The limiting distances for multihop modes are these distances multiplied by the number of hops.

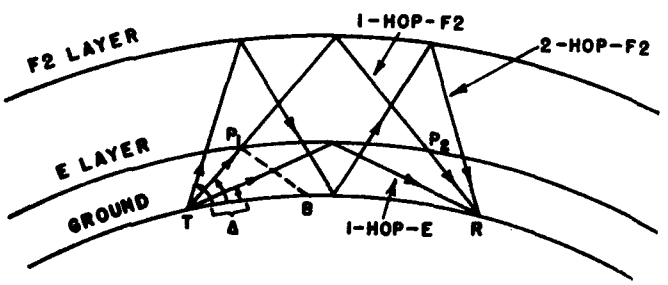


FIGURE 7.1. *Modes of propagation.*

The radiation angle, Δ, is plotted as a function of distance for various modes of propagation in figure 7.3. These angles are based upon virtual E- and $F2$-layer heights of 105 and 320 km, respectively.

b. Active Modes of Propagation

A mode of propagation is said to be "active" if the path length is less than the limiting distance for the mode, and the wave frequency is less than the muf for the mode, and greater than the "E-layer cut-off frequency" in the case of $F2$-layer modes. Procedures for determination of the muf for single-hop modes are discussed in section 6.6. The muf for a two-hop mode is simply the lower of the two mufs obtained for each hop separately. The E-layer cut-off frequency is discussed below.

c. E-Layer Cut-Off Frequency for $F2$-Layer Modes

Since the wave path for any mode involving $F2$-layer reflections penetrates the E layer at one or more pairs of points (see fig. 7.1), in order for a wave to traverse the entire path, the frequency must be high enough to penetrate the E layer at each of these points. The lowest such frequency is called the E-layer cut-off frequency.

Referring to figure 7.1, a wave traveling along the 1-hop-$F2$ wave path, starting at T, first strikes the E layer at P_1. If the frequency is below the E-layer penetration frequency, the wave is reflected and travels along the dotted line, returning to earth at the point B. This constitutes a 1-hop-E mode for the path TB, the radiation angle of which is the same as that for the 1-hop-$F2$ mode. The penetration frequency for the point P_1 is therefore equal to the E-layer muf for the path TB.

To find the cut-off frequency for a particular $F2$-layer mode, proceed as follows:

1. Determine the radiation angle for the $F2$ mode considered by using the chart in figure 7.3.
2. On the same chart, and at the same radiation angle, read the distance corresponding to the 1-hop-E mode. Call this distance "x". (In fig. 7.1, $x=TB$ for the 1-hop-$F2$ mode.)
3. Mark the E-layer penetration points on a great-circle curve constructed for the path in the manner described in section 6.6, in connection with muf calculations. These points are located a distance $x/2$ from each end of the path and on both sides of each ground reflection point.
4. Following the procedure in section 6.6, determine the E-layer muf corresponding to a

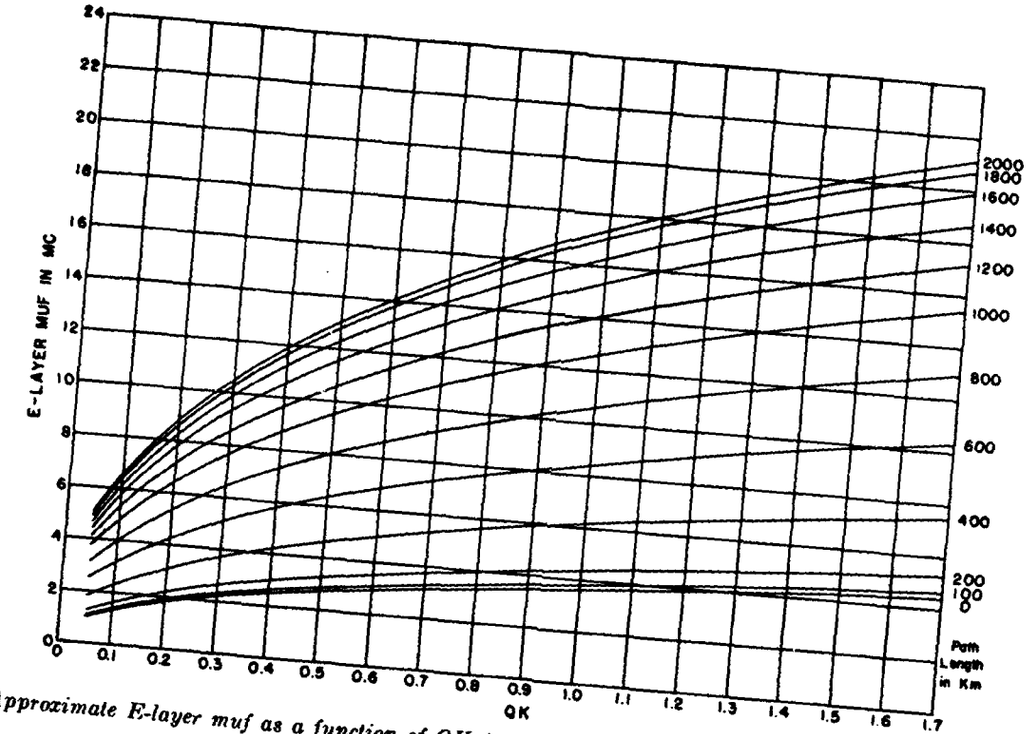


FIGURE 7.2. *Approximate E-layer muf as a function of QK (product of solar-cycle variation factor Q and diurnal variation factor K) and the path length.*

distance "x" for each penetration point. The highest of these is the cut-off frequency for the mode of propagation considered.

For paths less than about 4,000 km in length, it is usually sufficiently accurate to find the E-layer muf corresponding to the distance "x" for the midpoint of the path, and to use this value as the cut-off frequency.

The E-layer muf for a distance "x" may be expressed to a close approximation as a function of the product QK, K and Q being the diurnal and solar cycle absorption variation factors discussed in sections 7.5, d, and 7.5, e. Curves representing this function are presented in figure 7.2.

7.2. Radiated Power

a. Sky-Wave Radiation

In traversing a nonionized region of the atmosphere, practically no energy is lost from the wave, and the only decrease in field intensity is that due to the spreading out of the wave front, the "inverse distance attenuation." The field intensity along a path encountering no obstacles (neither large masses nor ions) and no interfering wave trains, varies inversely as the distance from the emitting source; the energy density in the waves, which is proportional to the square of the field intensity, varies inversely as the square of the distance (the familiar "inverse-square law").

The field intensity produced in free space by an isotropic radiator (one which radiates equally in all directions) radiating 1 kw of power, is given, in volts per meter, by

$$E=\frac{173.2}{d},$$

where d is the distance from the radiator in meters. This value is obtained from the energy density, P_1, at that distance (1,000 watts divided by $4\pi d^2$ square meters) by using the relation $E=\sqrt{377P_1}$. The field intensity produced in free space by any other antenna radiating in free space can be expressed in volts per meter as

$$E_\theta=\frac{173.2}{d}\sqrt{G_{T_\theta}P_0},$$

where P_0 is the radiated power in kilowatts, and G_{T_θ} is the free space gain of the antenna in a given direction over the isotropic radiator. G_{T_θ} can be calculated from the antenna dimensions. Note that P_0 is the actual radiated power (current squared times radiation resistance) and not the power output of the transmitter or the power input to the antenna.

For a short dipole (less than 0.1 wavelength) in free space, $G_{T_\theta}=1.5\sin^2\theta$, where θ is the direction with the axis of the dipole, and thus the field intensity, in volts per meter, produced by such an antenna in the direction θ is

$$E_\theta=212\frac{\sqrt{P_0}}{d}\sin\theta.$$

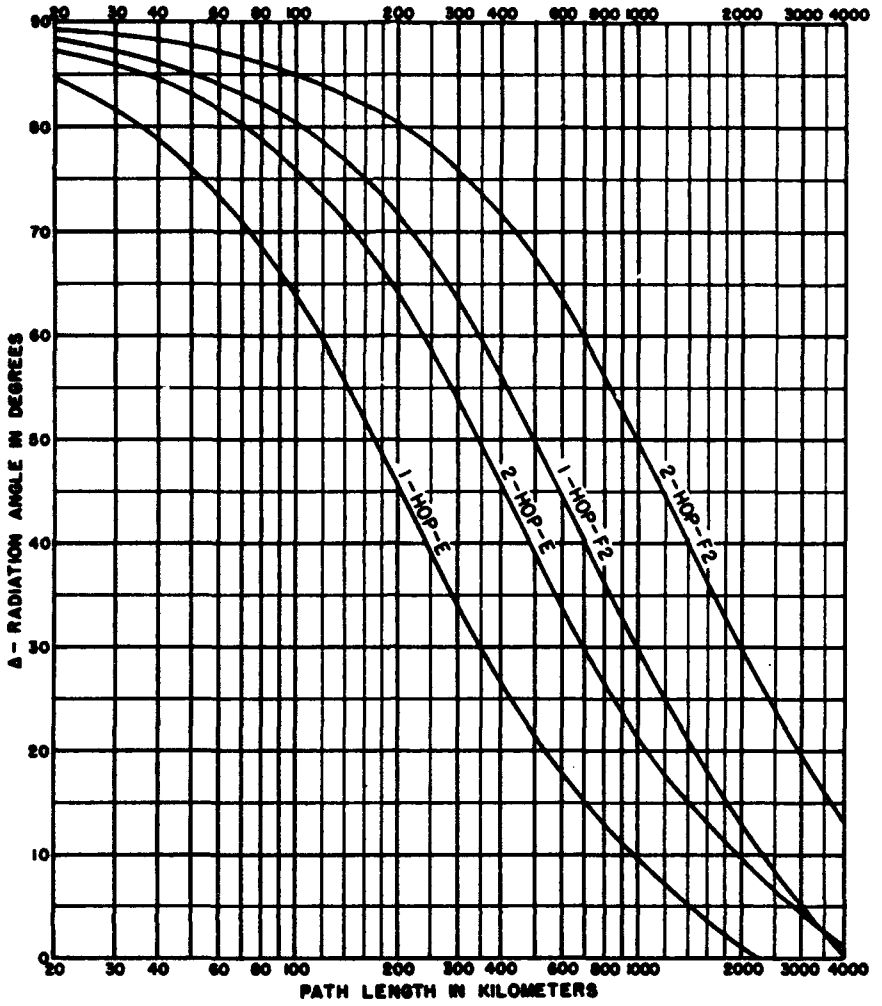


FIGURE 7.3. *Radiation angle versus path length, based on virtual reflection heights, E layer—105 km, F2 layer—320 km.*

The field intensity produced by other antennas is similar in magnitude, but different in detail as to directional pattern, etc. For example, the field intensity at 1 km, due to a short vertical wire, one end of which is on the ground, is given, in microvolts per meter, by

$$E_\theta = 300{,}000\sqrt{P_0}\sin\theta,$$

for perfectly conducting ground. E is somewhat less than this for average ground. Part of the field intensity near the ground (near $\theta=90°$) is due to the surface wave, which is negligible when the wave reaches the ionosphere. Consequently, the vertical directional pattern of an antenna, for sky-wave calculations, is considerably different from the pattern as measured within a kilometer or so of the antenna.

The gain of an antenna for sky-wave radiation is obtained by reducing the gain over a perfect earth by the amount of the ground-wave radiation. To a first approximation, the sky-wave gain G_{T_S} of an antenna whose free space gain is G_T, located at a height h above the surface of the earth, is

$$G_{T_S} = G_{T_\theta}\left(1+Re^{-j\frac{4\pi h}{\lambda}\sin\Delta}\right)^2,$$

where R is the complex plane-wave reflection coefficient of the ground, given by

$$R=\frac{\sin\Delta-u\sqrt{1-u^2\cos^2\Delta}}{\sin\Delta+u\sqrt{1-u^2\cos^2\Delta}} \text{ for vertical antennas}$$

$$R=\frac{u\sin\Delta-\sqrt{1-u^2\cos^2\Delta}}{u\sin\Delta+\sqrt{1-u^2\cos^2\Delta}} \text{ for horizontal antennas}$$

Δ=radiation angle

$$u=\sqrt{\frac{1}{\epsilon-jx}}$$

ϵ=specific inductive capacity of ground

$$x=\frac{18000\sigma}{f}$$

σ=conductivity of ground in mhos per meter

f=frequency in megacycles

$j=\sqrt{-1}$.

For a vertical antenna of length l with one end on the ground G_{T_θ} is that for a dipole of length $2l$ and $h=0$; for any other antenna, G_{T_θ} is that for the antenna in free space and h is the height of the center of the antenna.

Further discussion of R and description of antenna radiation patterns is outside the scope of this book.

b. Effective Radiated Power P

The effective radiated power, P, radiated in a particular direction from a transmitting antenna is, for the purpose of calculations in this chapter, defined as $1/3\ G_T P_0$.

The factor 1/3 is introduced for the sake of convention, so that 1 kw of effective radiated power corresponds to a field intensity of 300,000 μv/m at 1 km, which is the field produced at 1 km in the horizontal direction by a short vertical antenna over a perfect earth, with one end on the ground, radiating 1 kw total power. The gain G_T of such an antenna in the horizontal plane is 3.

The effective radiated sky-wave power determines the field intensity for sky-wave propagation. Calculations of incident field intensity described later in this chapter are based upon a standard effective radiated power of 1 kw, i. e., a field intensity of 300,000 μv/m at a distance of 1 km, which, as seen from the above, corresponds to an isotropic antenna radiating 3 kw total power in free space.

c. Antenna Gain Factor, G

The gain factor, G, of the antenna for radiation in a given direction is the ratio of the effective radiated power P as defined above, corresponding to radiation in that direction, to the total radiated power. Defined thus, the gain factor G is equal to $1/3\ G_T$, and is also equal to the square of the ratio of the actual field intensity produced at 1 km, to the standard field intensity (300,000 μv/m), divided by the total radiated power expressed in kilowatts.

d. Incident Field Intensity

Just as the sky-wave field intensity produced at a given distance from a transmitting antenna is the resultant of the field intensities of the direct- and ground-reflected waves, the field intensity which affects a receiving antenna due to a sky wave arriving via a particular mode of propagation is the resultant of the field intensities of the incident and ground-reflected waves. This resultant depends upon the angle of arrival of the sky wave, the ground-reflection coefficient, height above ground and wavelength. In the remainder of this chapter reference will usually be made to the incident field intensity.

7.3. Fading

When radio transmission takes place by way of sky waves, the received intensity is not constant but varies with time because of fluctuations in ionospheric conditions. The term "fading" refers to relatively rapid variations which take place in a space of minutes, seconds, or even fractions of a second duration. In general, fading is more rapid on high than on low frequencies.

Besides variations in intensity, fading also causes distortion of radiotelephone signals. This is because at any instant the fading is different on different frequencies, and therefore affects differently the sidebands and the carrier wave. This is called "selective fading" because the fading is thought of as "selecting" some frequencies rather than others.

Fading, which is usually a nuisance, may be reduced by different methods, such as automatic volume control, suppressed carrier transmission, and diversity reception. Discussion of these methods is outside the scope of this book.

The most convenient value of a fluctuating field intensity to consider is the median value, i. e., the value which the instantaneous field intensity exceeds 50 percent of the time. This is the quantity on which most sky-wave calculations are based. The rms (root-mean-square) value of the field intensity is the same value that would be observed if there were no fading. The relation of the median to the rms value depends upon the time distribution of the instantaneous values of field intensity, which in turn depends upon the type or character of the fading.

There are many types of fading, which fall into four principal classes: (a) interference fading, (b) polarization fading, (c) absorption fading, and (d) skip fading. Most of the rapid fading in the input to a receiving set is usually a combination of the first two types.

a. Interference Fading

Interference fading is caused by phase interference of two or more waves from the same source arriving at the receiver over slightly different paths. If the paths are of different lengths, and their relative lengths vary, because, say, of fluctuations in the height of the ionosphere layers, the relative phases of the waves arriving over the

various paths vary with time also, causing alternate reinforcement and cancellation of the field intensity. Because of irregularities in the ionosphere, a single arriving sky wave is really the summation of a great number of waves of small intensity and of random relative phases. The resultant field intensity can vary over wide limits, the maximum possible value tending toward infinity, when all the components are in phase. The rms value of the fading intensity is equal to the "homogeneous field," i. e., the steady value of the field that would have existed had the ionosphere not broken it up into many components,

The time distribution of the instantaneous field intensity resulting from the combination of a large number of waves of random phases and of nearly the same amplitude has been studied and found to be represented by the following formula:

$$T=\epsilon^{-0.693(E/E_m)^2},$$

where T is the fraction of the time the instantaneous field intensity exceeds the value E, and E_m is the median value of the field intensity. This is called the "Rayleigh" distribution. For such a distribution, the median field intensity is equal to $\sqrt{0.693}$ times the rms value, i. e., 0.832 times the intensity of the homogeneous or nonfading field. For such a distribution the lower decile value, or value of field intensity exceeded 90 percent of the time by the instantaneous values, is 0.39 times the median value. The upper decile value, or value exceeded only 10 percent of the time by the instantaneous values, is 1.8 times the median value.

Very bad interference fading is experienced in cases where the sky wave returns to earth at a distance from the transmitter such that the ground wave is of comparable amplitude. The combination of a randomly fluctuating sky wave with the steady ground wave produces much more severe fading than is commonly experienced with the sky wave alone.

The "flutter" fading, or very rapid fluctuations of intensity, associated usually with ionospheric disturbances on paths passing in or near the auroral zone, is another type of interference fading. It is caused by the combination of a large number of wave components which have traversed paths differing appreciably (possibly several wavelengths) in length. This type of fading is also associated with transmission by scattered reflections, for the scattering centers, themselves, are fairly widely separated, and the waves travel different distances in going to and from the scattering centers.

Another type of interference fading is experienced primarily at low frequencies, where radio transmission is relatively stable. Near sunrise and sunset the heights of the ionosphere layers change rather rapidly, and the sky waves arrive alternately in and out of phase with the ground waves, producing interference fading with a relatively long period.

b. Polarization Fading

Whereas interference fading is regarded as a characteristic of the incident wave, additional variation in the field intensity affecting the receiving antenna, called polarization fading, occurs as a result of changes in the state of polarization of the incident wave relative to the orientation of the receiving antenna. When the polarization is such that the electric force in the wave has a large component in the direction of the receiving antenna, the resultant voltage induced in the antenna is large; when that component is small, the induced voltage is also small. For a linearly polarized receiving antenna and random polarization fading, the median effective field intensity is $1/\sqrt{2}=0.707$ times the intensity of the homogeneous field.

In general, the state of polarization of the downcoming sky wave is constantly changing. This is due mainly to the combination, with random amplitudes and phases, of the two oppositely polarized magneto-ionic components, the ordinary wave and the extraordinary wave. The state of polarization of the downcoming sky wave is in general elliptical, with either direction of rotation, and with random and constantly changing values of the dimensions and orientation of the ellipse with respect to the receiving antenna. The state of polarization of sky waves varies more rapidly the higher the frequency, which accounts in part for more rapid fading on the higher frequencies.

c. Absorption Fading

Absorption fading is caused by short-time variations in the amount of energy lost from the wave because of absorption in the ionosphere. In general the period of this type of fading is much longer than for the other two types, since the ionospheric absorption usually changes only slowly. The sudden ionosphere disturbance is an extreme case of this type of fading, although it is usually classified as an anomaly rather than as fading.

Somewhat analogous to this type of fading, although the cause is not in the ionosphere but in reflections and absorption in objects close to the receiver, is the type of fading experienced in receiving a signal while moving along in an automobile. The fading out of a signal when the receiver is passing under a bridge or near a steel structure is caused by absorption of the wave's energy by the structure. Effects of this sort are involved in so-called "dead spots" or places where radio reception is especially difficult. Also, reradiation from wires, fences, and steel structures can cause an interference pattern which is rela-

ively fixed in space, and can be noticed on moving the receiving equipment around. Where there are nearby structures which can cause these effects, care must be exercised in selecting a receiving site.

d. Skip Fading

Skip fading is observed at places near the skip distance, and is caused by the waves alternately skipping and returning to earth. Near sunrise and sunset, when the ionization density of the ionosphere is changing, it may happen that the muf for a given transmission path oscillates about the actual frequency. When the skip distance moves out past the receiving station (sometimes called "going into the skip") the received intensity abruptly drops by a factor of 100 or more, and just as abruptly increases again when the skip distance moves in again. This may take place many times before steady conditions of transmission or skip are established.

7.4. The Unabsorbed Field Intensity

The unabsorbed field intensity of a sky wave at a given distance from the transmitting station is defined as the median incident field intensity that would be observed using an antenna of fixed linear polarization if no absorption were introduced by the ionosphere.

The unabsorbed field intensity, so defined, differs from that which would result from distance attenuation alone because of:

(1) focusing caused by the curvature of the ionosphere reflecting layer;

(2) interference and polarization fading;

(3) loss of energy upon reflection at the ground between hops.

Focusing tends to increase the field intensity but is relatively unimportant except at low angles of takeoff and arrival, and even here the effect tends to be vitiated by irregularities in the reflecting layer.

As pointed out in the preceding section, random interference and polarization fading both reduce the median field intensity received on a linearly polarized antenna below that of the homogeneous wave. The factors involved are 0.832 (−1.6 db) and 0.707 (−3.0 db), respectivelv

The loss of energy at ground reflection between hops depends upon the frequency, polarization, angle of incidence at the ground, and the electrical characteristics of the ground.

Curves in figure 7.4 and 7.5 give unabsorbed field intensities for 1 kw effective radiated power as a function of distance for 1-hop-*E* and 1-hop-*F*2 modes of propagation, assuming virtual reflection heights of 105 and 320 km, respectively. These curves take account of fading effects but not of focussing. However, as most practical antennas

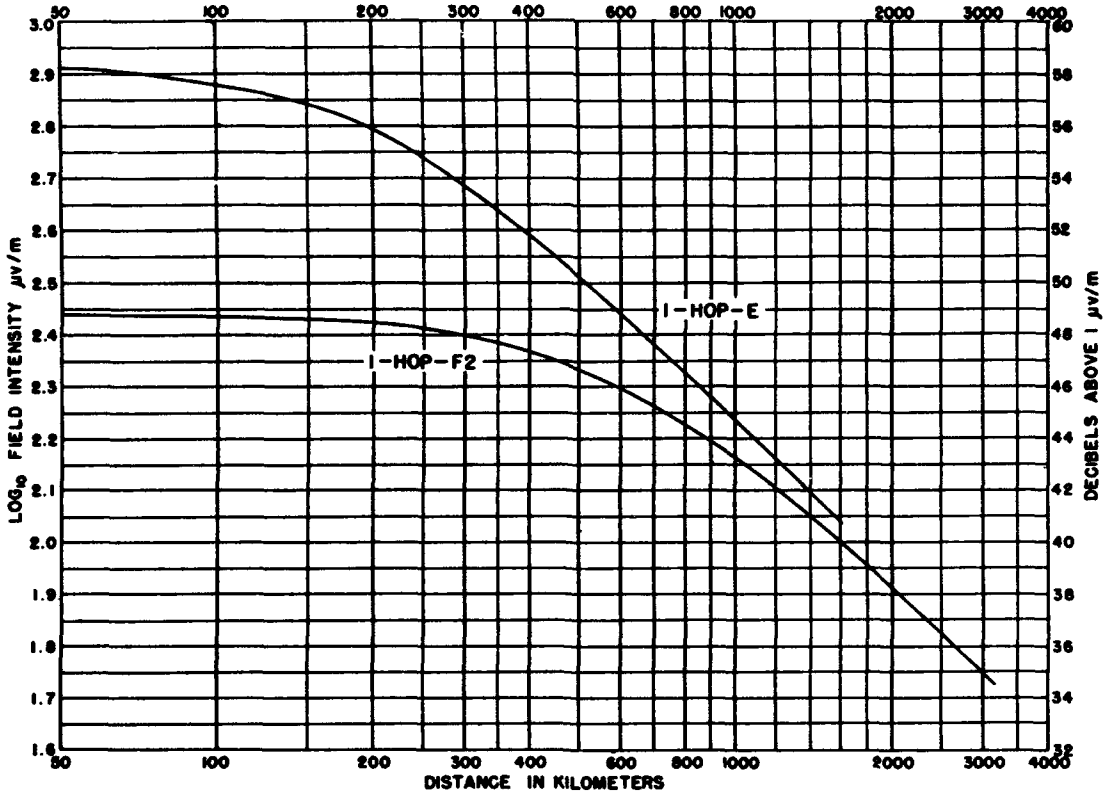


FIGURE 7.4. *Median incident unabsorbed field intensity for 1 kw effective radiated power based on virtual reflection heights, E layer—105 km, F2 layer—320 km.*

935333 O - 51 - 8

FIGURE 7.5. *Median incident unabsorbed field intensity for 1 kw effective radiated power, F2-layer propagation distances over 3000 km.*

are very inefficient at low angles of radiation and arrival, accurate knoweledge of the unabsorbed field intensity for low angles is usually unnecessary.

The unabsorbed field intensity $E_0(nx)$ for a mode consisting of n hops each of length x, is given by

$$E_0(nx)=\frac{(0.63)^{n-1}}{n}E_0(x),$$

where 0.63 is the assumed loss factor for each ground reflection, and $E_0(x)$ is the 1-hop unabsorbed field intensity for the distance x. The logarithm of the unabsorbed field intensity, $F_0(nx)$, is given by

$$F_0(nx)=F_0(x)-(n-1)0.2-\log n.$$

The curve in figure 7.5 represents the "average" unabsorbed field intensity for distances greater than 3,200 km, assuming $F2$-layer propagation and 1 kw effective radiated power at fairly low radiation angles, i. e., between, say, 5° and 25°.

7.5. Interpretation of Ionospheric Absorption Measurements

a. General

As pointed out in chapter 2, the presence of electrons in the upper atmosphere not only causes bending and return to earth of a radio wave of sufficiently low frequency, but also, because of the collisions of the electrons with neighboring air molecules, causes part of the wave energy to be dissipated, thus reducing the intensity of the radio wave below that resulting from the normal spreading out of the wave front as it recedes from the transmitter. This absorption process is of great importance in the practical use of ionospheric radio transmission.

During daylight hours, absorption takes place principally in the D region of the ionosphere. Here electron densities are considerably less than in higher regions but the increased density of air molecules results in an increased collisional frequency, which more than compensates for the scarcity of electrons.

During the night, ionization and absorption in the D region become negligible. However, there is still some absorption for frequencies near the maximum usable frequency of the $F2$ layer because waves at such frequencies are abnormally retarded and there is sufficient time for appreciable energy loss to take place in spite of the relatively small collision frequency. Such absorption is called "deviative absorption" because it occurs in conjunction with retardation which also causes bending of the waves. Absorption which takes place even when the wave is not appreciably retarded is called "nondeviative" absorption (chapter 2). Absorption in the D region is usually largely nondeviative.

As ionization in the D region is caused principally by ultraviolet radiation from the sun, and the region is low enough in the atmosphere for rapid recombination to occur, the density of ionization in this region, and hence the absorption, varies practically in synchronism with the elevation of the sun over the horizon. D-region absorption is thus greatest at noon at latitudes directly beneath the sun.

b. The Absorption Index

The ionospheric absorption index α is defined as the logarithm of the ratio of the unabsorbed field intensity E_0 to the absorbed field intensity E, E_0 being the intensity of the sky wave reflected from a given height if there were no absorption, and E being the actual intensity of the reflected wave. It may be written:

$$\alpha = \log\,(E_0/E) = F_0 - F,$$

where $F_0 = \log E_0$ and $F = \log E$. Thus, we can write $E = E_0 \times 10^{-\alpha}$ or $F = F_0 - \alpha$ as the basic field intensity formula.

In using and measuring absorption indexes, it is necessary to make allowance for the differential absorption of the ordinary and extraordinary components of the received wave. The relative intensities of these components depend upon the frequency and the degree of absorption.

The quantity actually measured in most cases is the relative field intensity, i. e., the voltage delivered by the antenna to the receiving set. To obtain the actual field intensity in microvolts per meter, the receiver input voltage must be multiplied by a calibrating factor depending upon the characteristics of the receiving set and the antenna system. The unabsorbed field intensity may be calculated if the effective radiated power and height of reflection are known. It is usually impracticable to obtain the factors necessary for these calculations. However, if reflections can be obtained at night when absorption is negligible, the field intensity measured at such time, corrected for any difference in the height of reflection, is the unabsorbed field intensity. In CW measurements, the unabsorbed field intensity obtained in this way should also be corrected for the presence of multiple reflections.

In measurements using pulse technics, the absorption may be obtained directly by comparing the amplitudes of successive multiple reflections of the pulse. The absorption index is equal to

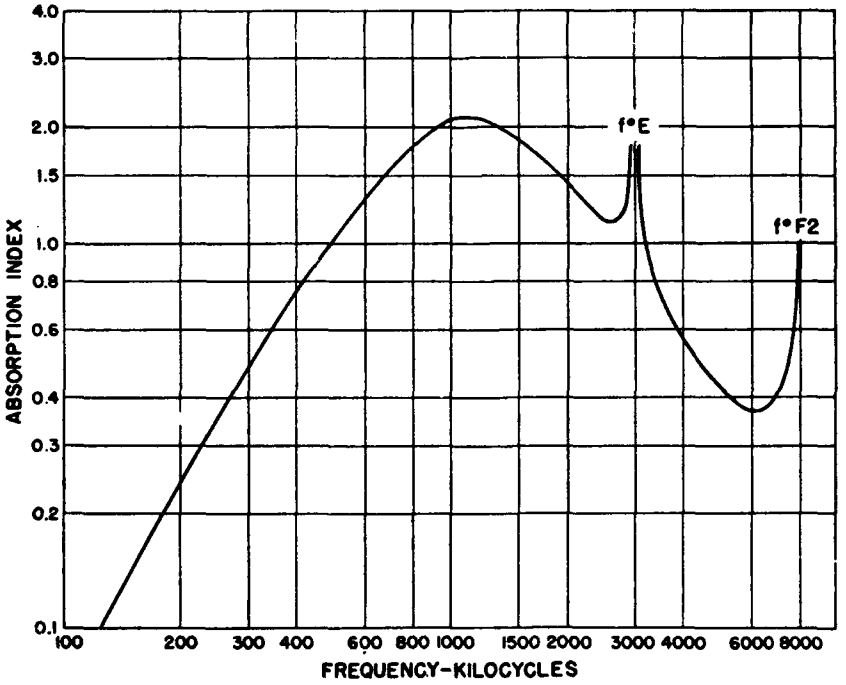


FIGURE 7.6. *Variation of absorption of the ordinary wave with frequency, noon, January 1942, at Washington, D. C.*
Absorption is given as log (incident field intensity/reflected field intensity). Cusps on curve correspond to indicated critical frequencies.

the logarithm of the ratio of the intensities of successive pulses corrected for loss occurring at the intervening ground reflection.

c. Variation with Frequency

The variation of absorption with the frequency of waves incident vertically on the ionosphere at [illegible] in January 1942, as deduced from oblique- and vertical-incidence data of the National Bureau of Standards, is shown in figure 7.6. The cusps in the curve labeled f^0E and f^0F2 correspond to deviative absorption taking place in the vicinity of the E- and $F2$-layer critical frequencies, respectively. Except for these cusps, absorption at frequencies above about 2,000 kc follows the law for quasi-longitudinal propagation (S.K. Mitra, see references, p. 121) through the absorbing region, to a good approximation. That is,

$$\alpha=\alpha_1/(f\pm f_H\cos\theta)^2,$$

where α is the absorption for a frequency f, f_H is the gyrofrequency, θ is the angle between the direction of the earth's magnetic field and the direction of propagation, and α_1 is a constant. The plus sign refers to the ordinary component and the negative sign refers to the extraordinary component of the wave.

Below 2,000 kc, reflection takes place before the wave has penetrated entirely through the absorbing region, with the result that the absorption is a maximum at about 1,200 kc, and decreases steadily toward lower frequencies.

At oblique incidence, the variation with frequency is similar except that the whole curve is raised in proportion to the secant of the angle of incidence upon the D region.

d. Diurnal and seasonal variation

A formula frequently used to correlate the nondeviative absorption coefficient with the sun's zenith angle χ is

$$\alpha=\alpha_0(\cos\chi)^n,$$

where α_0 is the value of the absorption when $\chi=0$ (sun overhead), and n is an exponent determined empirically from the data. Taking the logarithm of both sides of this equation, we obtain

$$\log\alpha=\log\alpha_0+n\log\cos\chi.$$

For equivalent vertical-incidence frequencies above about 1 Mc, values of log α plotted against log cos χ tend to fall on a straight line, the slope of which is the appropriate value of n. For a Chapman region (see chapter 4), n should have the value 1.5 for values of χ not too near 90°. Various workers have reported experimental values of n varying between 0.5 and 2.0. A significant difference between values of n for the diurnal and seasonal variations seems to exist, those for the diurnal variation being somewhat the higher.

The value $n=1$ represents the diurnal variation fairly well on the average, and has the advantage of greatly simplifying calculations of the average zenith-angle over a long transmission path.

The above formula does not, however, represent the absorption very well for values of the zenith angle near 90°. Actually, there is appreciable absorption at times of ground sunrise and sunset, because of ionization of the D region at these times, due to diffusion and other effects. The diurnal variation, including times near sunrise and sunset, is fairly well represented by the following empirical expression,

$$K=0.142+0.858\cos\chi,$$

in which K denotes the diurnal variation factor.

TABLE 7.1. *Seasonal variation factor, J*

Month	Both terminals of transmission path		One terminal of transmission path north and the other south of equator
	North of equator	South of equator	
Jan.	1.3	1.0	1.15
Feb.	1.3	1.0	1.15
Mar.	1.15	1.15	1.15
Apr.	1.15	1.15	1.15
May	1.0	1.3	1.15
June	1.0	1.3	1.15
July	1.0	1.3	1.15
Aug.	1.0	1.3	1.15
Sept.	1.15	1.15	1.15
Oct.	1.15	1.15	1.15
Nov.	1.3	1.0	1.15
Dec.	1.3	1.0	1.15

A residual seasonal variation beyond that involved in the above K factor, as determined from measurements made in middle latitudes in the northern hemisphere, is given by factors listed in table 7.1. The factor was arbitrarily chosen to be unity for summer months. Values for the other months of the year were adjusted accordingly. The symbol J is used to denote the seasonal variation factor.

e. Variation with Solar Activity

In addition to the diurnal and seasonal variations with the sun's zenith angle, there is a variation with solar activity manifest in a 27-day cycle coinciding with the rotation period of the sun (see chapter 5), and a secular variation following the 11-year sunspot cycle. Various measurements indicate that the absorption for waves which penetrate entirely through the D region increases by a factor of approximately 1.5 between sunspot num-

ber zero and 100. Assuming a linear variation, the ratio Q of the absorption at sunspot number R to the absorption at zero sunspot number is thus

$$Q=1+0.005R \qquad Q_1=0.005R.$$

See Errata in front of book

Values of R predicted 3 months in advance are given in each issue of the CRPL–D series, beginning with CRPL–D32.

f. Summary of Regular Absorption Variations

Combining the above variation factors, we can write the following expression for the nondeviative absorption when the equivalent vertical-incidence frequency (sec. 6.2, a) is above about 1 Mc.

$$\alpha=JQ\overline{K}S=AS,$$

where J is the seasonal variation factor, Q is the solar cycle variation factor, $\overline{K}$ is the average value of K for the transmission path, and S is a function of the frequency and the distance traveled by the wave in the absorbing region.

g. Normal Day-to-Day Variations

In the above summary of regular absorption variations, no account was taken of the normal but irregular variations which take place from day-to-day. These, together with day-to-day changes in the unabsorbed field intensity, cause variations in the incident field intensity. The daily median values of incident field intensity are distributed approximately symmetrically about the monthly median value on a decibel or logarithmic scale, such that 90 percent are less than about twice (6 db above) the monthly median and 90 percent are greater than about one-half (6 db below) the monthly median. To a first approximation, these limits are also independent of the frequency, path length, season, and time of day.

h. Absorption in Oblique-Incidence Propagation

A wave which passes through the D region obliquely traverses a greater distance in this region than one which penetrates at vertical incidence. Specifically, if the oblique path makes an angle ϕ_D with the vertical, the ratio of these distances is sec ϕ_D.

Thus neglecting the effect of varying the angle between the direction of propagation and the direction of the earth's magnetic field, the ratio of the nondeviative absorption α' for oblique-incidence propagation to α for vertical-incidence propagation is

$$\alpha'/\alpha=\sec\phi_D,$$

provided complete penetration of the D region takes place at both angles of incidence, or the same fraction of the D region is traversed in each case. These conditions are satisfied for $F2$-layer propagation, but not in general for E-layer propagation.

In the presence of a magnetic field, if the propagation may be considered as quasi-longitudinal, the above relation becomes, for the ordinary component

$$\alpha'=\alpha\sec\phi_D(f+f_H\cos\theta)^2\,\tfrac{1}{2}\left[\frac{1}{(f+f_H\cos\theta_1)^2}+\frac{1}{(f+f_H\cos\theta_2)^2}\right];$$

where f is the wave frequency, f_H is the gyrofrequency, and θ, θ_1, and θ_2 are, respectively, the angles between the direction of the earth's magnetic field and the directions of propagation of the vertical wave, the incident oblique wave, and the reflected oblique wave, in the D region.

Replacing the factor

$$\tfrac{1}{2}\left[\frac{1}{(f+f_H\cos\theta_1)^2}+\frac{1}{(f+f_H\cos\theta_2)^2}\right],$$

by

$$\frac{1}{(f+f_H\cos\theta')^2}$$

where θ' lies between θ_1 and θ_2, the formula becomes

$$\alpha'=\alpha\sec\phi_D\left(\frac{f+f_H\cos\theta}{f+f_H\cos\theta'}\right)^2.$$

The value of f_H varies from about 0.7 Mc at the geomagnetic equator to about 1.6 Mc at the geomagnetic poles. Thus for large values of f, the above expression becomes approximately $\alpha'=\alpha\sec\phi_D$.

As the D region extends somewhat into the E layer, and waves of the same frequency propagated obliquely and vertically are reflected from different levels, the fraction of the D region traversed varies with the angle of incidence. The above formula is therefore not valid in general for E-layer propagation. The corresponding formula for E-layer propagation is

$$\alpha'=\alpha\sec\phi_D\left(\frac{f+f_H\cos\theta}{f'+f_H\cos\theta'}\right)^2,$$

where now α' is the absorption for the frequency f' propagated obliquely and reflected at the angle ϕ_E from the E layer, and α is the vertical incidence absorption for the equivalent vertical incidence frequency $f=f'\cos\phi_E$, i. e., the frequency reflected at vertical incidence at the same level as f'.

i. Absorption for long paths

Beyond about 3,000 km attenuation is so great for frequencies propagated by the E layer, down to about 1 Mc, that field intensities produced with reasonable values of radiated power are usually too weak to be useful. Discussion of absorption for such long paths may reasonably be limited to $F2$-layer propagation.

It is useful to note that, in the case of $F2$ layer propagation, sec ϕ_D is approximately linearly proportional to the length of a single hop for hop lengths greater than about 1,000 km. This means that for a given path the absorption is approximately the same for all $F2$ modes for which the length of a single hop is greater than 1,000 km. For modes whose hop length is less than 1,000 km the absorption is greater.

Thus the absorption for long paths may be written

$$\alpha' = Sd,$$

where d is the path length, and S is the average absorption per unit distance along the path.

Introducing the factor $A = JQ\overline{K}$, mentioned in section 7.5, f, which expresses the regular variations, we have $\alpha' = S_0 Ad = S_0 JQ\overline{K}d$, in which S_0 is a function of frequency only.

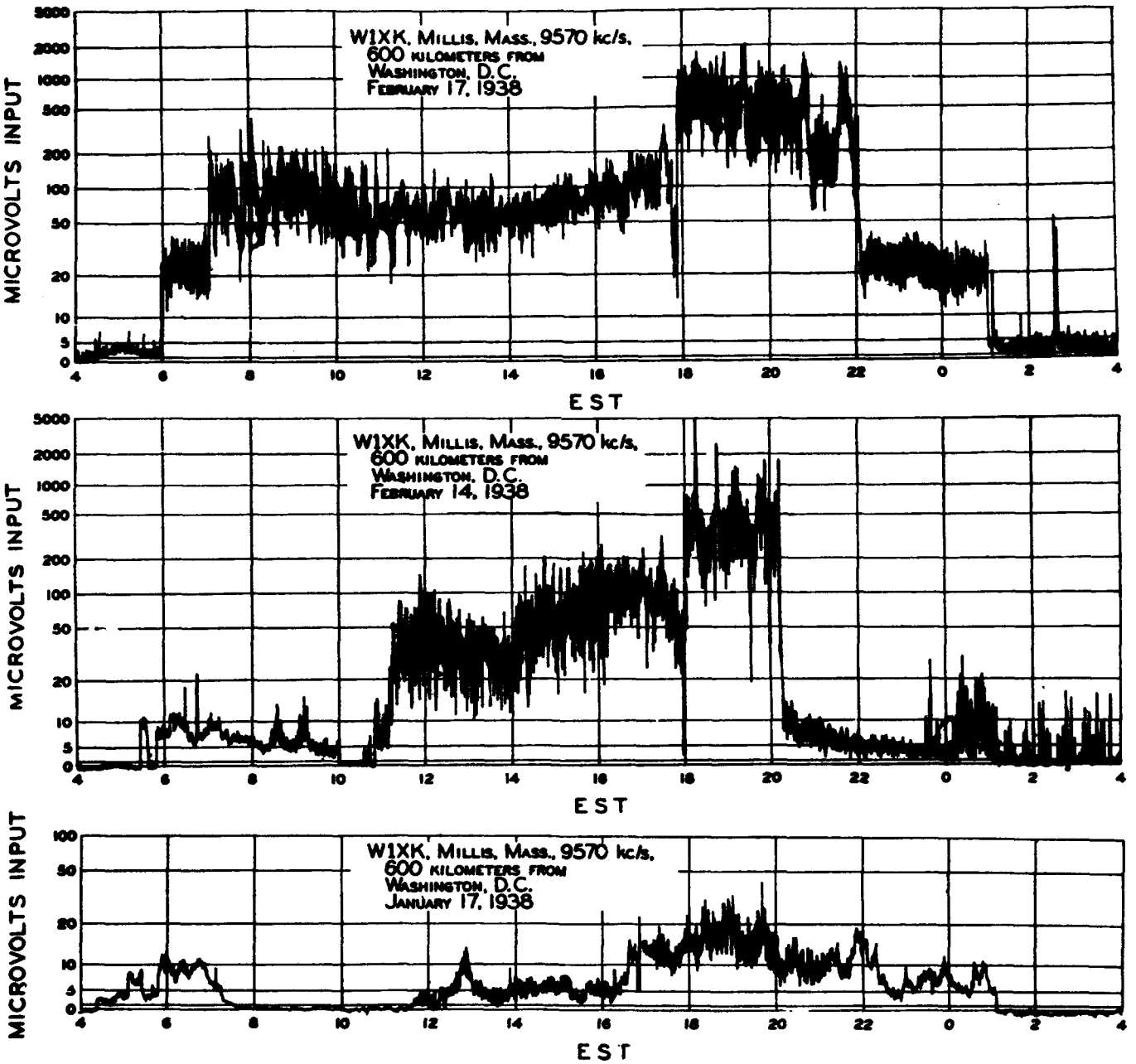


FIGURE 7.7. *Field-intensity measured over the same path, on the same frequency, for a quiet day (February 17, 1938), a day of moderate storm (February 14, 1938), and a day of severe storm (January 17, 1938).*

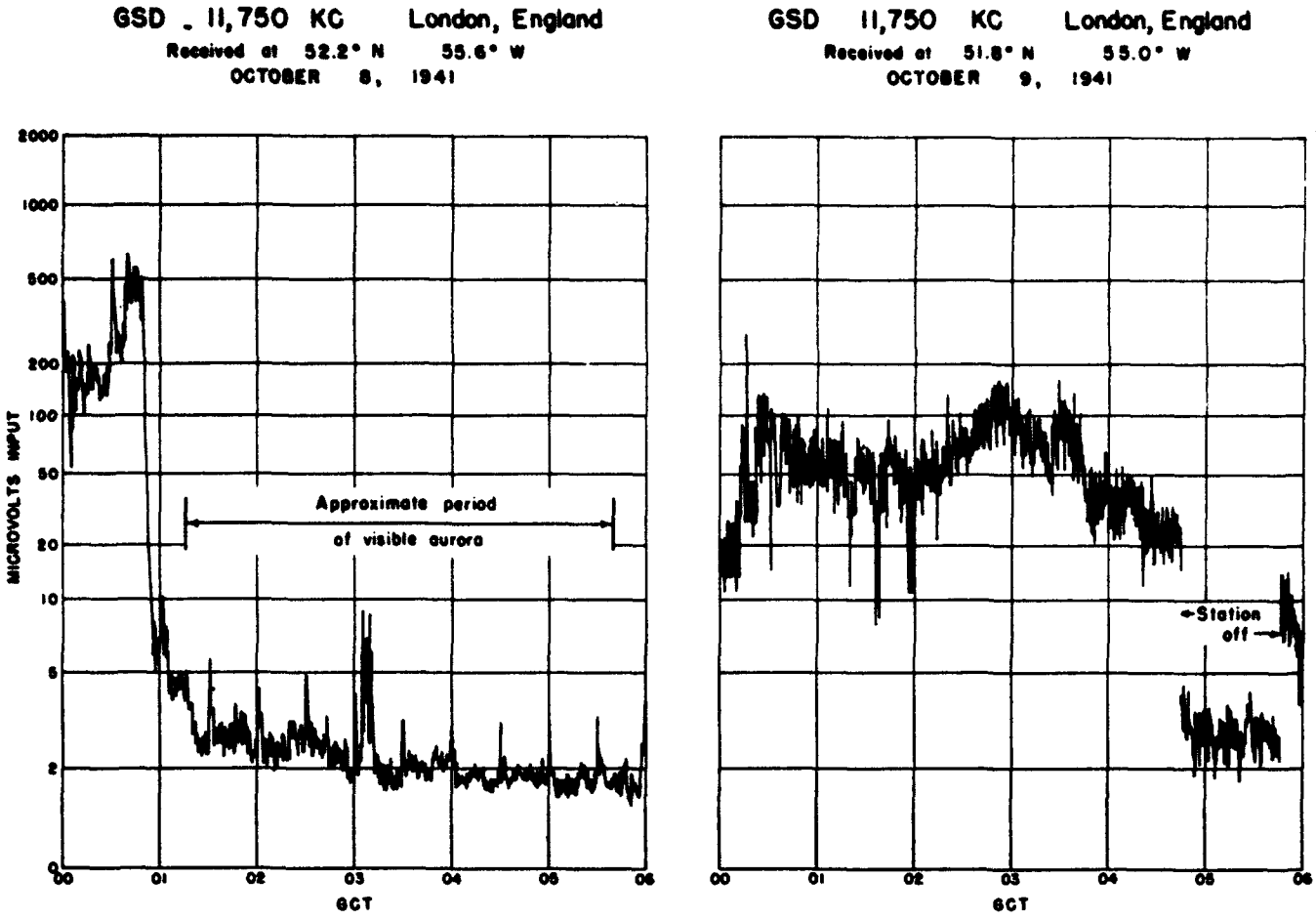


FIGURE 7.8. *Field-intensity variation during and after visible aurora of October 8, 1941.*

Note depression of field intensity as compared with following day.

7.6. Abnormal Absorption

Besides the regular absorption effects discussed above, which are believed to be caused by relatively steady ultraviolet radiation from the sun, there are effects due to relatively sudden and irregular outbursts of solar activity and to solar particle emanations. These effects include the absorption effects manifested during ionosphere storms and "sudden ionosphere disturbances," and the absorption continually present in the auroral belts.

Phenomena in the auroral belts, (chapter 5), are believed to be due to electrified solar particles guided into the belts by the earth's magnetic field. Ionization produced by these particles extends into low levels of the atmosphere and results in pronounced absorption at nearly all times, day and night.

Frequently coincident with the passage of large sunspots or groups of spots across the sun, the streams of solar particles are greatly intensified, and the auroral belts spread out, causing world-wide ionosphere-storm effects, the intensities of which decrease with distance from the auroral belts. During such storms, ionization becomes diffuse, and sometimes the normal stratification of the ionosphere disappears. At these times far more ionization exists at low levels than ordinarily, especially at night, and absorption is abnormally high and irregular in nature.

Figure 7.7 shows, for comparison, the records of field intensities, measured on the same frequency, for an ionospherically quiet day, and during days of moderate and severe storms. Figure 7.8 presents field-intensity measurements showing the sudden decrease in intensity with the onset of an auroral display that was visible at the receiving location.

Coincident with the appearance of solar flares, or bright eruptions on the sun, usually in the vicinity of a large sunspot or group of spots, there often occur periods of abnormally high *D*-region ionization, lasting from a few minutes to an hour or more, and occurring only during daylight hours. These are the "sudden ionosphere

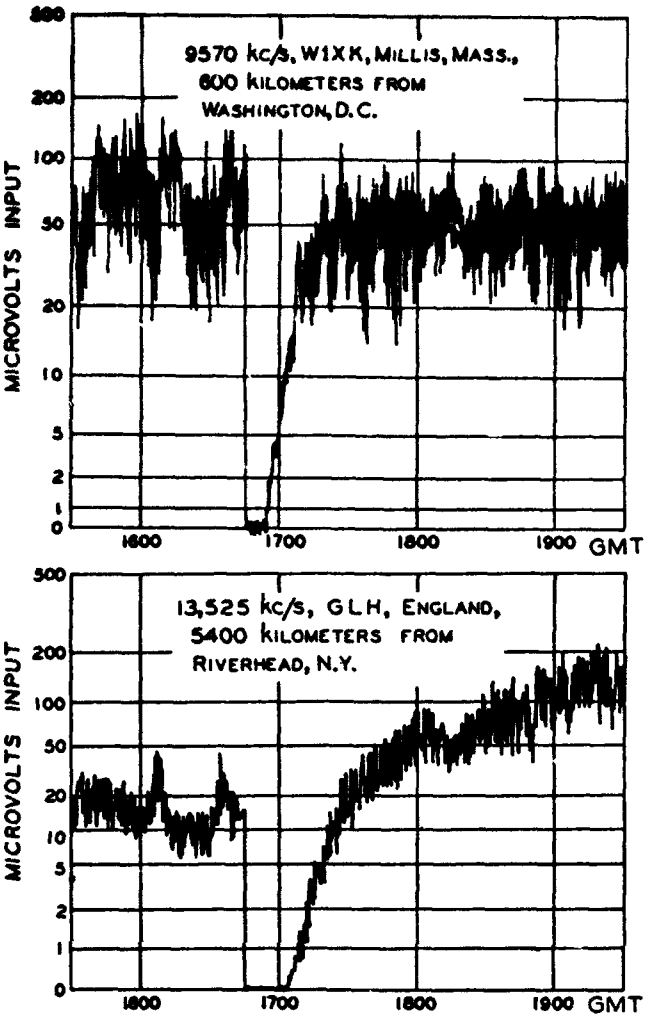


FIGURE 7.9. *Sudden disturbance of the ionosphere on April 8, 1936, as revealed in the field-intensity records of widely separated transmitting stations.*

disturbances", previously mentioned in chapter 5, caused by sudden intense bursts of utltaviolet light emanating from the solar flares.

This increase in *D*-region ionization makes the region a better conductor, so that the intensities of waves of frequencies below about 500 kc, which do not penetrate through the region, are actually increased. Even waves that normally penetrate the *D* region, as in the medium-frequency broadcast band, may be reflected below the *D* region, and exhibit increased intensities. However, absorption of waves of higher frequency, which continue to penetrate the *D* region, is so strong as to cause cessation of all useful high-frequency transmission, except possibly on frequencies very near the muf limit.

Figure 7.9 shows field-intensity records of signals received at widely separated stations over quite different paths during a day on which one of these sudden disturbances occurred. It will be seen that the striking sudden drop in field intensity to practically zero occurs simultaneously.

7.7. Calculation of Incident Field Intensity

a. General

The procedures described below enable rapid estimation of the incident field intensity for a given path, frequency, mode of propagation, and effective radiated power.

The field intensity obtained for a particular month and time of day corresponds to the median value for normal (undisturbed) days of the month. In section 7.5, g, it was pointed out that the distribution of daily median values of the incident field intensity is such that 90 percent are greater than about one-half (6 db below) the monthly median, and that 90 percent are less than about twice (6 db above) the monthly median value. It was also pointed out in section 7.3 that in a normally fading wave, the distribution (Rayleigh) of instantaneous values of field intensity is such that 90 percent are above the lower decile value, which is equal to 0.39 times (8.2 db below) the median value, while 90 percent are below the upper decile value equal to 1.8 times (5.1 db above) the median value.

In most radio transmission problems, the lower decile value is usually of greater interest than the upper decile value. Also, the value of field intensity which the daily lower decile values exceed on 90 percent of the days of the month may be of interest. This value is obtained by calculating the incident field intensity according to the methods given in the following sections and then multiplying this field intensity by $0.5 \times 0.39 = 0.195$ (decrease by 14 db).

The value of field intensity which is exceeded on only 10 percent of the days by the daily lower decile field intensity is obtained by multiplying the median value by $2 \times 0.39 = 0.78$ (decrease by 2.2 db).

The results obtained in a particular calculation may be in error due to incompleteness of the basic measurement data, and to simplifying assumptions made in reducing the measurement data and applying the results to calculation procedures. In general, the errors tend to be greatest for extreme values of the frequency, distance, and absorption factor *A*. However, the calculated field intensity will tend to be too high for any values of these quantities if deviative absorption is present in appreciable quantity, or if solar activity is unusually high during the month in question, and the results will tend to be too low if solar activity is unusually low. Other effects, such as sporadic-*E* ionization, may cause errors in either direction.

It should be noted that the effect of the incident field intensity in producing voltage at the input of the receiving set varies with the angle of arrival of the wave. Thus the combined effect of waves arriving at different angles should be calculated

taking into account the effective length of the antenna for each angle of arrival. The resultant voltage, V, induced in the receiving antenna is given by

$$V=\sqrt{(L_1E_1)^2+(L_2E_2)^2+\ \ldots\ +(L_nE_n)^2},$$

where E_1, E_2 * * * E_n are the incident field intensities for the various active modes of propagation expressed in microvolts per meter, and L_1, L_2 * * * L_n are the corresponding effective lengths of the receiving antenna expressed in meters. The effective length is proportional to the square root of the gain factor. Further discussion of effective length is outside the scope of this book.

b. Calculation of incident Field Intensity for Short Paths, 0 to 400 Km

For sky-wave transmission over short paths, the absorption is nearly independent of the path length and the mode of propagation for 1-hop modes. Accordingly, the expression for the absorption given in section 7.5, f, may be written

$$\alpha=AS_1,$$

where $A=JQ\overline{K}$, and S_1 is a function of the frequency only. Also, for short distances, the unabsorbed field intensity is also approximately independent of the path length. If, further, we neglect the variation of the unabsorbed field intensity with mode of propagation, the logarithm of incident field intensity in microvolts per meter given by

$$F=F_0+\tfrac{1}{2}\log P-AS_1,$$

may be calculated by means of the nomogram in figure 7.10. If P and A are plotted on the left- and right-hand scales of the nomogram, respectively, and a straightedge laid between them, the logarithm of the incident field intensity in microvolts per meter or the field intensity in decibels above 1 μv/m, is read at the intersection of the straightedge with the desired frequency.

The value of P used should be that corresponding to 1-hop-E or 1-hop-$F2$, whichever is active. If both are active, use the higher value of P.

The appropriate value of $\overline{K}$ may be obtained from K charts, as explained in section 7.7, f.

Let us calculate, for example, the incident sky-wave field intensity produced at a distance of 100 km by a transmitter operating at 100-w effective radiated power on 2.2 Mc near Manila, P. I., in June 1947, at 1200 local time. Referring to table 7.1, $J=1.0$. The CRPL predicted sunspot number for June 1947, is 112 (publication of predicted sunspot numbers in the CRPL-D series begins with the prediction for July 1947). Substituting in the formula for Q in section 7.5, e, we obtain $Q=1.56$. The latitude of Manila is about 14° N. Referring to the K-chart for June, figure 7.39, we find for 1200 local time at 14° N., $\overline{K}=K=1.0$, approximately. Thus $A=JQ\overline{K}=1.56$.

Referring to figure 7.3, the radiation angle for 1-hop-$F2$ is approximately 80° and for 1-hop-E is approximately 65°. The distance "x" corresponding to 1-hop-E for $\Delta=80°$ is 35 km. Referring to figure 7.2, we find that 2.2 Mc is well below the E-layer muf for both 100 and 35 km, and is therefore below the muf for 1-hop-E and also below the E-layer cut-off frequency for 1-hop-$F2$. Therefore, the 1-hop-E mode is active, but the 1-hop-$F2$ mode is inactive at this frequency.

On the field-intensity nomogram, figure 7.10, $P=100$ w is plotted on the left-hand scale, and $A=1.56$ is plotted on the right-hand scale. A straightedge is laid between these two points and the incident field intensity is read at the intersection of the straightedge with the vertical line corresponding to 2.2 Mc. The field intensity so read is 0.1 log (μv/m) or 2 db above 1 μv/m.

Suppose, now, the operating frequency were 7 Mc instead of 2.2 Mc. It is evident from figure 7.2 that 7 Mc is well above both the 1-hop-E muf and the 1-hop-$F2$ cut-off frequency. Referring to CRPL-D31, the 1-hop-$F2$ muf is 13 Mc. Thus the 1-hop-$F2$ mode would be active, but the 1-hop-E mode would be inactive at 7 Mc. The logarithm of the field intensity in microvolts per meter for this frequency would be 1.7 (34 db above 1 μv/m.)

NOTE: figures 6.17, 6.18, 6.21 to 6.30 are taken from CRPL-D31.

c. Calculation of Incident Field Intensity for Paths of Intermediate Length, 400 to 3,200 Km

The incident field intensity in microvolts per meter and in decibels above 1 μv/m for 1-kw effective radiated power is plotted as a function of frequency for various values of the absorption factor $A=JQ\overline{K}$ in figures 7.11 to 7.30. Field intensities for different distances and modes of propagation are plotted on separate charts. Only 1-hop-$F2$, 2-hop-$F2$, and 1-hop-E and 2-hop-E modes of propagation are represented. Consideration of these modes is sufficient for most purposes.

These charts are based on an analysis made by the United States Army Signal Corps of field-intensity measurements for a group of paths of intermediate lengths (400 to 2,600 km) located near Washington, D. C., where the inclination of the earth's magnetic field is nearly vertical. Although specifically applicable in regions of similar magnetic inclination, excluding the auroral belts, the charts are probably of nearly equal validity in other regions especially at the higher frequencies, say, above about 10 Mc.

The approximate muf for E-layer modes and the approximate cut-off frequency for $F2$-layer modes are shown on these charts. Separate curves are given for $J=1.0$ and 1.3. Values for $J=1.15$ lie approximately half-way between these curves. The $F2$-muf cannot be expressed as a function of A. It must be determined separately for each particular problem.

If the length of the transmission path for which the incident field intensity is desired is not equal to one of the distances represented in the charts, read the field intensity on the chart for the nearest distance less than and greater than the path length, and interpolate linearly between the values of field intensity in decibels to obtain the field intensity corresponding to the path length.

The field intensity so found corresponds to 1 kw effective radiated power. For P kw effective radiated power, multiply this field intensity by $\sqrt{P}$ (or add P expressed in db above 1 kw to the field intensity in decibels above 1 μv/m). See section 7.7, f, for calculation of $\overline{K}$.

Consider, for example, transmission from Washington, D. C., to Miami, Fla., a distance of 1,500 km at 1200 local time in Miami in June 1947 on a frequency of 13 Mc. What would be the field intensity produced in Miami by a transmitter radiating 1,000 w effective power at Washington? Referring to table 7.1, $J=1.0$. The CRPL predicted sunspot number for June 1947, is 112 (publication of predicted sunspot numbers in the CRPL–D series begins with the July 1947, prediction). Substituting this value in the formula for Q in section 7.5, e, we obtain $Q=1.56$. The latitude of the midpoint of the path is about 32° N. The longitude is approximately equal to that of Miami; therefore, no time correction need be made. On the K-chart for June, figure 7.39, we find for 1200 local time at 32° N., $\overline{K}=K=1.0$, approximately. Thus $A=JQ\overline{K}=1.56$.

As 1,500 km lies between two of the distances for which the intermediate-distance field-intensity charts are constructed, charts for 1,200 and 1,600 km are selected, figures 7.18 through 7.25.

The 1-hop-E and 2-hop-E muf for this path, at the time considered, as obtained from CRPL–D31, are 18 and 11.5 Mc, respectively. Thus 1-hop-E is active, but 2-hop-E is inactive at 13 Mc. Referring to figure 7.18 the 1-hop-E field intensity chart for 1,200 km, we find that the incident field intensity for 13 Mc at $A=1.56$ is 9 μv/m (19 db above 1 μv/m). On figure 7.22, the 1-hop-E chart for 1,600 km, the incident field intensity at 13 Mc and $A=1.56$ is 5.6 μv/m (15 db above 1 μv/m). Interpolating linearly, the 1-hop-E field intensity for 1,500 km is approximately 6.3 μv/m (16 db above 1 μv/m).

Referring to figures 7.20, 7.21, 7.24, and 7.25, it appears that 13 Mc is above the E-layer cut-off frequency at 1,200 and 1,600 km for both 1-hop-$F2$ and 2-hop-$F2$ modes. The muf for these modes, as obtained from CRPL–D31, are 15.3 Mc and 10.3 Mc, respectively. Thus 1-hop-$F2$ is active, but 2-hop-$F2$ is inactive.

In figures 7.20 and 7.24, the 1-hop-$F2$ incident field intensities for 1,200 and 1,600 km at 13 Mc and $A=1.56$, are 35 and 20 μv/m (31 and 26 db above 1μv/m), respectively. Interpolating linearly, the field intensity for 1,500 km is approximately 24 μv/m (27 db above 1 μv/m).

For equal effective radiated power the 1-hop-$F2$ field intensity is 3.3 times (10.4 db above) the 1-hop-E field intensity. Referring to figure 7.3, the radiation angle for 1,500 km is about 20° for 1-hop-$F2$, but only 5° for 1-hop-E. Most antennas radiate very poorly at 5°. In practice, therefore, the difference between the 1-hop-$F2$ and 1-hop-E field intensities would probably be more than the factor 3.3.

Suppose the effective radiated power in the above example were 500 w for the 1-hop-$F2$ radiation angle (20°) instead of 1,000 w. The 1-hop-$F2$ field intensity would then be $24\times\sqrt{(500/1000)}=17$ μv/m.

d. Calculation of Incident Field Intensity for Long Paths, over 3,200 Km

Utilizing the expression for absorption for long paths given in section 7.5, i, the incident field intensity in log microvolts per meter for long paths is

$$F=F_0+\tfrac{1}{2}\log P-S_0AD,$$

where F is for 1-kw effective radiated power and $A=JQ\overline{K}$. The incident field intensity may be calculated by direct substitution in this formula. The quantity S_0, as determined from analysis of long-path field-intensity measurements, is plotted as a function of frequency in fig. 7.31. F_0 is given by figure 7.4. A procedure for calculating $\overline{Kd}$ is given in section 7.7, e. Alternatively the incident field intensity may be obtained from the nomograms in figures 7.32 and 7.33. The nomogram in figure 7.32 is for use with small values of Ad, and that in figure 7.33 with large values of Ad. If d and Ad are plotted on the left- and right-hand scales, respectively, the logarithm of the incident field intensity in microvolts per meter, or the field intensity in decibels above 1 μv/m, corresponding to 1 kw effective radiated power, is read at the intersection of a straightedge laid between d and Ad with the desired frequency. To obtain the logarithm of the incident field intensity corresponding to P kw, add $\frac{1}{2}\log P$ to the logarithm of the field intensity for 1 kw. To obtain the field intensity in decibels above 1 μv/m, add P expressed in decibels above 1 kw to the field intensity for 1 kw expressed in decibels above 1 μv/m. If the field intensity is expressed in

microvolts per meter, multiply by the square root of P, expressed in kilowatts.

For long-path transmission, the effective radiated power, P, may be considered equal to the average value of P for radiation angles between 0° and 30°.

For calculation of input to the receiver, use the average effective length of the receiving antenna for angles of arrival between 0° and 30°.

The appropriate value of $\overline{K}d$ is calculated as explained in section 7.7, e.

For example, let us find the incident field intensity produced at 1200 local time in Trieste, Italy, in June 1947, by a transmitter in Washington, D. C., operating on 15 Mc and radiating an average effective power of 1 kw between 0° and 30° radiation angle. The length of this path is 7,100 km and $\overline{K}d=4.9$ (see sample calculation in section 7.7, e. Referring to table 7.1, $J=1.0$. The CRPL predicted sunspot number for June 1947, is 112. Substituting this value in the formula for Q, section 7.5, e), we obtain $Q=1.56$. Thus $Ad=7.6$ We choose the nomogram of figure 7.33 as the one appropriate to the magnitude of Ad. Laying a straightedge between 7,100 km on the left-hand distance scale, and 7.6 on the right-hand Ad scale, the logarithm of the field intensity in microvolts per meter read at 15 Mc is 0.15 (3 db above 1 μv/m).

For an effective radiated power of 4 kw (6 db above 1 kw), the logarithm of the incident field intensity would be $0.15+\frac{1}{2}$ log $4=0.45$. Using decibels, the field intensity would be $3+6=9$ db above μv/m.

e. Calculation of $\overline{K}d$

Figures 7.34 through 7.45 are charts of the diurnal variation factor $K=0.142+0.858\cos\chi$ for each month of the year. To obtain $\overline{K}d$, proceed as follows:

1. Construct a great-circle curve for the path on a piece of transparent paper exactly as in steps 1 and 2 in the procedure given in 6.6, a, chapter 6. Be sure to mark the meridian whose local times are to be used as the times for the calculations. Call this the "reference" meridian.

2. Before removing the paper from the great-circle chart of figure 6.30 (chapter 6) at the end of step 2 mentioned above, mark distance intervals of 500 km along the curve, using as guides the numbered dot-dash lines and the dotted lines which are spaced at 500-km intervals along the great-circle curves.

3. Place the transparent paper over the K chart, figures 7.34 through 7.45 for the desired month, and aline the reference meridian marked in step 1 with the time for which the calculation is to be made.

4. Read K_1 and K_2, the values of K at the terminal points of the path. Record 0 for K_1 or K_2 if the corresponding terminal lies outside the $K=0$ curve.

5. Read the length of the path in thousands of kilometers, which lies in the region where K is not zero. Call this length D'

6. The quantity $\overline{K}d$ in thousands of kilometers units may now be calculated by using the formula

$$\overline{K}d=0.142D'+(K_1+K_2-0.284)\tan(D'/2R),$$

where R is the radius of the earth expressed in the same units as D', i. e., thousands of kilometers. The angle $D'/2R$ is in radians.

Alternatively, $\overline{K}d$ may be obtained by entering the values of K_1+K_2, and D' in the appropriate nomogram, figures 7.46 and 7.47. These nomograms are not accurate for D' near 20,000 km. In such cases the path should be divided into two parts, $\overline{K}d$ calculated for each part separately, and the results added to obtain $\overline{K}d$ for the entire path.

7. Repeat steps 3 through 6 for each time for which $\overline{K}d$ is desired.

For example, what is $\overline{K}d$ for the path between Washington, D. C., and Trieste, Italy, in June, at 1200 local time at Trieste? Following the above procedure, we obtain K_1 (Washington)$=0.33$, K_2 (Trieste)$=0.93$, and $D'=d=7{,}100$ km. Entering D' and K_1+K_2 in the nomogram of figure 7.46, we obtain $\overline{K}d=4.9$.

f. Calculation of $\overline{K}$ for Short or Intermediate Paths

If the path is of short or intermediate length, and does not cross the $K=0$ curve, it is usually sufficiently accurate, after step 3 above, to take $\overline{K}$ equal to the value of K at the midpoint of the path. If the path straddles the $K=0$ line, read K at the midpoint of the portion that lies on the side where K is greater than zero. $\overline{K}$ is equal to this value multiplied by D'/d.

Otherwise, find $\overline{K}d$ and divide by d expressed in thousands of kilometers to obtain $\overline{K}$.

g. Paths Passing Through the Auroral Zone

Transmissions over paths that lie, even in part, in the auroral zones, are subject to a greater degree of irregularity and erratic performance than are transmissions over other paths.

The absorption is very great, especially during daylight conditions. For example, stations operating at less than about 6,000 kc usually cannot be heard during daylight hours over any great distance (2,000 km or more) in these regions.

Severe and prolonged ionosphere storms occur frequently, often developing suddenly in the course of a few minutes. They are manifested by greatly increased absorption and by a drop in the

ionization density of the higher reflecting layers, which lowers the maximum usable frequency. The result is the narrowing or complete disappearance of the band of useful frequencies. It is not unusual for long-distance transmission to be impossible on all high frequencies for a day or more at a time, and to be erratic and only partially recovered on a small portion of the high-frequency spectrum for as much as a week. There have been instances of ionospheric storminess lasting almost continuously for a month.

Frequently in auroral zones during ionosphere storms, there also appears strong, widespread, and continuous intense sporadic-E transmission lasting for many hours. This may considerably improve radio reception in certain directions and over some paths while it lasts, but there is no way of predicting it.

During the polar winter night, except during ionosphere storms, good radio transmission is often experienced on all high frequencies up to near the maximum usable frequency.

Some paths that are similar except for direction seem to display different propagation characteristics. For example, in parts of Greenland, the European high-frequency stations on about 9 to 15 Mc are heard much better than United States stations at similar distances and frequencies. Also, while transmission across the auroral zone between the United States and Greenland is unfavorable for the broadcast frequencies, 550 to 1,500 kc, United States stations on these frequencies are received well in northern Canada and Greenland, during the winter night. Not enough is yet known about the auroral zone to explain such effects fully.

Observations indicate that the region of severe absorption associated with the auroral zone is a relatively narrow band located slightly equatorward from the line of maximum auroral frequency. Transmissions over paths just outside this band are affected to a much smaller extent by the auroral absorption.

Although very little is known, quantitatively, about auroral-zone absorption, the amount and distribution for relatively undisturbed conditions has been estimated on the basis of meager available data, and the following procedure is offered to serve as a rough guide when it is necessary to estimate the additional absorption encountered by waves traversing these regions. On the map of figure 7.48, contours have been drawn following observed lines of equal auroral frequency. The numbers on these curves are proportional to the amount K' to be added to $\overline{K}$ for paths that intersect the curves.

0 to 400 km. Plot the terminal points of the transmission path on the map of figure 7.48 and read K' at the midpoint. This is the amount to be added to $\overline{K}$ in the calculation of incident field intensity by the method of section 7.7, b.

400 to 3,200 km. Sketch the great-circle path between the terminal points on a piece of transparent paper. Determine the active modes of propagation. The mufs for E-layer modes and the cut-off frequencies for E-layer modes may be higher than normal because of sporadic-E. Call the length of each hop d_n, where n denotes the number of hops. Locate and mark the ionosphere reflection point at the midpoint of each hop. In the case of F-layer modes also mark points one-third of d_n on each side of each ionosphere reflection point. These are the D-region points. In the case of E-layer modes, the E-layer reflection points may be regarded as D-region points.

Place the transparent sheet on the map of figure 7.48 and note the value of K' at each D-region point. In the case of E-layer modes, add together the values of K' for all D-region points and divide by the number of hops n. In the case of F-layer modes, add together the values of K' for all D-region points and divide by $2n$. In either case, the average value of K' so obtained is the amount to be added to $\overline{K}$ in the calculation of incident field intensity by the method and charts described in section 7. 7,c. Note that the curves for E-layer cut-off and E-layer muf on these charts are valid only for $A=JQ\overline{K}$, and not when K' has been included.

Paths longer than 3,000 km. Draw the great-circle path on a transparent sheet and place marks at 500-km intervals along the path. Estimate the length of path in kilometers lying in each of the subzones between adjacent K' contours in figure 7.48. Multiply each of these distances by the estimated average values of K' for the corresponding subzones. For example, the average value of K' for the subzone between K'—0 and K'—0.1 is approximately 0.05. The sum of all these products is the amount to be added to $\overline{K}d$ in the calculation of incident field intensity by the method of section 7.7, d.

7.8. References

Radiation from antennas in the 1.5 to 20 Mc band, RPU–96, Radio Propagation Unit Technical Report No. 2, April 1945. Prepared under the direction of the Chief Signal Officer by the Radio Propagation Unit (9463d TSU), Holabird Signal Depot, Baltimore, Md.

Space-wave patterns of ground based antennas in the 2–18 Mc range. Ohio State Univ. Research Foundation (1946).

Lord Rayleigh, On the resultant of a large number of vibrations of the same pitch and arbitrary phase. Phil. Mag. **10**, 73 (1880); also "Theory of sound", 2d ed. paragraph 42a, (1894) (Also Dover Publications, New York, N. Y., 1945).

K. A. Norton, The polarization of downcoming ionospheric radio waves, Federal Communications Commission, Mimeograph No. 60047 (May 1942).

K. A. Norton, The nature of very high frequency ionospheric wave propagation, Federal Communications Commission, Mimeograph No. 81009 (Sept. 28, 1944).

D. F. Martyn, The propagation of medium radio waves in the ionosphere, Proc. Phys. Soc. (London) **47**, 323 (1935).

E. V. Appleton, Regularities and irregularities in the ionosphere. Proc. Roy. Soc. [A] **162**, 451 (1936).

E. V. Appleton and J. H. Piddington, The reflection coefficients of ionospheric regions, Proc. Roy. Soc. [A] **164**, 467 (1938).

J. E. Best and J. A. Ratcliffe, The diurnal variation of the ionospheric absorption of wireless waves, Proc. Phys. Soc. (London) **50**, 233 (1938).

F. W. G. White and T. W. Straker, The diurnal variation of absorption of wireless waves, Proc. Phys. Soc. (London) **51**, 865 (1939).

F. T. Farmer and J. A. Ratcliffe, Wireless waves reflected from the ionosphere at oblique incidence, Proc. Phys. Soc. (London) **48**, 839 (1936).

C. R. Burrows, The propagation of short radio waves over the North Atlantic, Proc. Inst. Radio Engrs. **19**, 1634 (1931).

J. H. Dellinger, Sudden disturbances of the ionosphere. J. Research NBS **19**, 111 (1937) RP 10106.

Intermediate distance ky-wav e field intensities. RPU-110, Radio Propagation Unit Technical Report Procedures Technical Report No. 6 (Feb. 1946). Prepared under the direction of the Chief Signal Officer by the Radio Propagation Unit (9463d TSU), Holabird Signal Depot, Baltimore, Md.

Calculation of sky-wave field intensities, maximum usable frequencies and lowest useful high frequencies. RPU-144, Radio Propagation Unit Technical Report No. 6 (March 1947). Prepared under the direction of the Chief Signal Officer by the Radio Propagation Unit (9463d TSU), Holabird Signal Depot, Baltimore, Md.

S. K. Mitra, Report on the present state of our knowledge of the inosphere, Proc. Ntl. Inst. Sci. of India, **1**, No. 3, 131 (1935).

H. R. Mimno, The physics of the ionosphere, Rev. Modern Phys. **9**, 1 (1937).

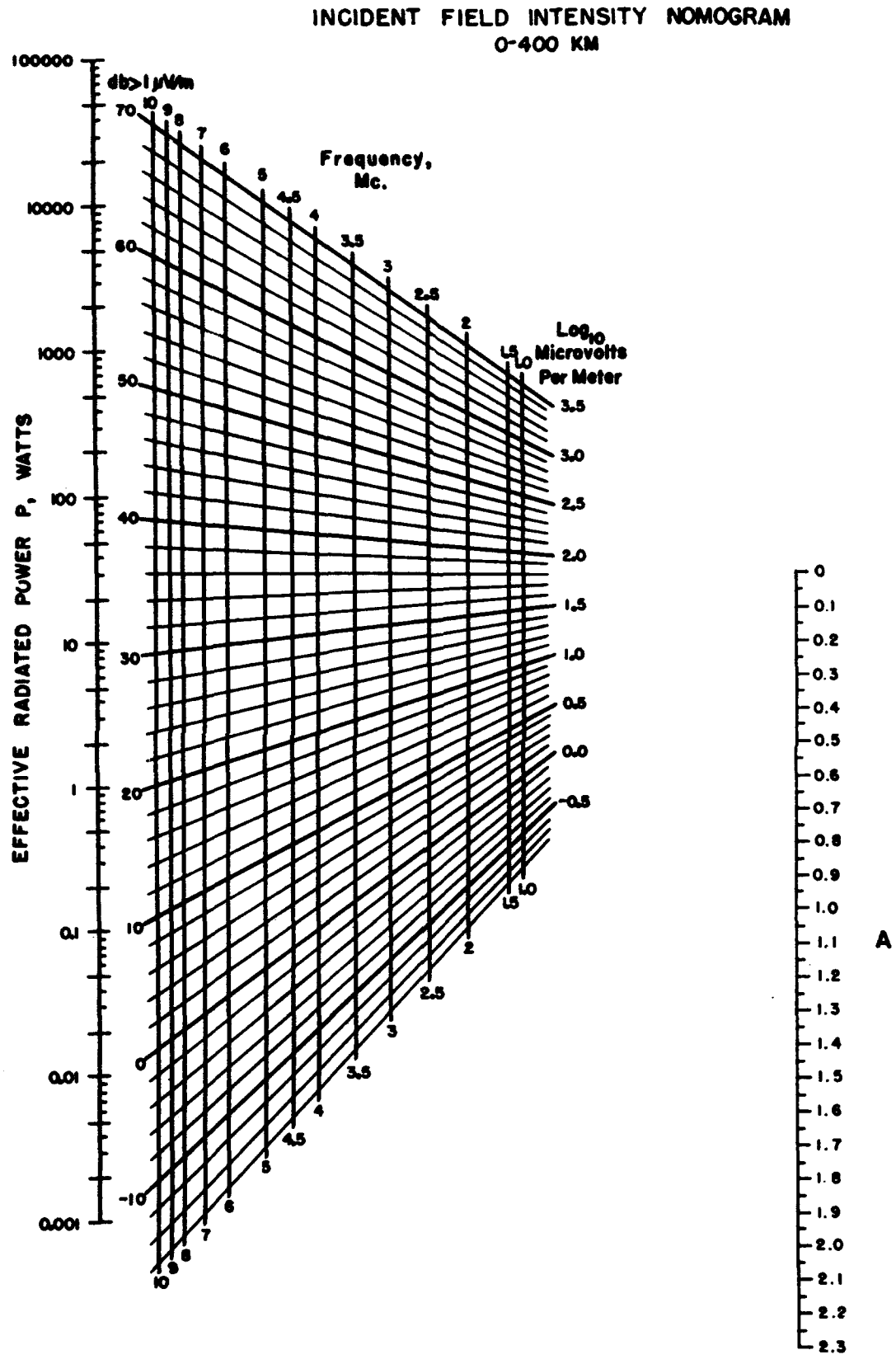
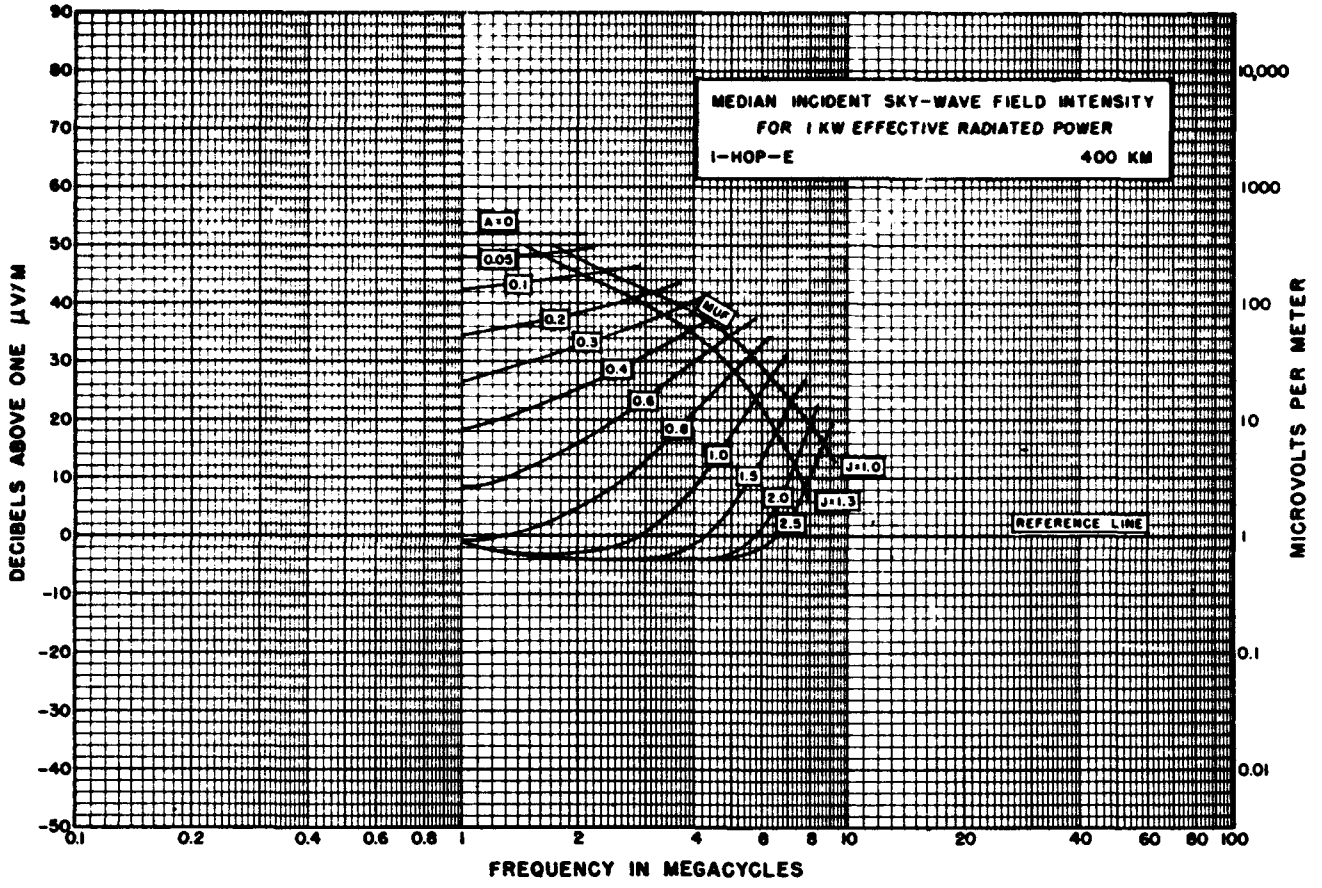
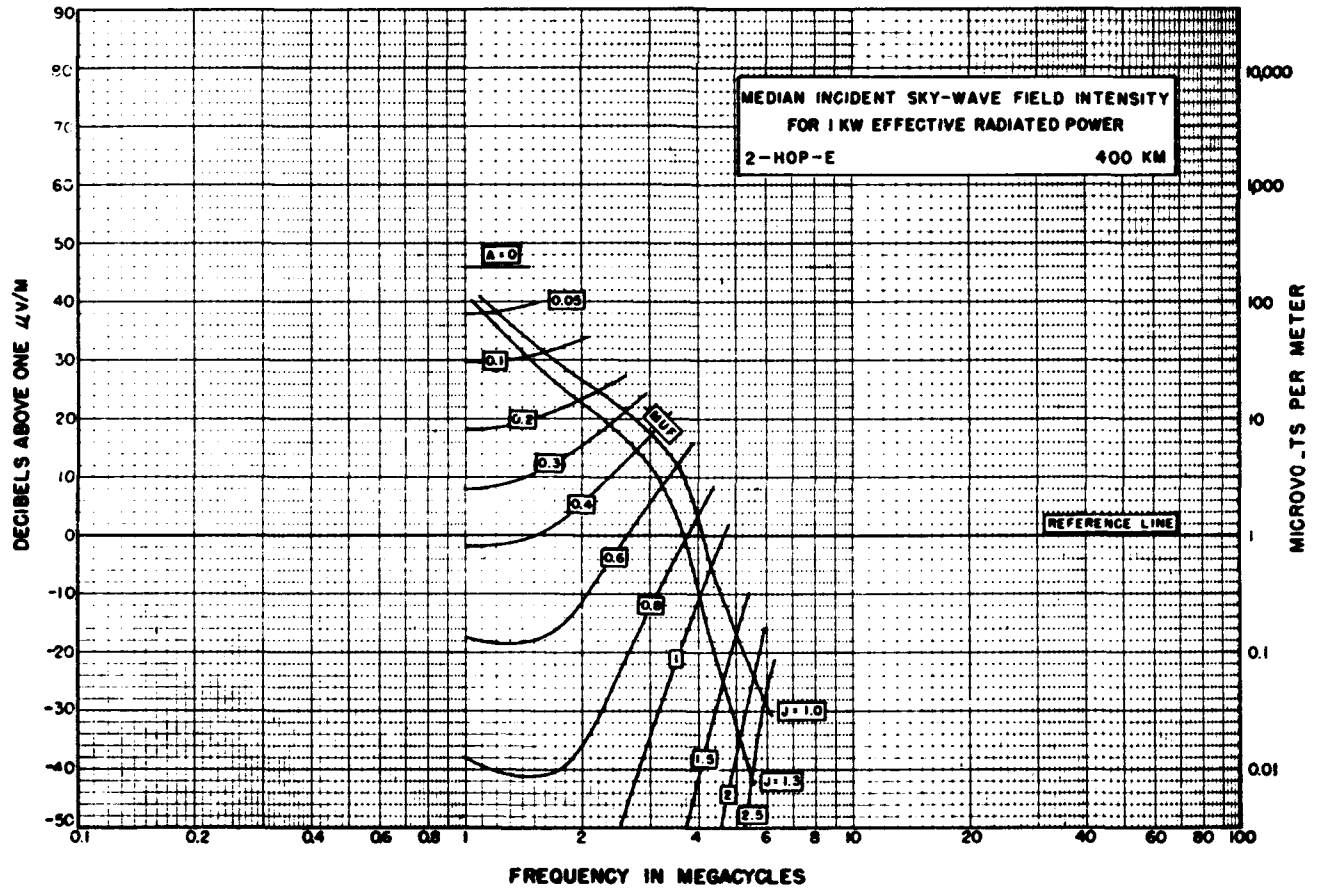
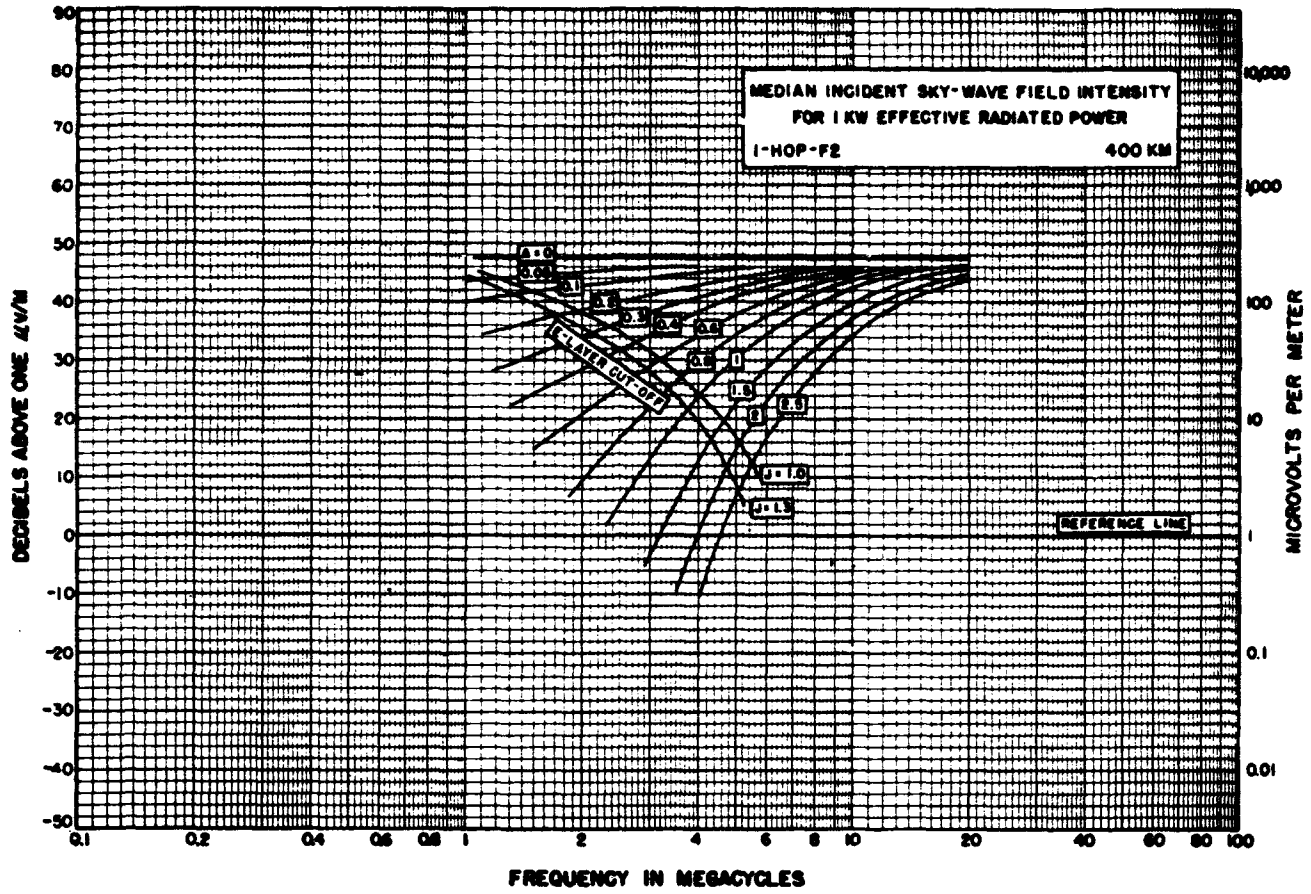

INCIDENT FIELD INTENSITY NOMOGRAM
0-400 KM
EFFECTIVE RADIATED POWER P, WATTS
100000
10000
1000
100
10
1
0.1
0.01
0.001
db>1 μV/m
70
60
50
40
30
20
10
0
-10
Frequency, Mc.
10
9
8
7
6
5
4.5
4
3.5
3
2.5
2
1.5
1.0
Log10 Microvolts Per Meter
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
A
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
1.3
1.4
1.5
1.6
1.7
1.8
1.9
2.0
2.1
2.2
2.3

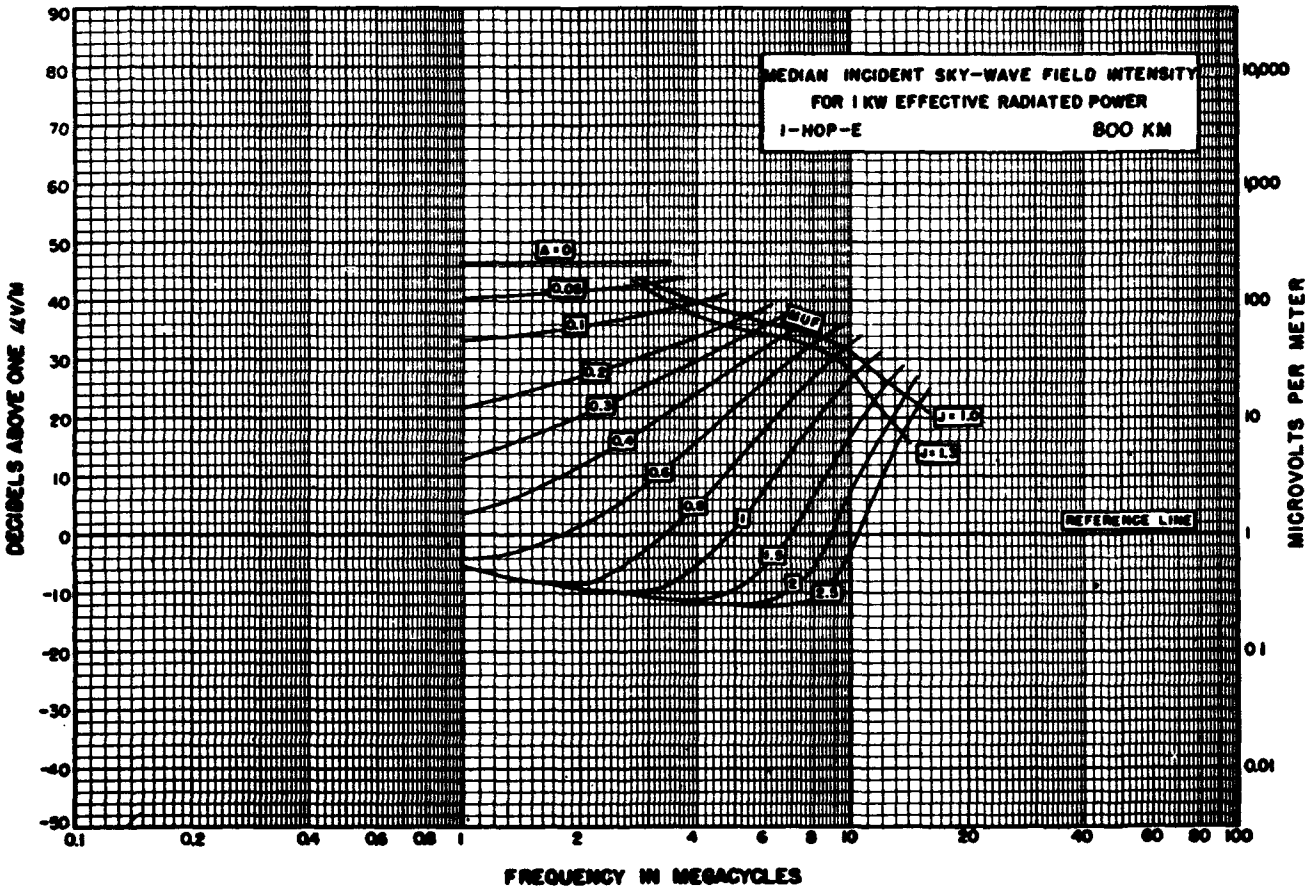
FIGURE 7.10



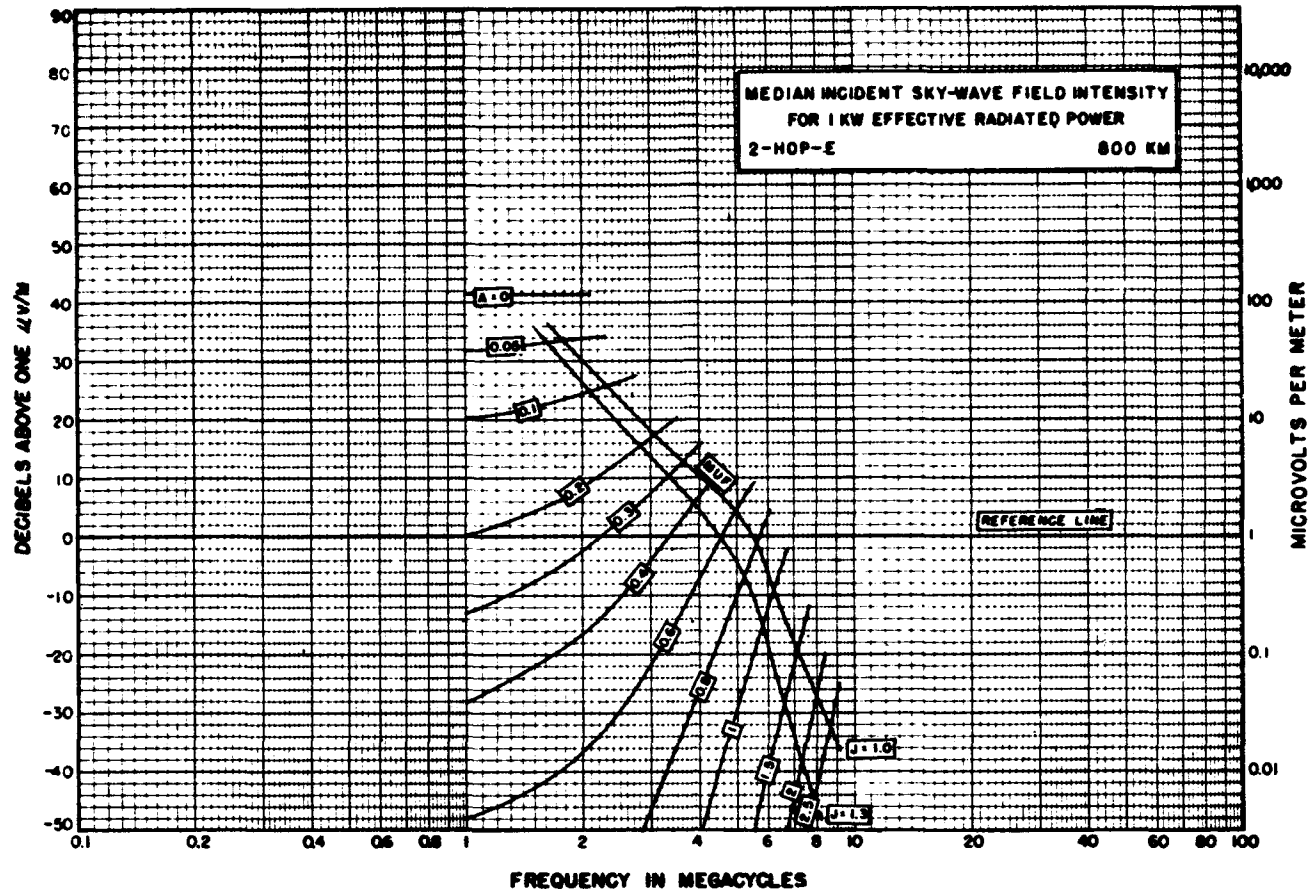


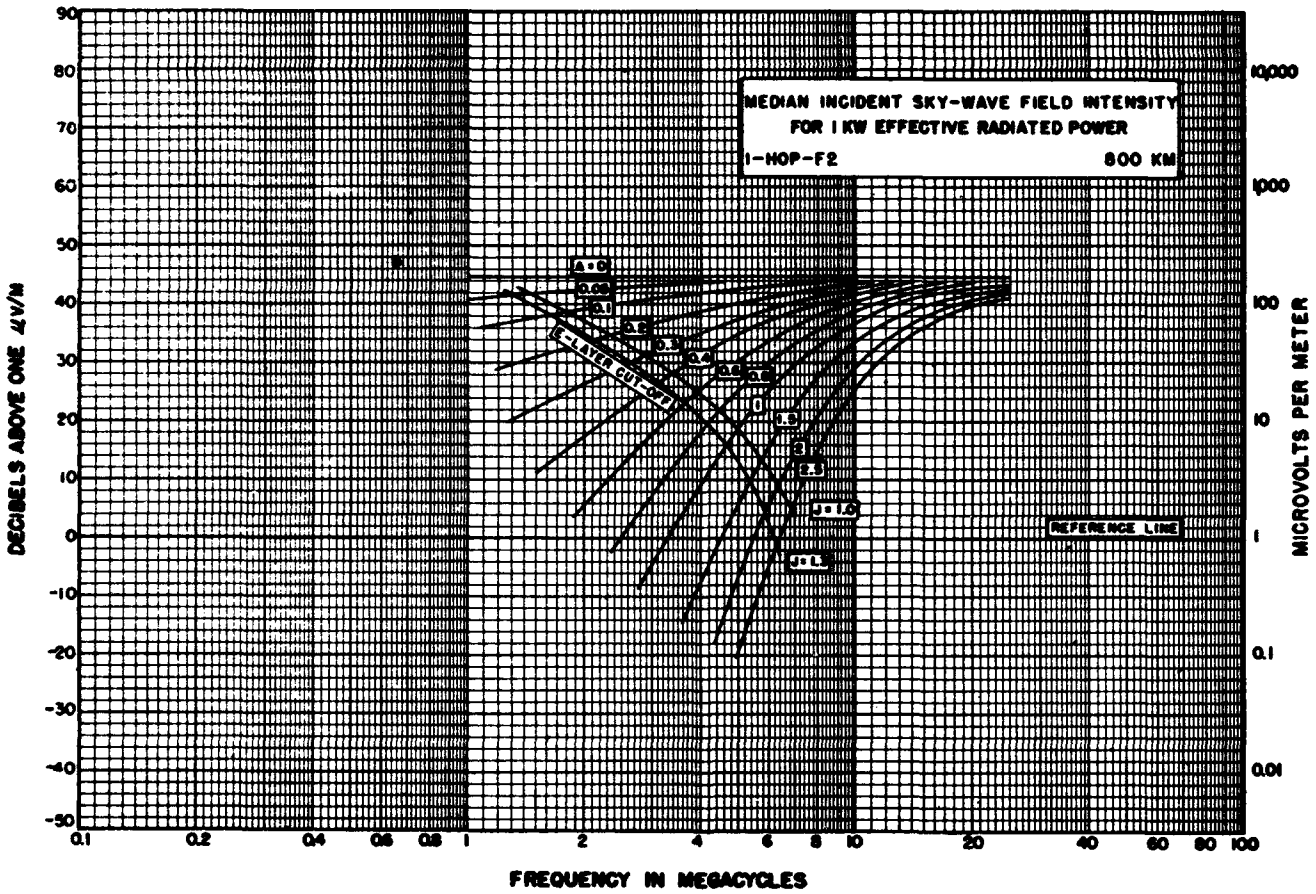
FIGURES 7.11 and 7.12.





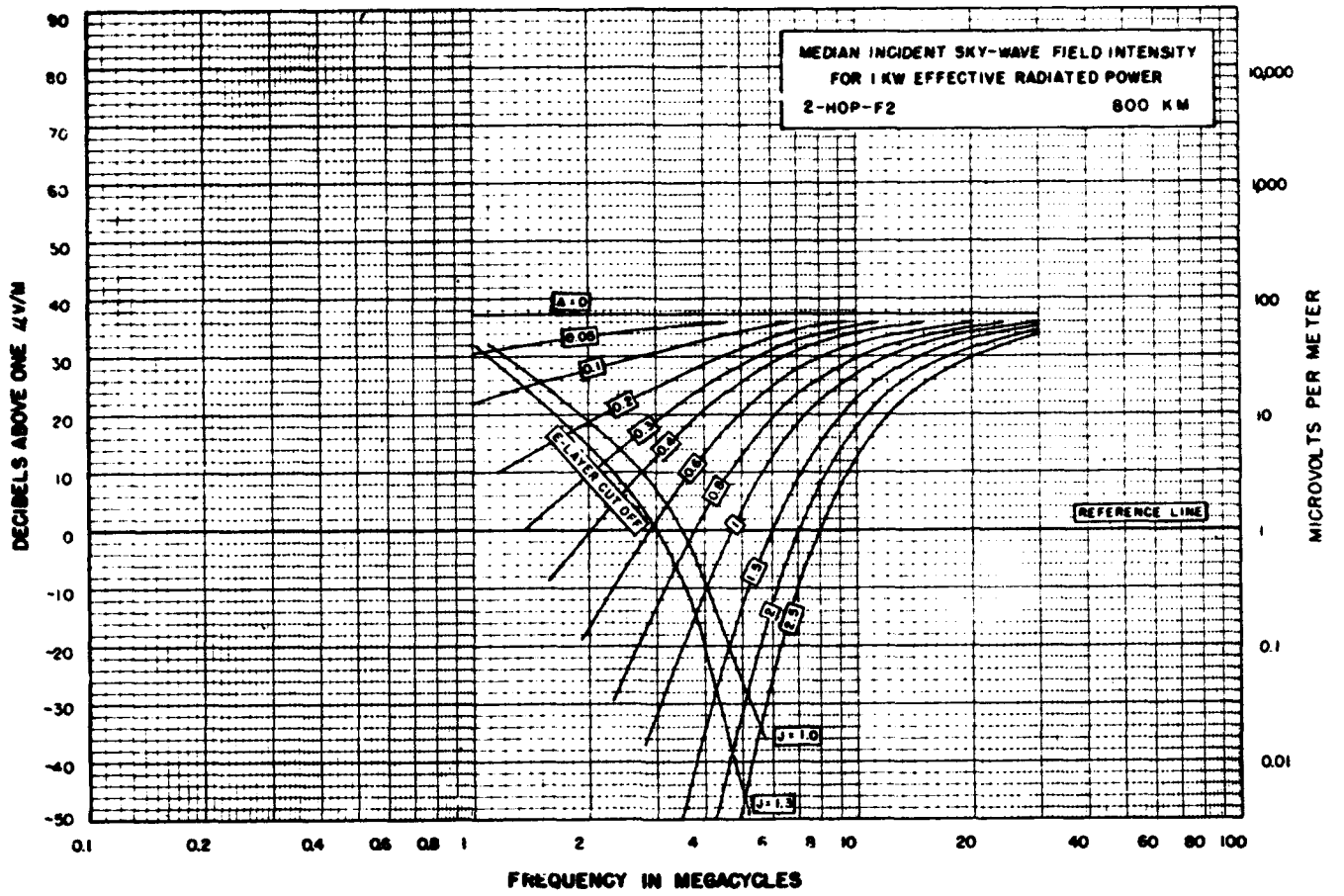
FIGURES 7.13 and 7.14.

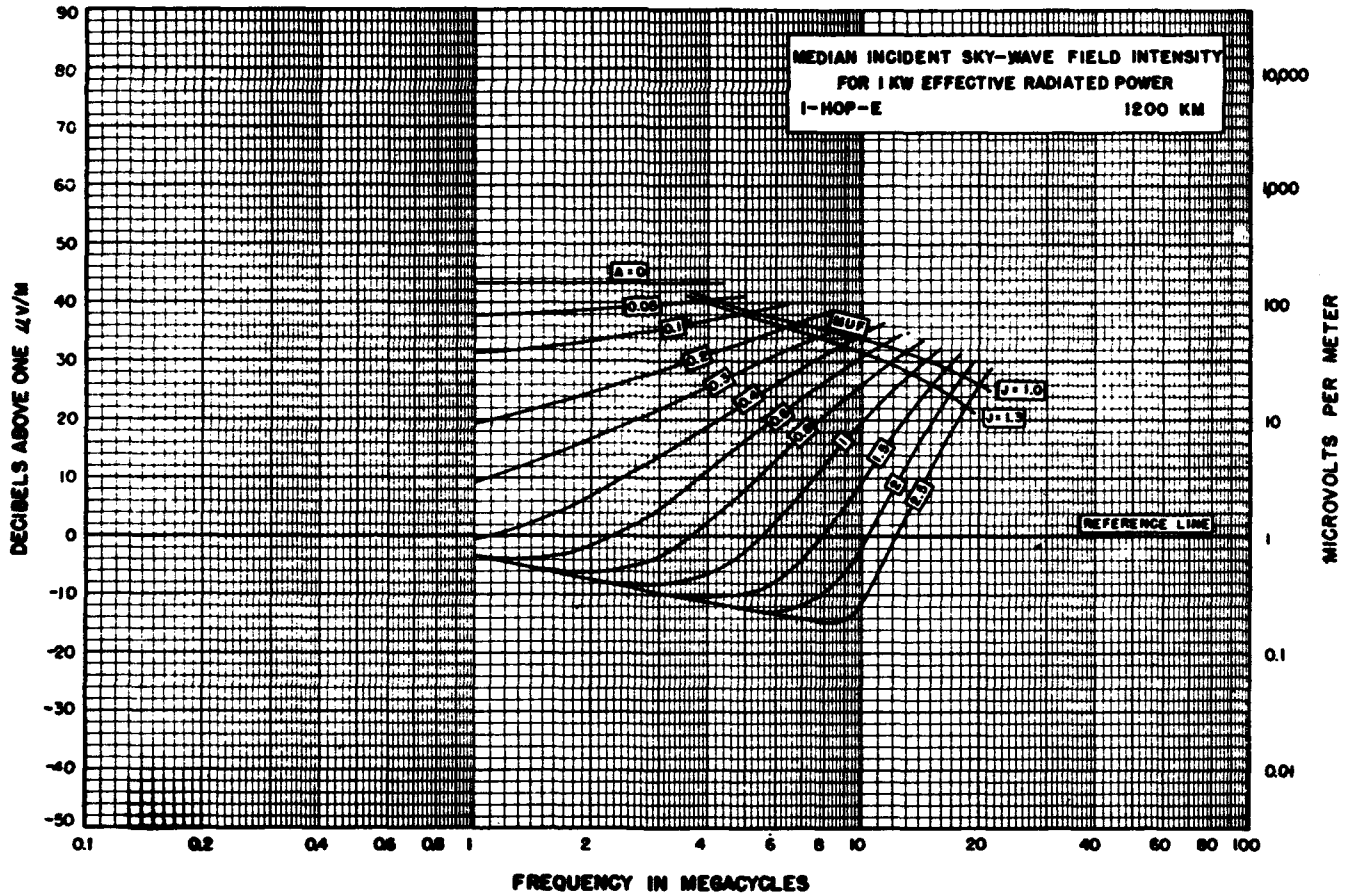




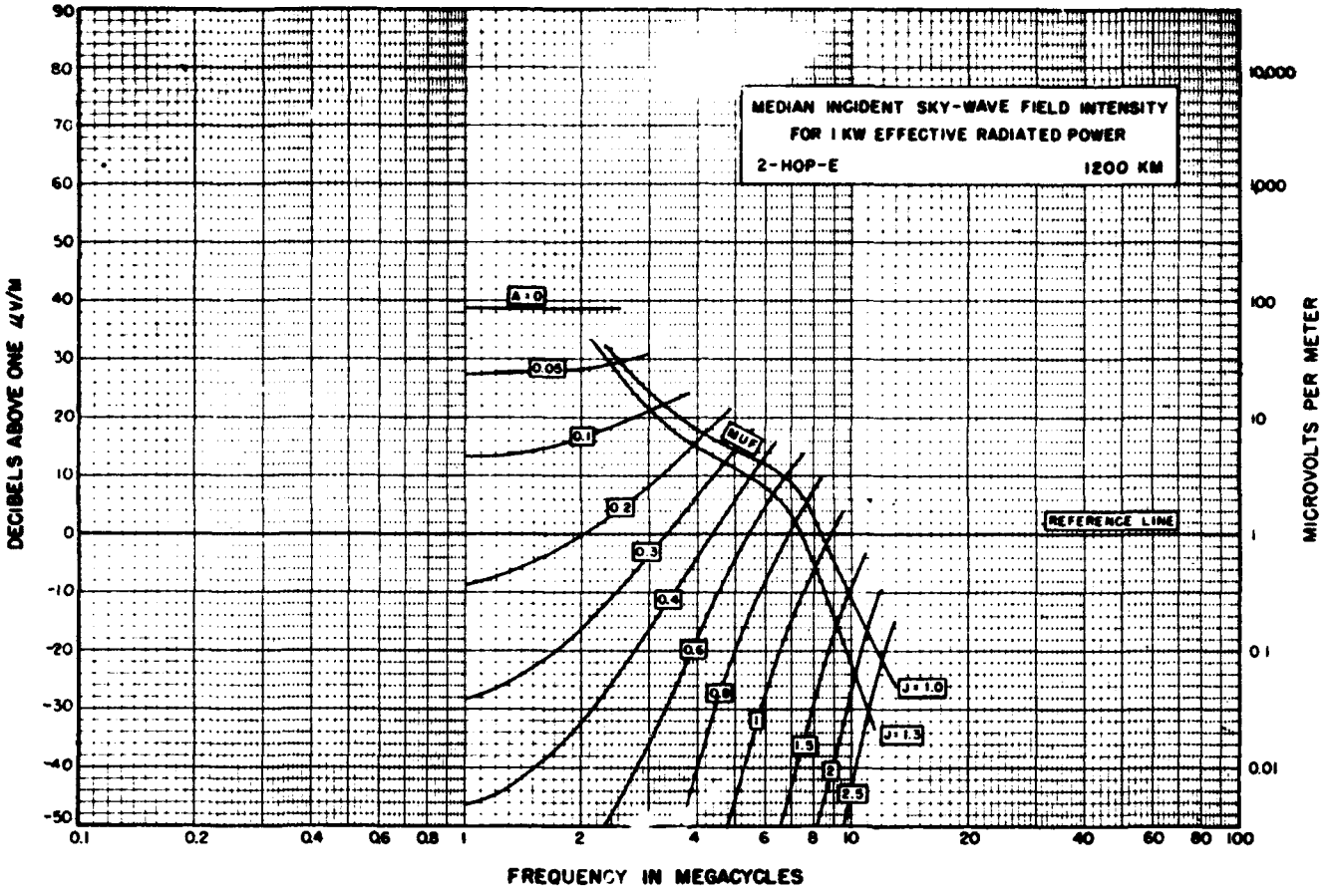
FIGURES 7.15 and 7.16.

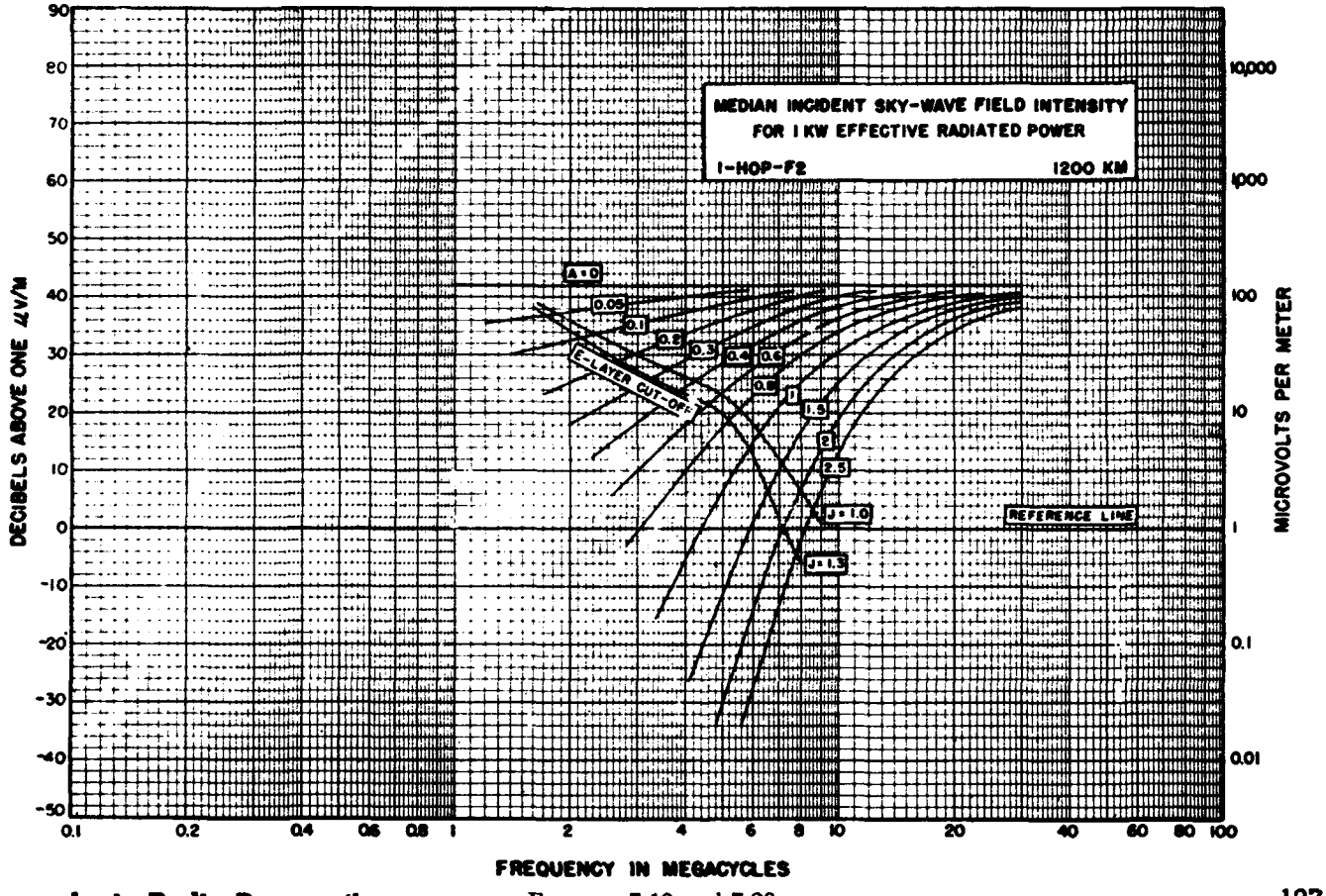
935333 O - 51 - 9



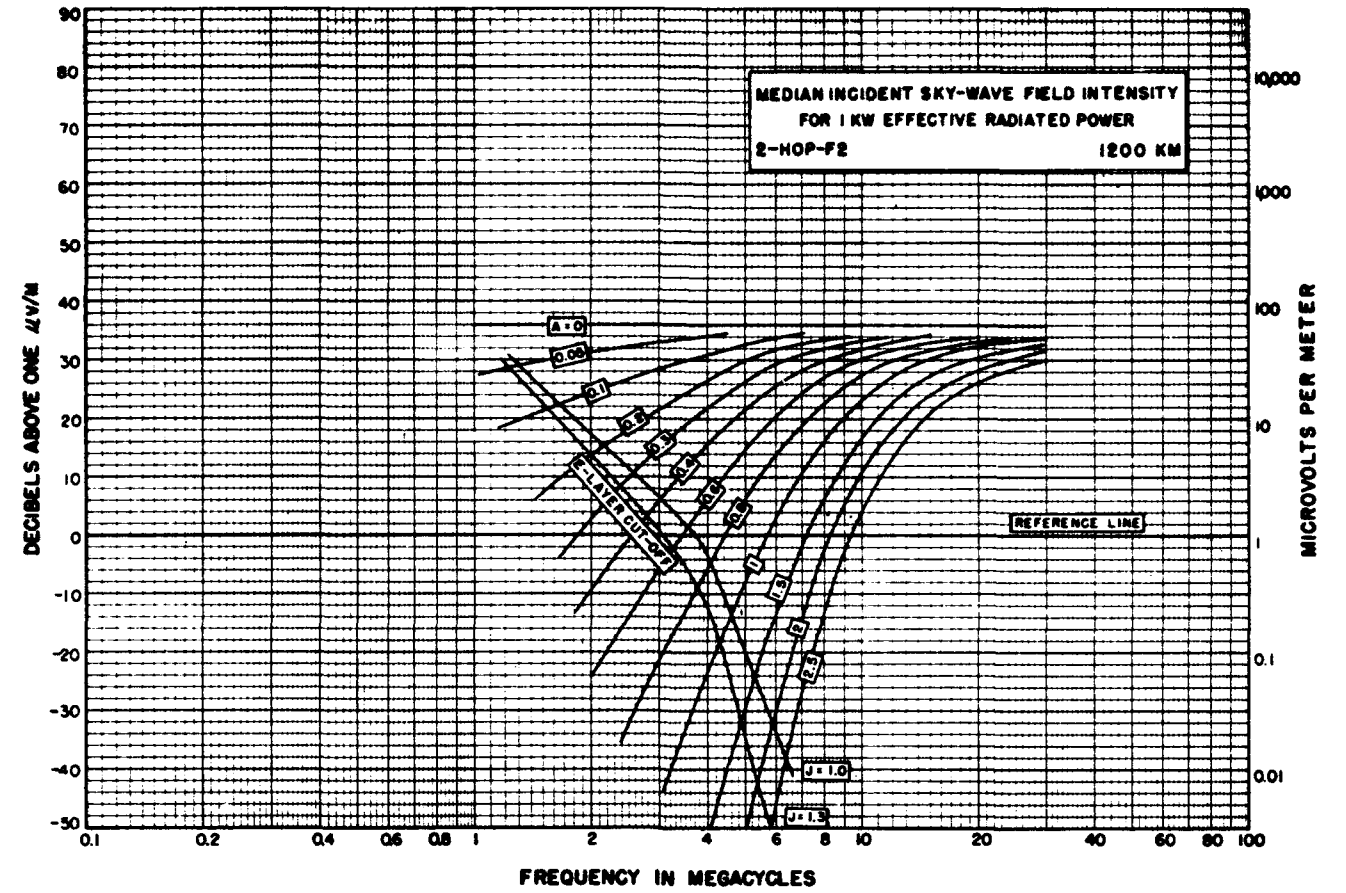


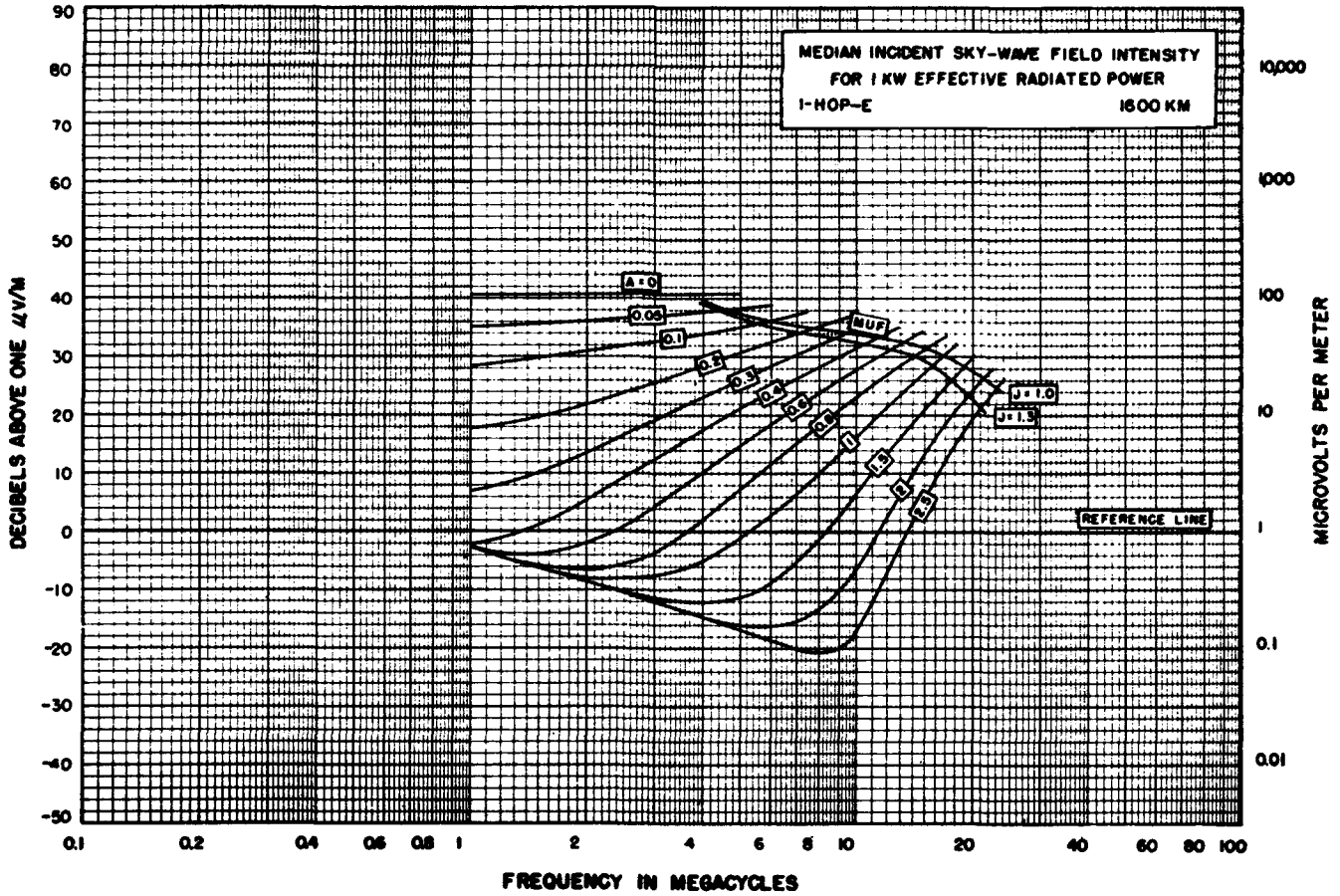
FIGURES 7.17 and 7.18.



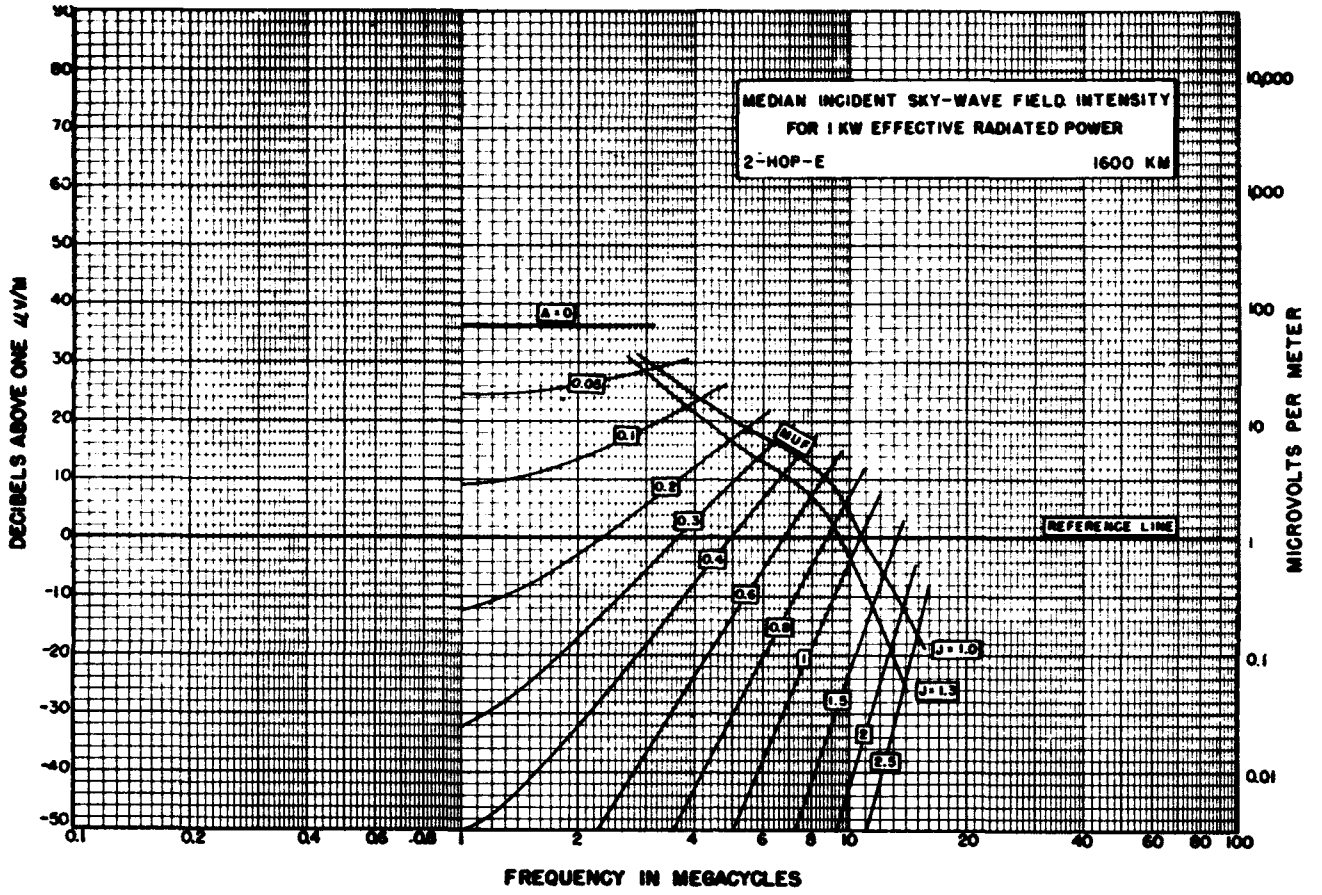


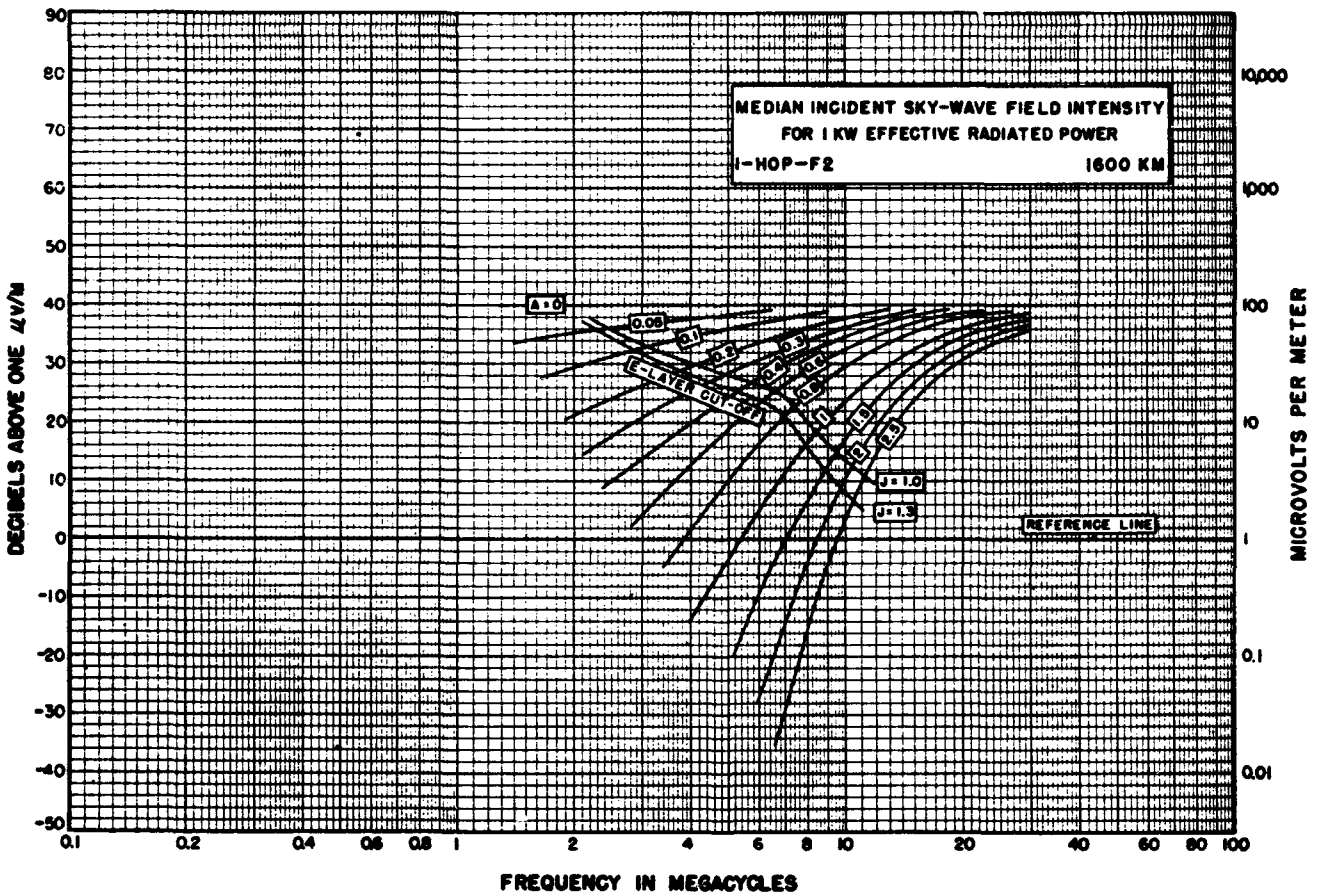
FIGURES 7.19 and 7.20.



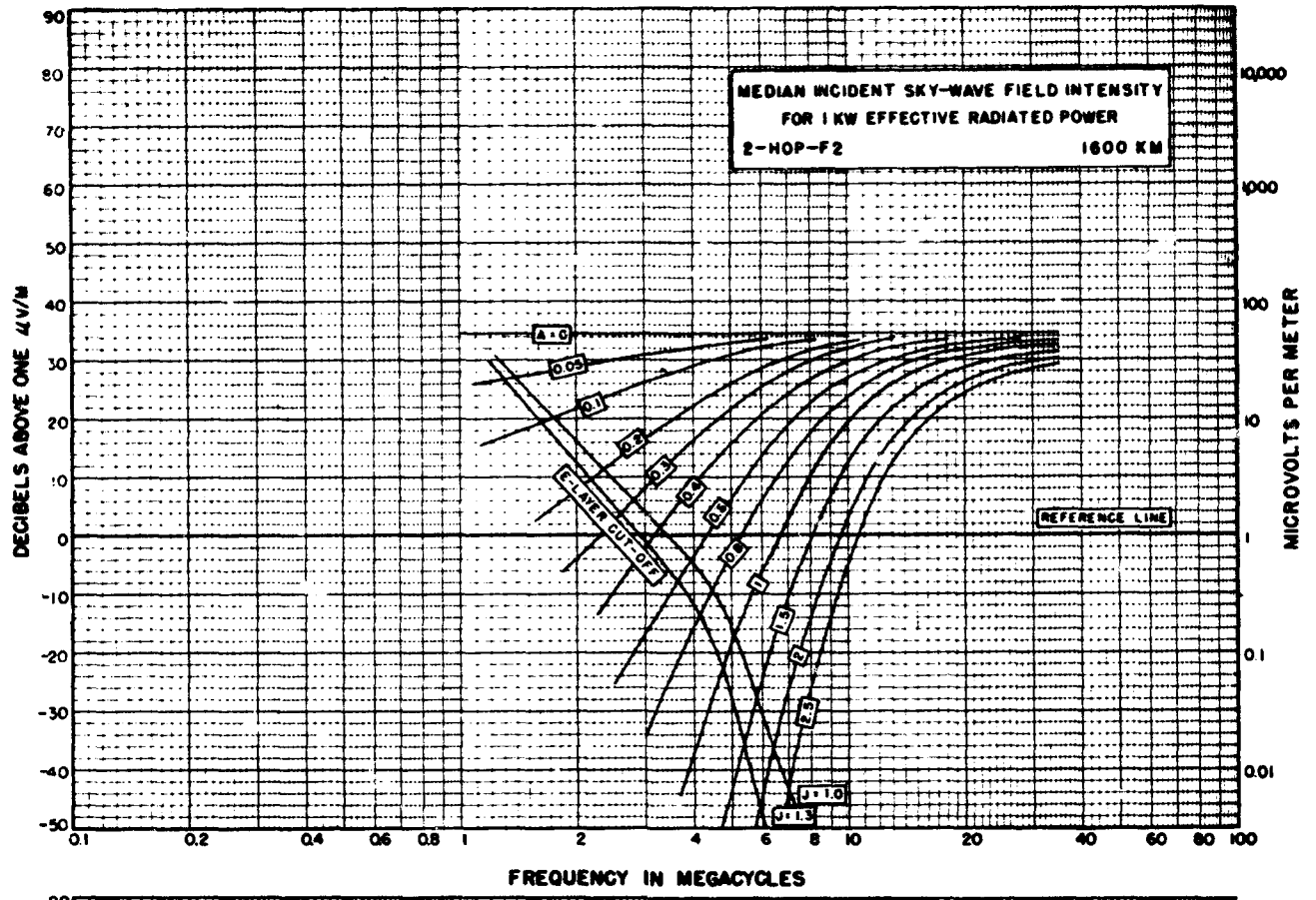


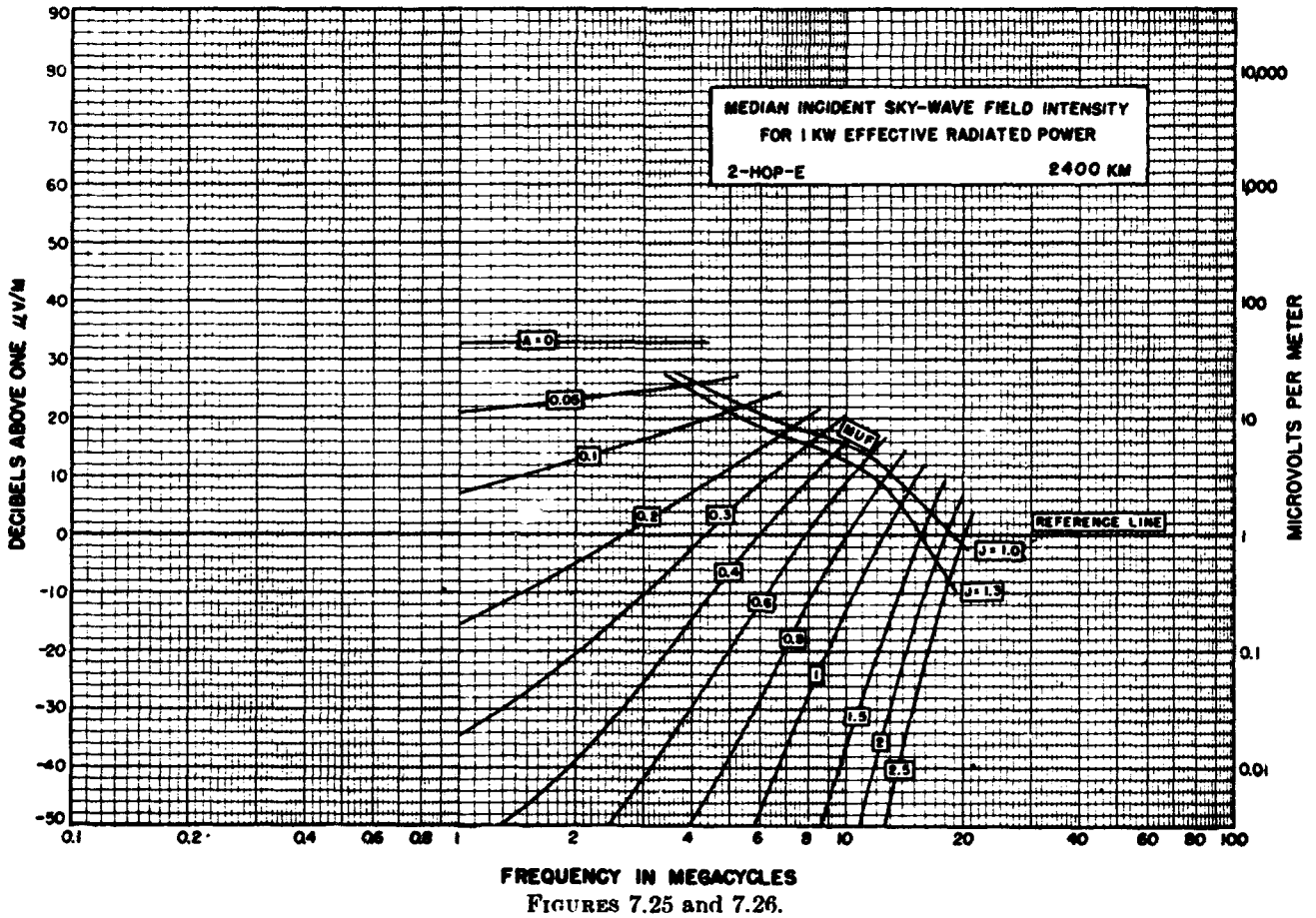
FIGURES 7.21 and 7.22.



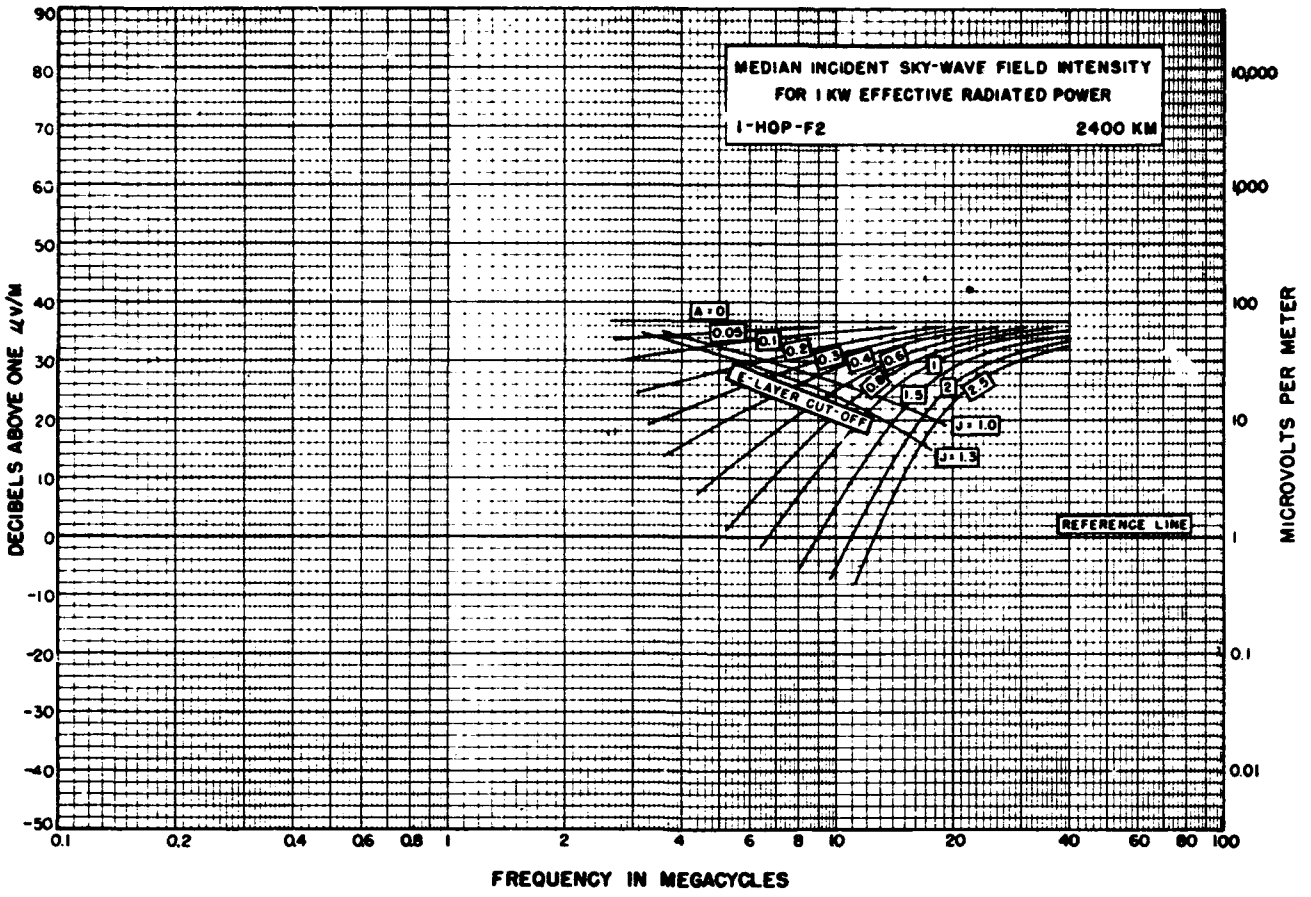


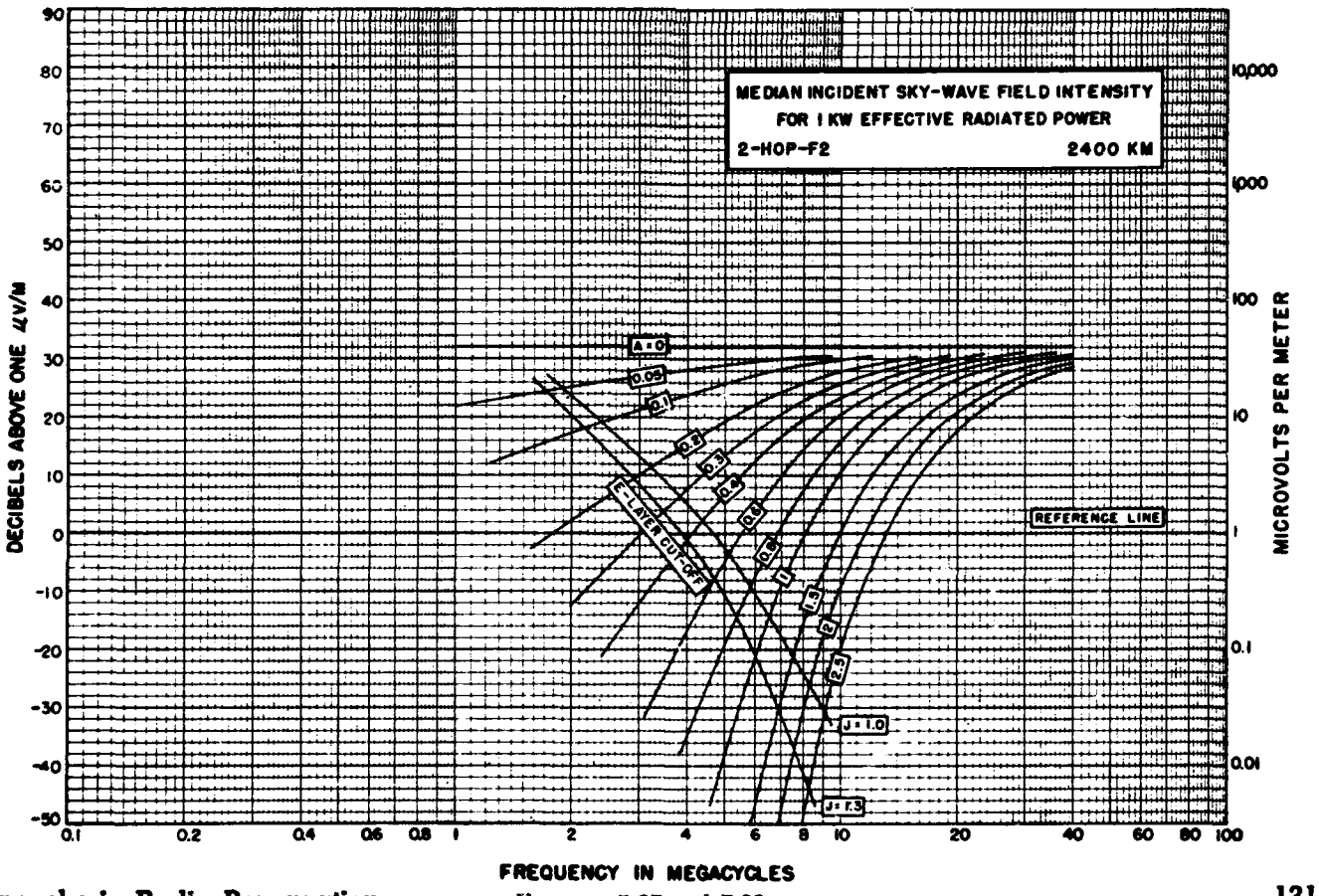
FIGURES 7.23 and 7.24.





FIGURES 7.25 and 7.26.





FIGURES 7.27 and 7.28.

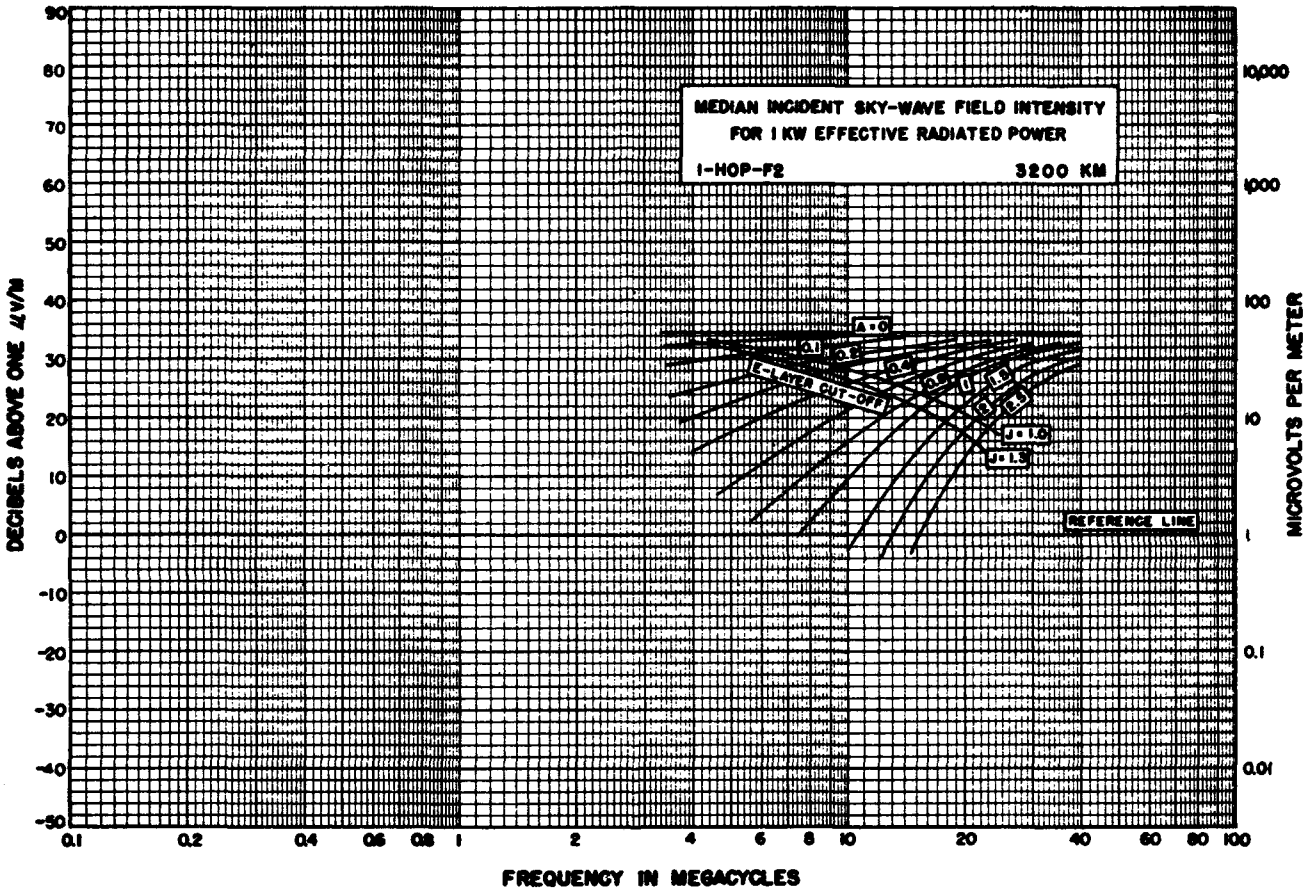

MEDIAN INCIDENT SKY-WAVE FIELD INTENSITY
FOR 1 KW EFFECTIVE RADIATED POWER
1-HOP-F2
3200 KM
E-LAYER CUT-OFF
REFERENCE LINE
DECIBELS ABOVE ONE μV/M
MICROVOLTS PER METER
FREQUENCY IN MEGACYCLES

FIGURE 7.29

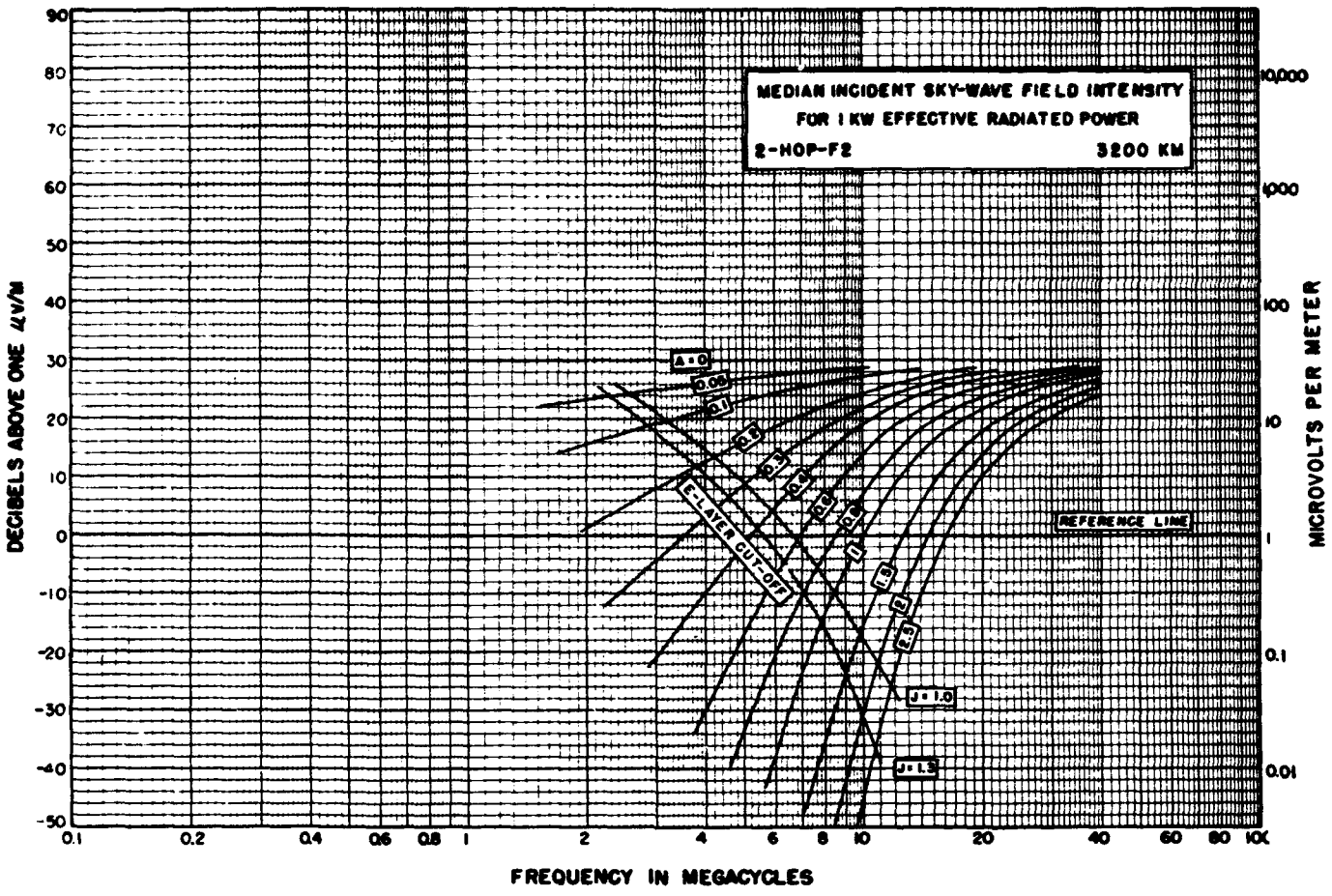


FIGURE 7.30. See Errata in front of book

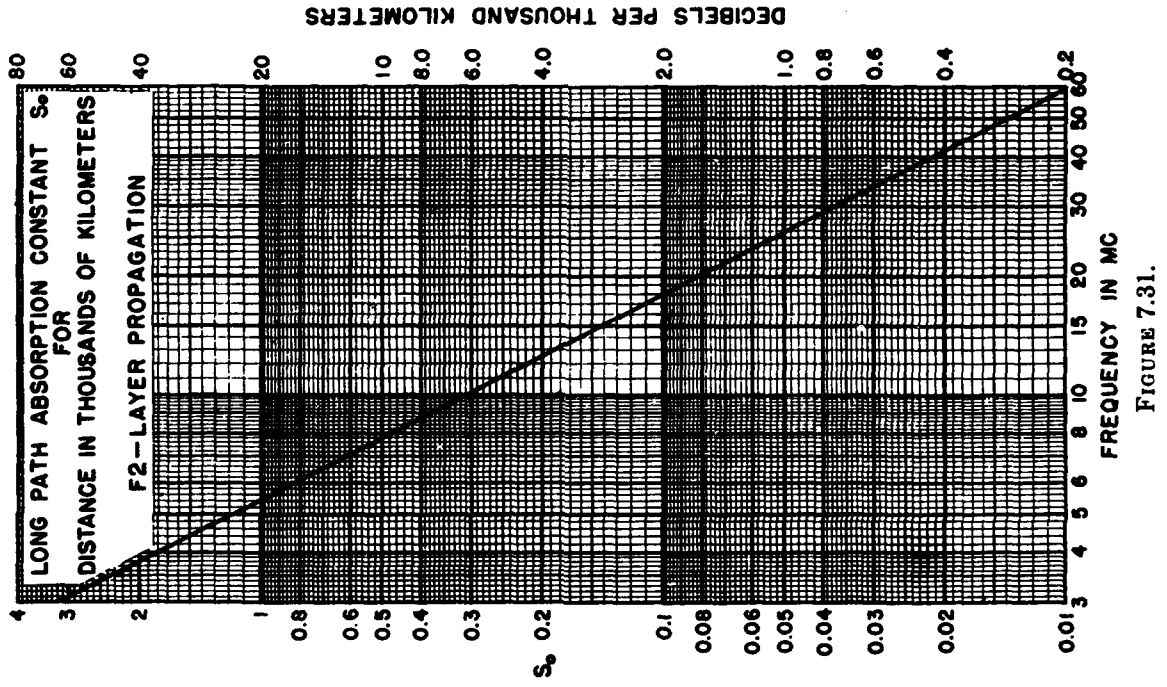


FIGURE 7.31.

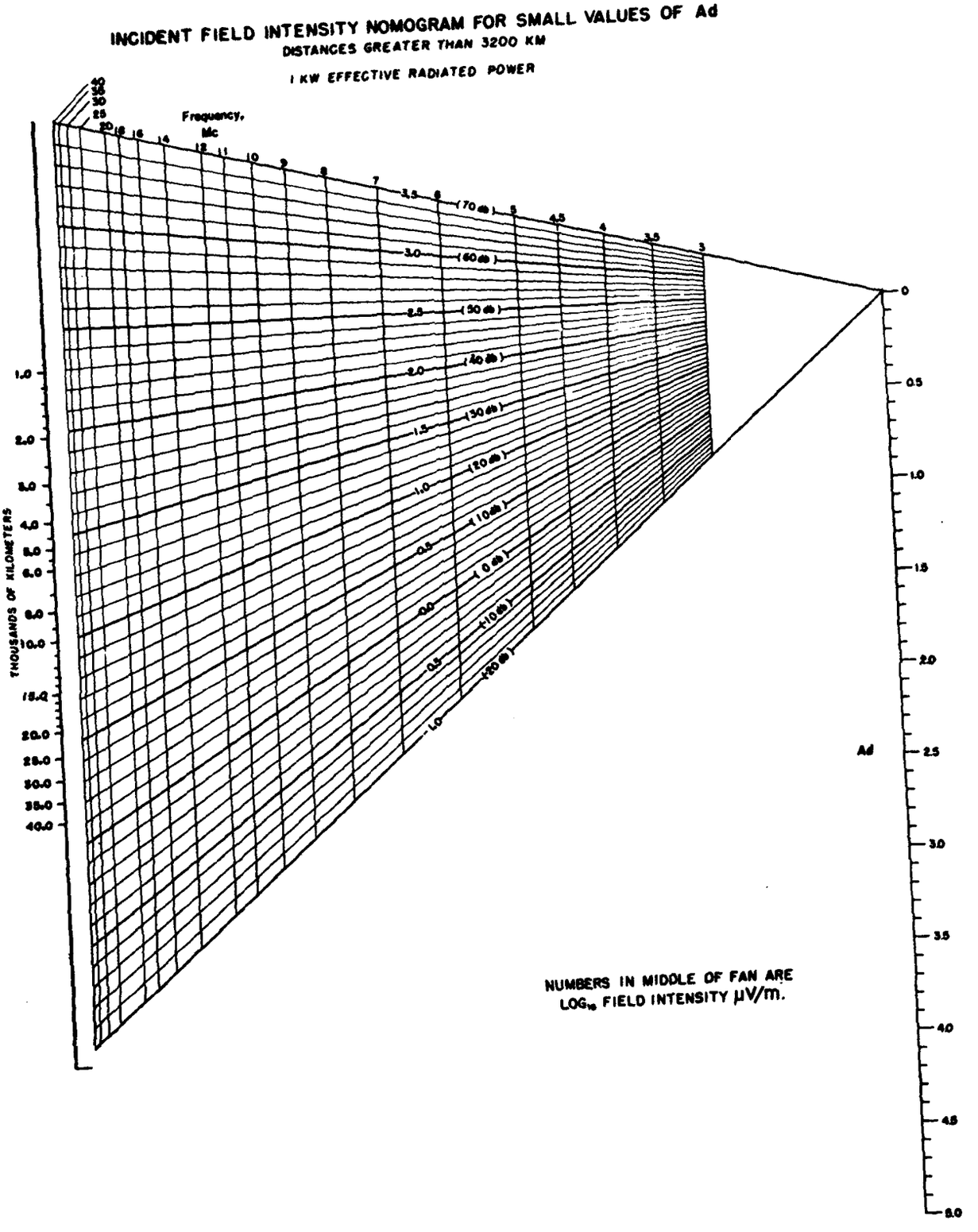

INCIDENT FIELD INTENSITY NOMOGRAM FOR SMALL VALUES OF Ad
DISTANCES GREATER THAN 3200 KM
1 KW EFFECTIVE RADIATED POWER
Frequency, Mc
THOUSANDS OF KILOMETERS
Ad
NUMBERS IN MIDDLE OF FAN ARE
LOG10 FIELD INTENSITY μV/m.

FIGURE 7.32.

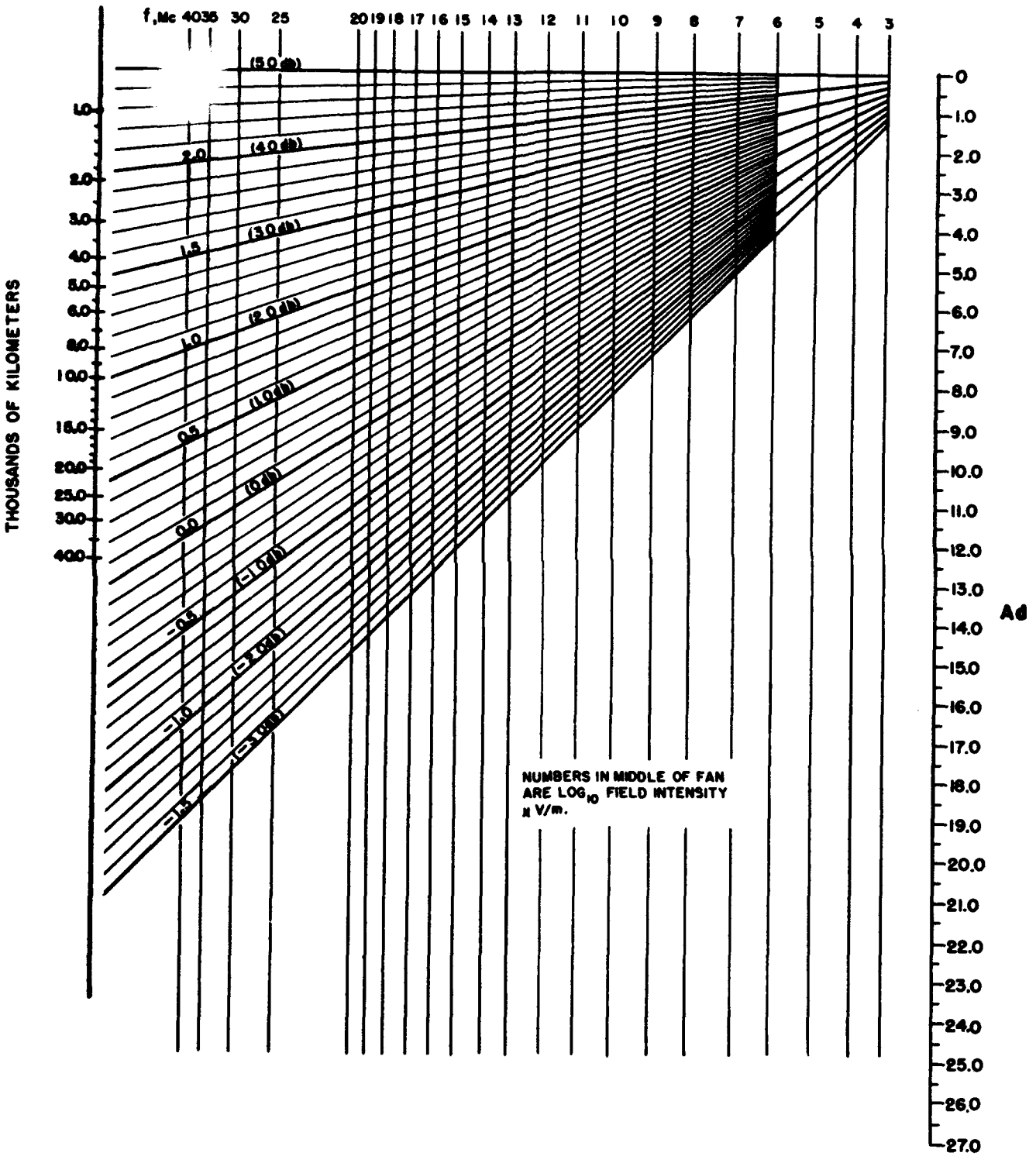

INCIDENT FIELD INTENSITY NOMOGRAM FOR LARGE VALUES OF Ad
DISTANCES GREATER THAN 3200 KM
1 KW EFFECTIVE RAIDIATED POWER
f, Mc 40 35 30 25 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3
(5.0 db)
(4.0 db)
(3.0 db)
(2.0 db)
(1.0 db)
(0 db)
(-1.0 db)
(-2.0 db)
(-3.0 db)
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
THOUSANDS OF KILOMETERS
5
2.0
3.0
4.0
5.0
6.0
8.0
10.0
15.0
20.0
25.0
30.0
40.0
NUMBERS IN MIDDLE OF FAN ARE LOG10 FIELD INTENSITY μ V/m.
0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
11.0
12.0
13.0
14.0
15.0
16.0
17.0
18.0
19.0
20.0
21.0
22.0
23.0
24.0
25.0
26.0
27.0
Ad

FIGURE 7.33.

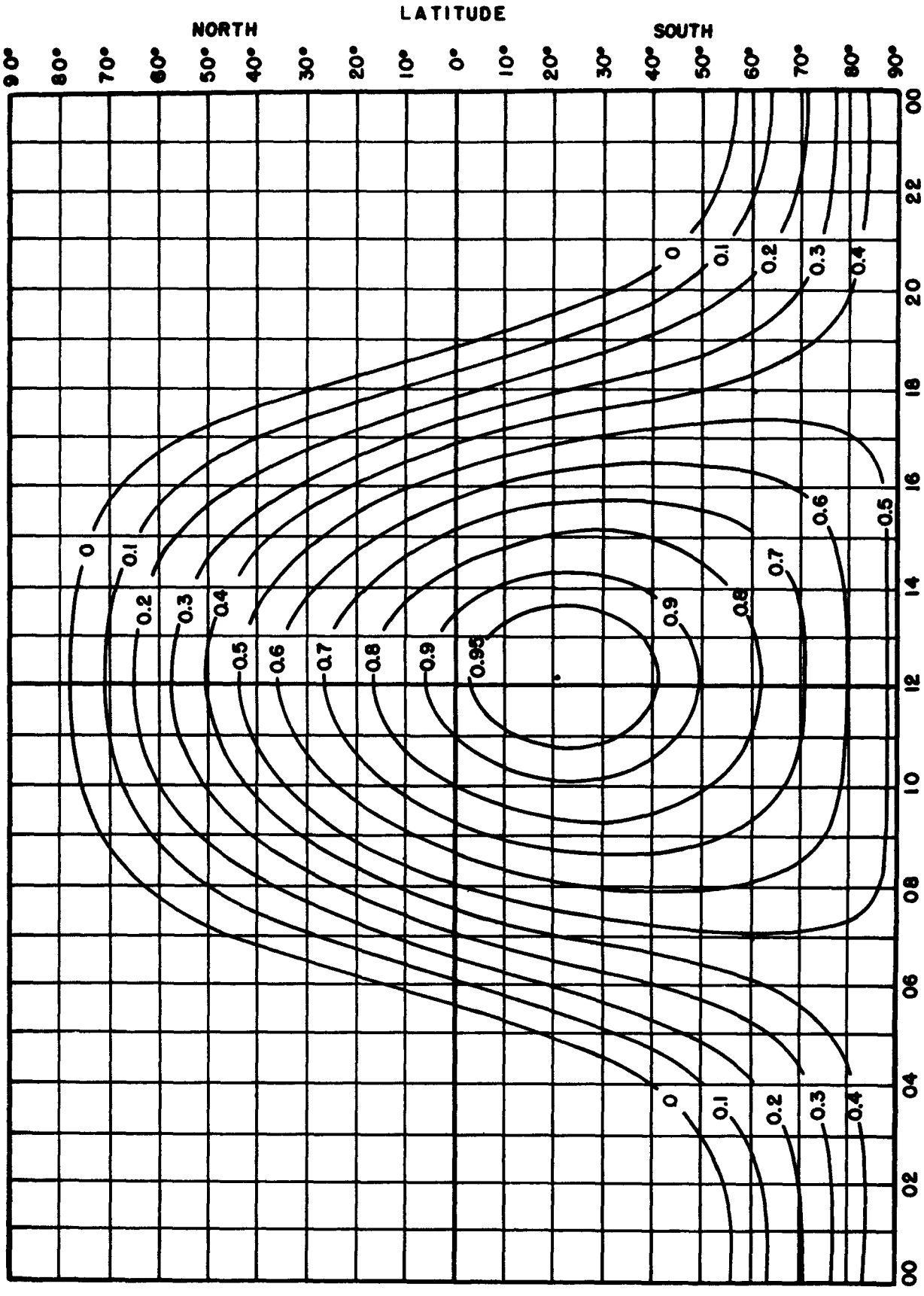


FIGURE 7.34. *Absorption factor, K, for January.*

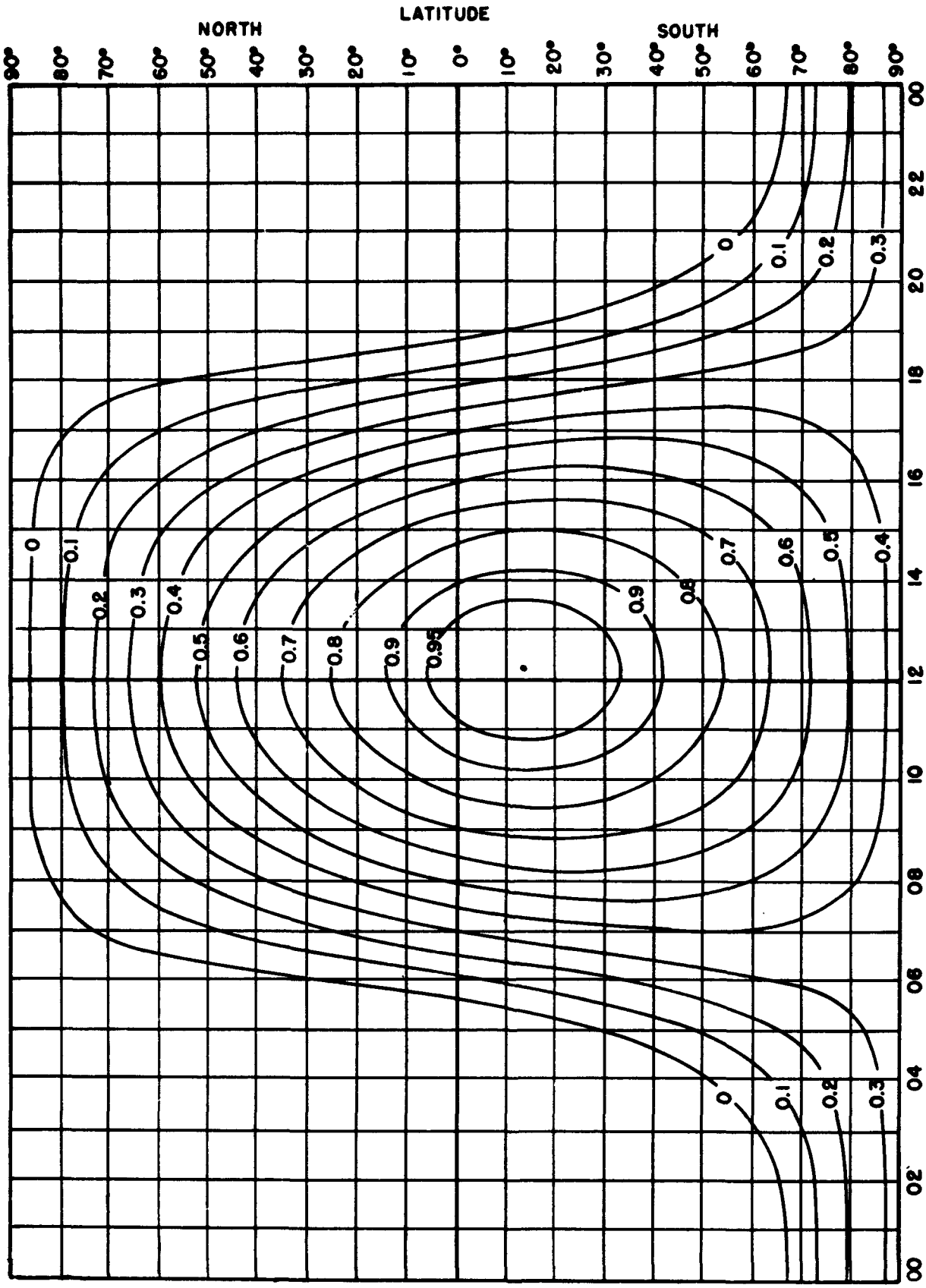


FIGURE 7.35. *Absorption factor, K, for February.*

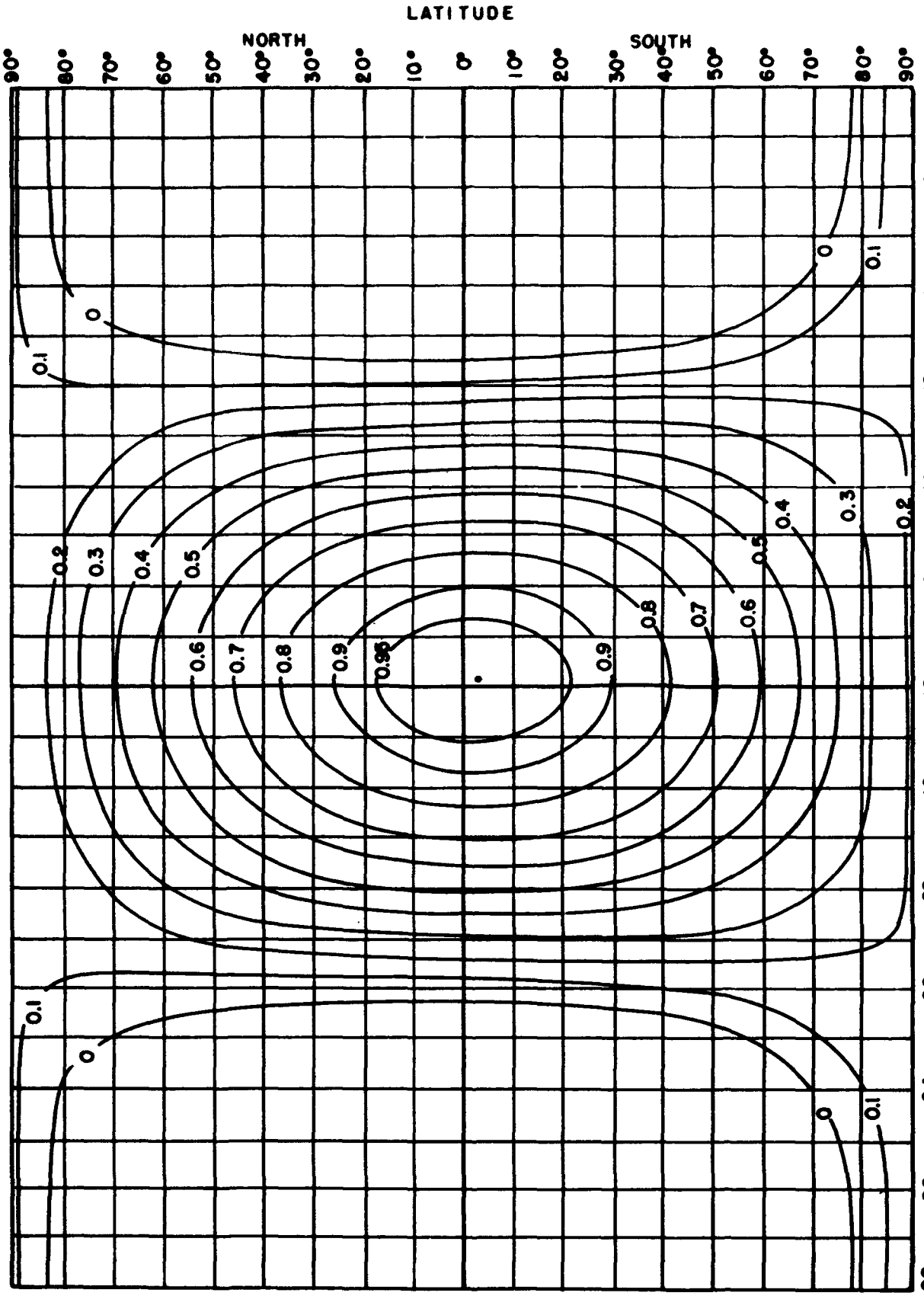


FIGURE 7.36. *Absorption factor, K, for March.*

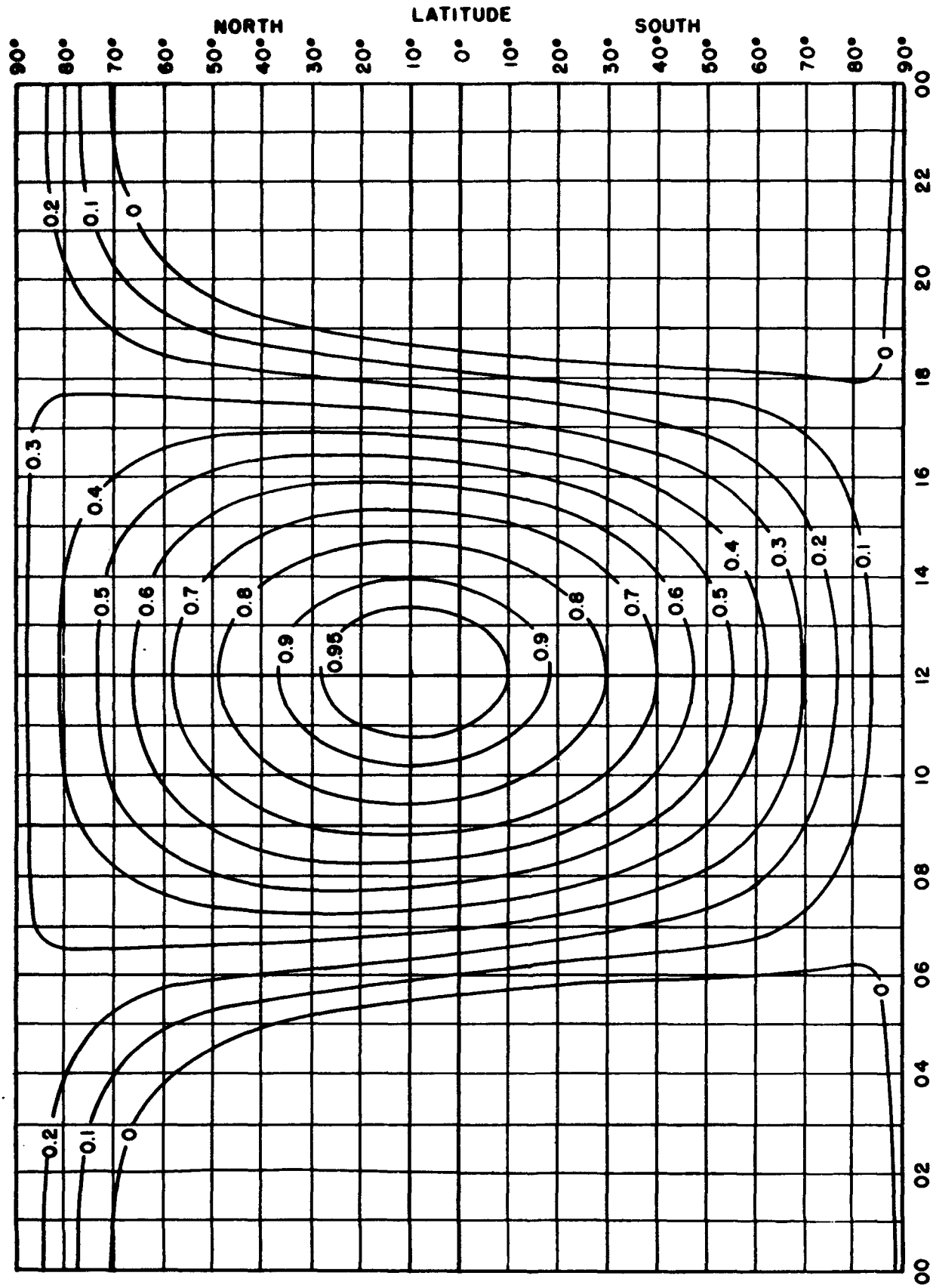


FIGURE 7.37. *Absorption factor, K, for April.*

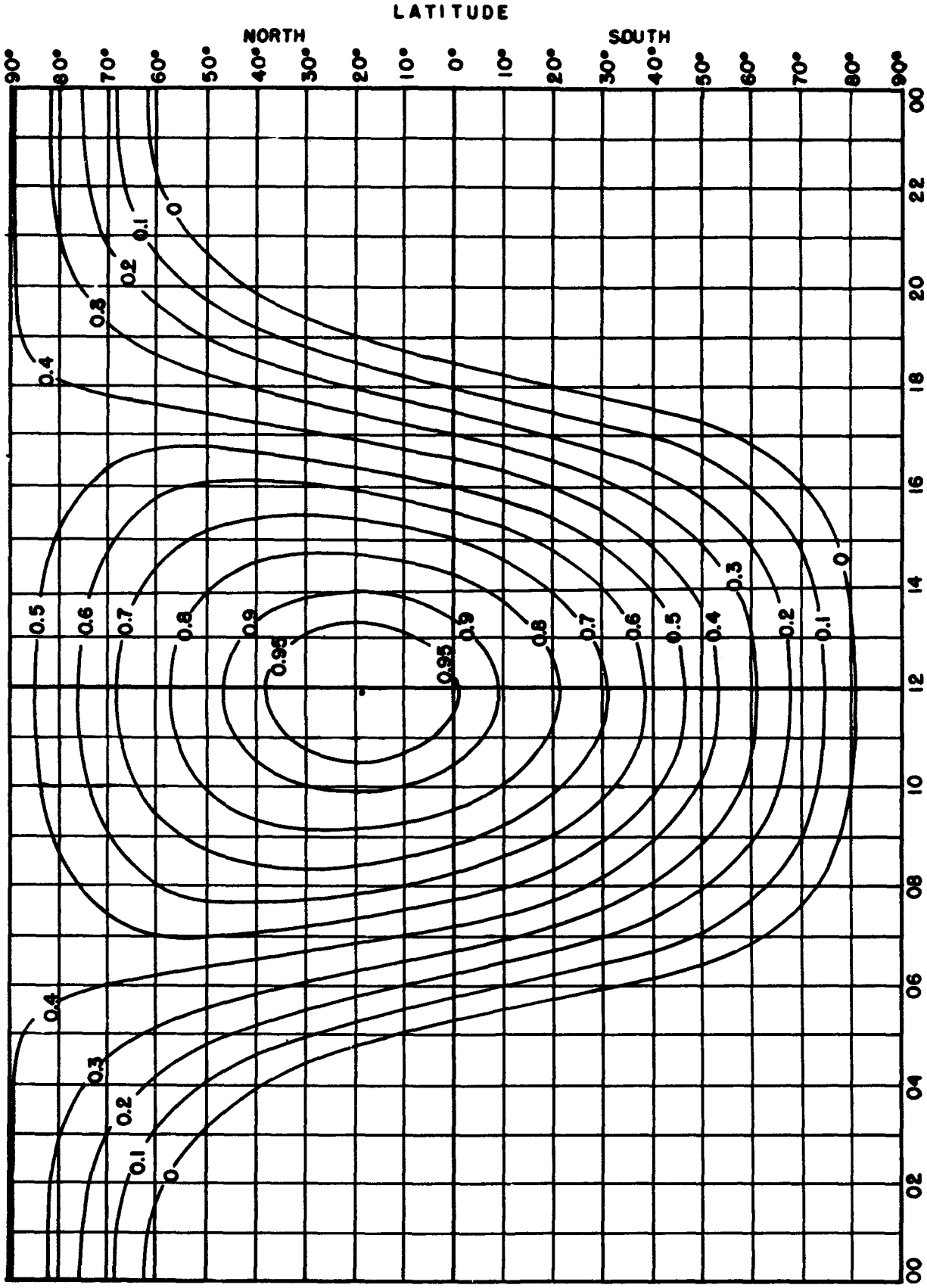


FIGURE 7.38. *Absorption factor, K, for May.*

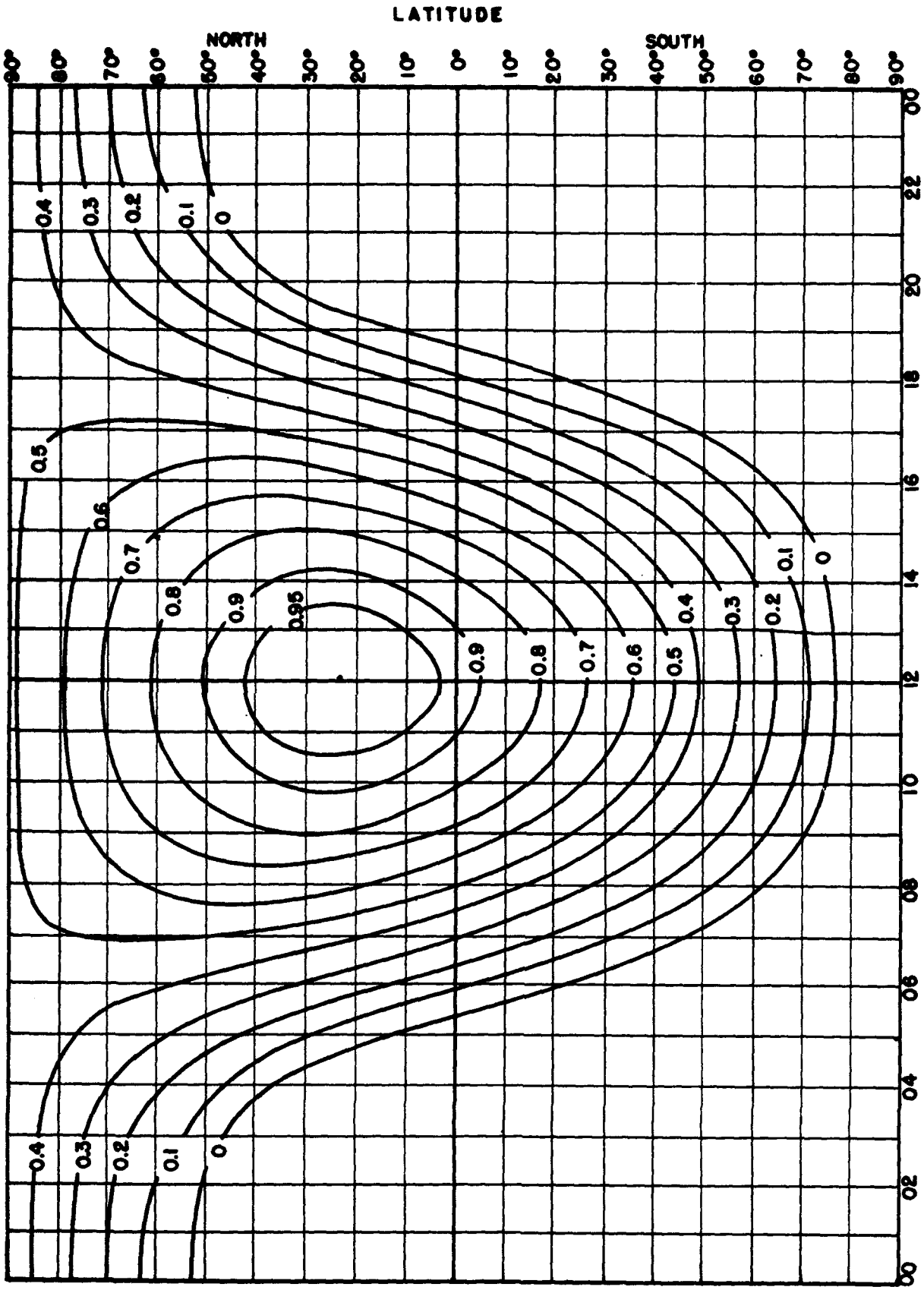


FIGURE 7.39. *Absorption factor, K, for June.*

935333 O - 51 - 10

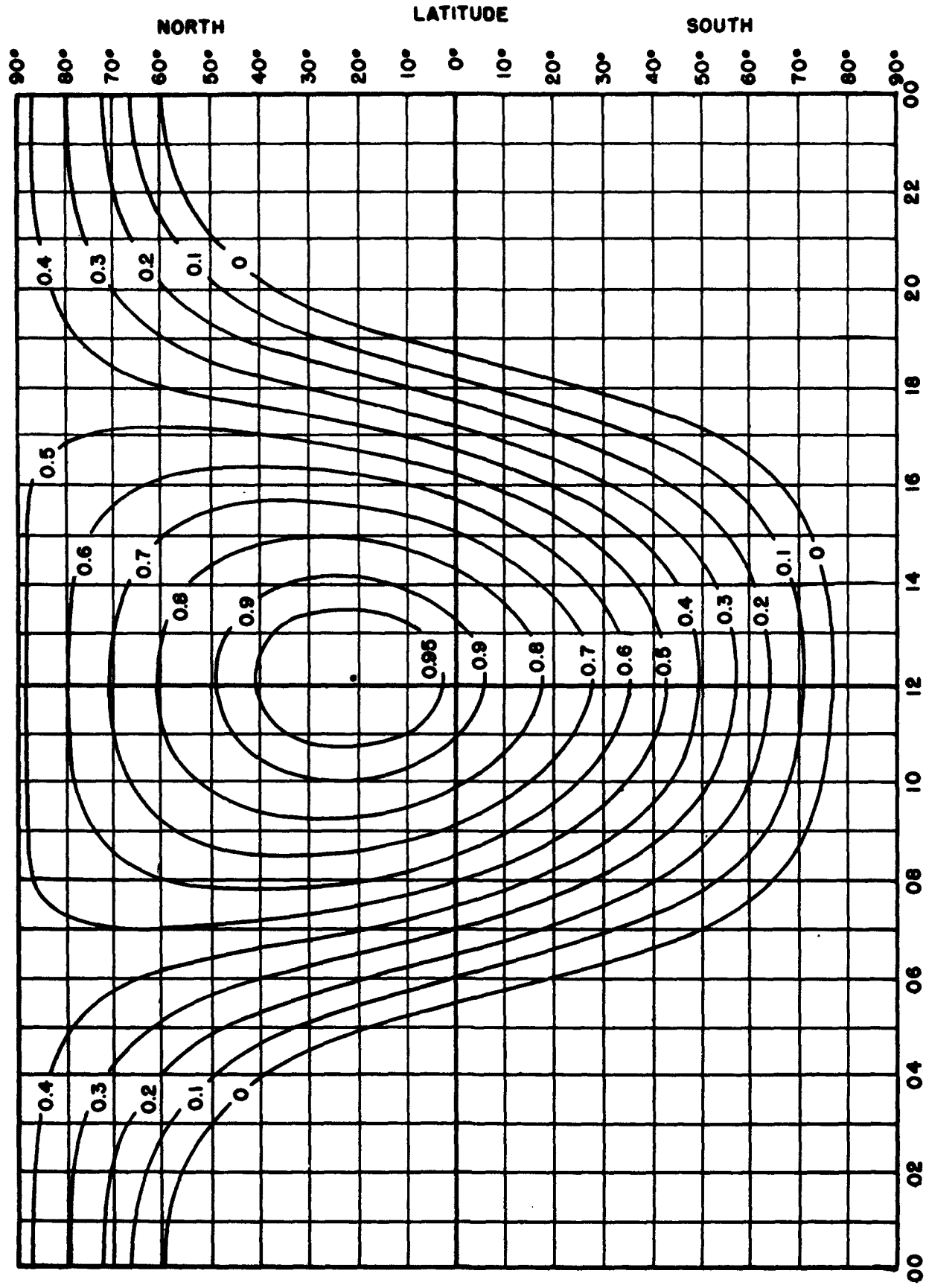


FIGURE 7.40. *Absorption factor, K, for July.*

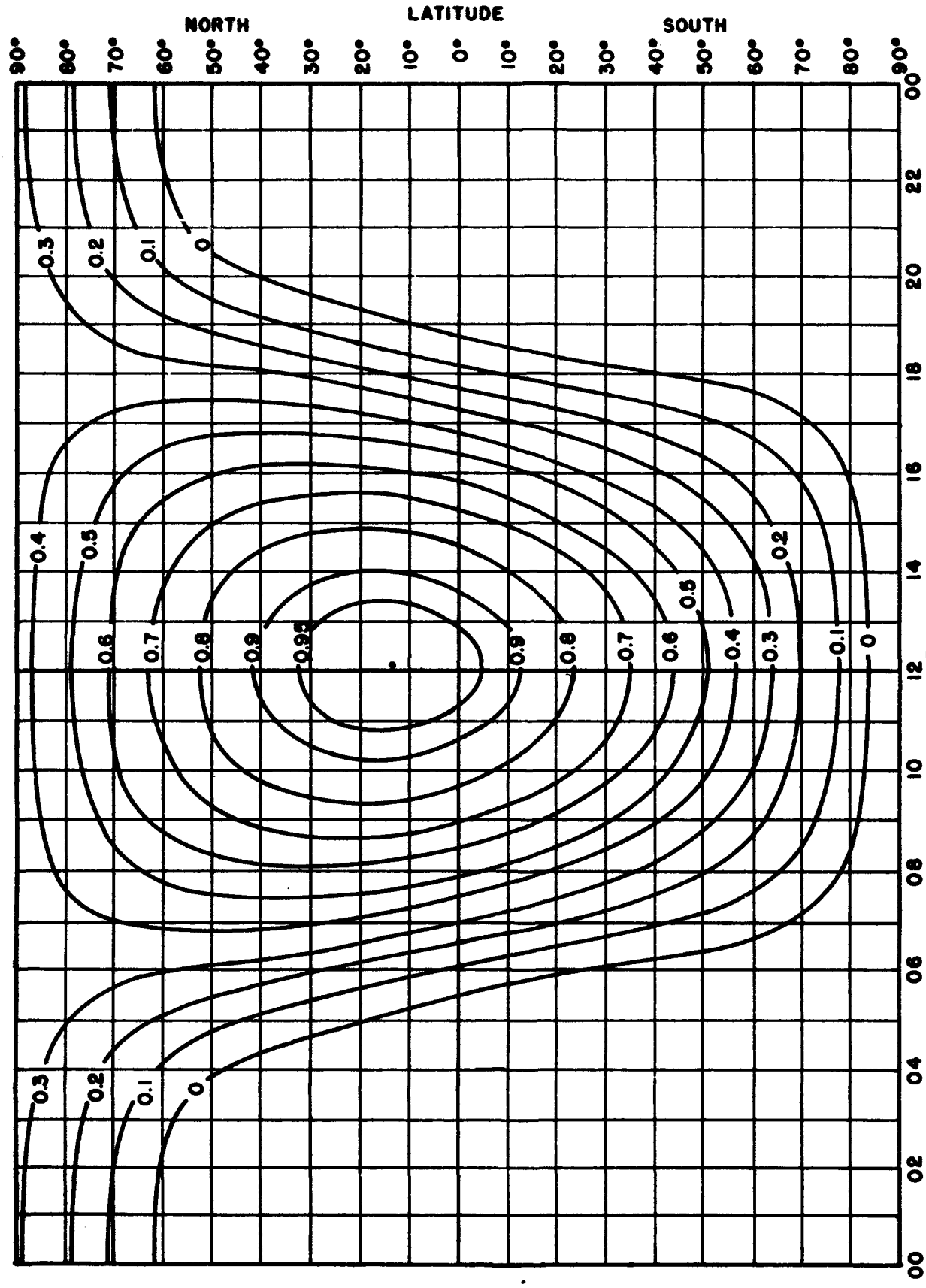


FIGURE 7.41. *Absorption factor, K, for August.*

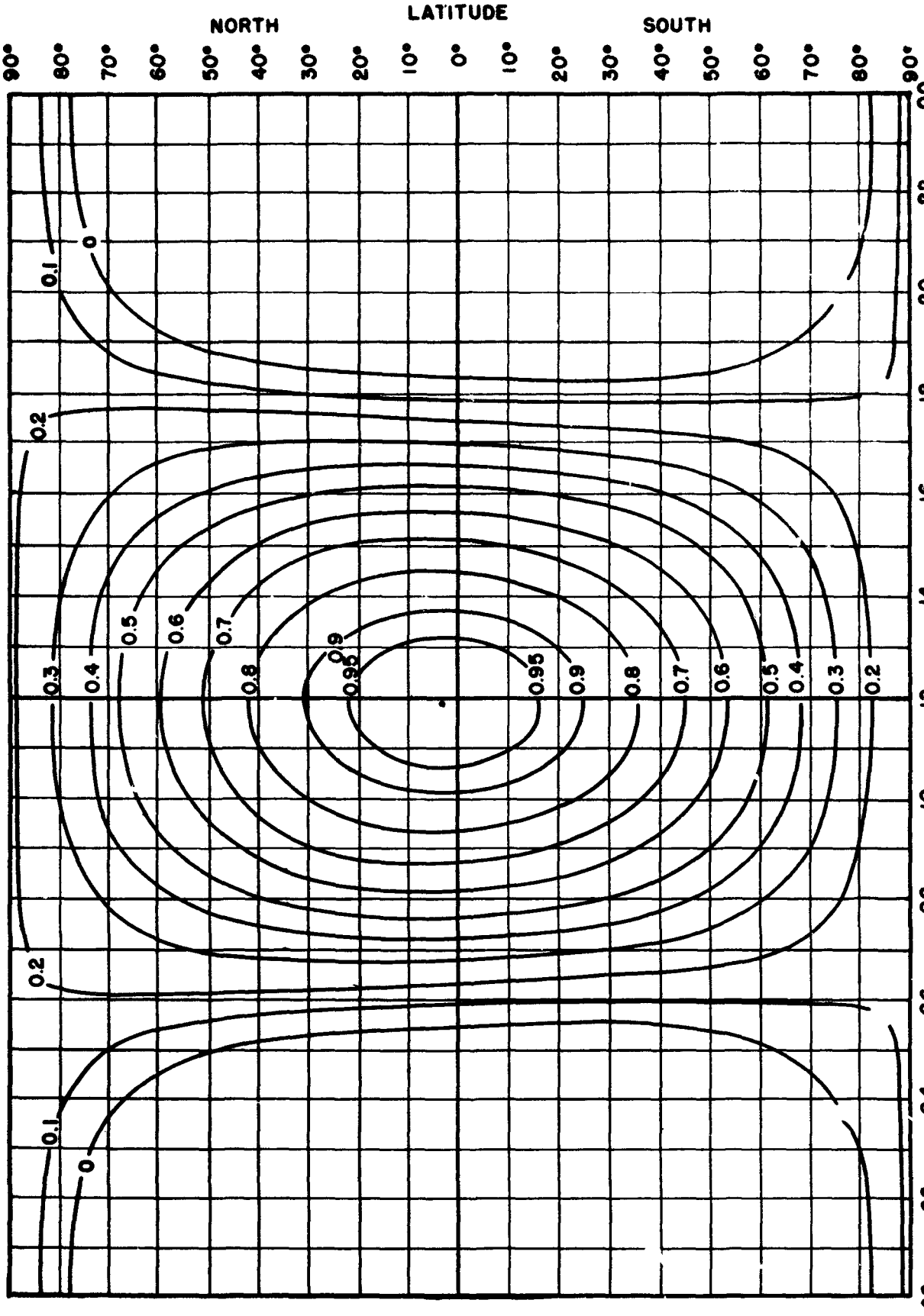


FIGURE 7.42. *Absorption factor, K, for September.*

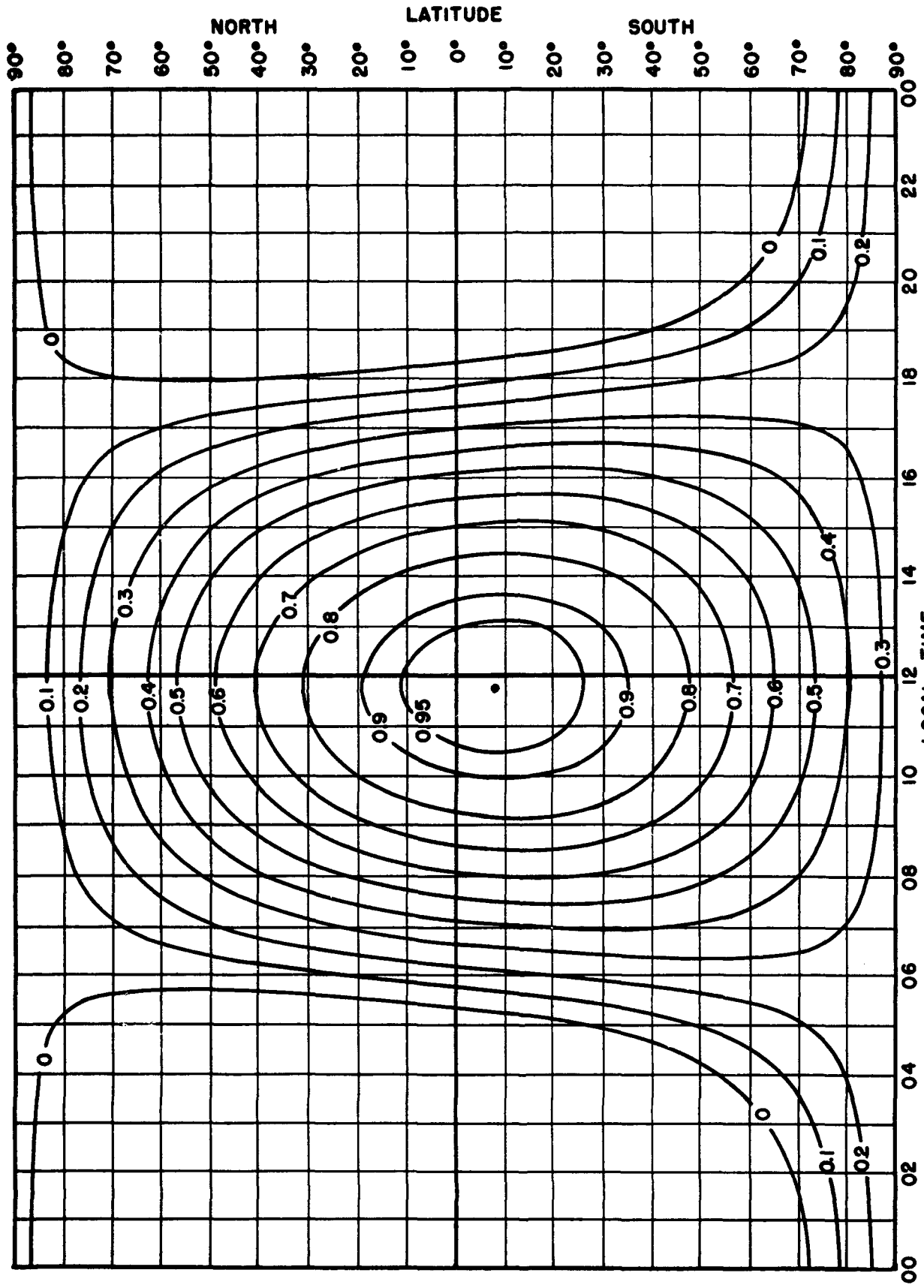


FIGURE 7.43. *Absorption factor, K, for October.*

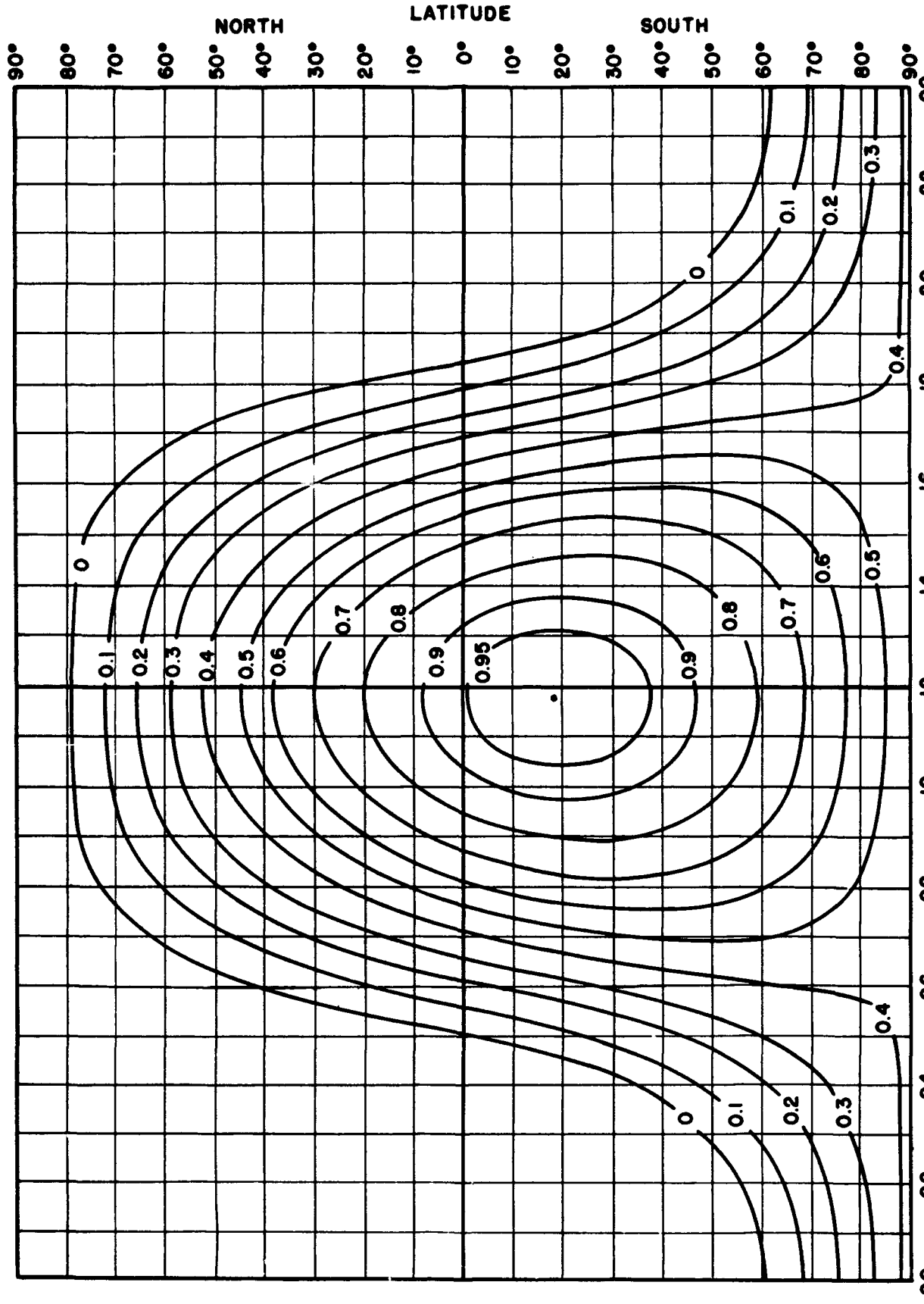


FIGURE 7.44. *Absorption factor, K, for November.*

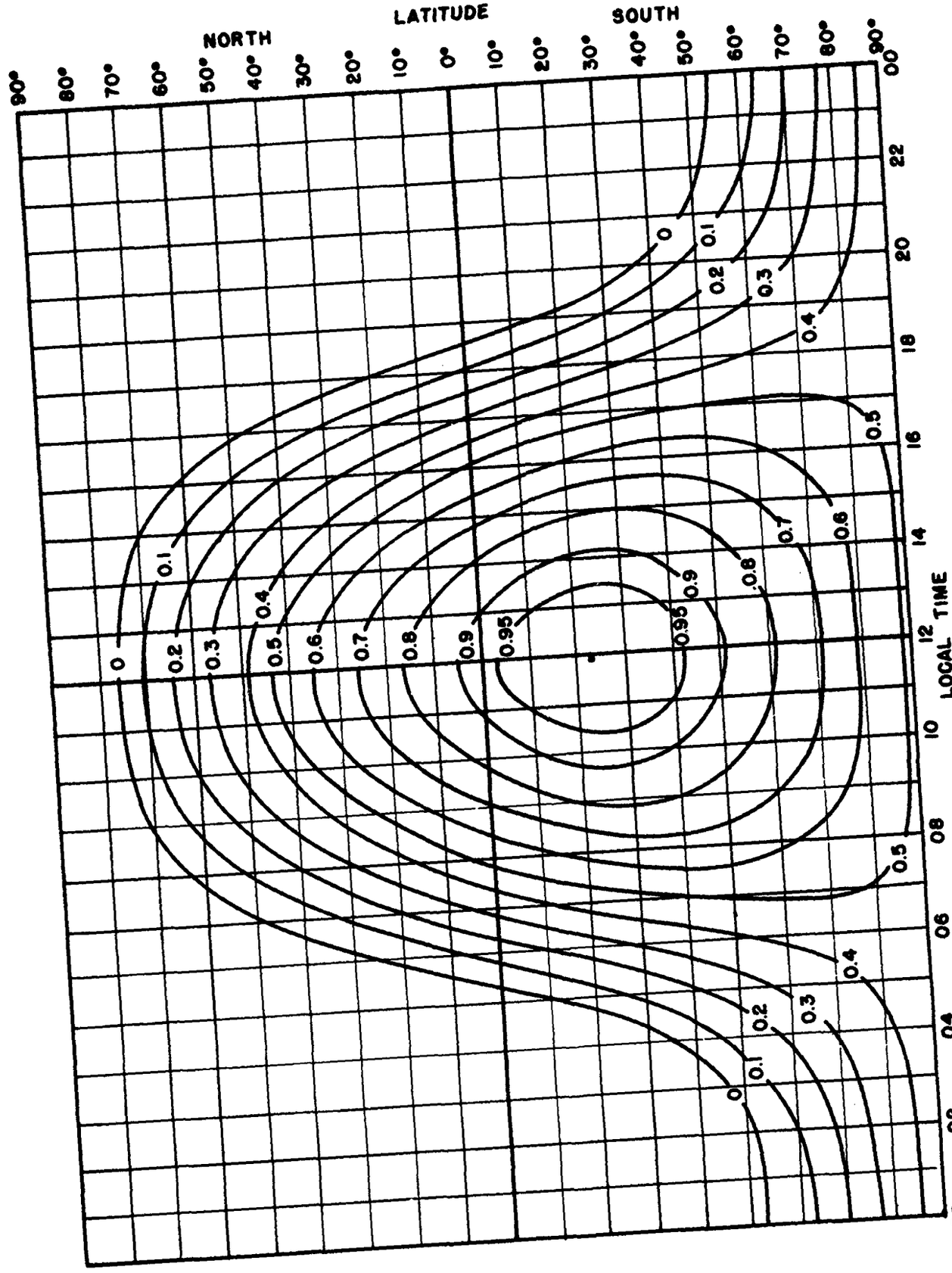


FIGURE 7.45. *Absorption factor, K, for December.*

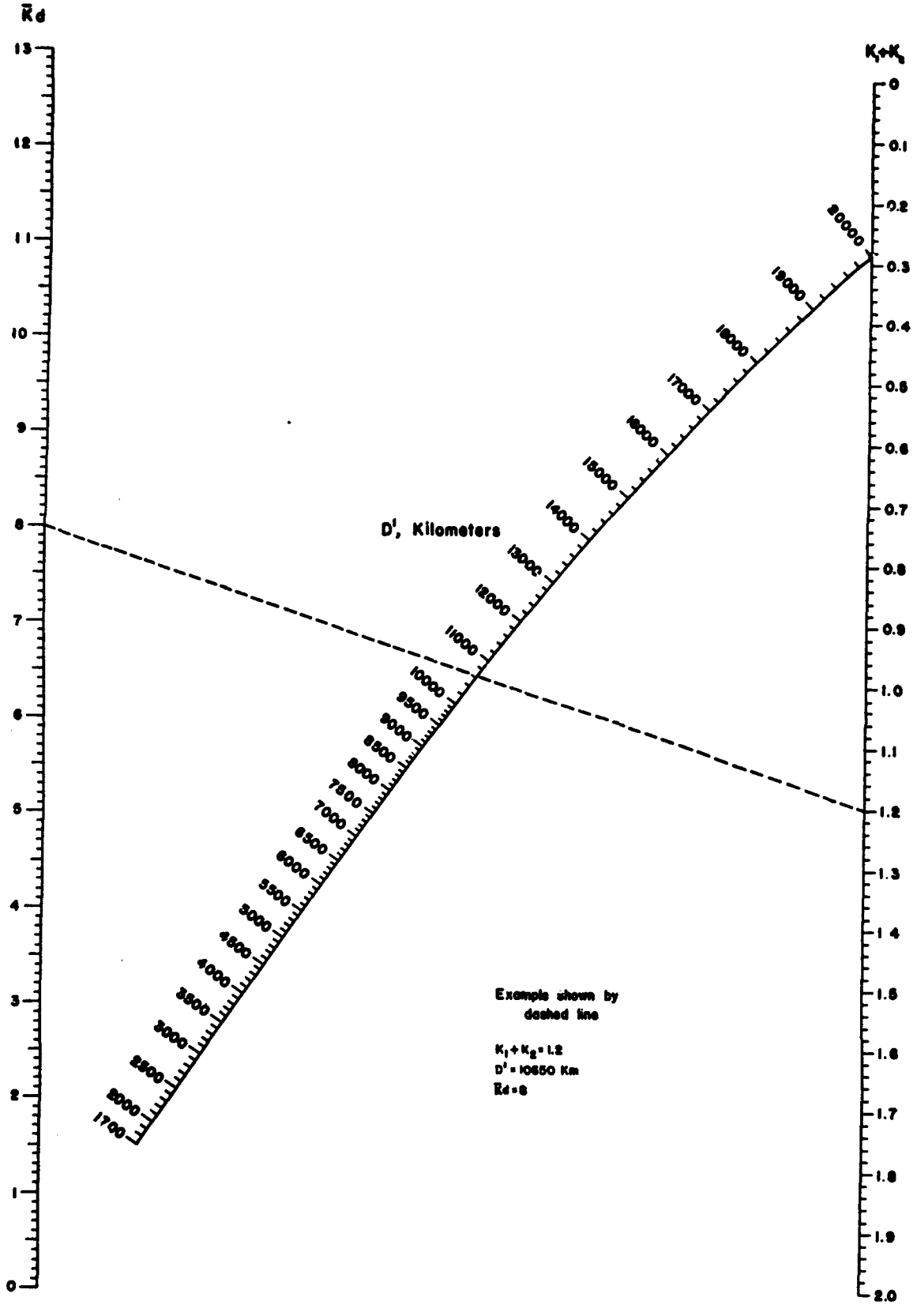


FIGURE 7.46. *Nomogram giving variation of* $\overline{K}d$ *with* K_1+K_2 *and* D' *(length of transmission path lying in the region where* $K>0$.).

Distance in kilometers, up to 20,000.

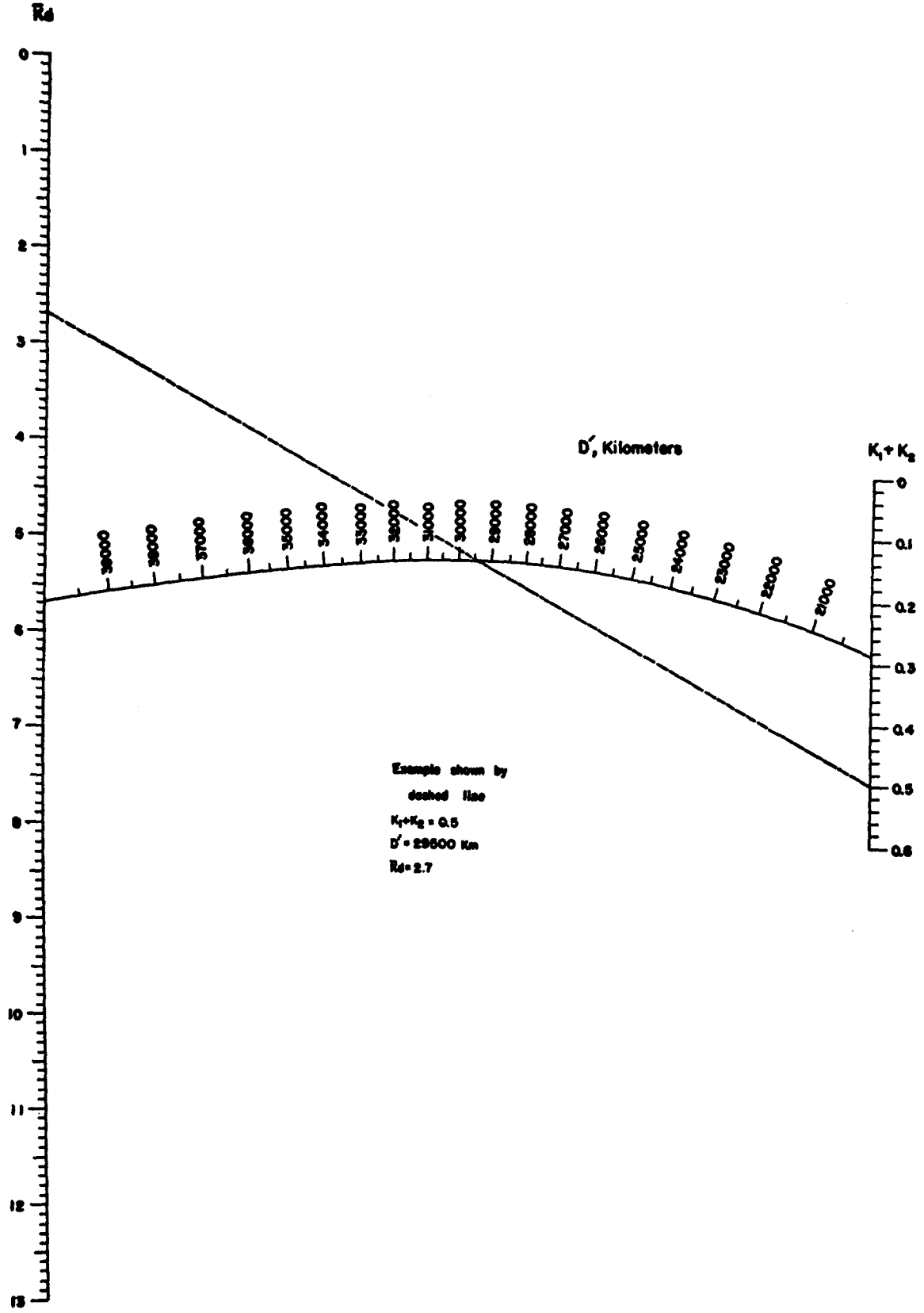


FIGURE 7.47. *Nomogram giving variation of Kd with K_1+K_2 and D' (length of transmission path lying in the region where $K>0$.).*

Distance in kilometers, 20,000 to 40,000.

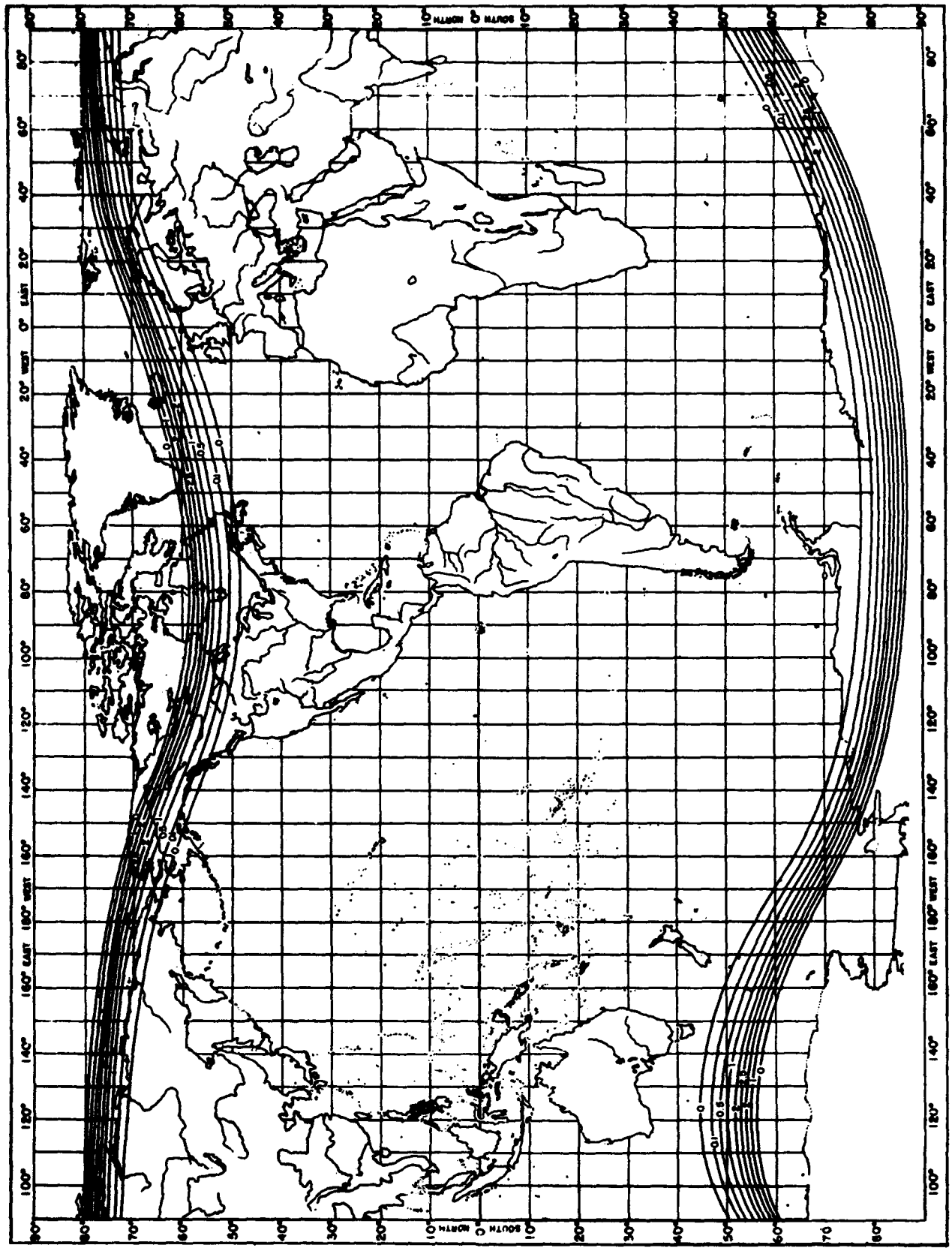


Fig. 7.48. AURORAL ZONE ABSORPTION MAP. NUMBERS ON CURVES ARE ABSORPTION FACTOR, K'

CHAPTER 8

RADIO NOISE AND REQUIRED FIELD INTENSITY

8.1. Introduction

As radio waves decrease in intensity on spreading out from the transmitter, and as they lose energy by absorption in either the ground or the ionosphere, they finally become so weak that the signals carried by these waves are no longer capable of conveying any intelligence to an observer when they are received in the presence of radio noise or other signals occupying the same part of the radio spectrum. In this chapter, the only undesired fields to be considered are those due to radio noise of various types; a discussion of the fields required in the presence of other radio transmissions is outside the scope of this handbook. The minimum radio field intensity needed to allow an intelligible signal of a particular type to be satisfactorily received in the presence of the radio noise at the receiving station is called the required field intensity for that type of service, and the distance from the transmitter to any point beyond which the radio field intensity is less than the required field intensity is the distance range.

The required field intensity is subject to wide variation. It depends upon the receiving antenna; the band width and noise figure of the receiving set; the local radio noise or "static"; the type of modulation of the radio wave, i. e., type of service; and the grade of service desired, e. g., barely intelligible, entirely satisfactory, etc. It also varies, with the radio noise, according to the time of day and season.

8.2. Types of Radio Noise

In general, radio noise may be defined as interference whose energy is not confined to a narrow band of frequencies. Two general types of radio noise may be distinguished: (a) impulse noise which is interference due to a single elementary disturbance, or to an aggregate of elementary disturbances with systematic relative phases and (b) random, or fluctuation, noise, which is the aggregate of a large number of elementary disturbances with random relative phases. In general, the peak value of the receiver response to impulse noise is proportional to the effective receiver band width (see chapter 3), the effective (rms) value is proportional to the square root of the effective band width, and the average value is independent of the effective band width. On the other hand, both the effective and average values of random noise are proportional to the square root of the effective band width, while the peak value has significance only to the extent that individual peaks of the random noise are isolated and considered as impulse noise.

From the above we see that a distinction between impulse and random noise is not always easy to make. However, electrical, or man-made, noise caused by the operation of electrical apparatus is usually of the impulse type, whereas atmospheric noise, originating in thunderstorms or due to other atmospheric conditions, may ordinarily be considered to have the bandwidth characteristics of random noise. The best example of random noise is the fluctuation noise originating in the resistance components of impedance elements in the receiver or due to the fluctuations of electrons in vacuum tubes. Another example of random noise is the cosmic noise from interstellar space as received on frequencies which are sufficiently high to penetrate the earth's atmosphere.

8.3. Atmospheric and Precipitation Radio Noise

At the frequencies under consideration in this book, atmospheric and precipitation noise are the most important types to be considered. Radio noise from electrical apparatus, such as the ignition systems of automobiles, may be very serious, but such noise is more or less, under the control of the observer and can be largely eliminated if necessary. Atmospheric and precipitation noise, on the other hand, since it originates in thunderstorms, or in rain, snow, or dust storms, cannot usually be eliminated and thus sets the limit to radio reception. Precipitation noise arising from an interaction between the receiving antenna and a rain, snow, or dust storm is not included in the graphs of required field intensity to be given later, although it may at times, depending on local considerations at particular receiving sites, be the principal source of radio noise.

Practically all atmospheric radio noise is considered to originate in the lightning flashes associated with thunderstorms. As received at a distant receiving point, this noise may be expected to display the same propagation characteristics, such as intensity variations, as do ordinary radio transmissions propagated from a distant source. In addition to these expected propagation variations, there are the variations to be expected as a result of different concentrations in thunderstorm activity in various parts of the world and the

variation, with frequency, in the noise power radiated from individual lightning flashes; the latter is known to be approximately inversely proportional to the square of the radio frequency for frequencies above about 10 to 20 kc. Thus, at a given location, the atmospheric radio noise level is made up of noise from nearby centers of noise generation such as local thunderstorms whose distances from the receiving set may vary from a few miles to hundreds of miles, plus noise which has been propagated to the receiving location from one or more of the principal centers of noise generation, such as the active thunderstorm areas in equatorial Africa, Central America, and the East Indies. The location and activity of the various noise centers vary with time of day and season, but probably not much with the sunspot cycle. Generally, thunderstorms occur much more frequently over the land than over the sea and are much more common at low than at high latitudes.

Figures 8.1 to 8.4, inclusive, are maps of the world divided into atmospheric radio noise zones. Figures 8.5 to 8.10, inclusive, are required field-intensity curves corresponding to the noise zones. The required field-intensity curves and noise maps were prepared by D. K. Bailey, and the noise curves were taken directly from a United States Signal Corps report (first reference, section 8.7).

It will be noted that the four noise maps correspond to four different periods of the year. Areas of the world in which thunderstorms are most frequent are indicated as zones 4 and 5. The areas most remote from the principal thunderstorm areas, and in which but little atmospheric radio noise may be expected, even by way of long-distance sky-wave propagation, are indicated as zone 1. The other zones are intermediate in radio noise expectation.

The required-field-intensity curves of figures 8.5 to 8.10, inclusive, are plotted as required incident field intensity for radio-telephone communication versus frequency, for six different times of day, for each particular noise grade. It will be noted that the night curves have an almost constant slope, the required intensity decreasing with increasing frequency. This represents roughly the frequency characteristic of the noise produced by the thunderstorms, as there is very little sky-wave absorption during night hours. Above the maximum usable frequency for long-distance transmission the noise cannot be propagated by sky-wave transmission, and hence decreases very rapidly with increasing frequency. The variation in the magnitude of the required-field-intensity curves for various hours during darkness is due to the average variation of thunderstorm activity throughout the night, with the maximum usable frequency effect mentioned in the preceding sentence also a factor at the higher frequencies. During daylight hours the absorption of the sky-wave propagated noise increases, and is greatest in the frequency range of 1.5 to 2.5 Mc. This is evident in the shape and magnitude of the noise curves for daylight hours. It is also pointed out that sky-wave propagated noise extends to higher frequencies in the daytime, because of the higher maximum usable frequencies occurring during those hours.

The required-incident-field-intensity curves are plotted for 90 percent intelligibility of 100-percent modulated radiotelephone service, assuming a nondirectional receiving antenna that responds equally to the noise and to the desired signal; for reception of other types of service the required field intensities need to be modified by the factors shown in table 8.1, taken from another United States Signal Corps report (second reference, section 8.7). It will be noted that the curves for the lower noise grades are plotted for summer (May, June, and July) and winter (November, December, and January), months. This variation with season is due to the variation in the length of daylight and darkness hours for north or south latitudes greater than 30°, and the consequent effect on the propagation of the noise. When the receiving location is at a latitude of less than 30° the required field-intensity curves for the equinox months apply for all months of the year. These noise curves represent a signal-to-average noise ratio of 5.5 (15 db), with a receiving-set band-width of 3 kc each side of the carrier frequency.

When the receiving antenna discriminates either for or against the desired signal with respect to the atmospheric noise, the required incident field intensity for a particular type of service needs to be modified accordingly. For instance, assume that most of the noise is arriving from a noise center to the south of a particular receiving location, and that a signal is being received from the east on a directive antenna. Assume further that the antenna gain in the direction of the arrival of the desired signal is three times the gain in the direction of the arrival of the noise. Then the required field intensity to overcome the atmospheric noise is one-third of the value indicated by the curves. It is to be remembered that the vertical angles of arrival of the signal and noise need to be considered, as well as the azimuth angles, in determining the relative gain of an antenna for the signal and for the noise. In this connection it should be pointed out that only meager data are available on vertical angles of arrival of atmospheric noise; an average angle of 20° above the horizontal has frequently been assumed for the intermediate- and low-noise grades.

Also it should be noted that the data on the basis of which these curves were drawn are frag-

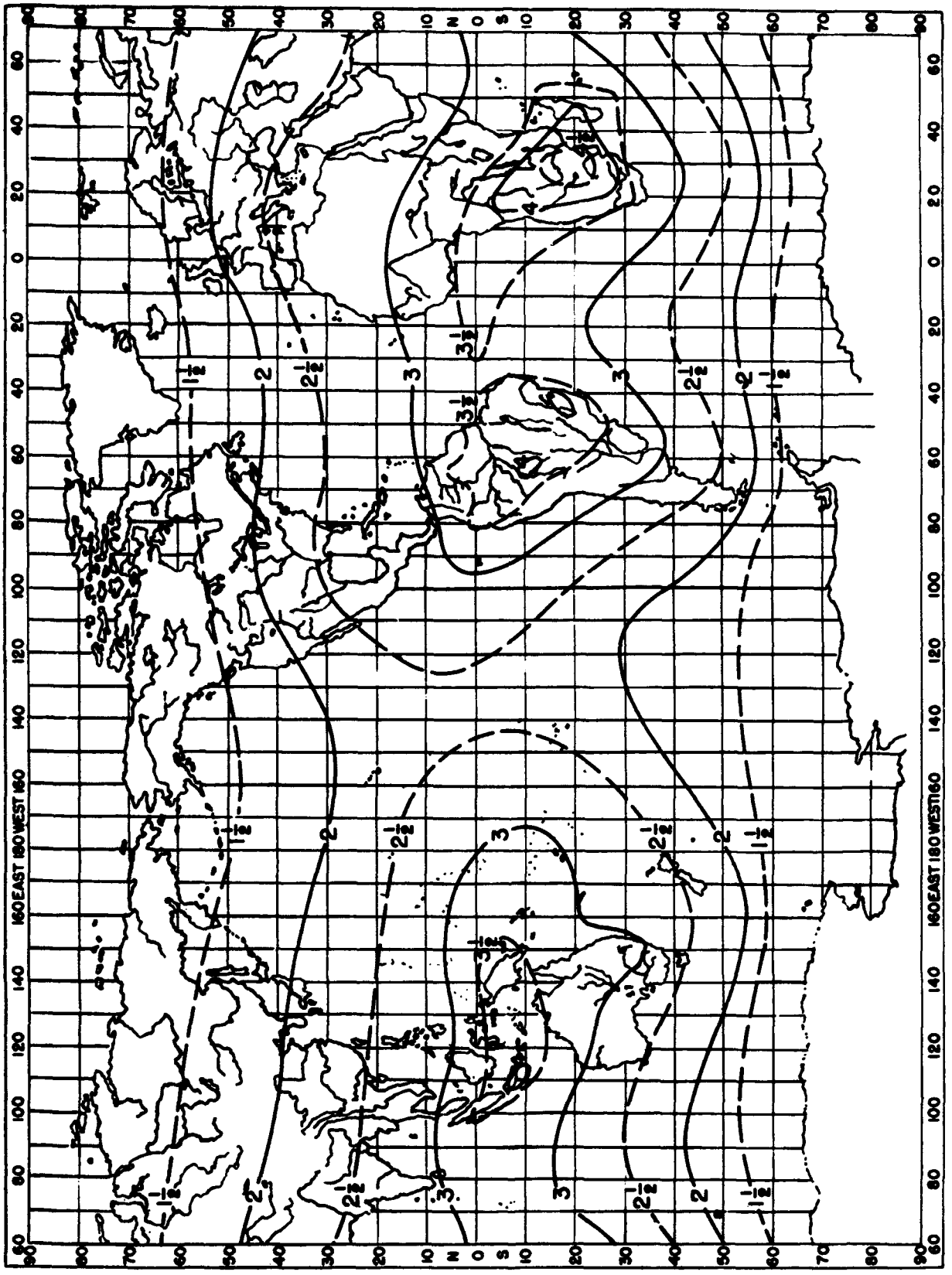


FIGURE 8.1. *Noise distribution for period December-January-February.*

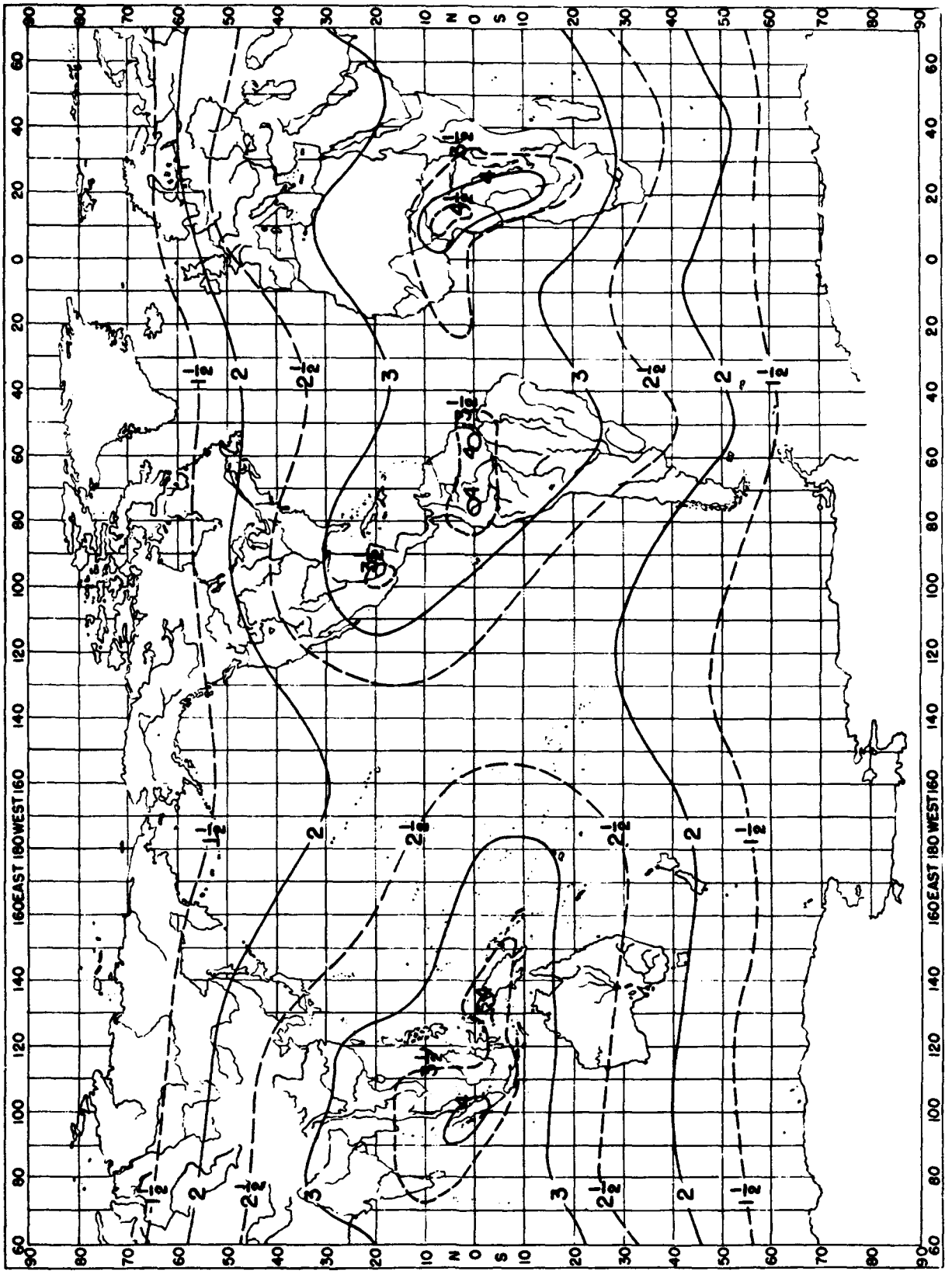


FIGURE 8.2. *Noise distribution for period March-April-May.*

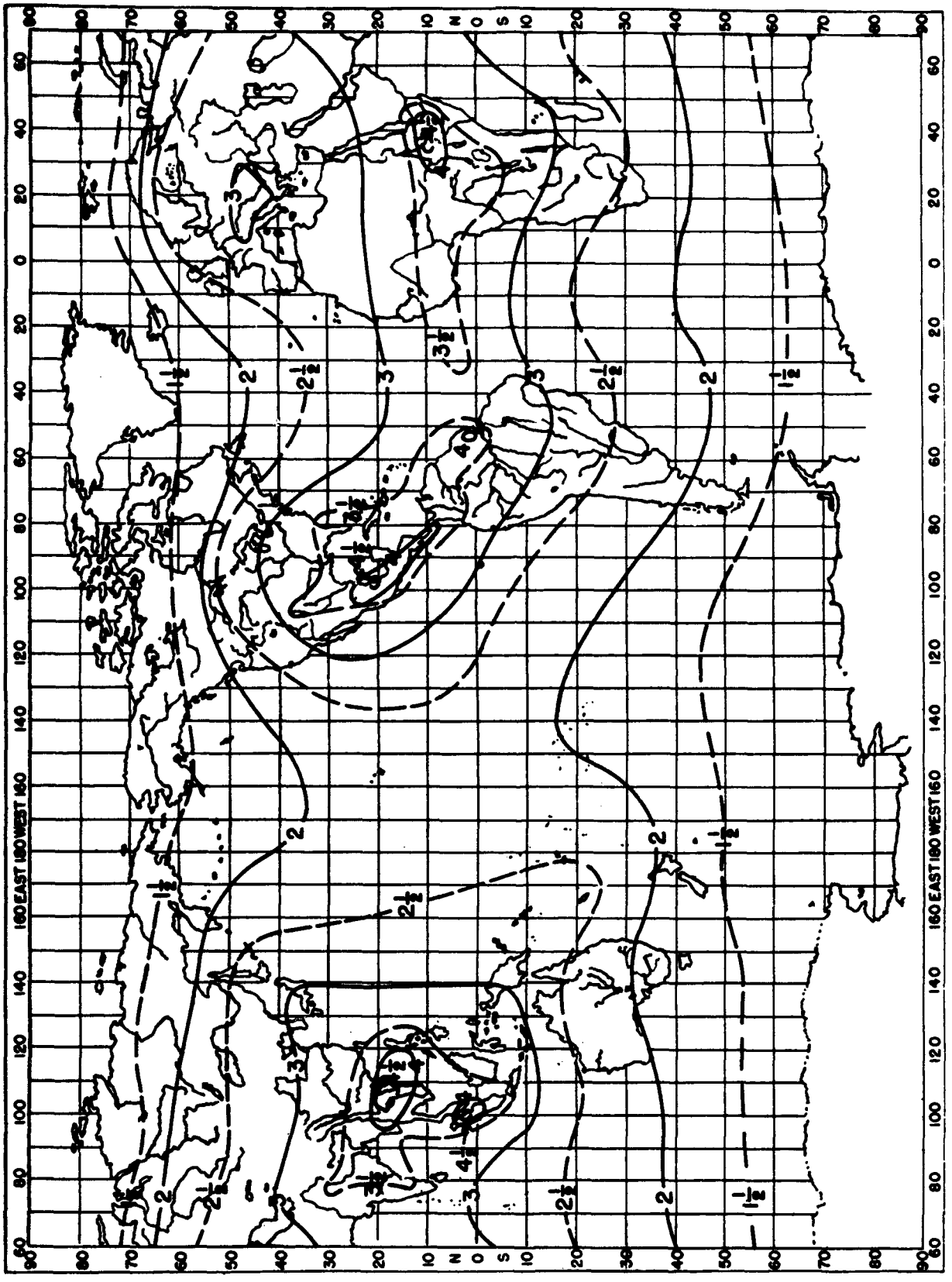


FIGURE 8.3. *Noise distribution for period June-July-August.*

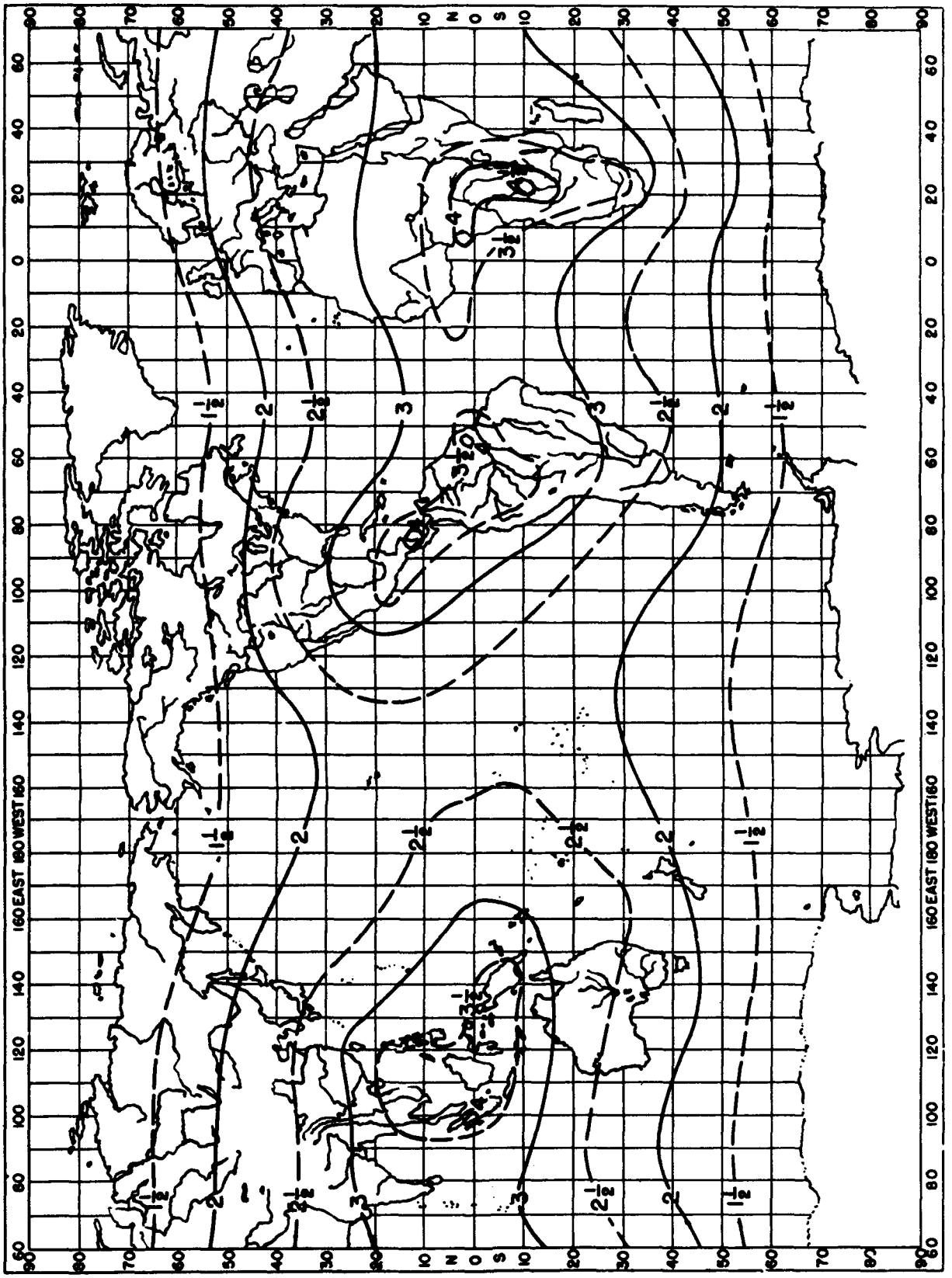
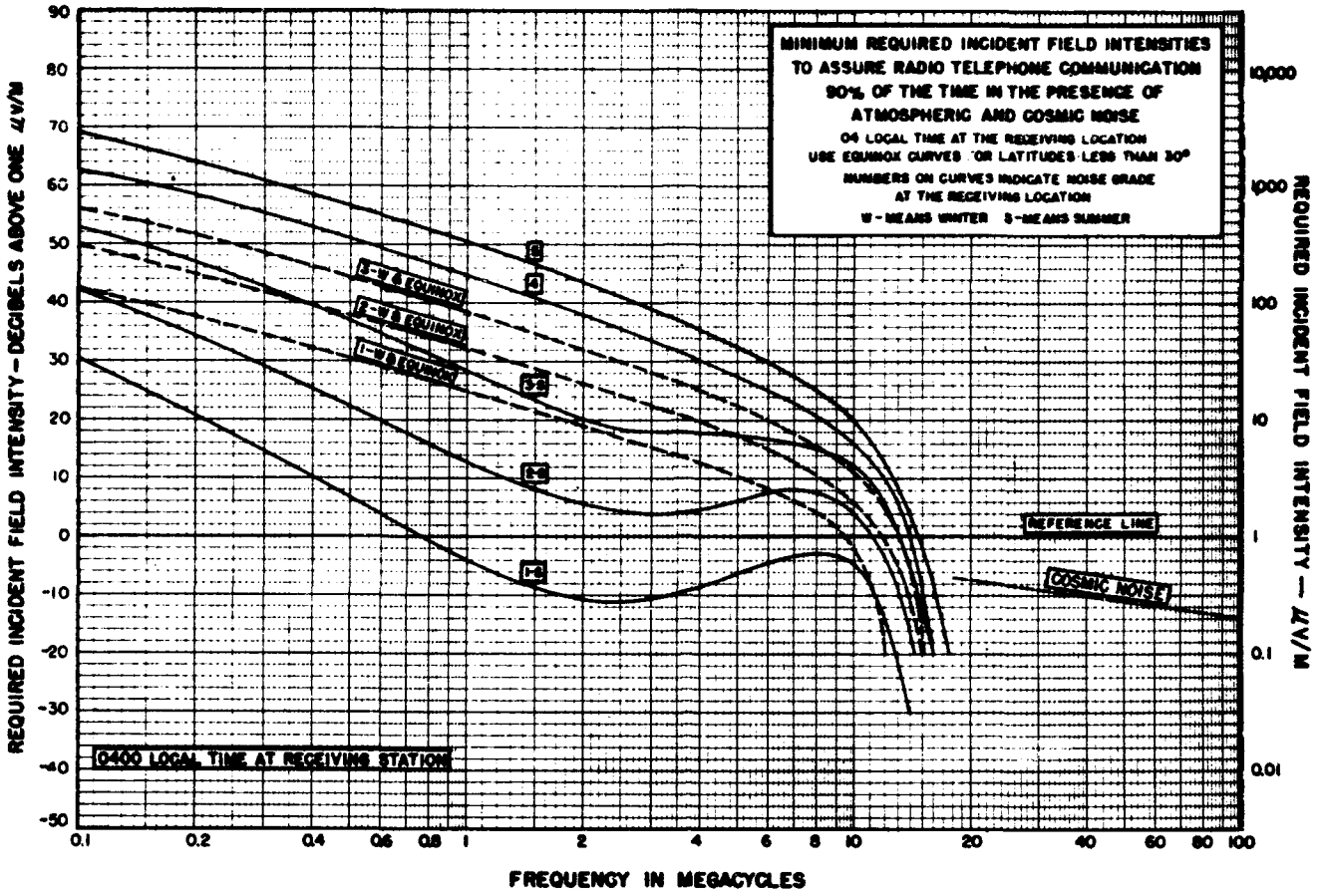
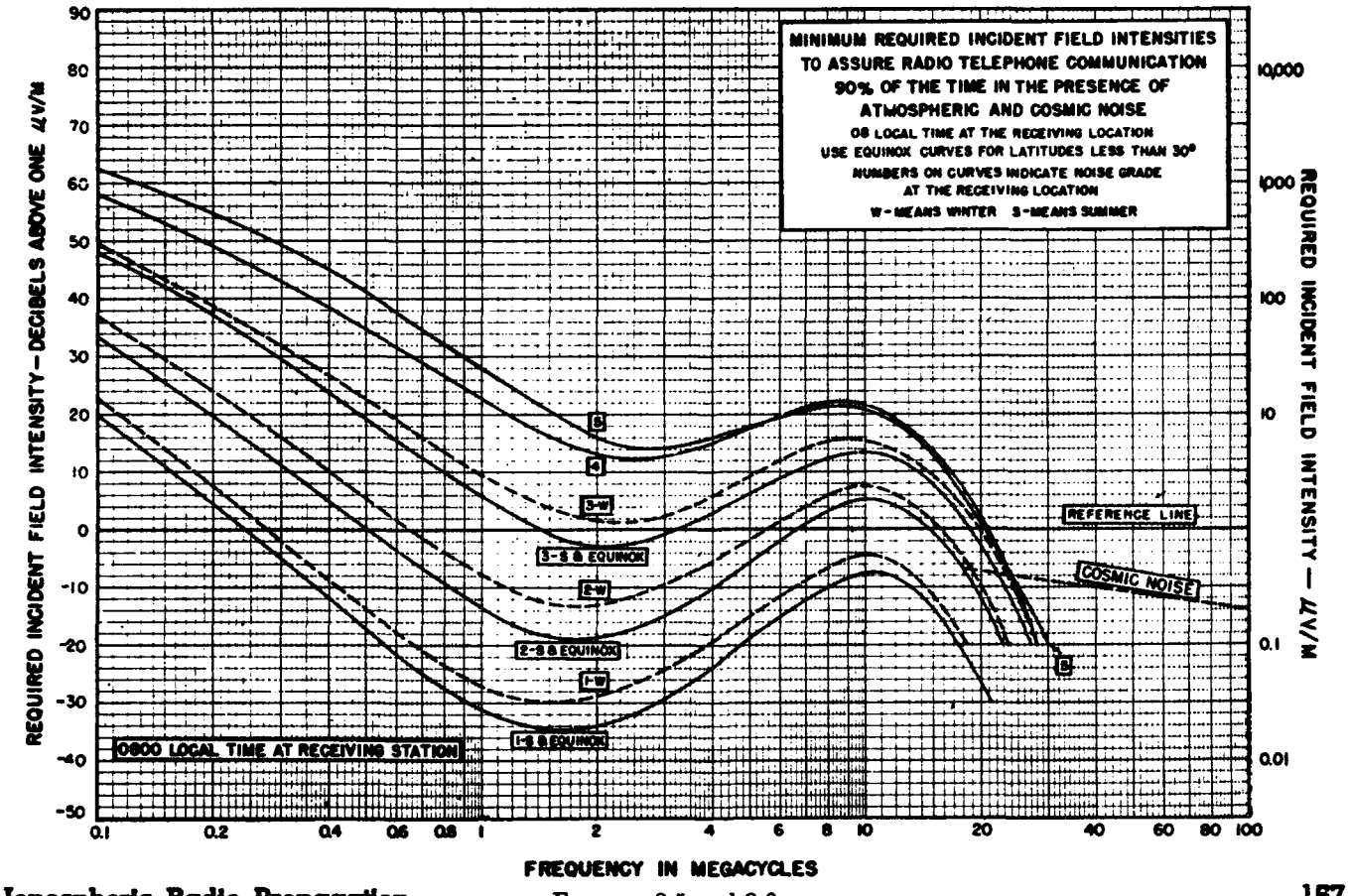


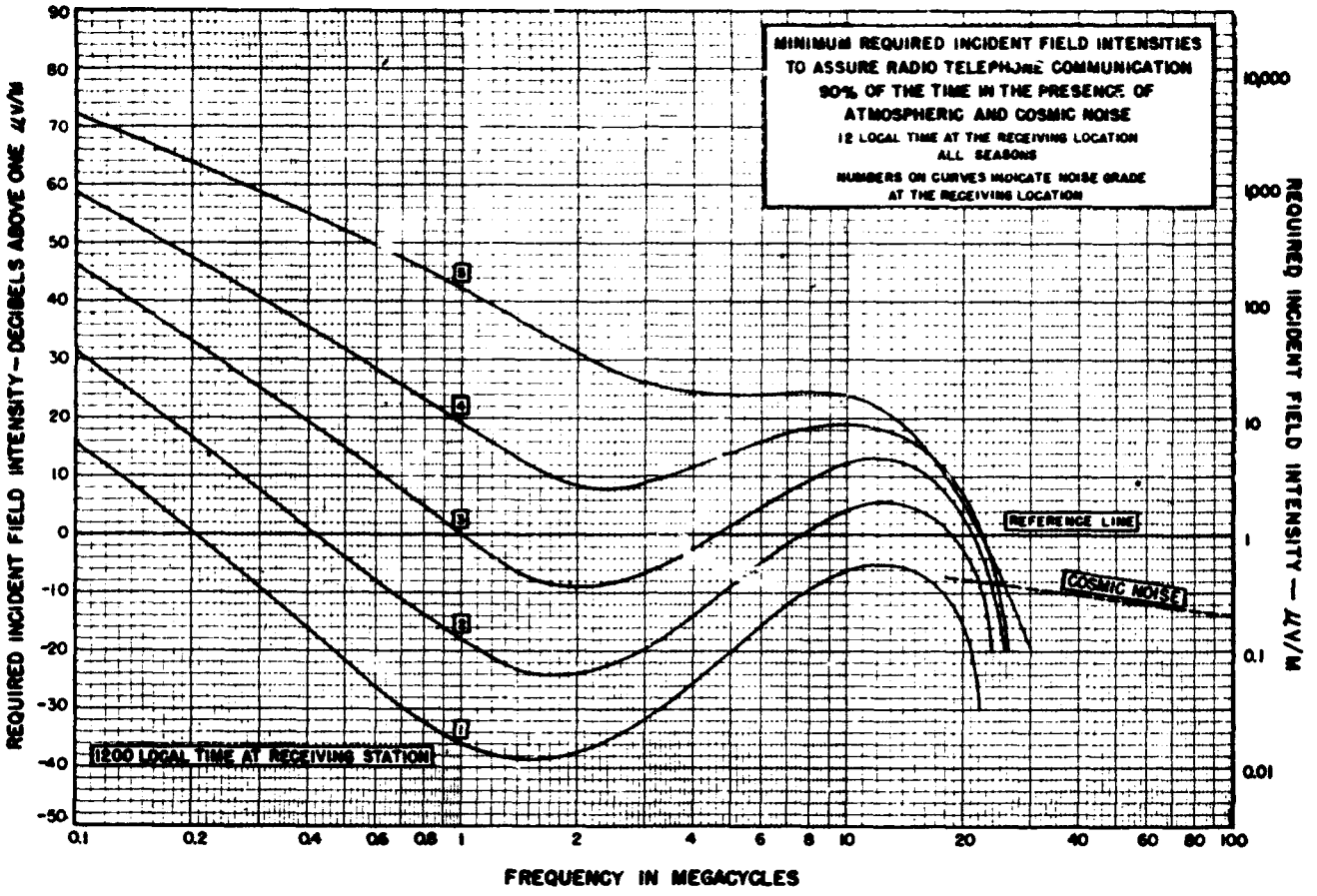
FIGURE 8.4. *Noise distribution for period September-October-November.*

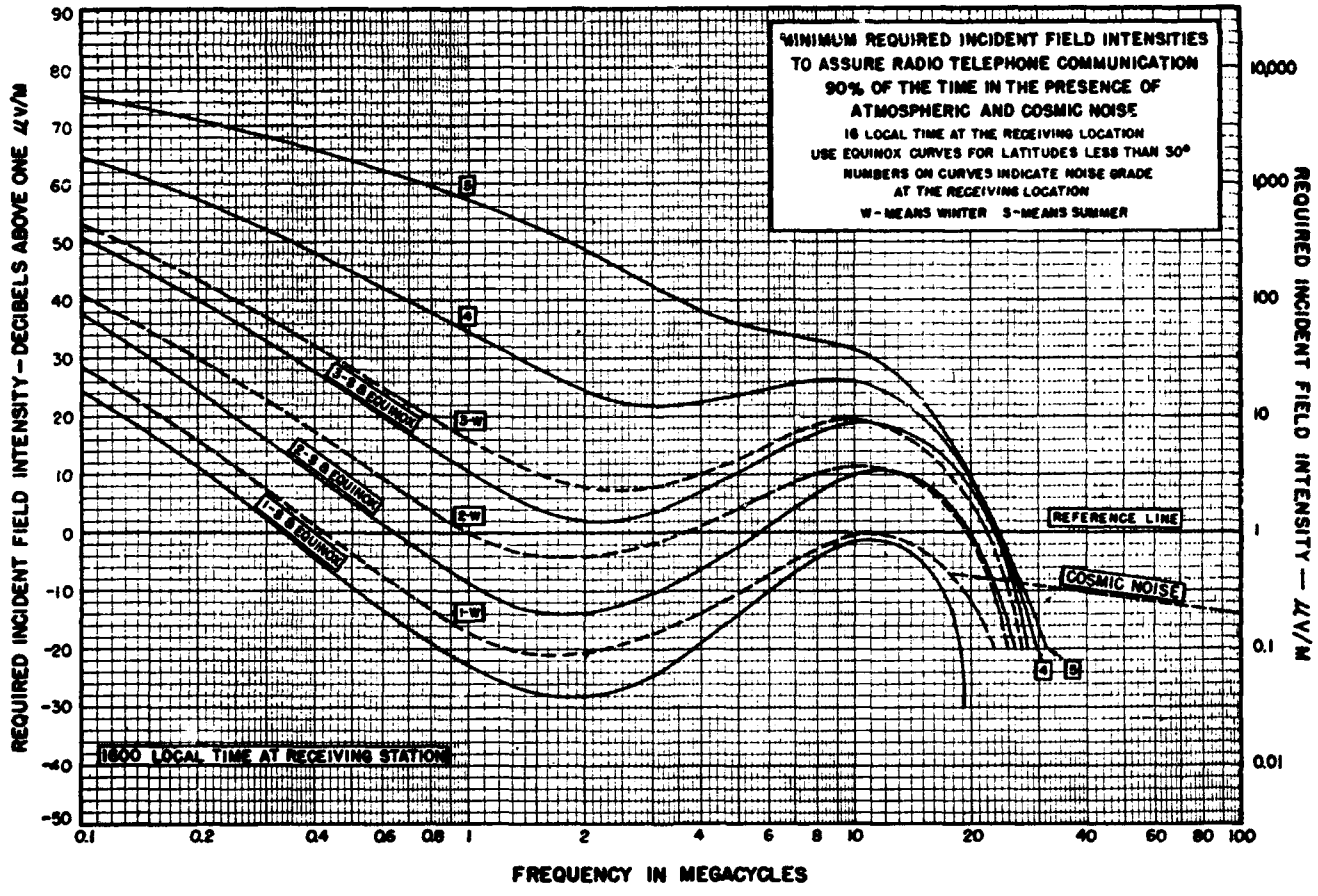




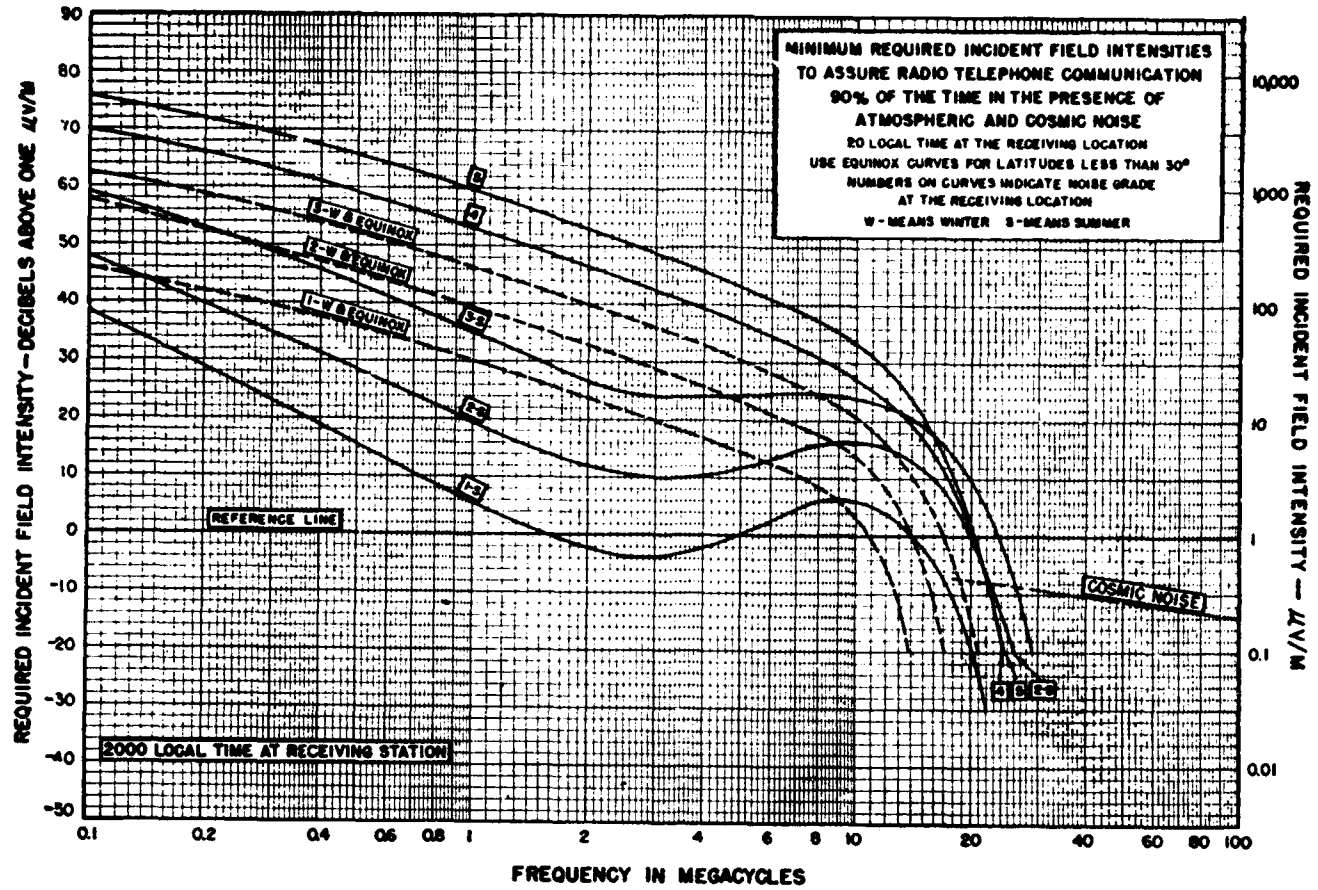
FIGURES 8.5 and 8.6.

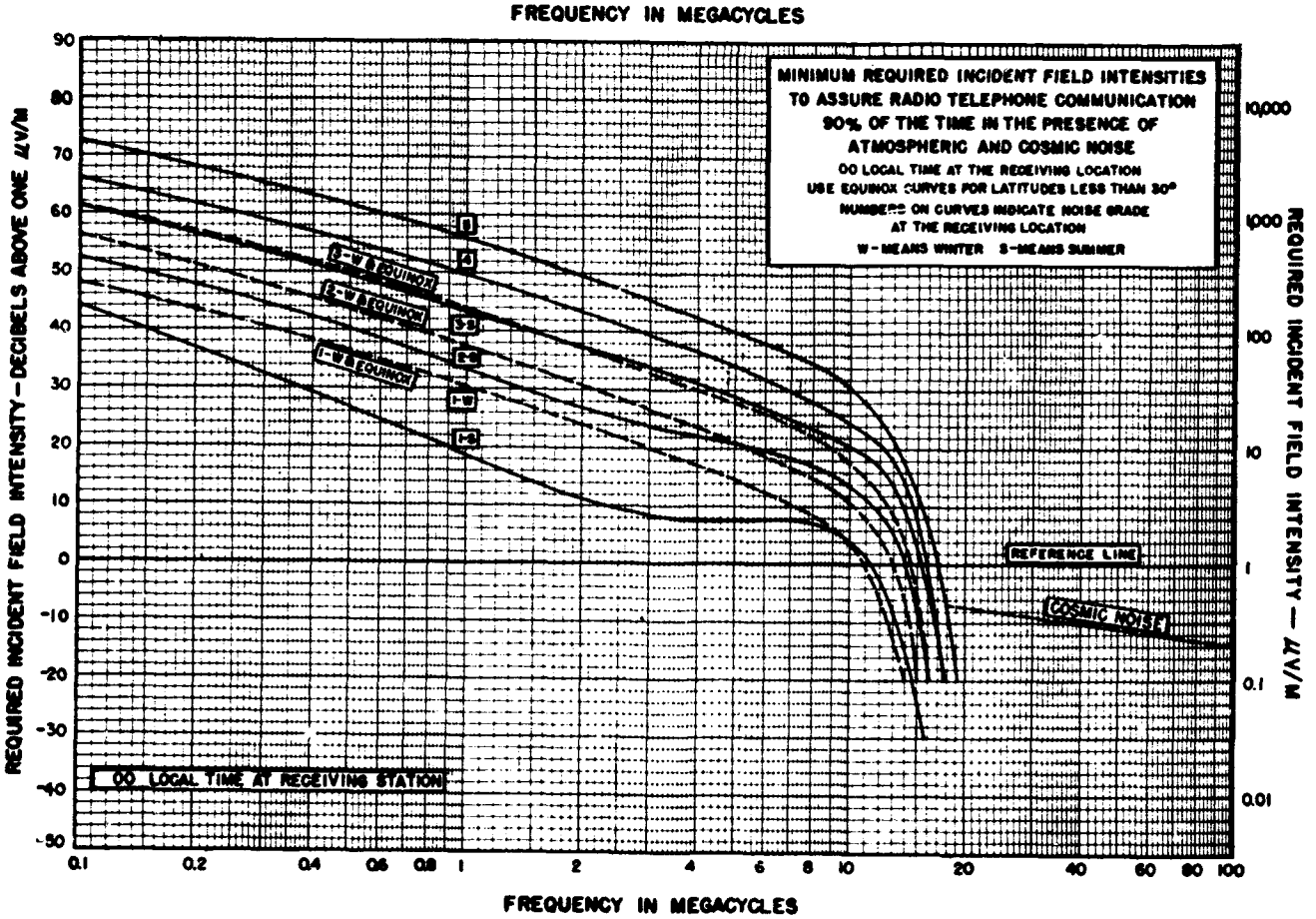
935333 O - 51 - 11





FIGURES 8.7 and 8.8.





FIGURES 8.9 and 8.10.

TABLE 8.1. *Factors needed to modify required field intensities for various types of services*

Type of radio communication service	Conditions of operation	Ratio of the field intensity required for intelligible reception to the field intensity required for radiotelephony		Type of service factor, T [1]
		Intensity ratio	Ratio, in decibels	
Double sideband radiotelephony	Receiver band width 3 kc each side of carrier; 100-percent modulation; intelligibility 90 percent.	1 (reference level	0	1
Single sideband radiotelephony	Carrier suppressed 10 to 25 db	0.50	−6	4.0
Standard broadcast	Receiver band width 6 kc each side of carrier; 38 db signal-to-noise ratio considered tolerable.	20	26	0.0025
International short-wave broadcast	Assume considerably more disturbance tolerable than on standard broadcast.	5.6	15	.032
Manual CW telegraphy	Receiver band width 3 kc each side of carrier; 90-percent intelligibility at 10 words (cipher groups) per minute.	0.14	−17	50
High-speed automatic CW telegraphy	80 to 120 words per minute: Diversity reception	.18	−15	32
	80 to 120 words per minute: Nondiversity reception	.28	−11	12.6
Interrupted carrier radioteletypewriter	60-words-per-minute radioteletypewriter; carrier on for marking, carrier off for spacing. Diversity reception	.25	−12	16
	60-words-per-minute radioteletypewriter; carrier on for marking, carrier off for spacing. Nondiversity reception	.40	−8	6.3
Carrier shift system—radioteletypewriter	60-words-per-minute radioteletypewriter; shift 425 cycles each side of nominal carrier frequency; diversity reception.	.10	−20	100
Carrier shift system—"Diplex" radioteletypewriter.	Two 60-words-per-minute radioteletypewriter channels operating simultaneously; diversity reception.	.18	−15	32
Modulated CW manual telegraphy	100-percent modulation (required field intensity, is inversely proportional to percentage modulation); signal received without use of BFO (use of BFO provides reception equal to or better than pure CW).	.18	−15	32
Double sideband 1-tone voice frequency radioteletypewriter system.	1-channel 60-words-per-minute radioteletypewriter; one tone on marking, tone off on spacing.	.50	−6	4
Double sideband two-tone voice frequency radioteletypewriter system.	1-channel 60-words-per-minute radioteletypewriter; one tone on marking, another tone on spacing.	.25	−12	16
Double sideband four-tone voice frequency radioteletypewriter system.	1-channel 60-words-per-minute radioteletypewriter; two tones on marking, another two tones on spacing.	.14	−17	50
Single sideband two-tone voice frequency radioteletypewriter system.	Each channel 60-words-per-minute radioteletypewriter; one tone on marking, another tone on spacing. 1-channel operation	.13	−18	63
	Each channel 60-words-per-minute radioteletypewriter; one tone on marking, another tone on spacing. 3-channel operation	.22	−13	20
	Each channel 60-words-per-minute radioteletypewriter; one tone on marking, another tone on spacing. 6-channel operation	.32	−10	10
Single sideband four-tone voice frequency radioteletypewriter system.	Each channel 60-words-per-minute radioteletypewriter; two tones on marking, another two tones on spacing. 1-channel operation	.071	−23	200
	Each channel 60-words-per-minute radioteletypewriter; two tones on marking, another two tones on spacing. 2-channel operation	.10	−20	100
	Each channel 60-words-per-minute radioteletypewriter; two tones on marking, another two tones on spacing. 3-channel operation	.13	−18	63
Single sideband telephoto system		.50	−6	4.0
Double sideband telephoto system	Diversity reception	.71	−3	2.0
	Nondiversity reception	1	0	1
Carrier shift telephoto system	Diversity reception	0.36	−9	7.9
	Nondiversity reception	.50	−6	4.0

[1] This column is used in chapter 9.

mentary and far from complete; a considerable portion of the information is based on estimates.

8.4. Cosmic and Solar Radio Noise

At frequencies between about 10 and 100 Mc, cosmic radio noise of extraterrestrial origin is known to be the principal source of interference to reception under many circumstances. As mentioned above, this noise has much the same characteristics as the fluctuation noise originating in the components of a receiving set. The sources of this cosmic noise are not uniformly distributed over the sky but tend to be concentrated in several regions on the celestial sphere, the principal source being in the region Scorpio-Sagittarius near the center of the galaxy. Consequently, when received on a directional antenna, the noise varies in a characteristic manner from hour to hour and from day to day.

The reason for the existence of this noise is not yet understood. Some investigators believe it is radio-frequency radiation from eruptions similar to spot eruptions on our sun occurring on all the stars in the galaxy, while others have considered it to originate in free-free electron transitions occurring in all of the space between the stars. Many observers throughout the world are now in the process of further research on the variations in the intensity of this noise with time of day, season, geographical location, radio frequency, and receiving antenna directivity. An analysis of cosmic noise data now available appears to indicate that, as received on an antenna with no more directivity than a half-wave dipole, the required incident field intensity for intelligible reception of radio telephony in the presence of cosmic noise may be expressed approximately:

$$E = 2/\sqrt{f_{Mc}} \text{ microvolts per meter}$$
$$(18 \text{ Mc} < f_{Mc} < 160 \text{ Mc})$$

f_{Mc} = frequency in megacycles per second.

This relation is shown as a dashed line between 18 and 100 Mc on figures 8.5 to 8.10, and it is clear from its relation to the corresponding atmospheric-noise curves that cosmic noise will often be the limiting factor in radio reception on frequencies between 18 and 100 Mc.

At the time cosmic radio noise was first discovered an attempt was made to determine if similar phenomena were observable in connection with the

sun. Negative results were obtained at that time. Relatively recently, however, measurements of radio noise at 200 Mc, made with a rather sensitive receiving system and a directive antenna with high gain, showed that the sun also acts as a radiator of radio noise. The theory of thermal radiation predicts that any black body will radiate a certain amount of energy at all wavelengths. If the Rayleigh-Jeans formula is applied to the sun, assuming that the sun is a black body at 6,000°K, the intensity of radiation at 200 Mc would be very much below the threshold of present-day receiving systems. The intensities which have actually been measured at 200 Mc are of magnitudes which correspond to a solar temperature of approximately a million degrees Kelvin. Furthermore, at times of ionosphere storms, associated with large sunspots, intensities of the radiation have been observed corresponding to temperatures of a thousand million degrees Kelvin. By the use of polarized Yagi directional antennas it has been observed that these bursts of radio noise are circularly polarized, either right-handed or left-handed depending on whether the sunspots were in the northern or southern hemisphere. This polarization was observed only for bursts of high intensity; normal radio noise coming from a quiescent sun appears to have random polarization. Measurements have also been made of solar radiation at frequencies of 3,000 to 24,000 Mc, and these also gave distinct evidence of radio noise coming from the sun. The intensity of this noise, however, corresponded rather closely to that which one would expect from thermal radiation, with effective black-body temperatures of the order of 20,000°K as compared with visual temperatures of 6,000°K and as against effective temperatures of 10^6 to 10^9 °K, measured at 100 to 200 Mc. Some plausible and complete theories have been formed regarding the nature of solar noise and the frequency distribution of its intensity, but only a few experimental measurements have been performed towards confirming it. Except at the time of large sunspot eruptions, solar noise is important only on very high frequencies when highly directional antennas are actually pointed at the sun; consequently, it usually need not be considered in relation to practical problems of propagation.

8.5. Receiving Set Noise

As previously discussed in chapter 3, noise generated internally in a receiver will exist because of the random motion of electrons in resistance components of impedance elements and the fluctuations of the electrons in vacuum tubes. In the absence of all external noise, signals must be strong enough to override this internal noise to be intelligible. The ability of a particular receiver to receive signals with only internal noise present may be expressed in terms of the receiver's "noise figure" as determined by the methods given in chapter 3.

Experimental determination of the receiver input terminal voltage required to override the internal receiver noise in typical Army communication receivers has resulted in a value of approximately 1 μv for 90 percent intelligibility of 100 percent modulated radiotelephony. This value is somewhat dependent on frequency, but is considered sufficiently accurate for all frequencies in the band [illegible] to 20 Mc.

Figu[illegible]3, inclusive, are a set of curves showing [illegible] incident field intensity versus frequency for various practical receiving antennas to deliver 1 μv across a 100-ohm load resistance, as given in the Signal Corps report mentioned above. The average receiver input impedance has been assumed to be 100 ohms, and it is assumed that the antenna reactance, if any, will be tuned out. In the calculations for the rhombic antennas it has been assumed that a 600- to 100-ohm radio-frequency transformer will be used between the antenna and receiver, and that a voltage of 1 μv is required across the 100-ohm side. Transmission-line and matching-network losses have been neglected in these curves. If these losses are known, the loss in decibels should be added to the required field intensity as read from the curves. Thus these curves show the required incident field intensity for 90 percent intelligibility of radiotelephony with the various receiving antennas, when set noise is the limiting factor.

In computing these receiving antenna curves, advantage has been taken of the reciprocity theorem for transmitting and receiving antennas, in that the directivity factors for a given antenna have been assumed to be the same in both cases. The formula used for the calculations is

$$E_{uv/m}=\frac{24.6\times f_{Mc}}{\sqrt{R_r}\times F}\times\frac{R_r+R_1+R_z}{R_z}$$

to produce 1 μv across R_z,

where

R_r = antenna radiation resistance
R_1 = antenna loss resistance
R_z = antenna load resistance
F = theoretical inverse distance field intensity in millivolts per meter produced by antenna at 1 mile in desired direction when used as a transmitting antenna, with 1.0 kw delivered to antenna, assuming zero antenna loss resistance.
f_{Mc} = frequency in megacycles per second.
$E_{uv/m}$ = required incident field intensity in microvolts per meter.

It is to be noted that curves have been drawn

for antennas located over both "good" and "poor" ground, and for various vertical angles of wave arrival.

8.6. Reliability of the Data and Methods

The noise-grade maps and required field-intensity charts described in section 8.3 above were prepared by using the actual measurements of required field intensity then available, together with the theories of atmospheric noise generation and propagation described in the preceding sections. Since then the additional measurements for CW signals described in section 3.8 have become available. These data are not considered a sufficient basis yet for a fundamental revision of the prediction data and method, but they do permit an estimate of the dav-to-day variability of noise conditions and the probable reliability of the predictions at these particular locations.

Analysis of the CW measurements indicates that during a given month at a given hour, 80 percent of the daily values of required field intensity fall within a range of approximately ± 6 db of their median values. This gives some idea of the variability of required field intensities.

A comparison has also been made between predictions as given in this text and the required field intensity measured at 17 stations ranging in latitude between 51° N and 43° S. The data at different stations cover periods ranging from a few months to 2 years, all of the data corresponding to years near sunspot minimum. Only measurements on 2.5 and 5 Mc for the night hours 1900 through 0500 local time were used for the comparison because of set-noise limitations in the measurements at other frequencies and at other hours of the day.

As the predictions refer to the minimum value of field intensity that insures intelligible reception on 90 percent of the days, comparison was made with an estimate of the 90 percent value obtained by taking the average of the third highest of the daily values observed during each of the 3 months included in the season, at each of the three consecutive hours centered on the hour, for which the prediction was made. For example, a prediction for the hour 2000 local time and the season, including March, April, and May, was compared with the average of the third highest values observed at 1900, 2000, and 2100 in each of these months. The third highest of 30 daily values during a month is the value at or below which 90 percent of the daily values fall.

The percentage distributions of deviations of predicted from observed values for 2.5 and 5 Mc are shown in figure 8.44, a). The ratio of the observed to the predicted values are divided into 10-db class intervals in this figure. The deviations were counted without regard to sign; they are, however, distributed nearly symmetrically about zero. Approximately 20 percent and 10 percent of the deviations are greater than ± 10 db for 2.5 and 5 Mc, respectively. Although not shown on the chart, only 45 percent and 63 percent, respectively, were less than ± 5 db. These deviations from the predictions are considered to be relatively small in view of the fact that the predictions for these 17 stations and 2 frequencies were extrapolations of data measured at other stations and on other frequencies.

In addition to the above indicated small inaccuracies of the predicted required field intensities themselves, other uncertainties may arise in connection with their application to the solution of communication problems.

8.7. References

Minimum required field intensities for intelligible reception of radiotelephony in presence of atmospherics or receiving set noise. RPU–140, Radio Propagation Unit Technical Report No. 5, revised, July 1947. Prepared under the direction of the Chief Signal Officer by the Radio Propagation Unit (9463d–TSU), Holabird Signal Depot, Baltimore, Md.

Relative sky-wave signal strengths required for intelligible reception of various types of radio communication service. RPU–106, Radio Propagation Unit Technical Report No. 4, August 1945. Prepared under the direction of the Chief Signal Officer by the Radio Propagation Unit (9463d TSU), Holabird Signal Depot, Baltimore, Md.

K. A. Norton, Frequency distribution of the intensities of radio atmospherics, Paper presented at the URSI meeting held in Washington, D. C., on April 27, 1934, and later that same year at the URSI meeting held in London.

R. K. Potter, An estimate of the frequency distribution of atmospheric noise. Proc. Inst. Radio Engrs. **20**: 1512 (1932).

R. K. Potter, High-frequency atmospheric noise, Proc. Proc. Inst. Radio Engrs. **19**: 1731 (1931).

H. A. Thomas and R. E. Burgess, Survey of existing information and data on atmospheric noise level over the frequency range 1 to 30 Mc/s. Paper No. RRB/C90, Department of Scientific and Industrial Research, Radio Research Board Report (Feb. 21, 1944).

K. G. Jansky, Minimum noise levels obtained on short-wave radio receiving systems, Proc. Inst. Radio Engrs. **25**: 1517 (1937).

Grote Reber, Cosmic static, Proc. Inst. Radio Engrs. **30**: 367 (1942).

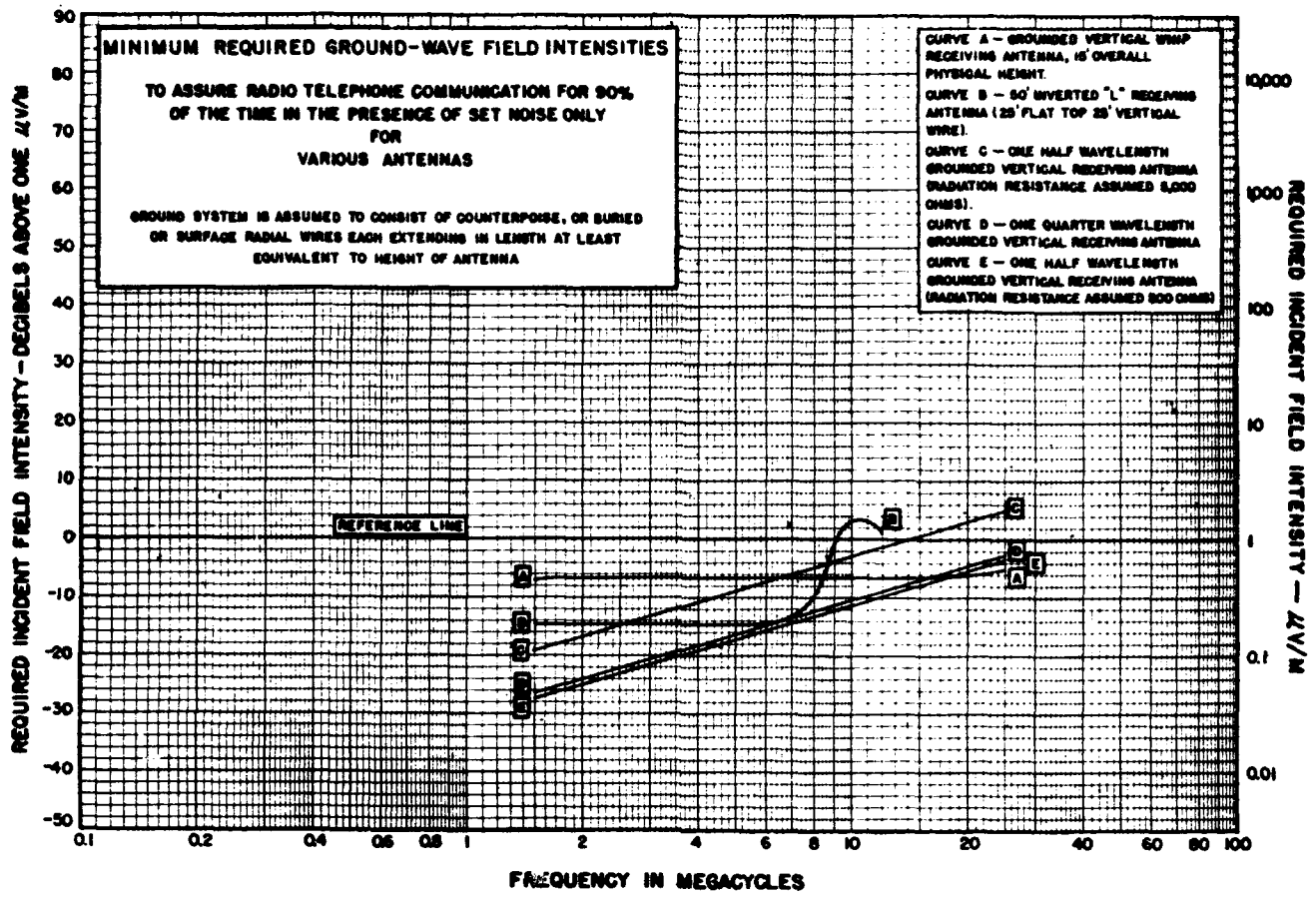
G. C. Southworth, Microwave radiation from the sun, J. Franklin Inst. **239**: 285 (1945).

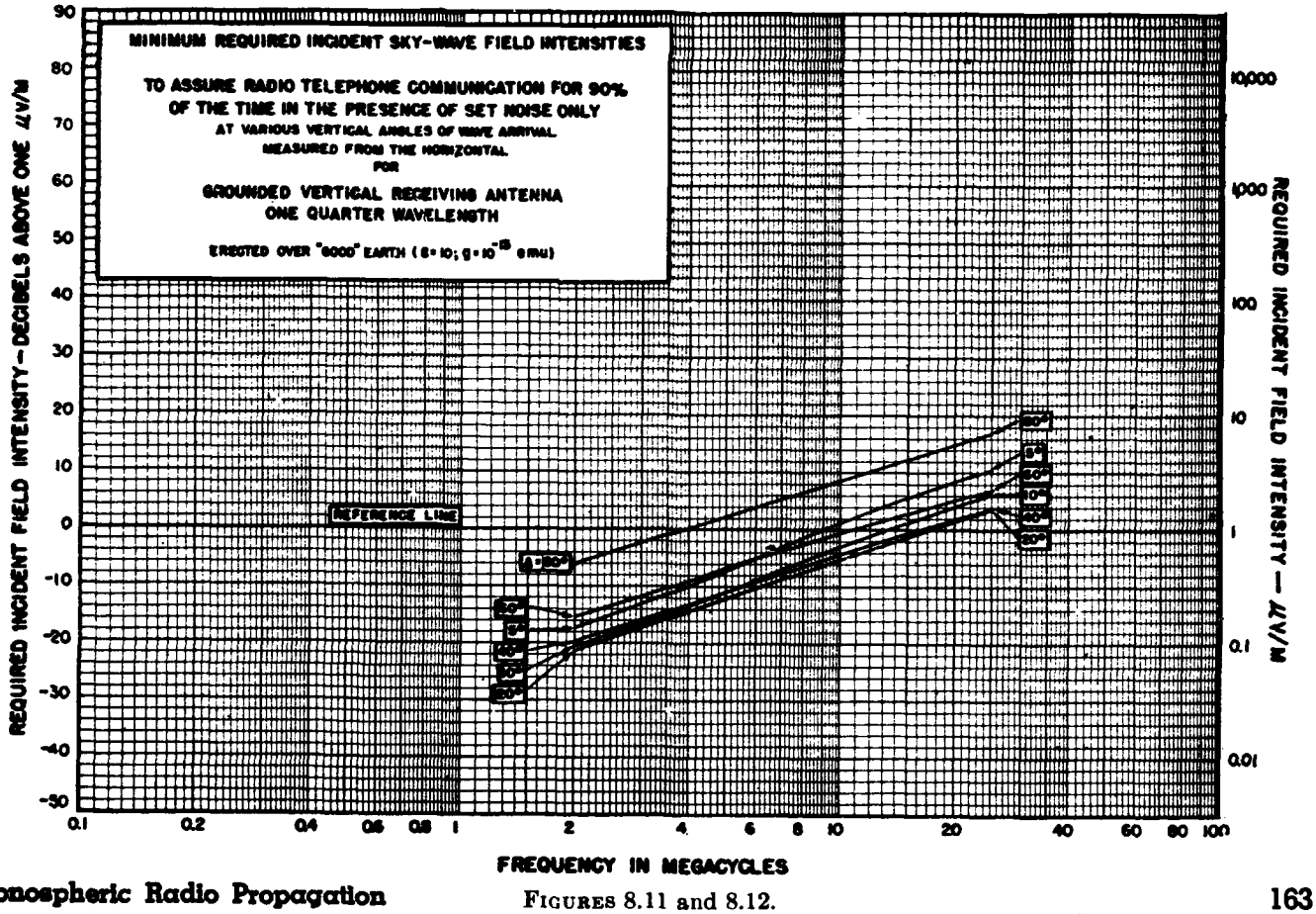
Robert H. Dicke and Robert Beringer, Microwave radiation from the sun and moon, Astrophys J. **103**: 375 (1946).

J. L. Pawsey, R. Payne-Scott, and L. L. McCready, Radio frequency energy from the sun, Nature **157**: 158 (Feb. 9, 1946).

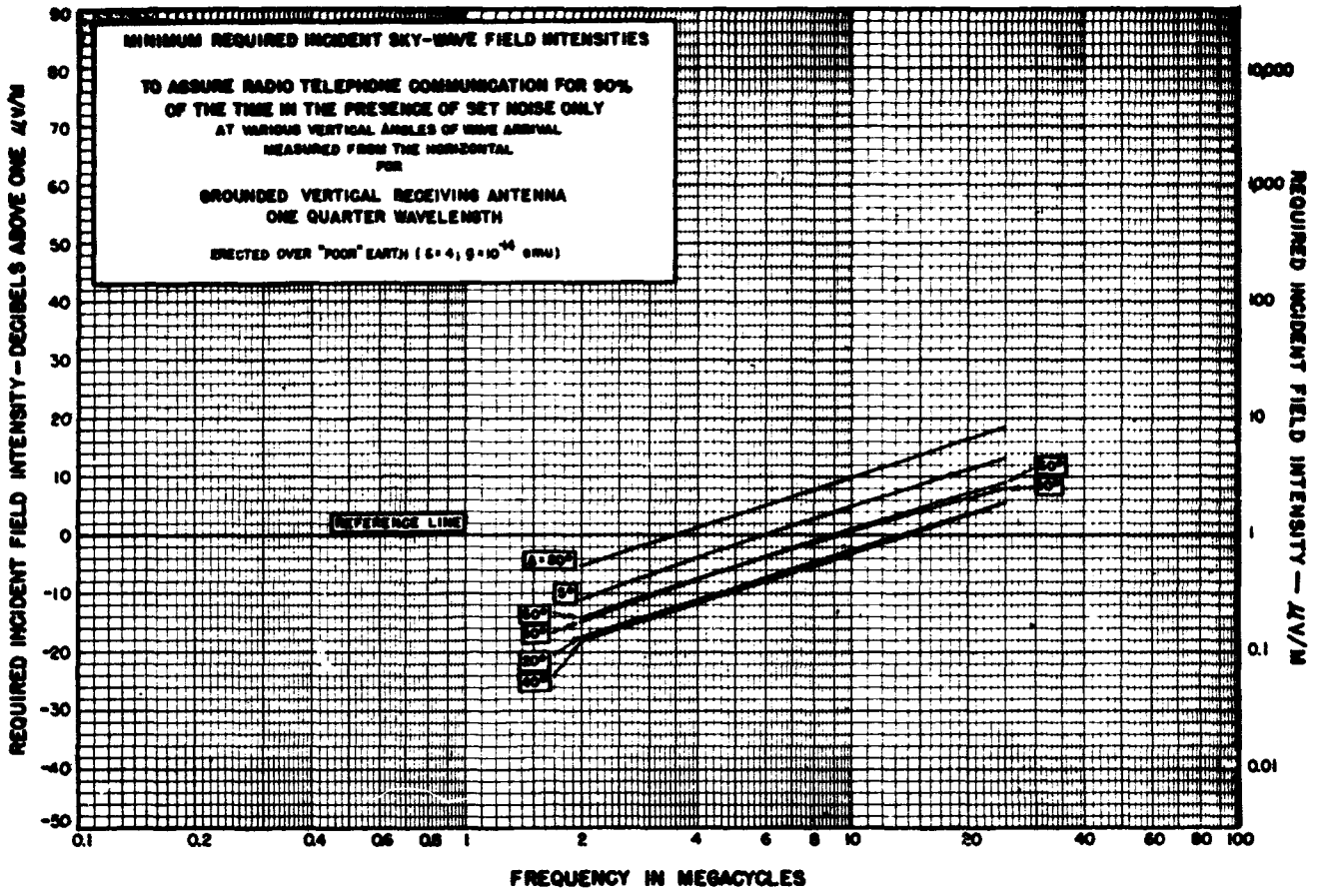
D. F. Martyn, Temperature radiation from the quiet sun in the radio spectrum, Nature **158**: 632 (Nov. 2, 1946).

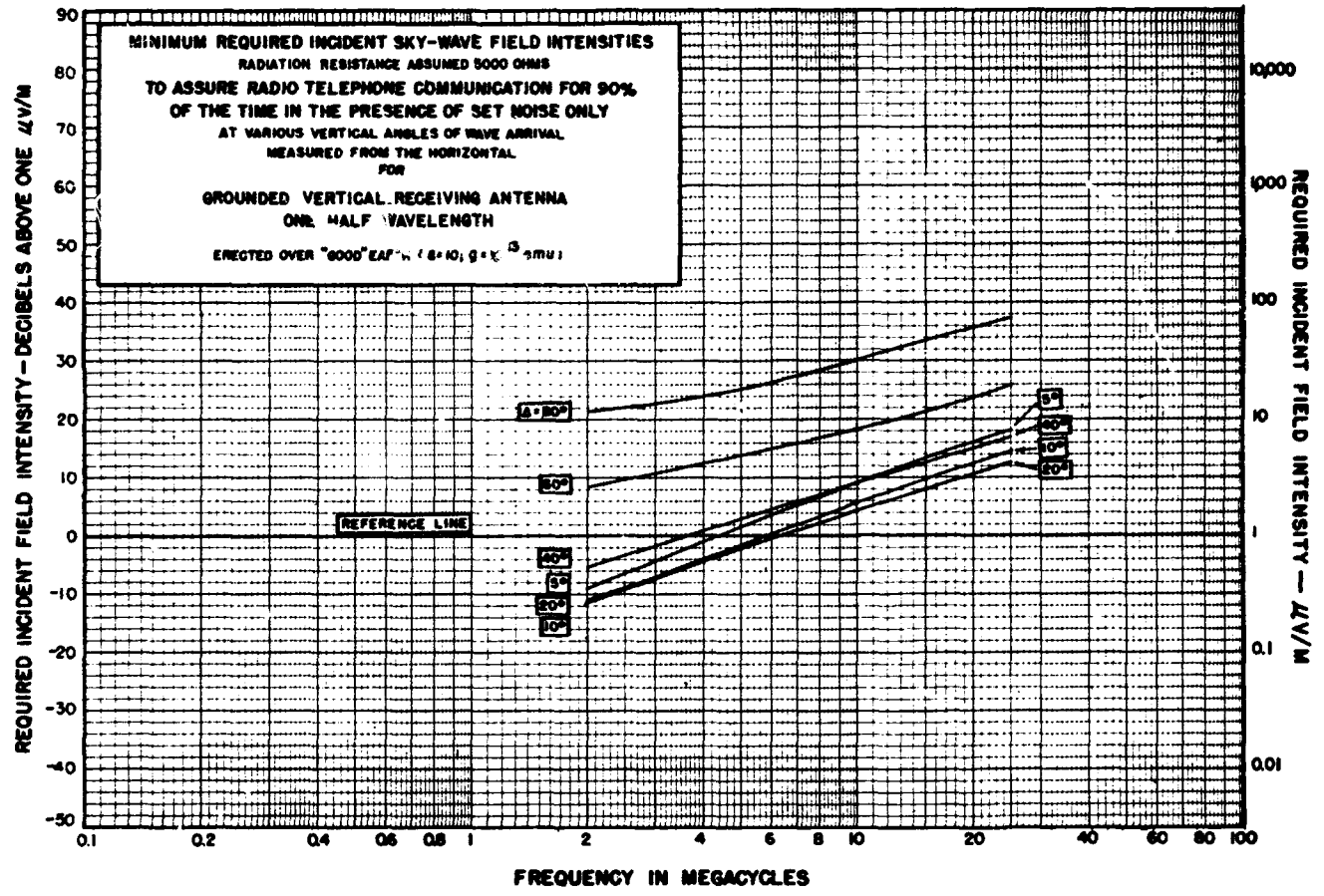
Charles Hard Townes, Interpretation of radio radiation from the Milky Way, Astrophys. J. **105**: (1947).



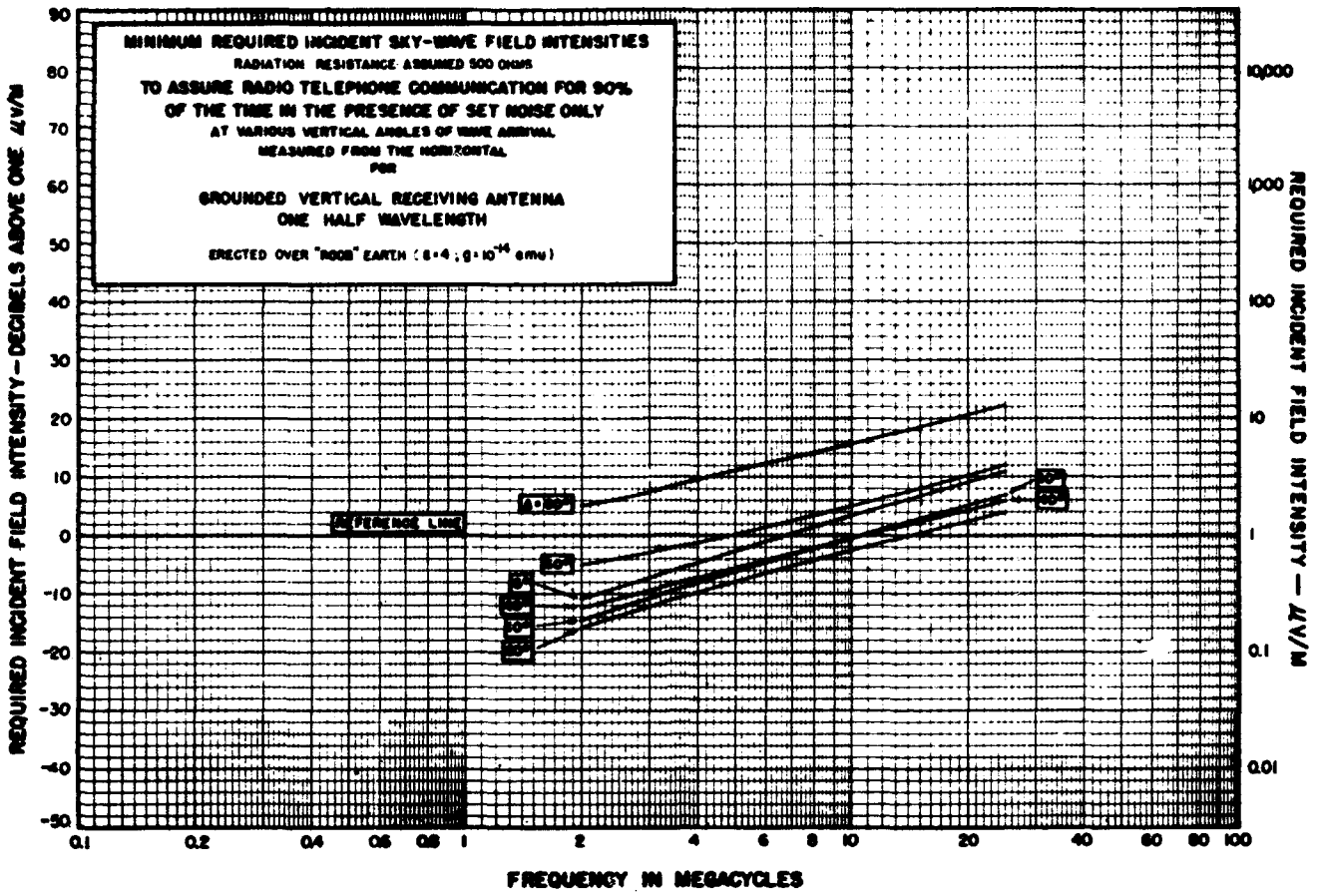


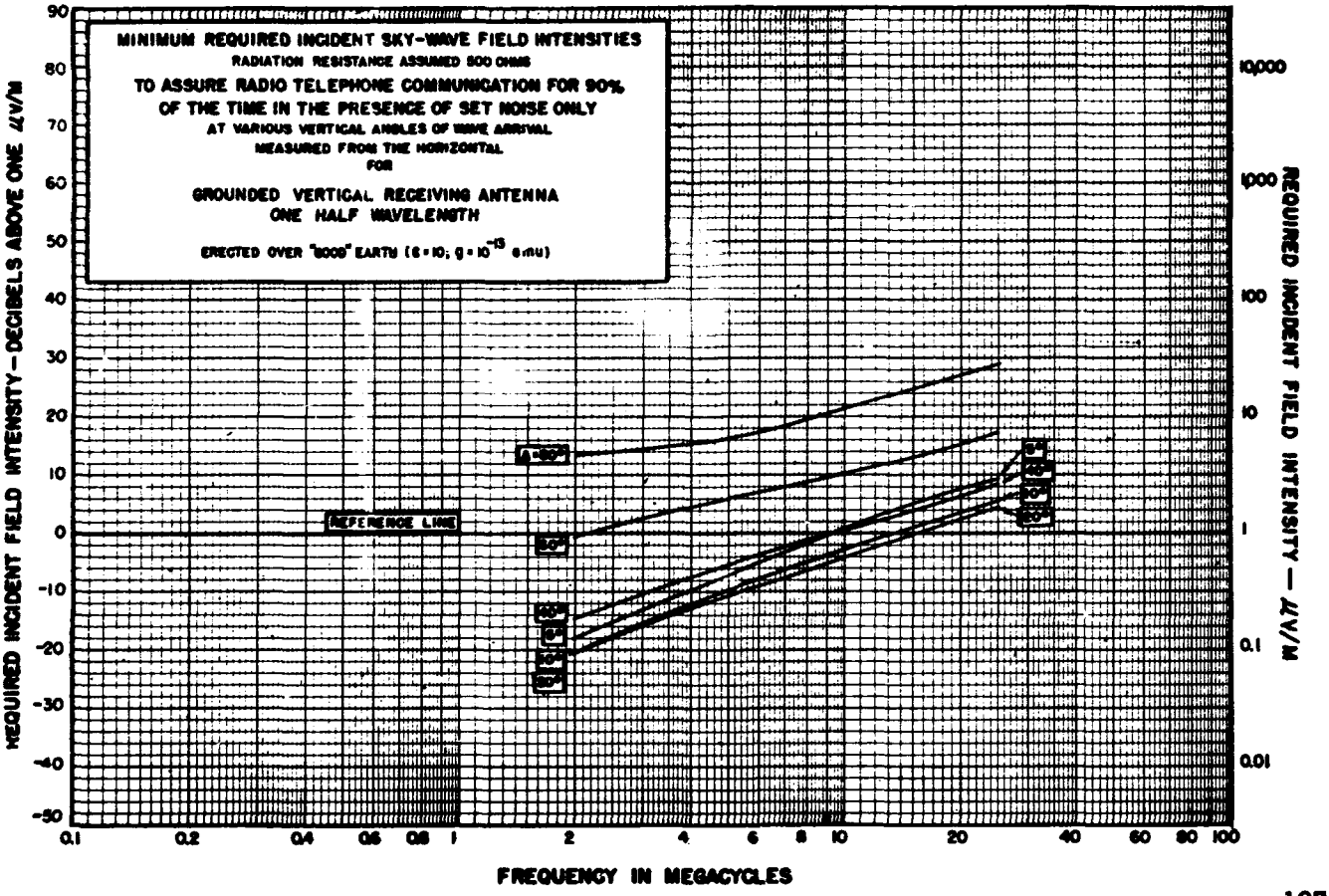
FIGURES 8.11 and 8.12.



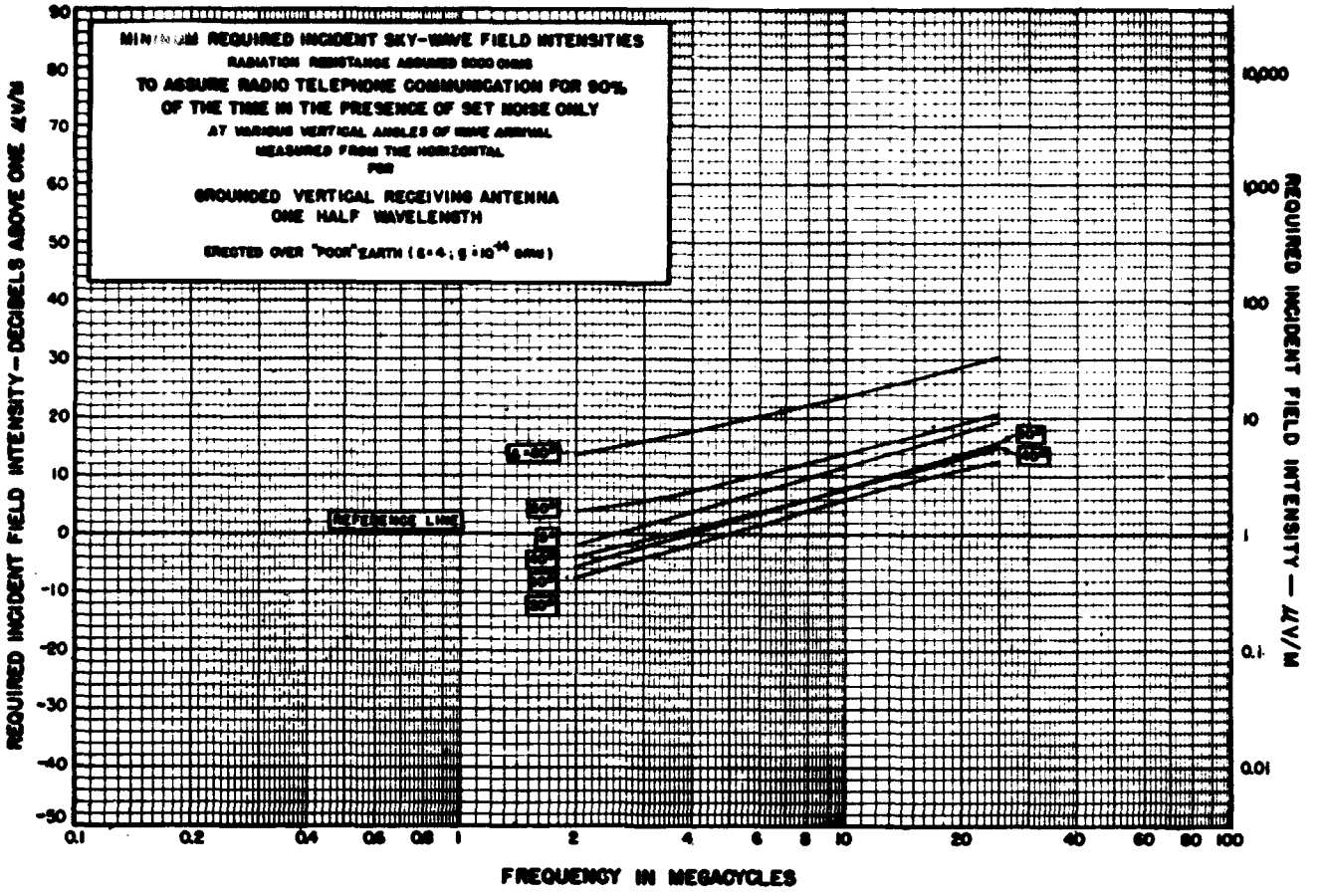


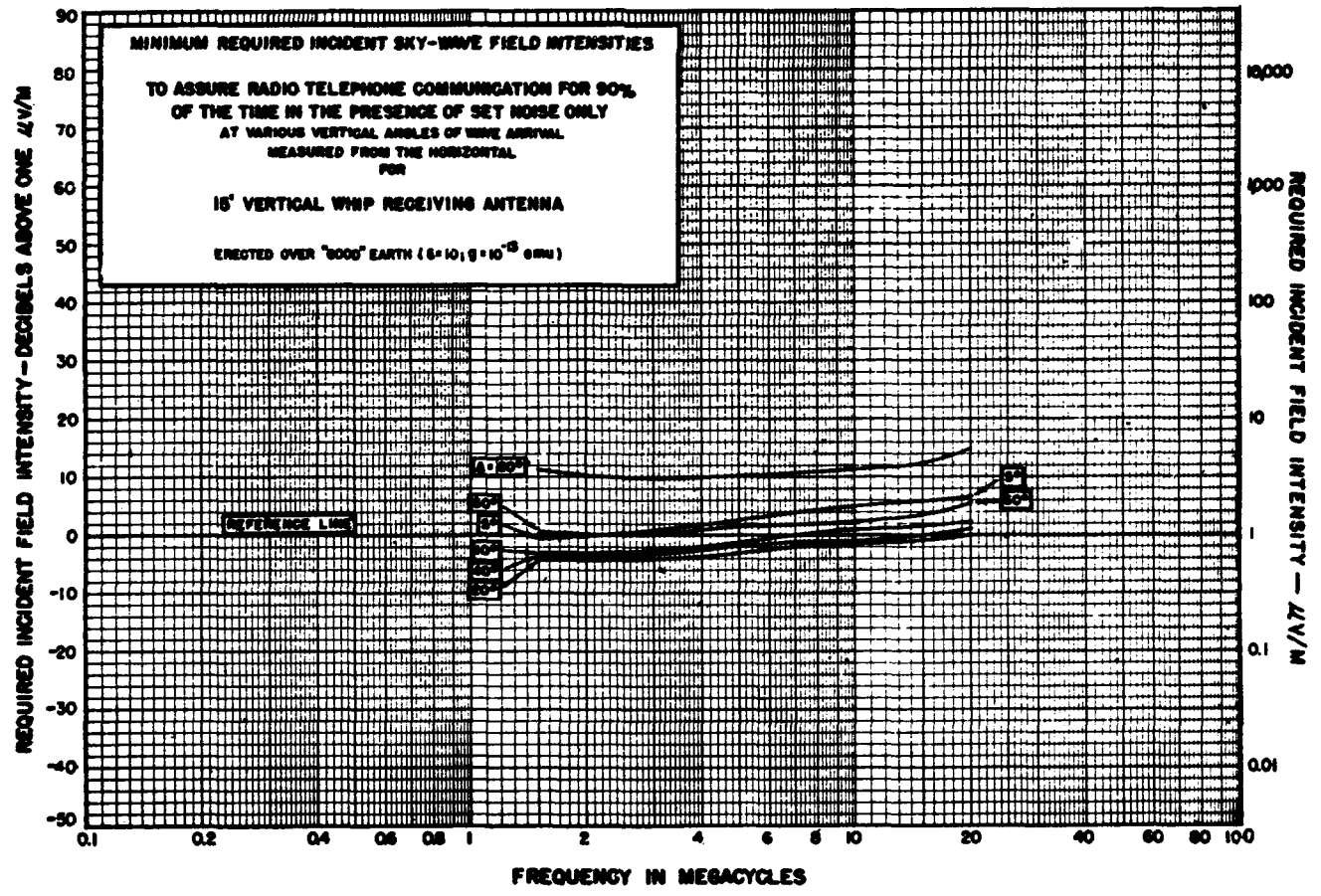
FIGURES 8.13 and 8.14.



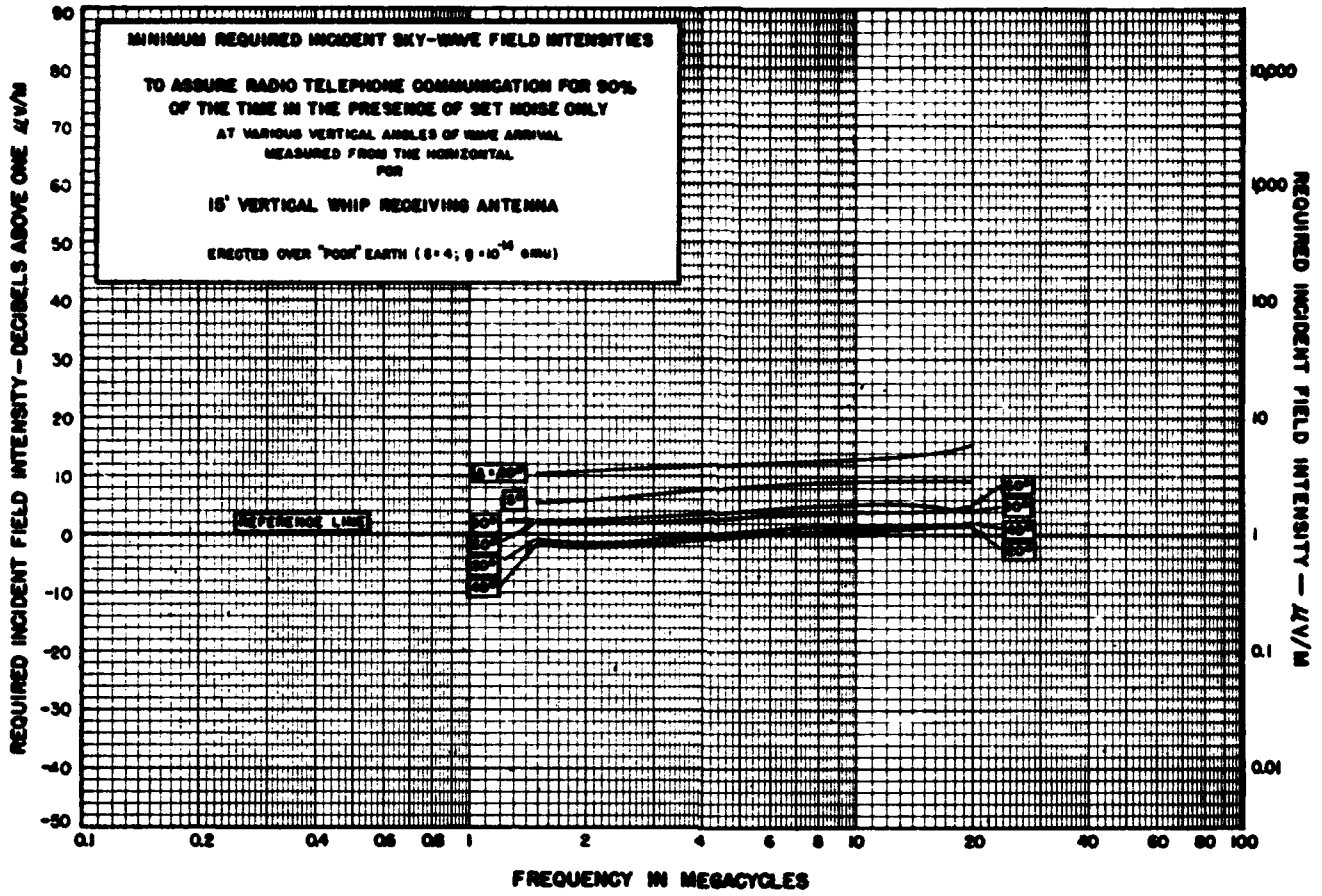


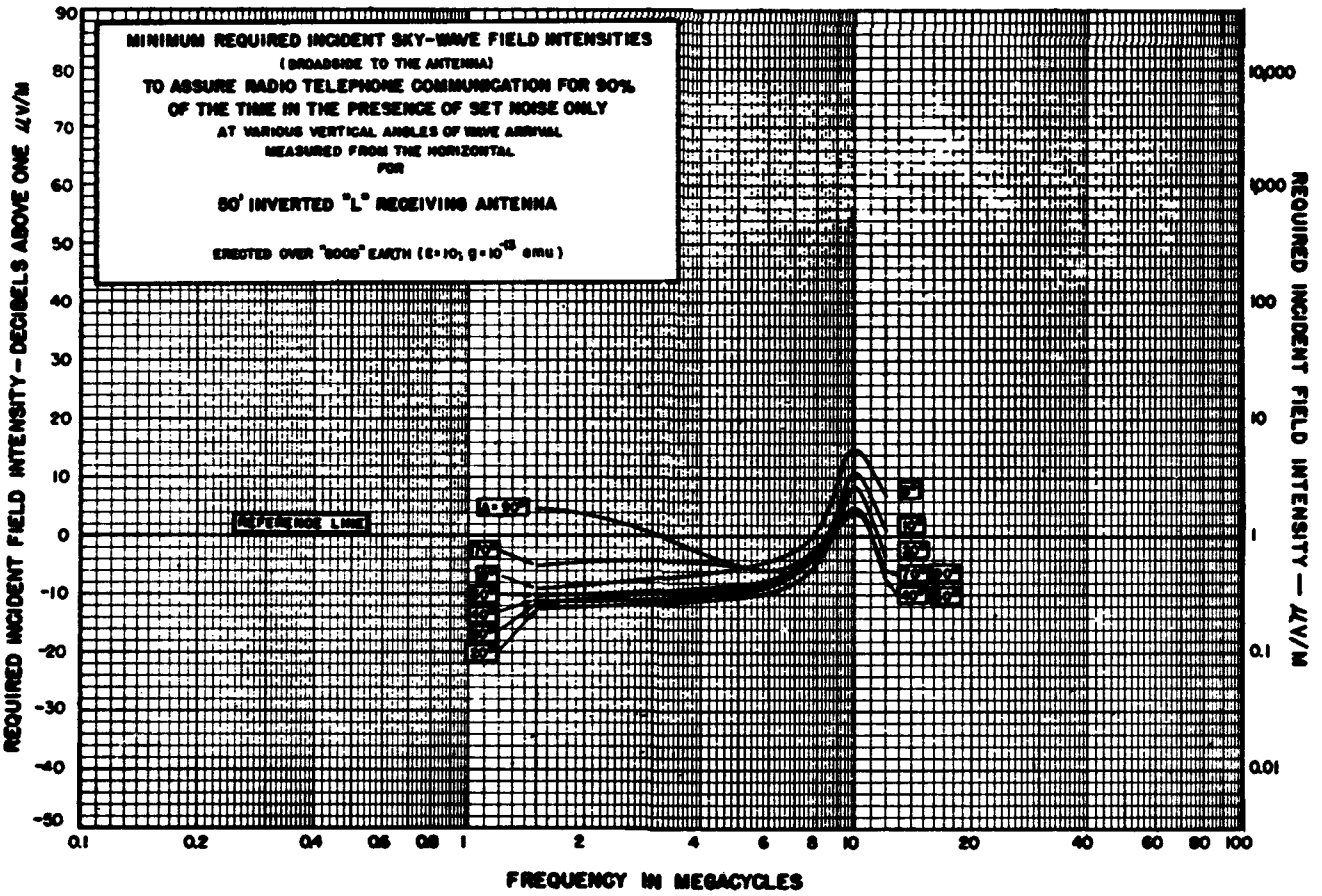
FIGURES 8.15 and 8.16.



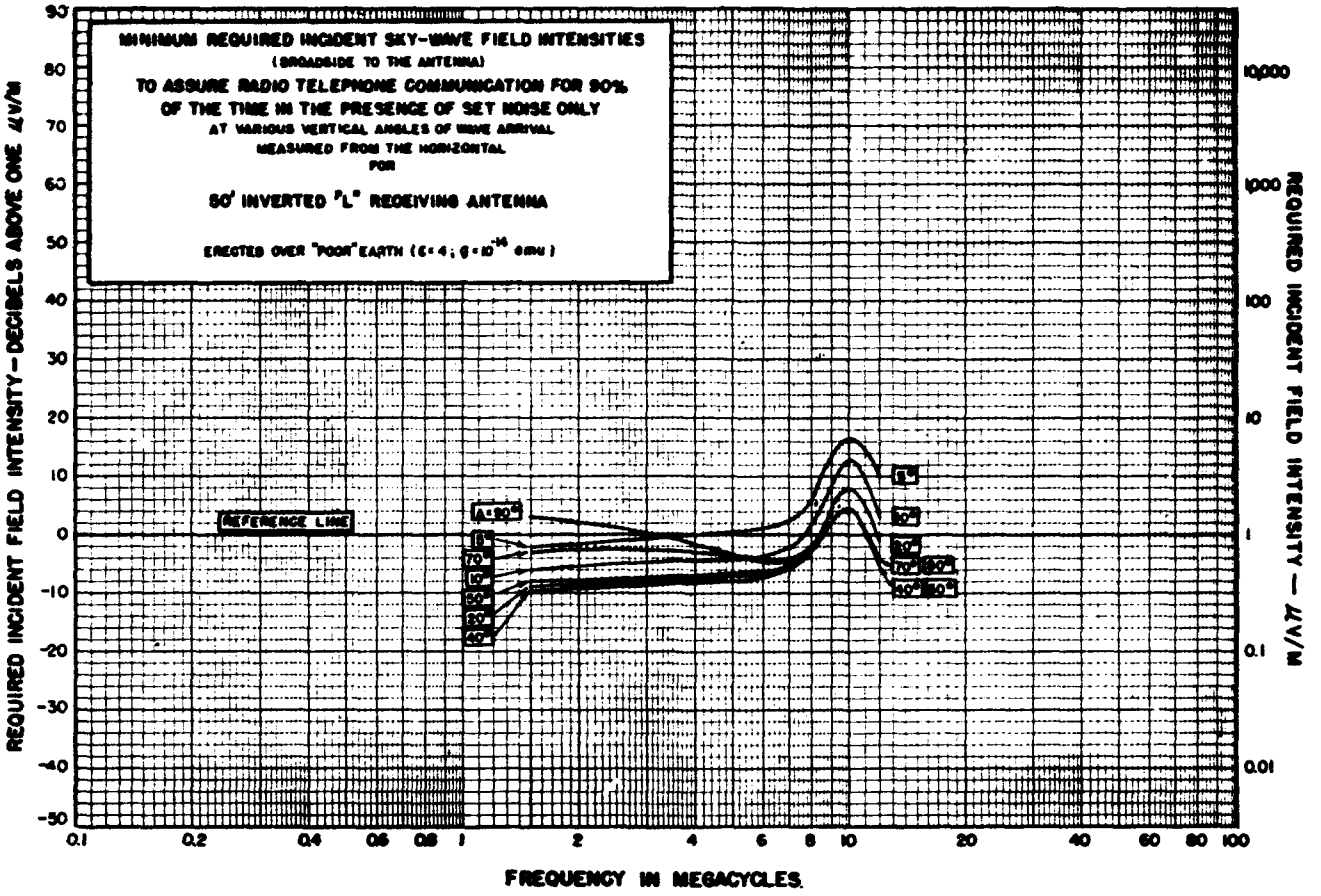


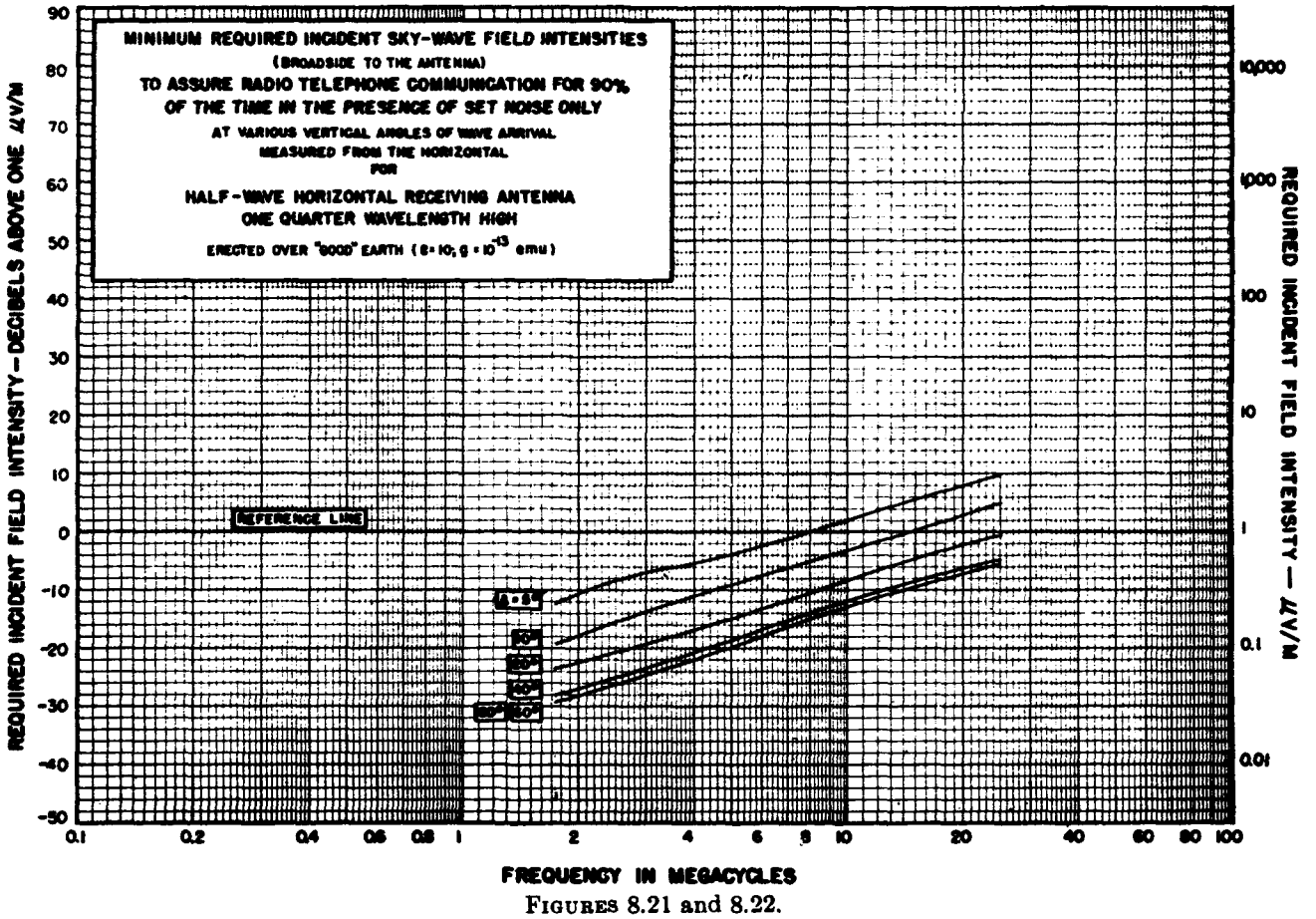
FIGURES 8.17 and 8.18.



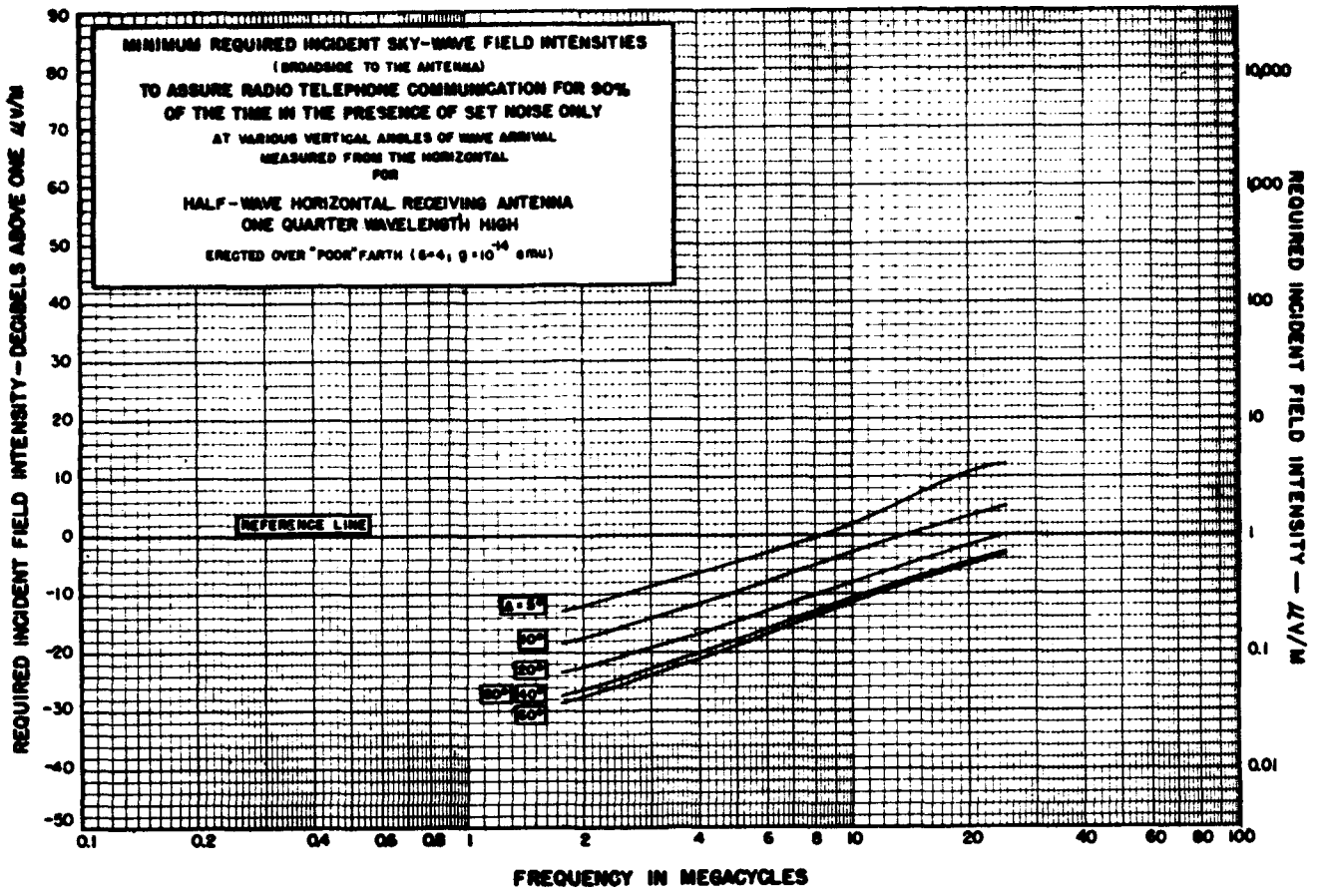


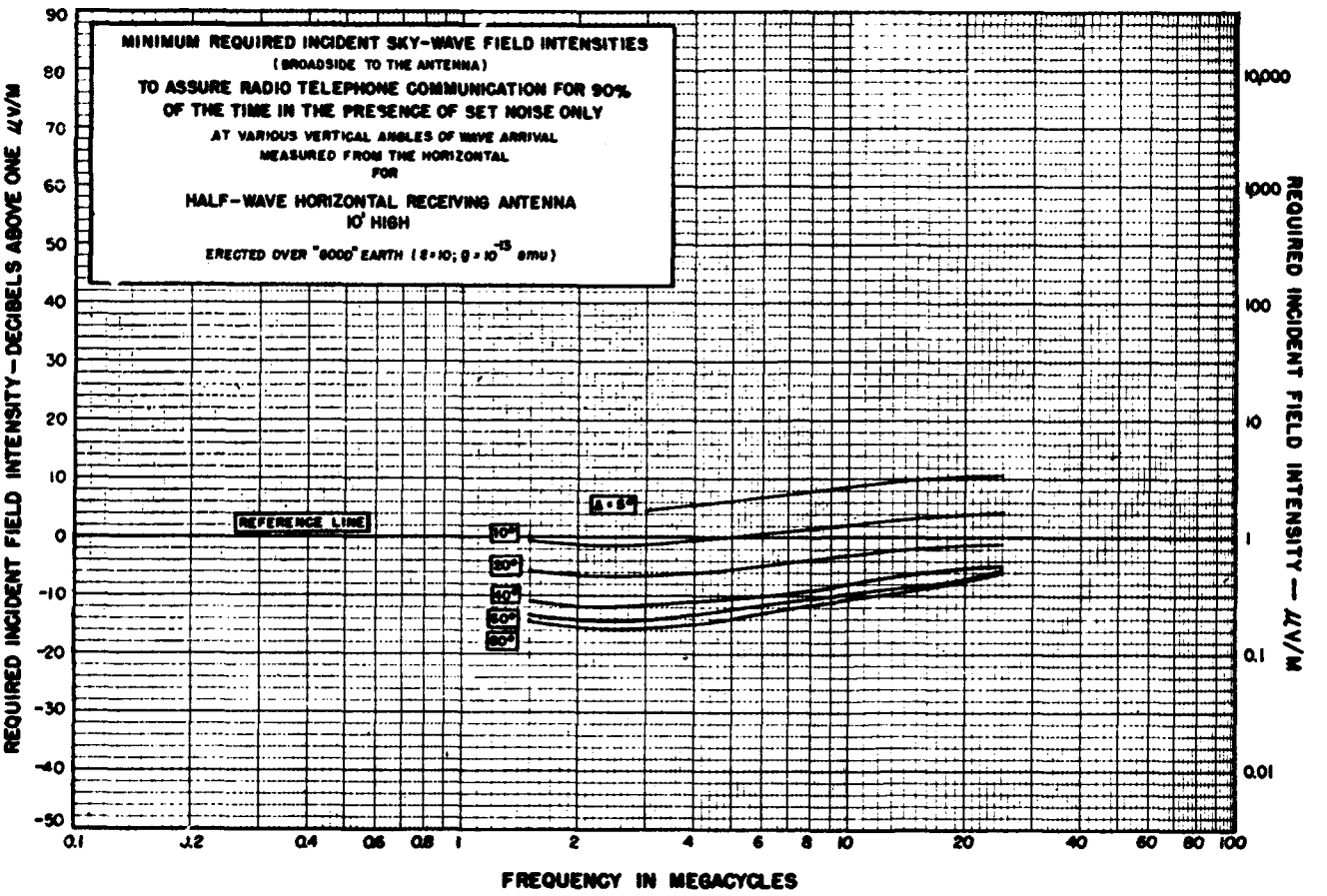
FIGURES 8.19 and 8.20.



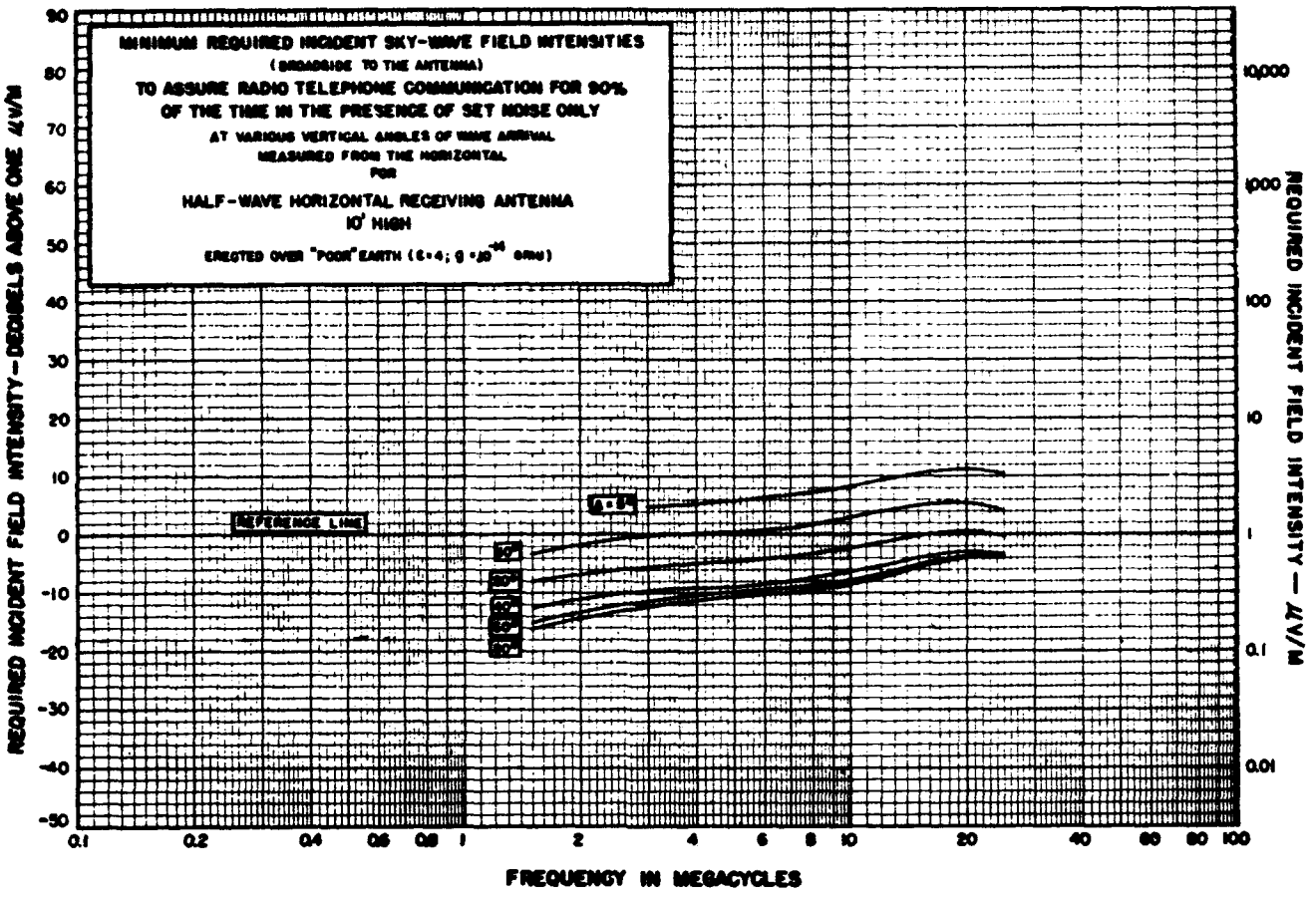


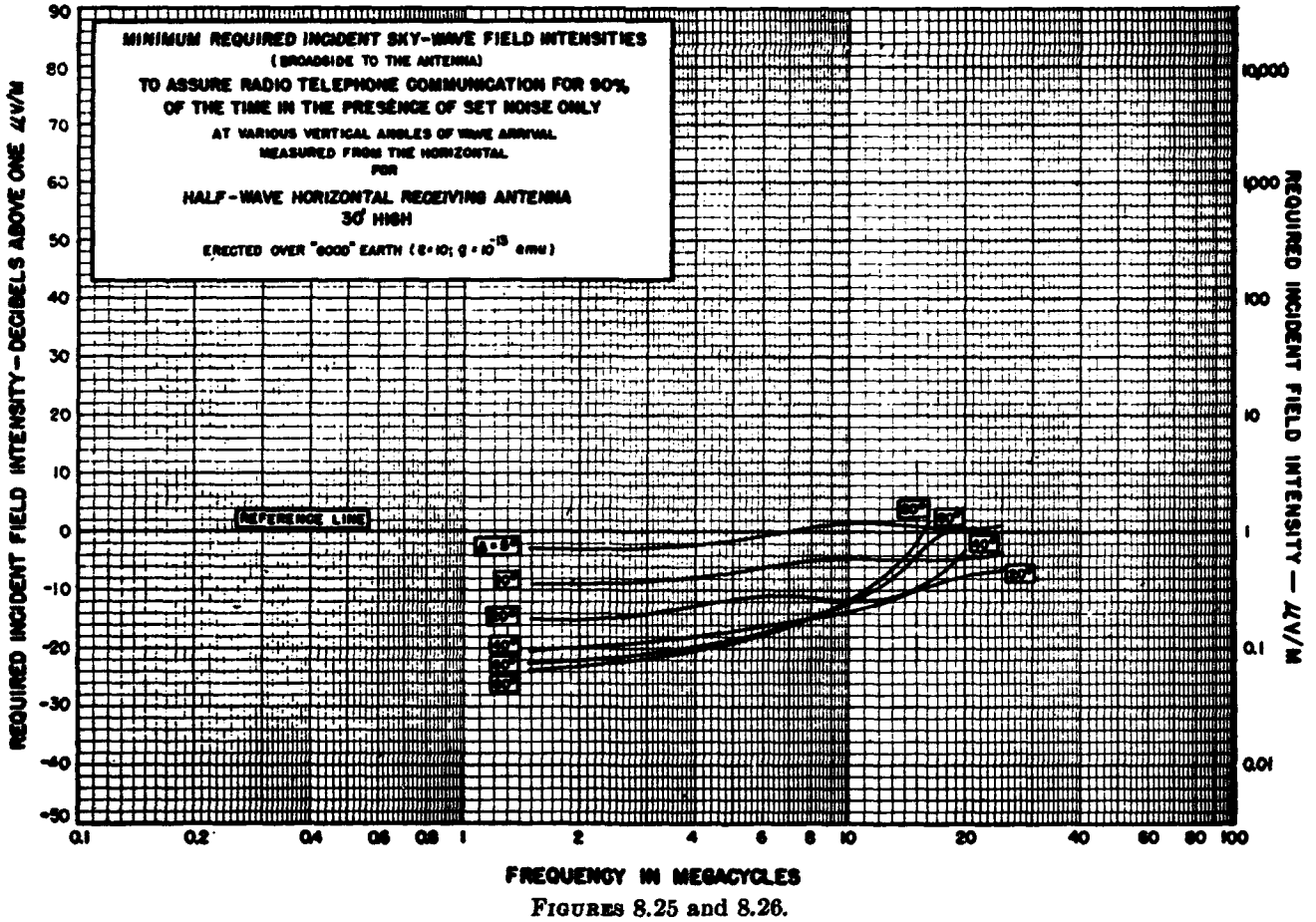
FIGURES 8.21 and 8.22.



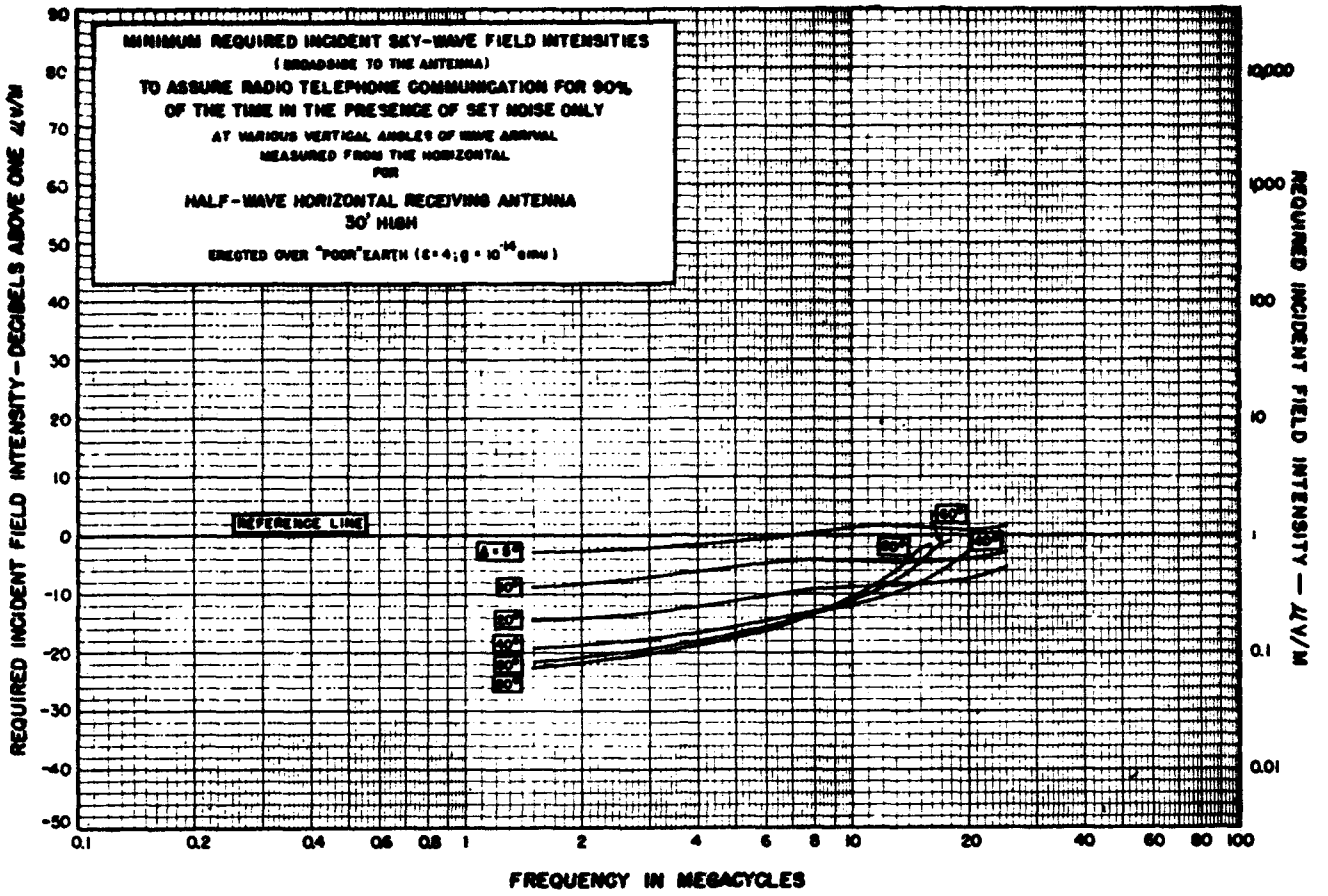


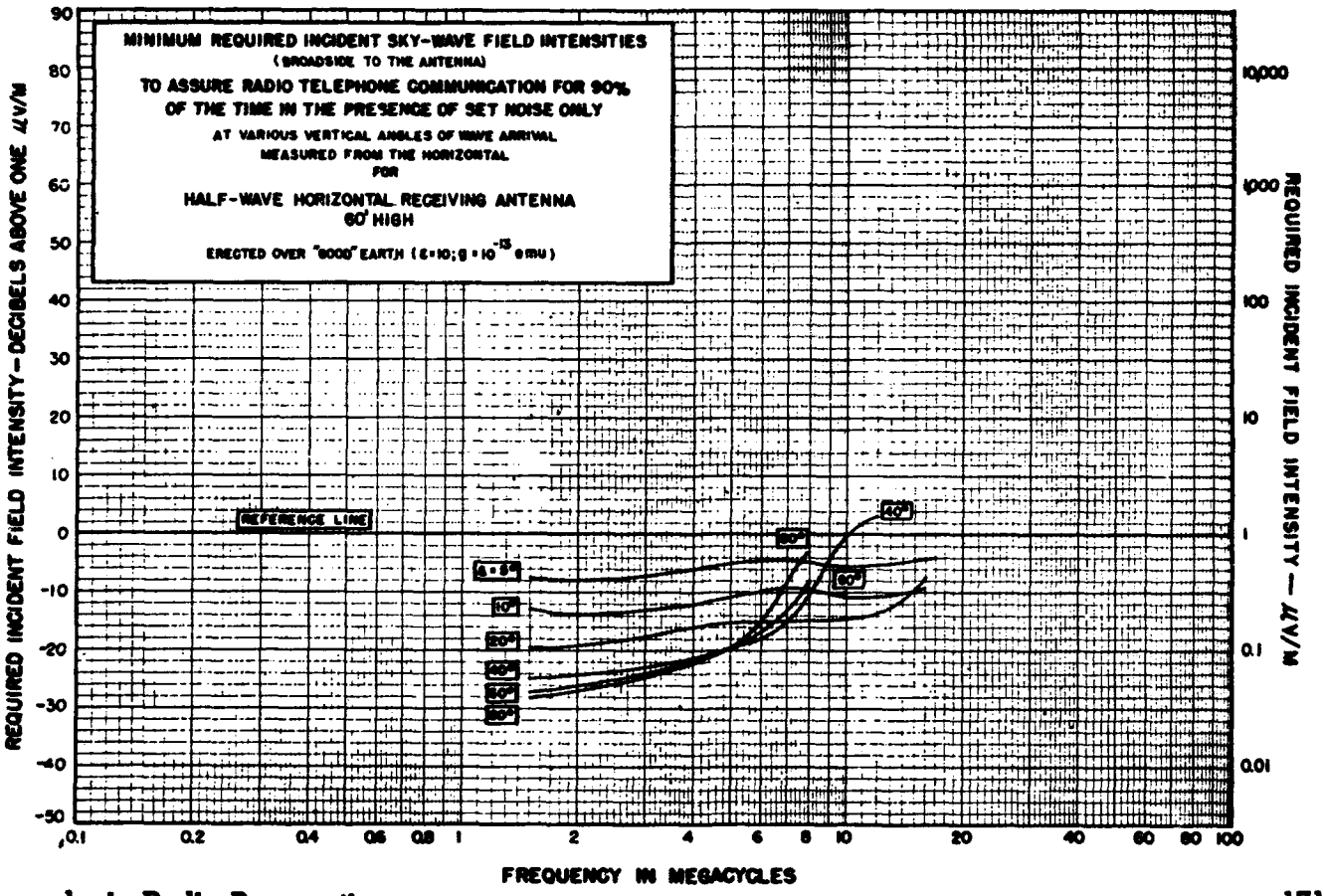
FIGURES 8.23 and 8.24.



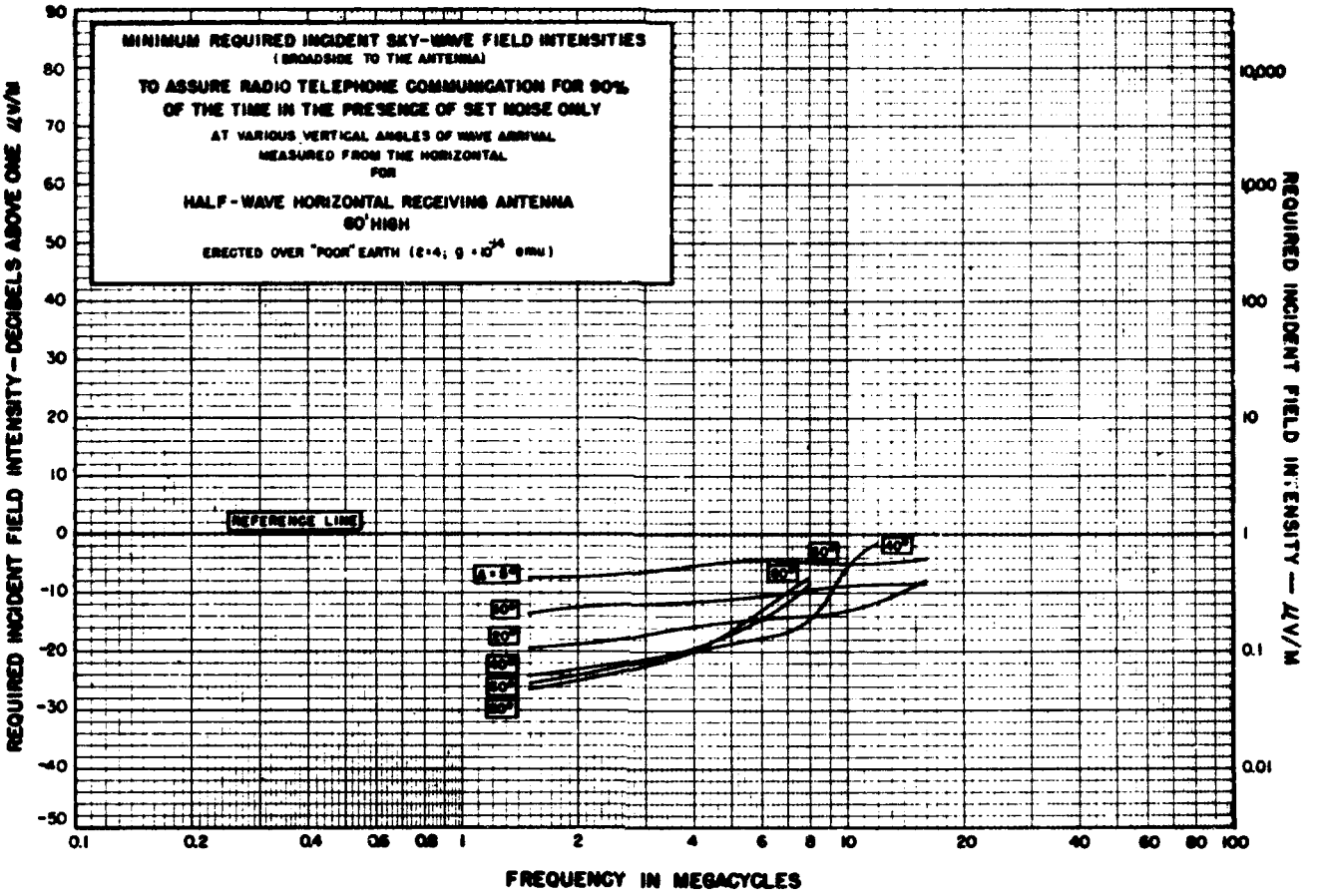


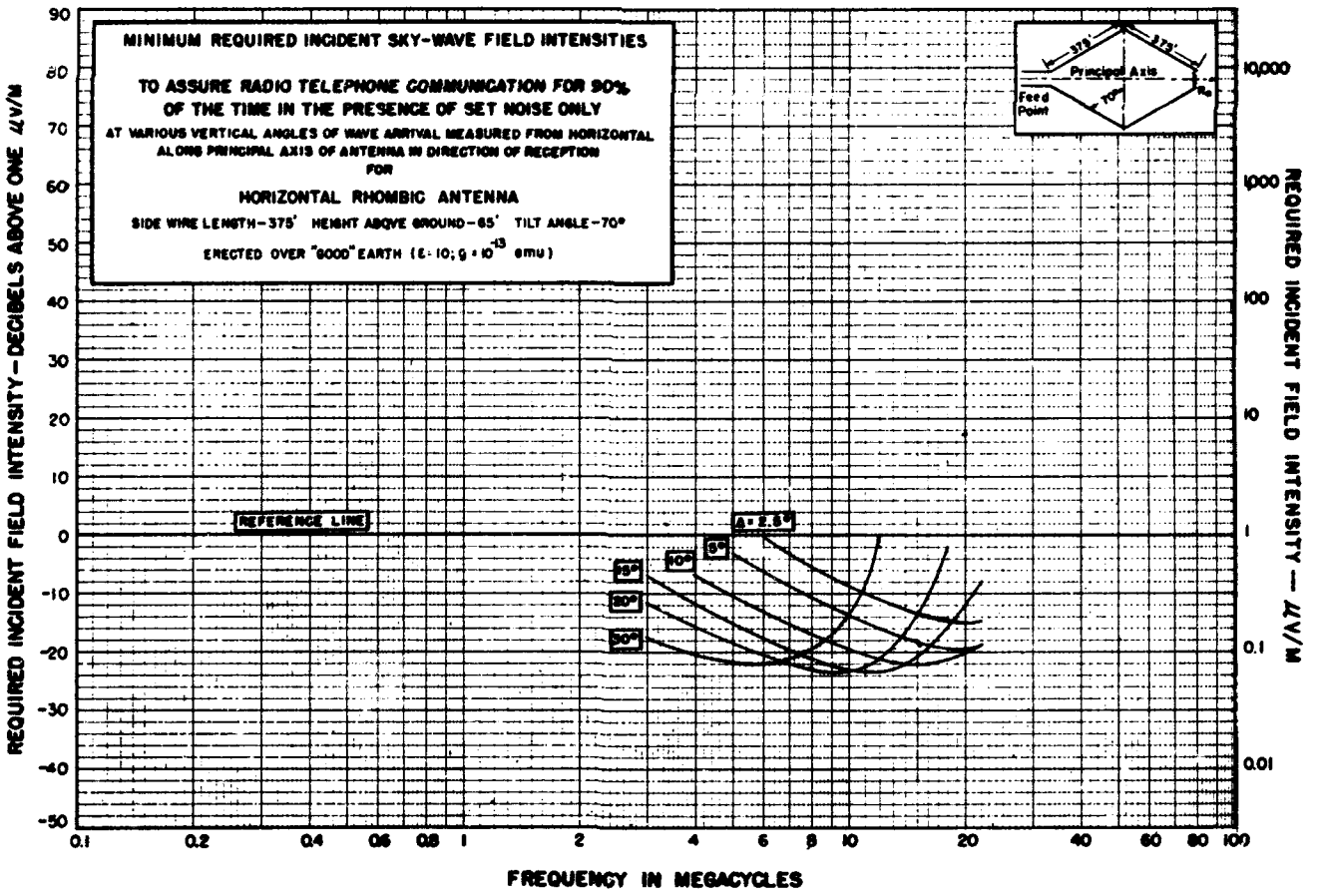
FIGURES 8.25 and 8.26.



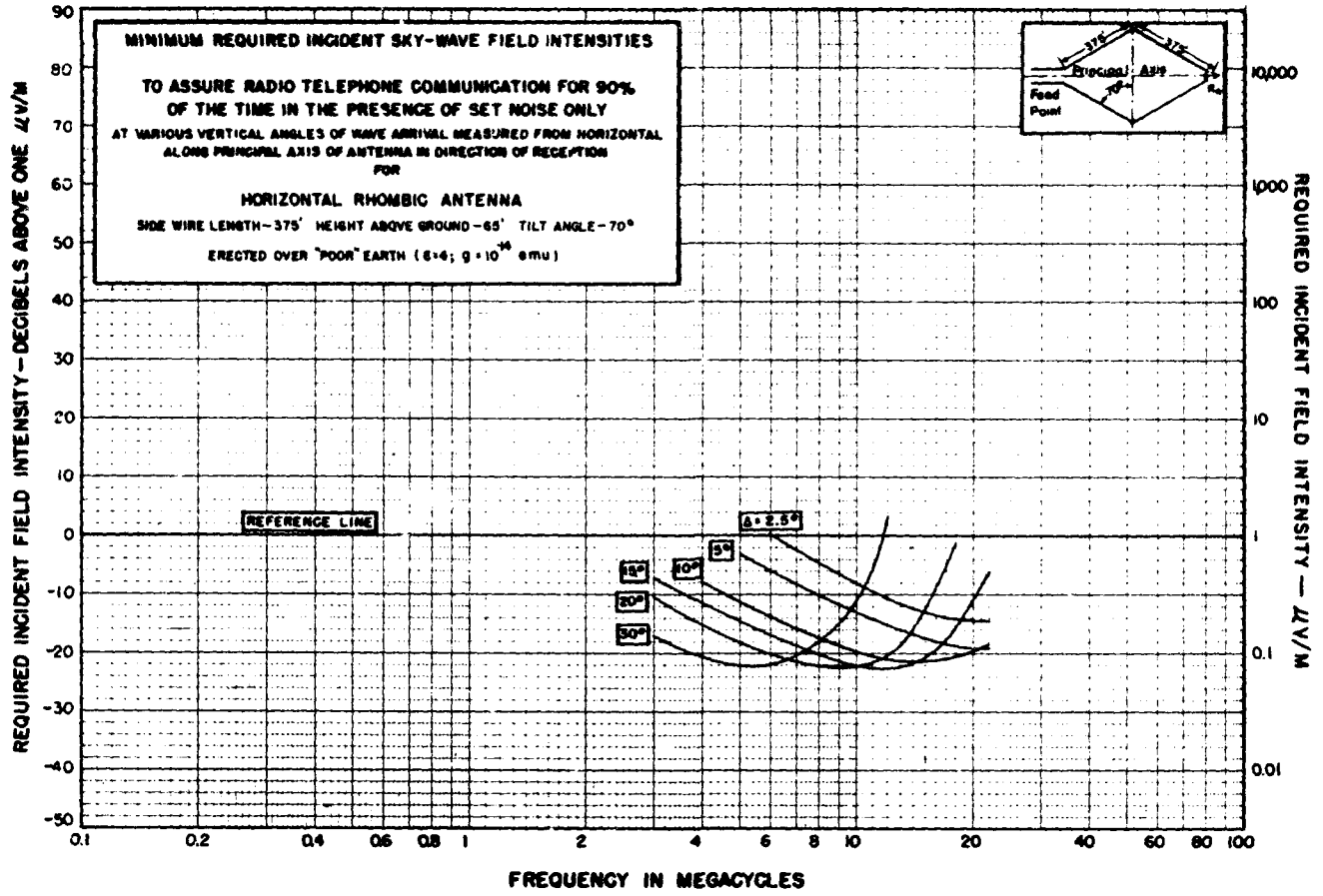


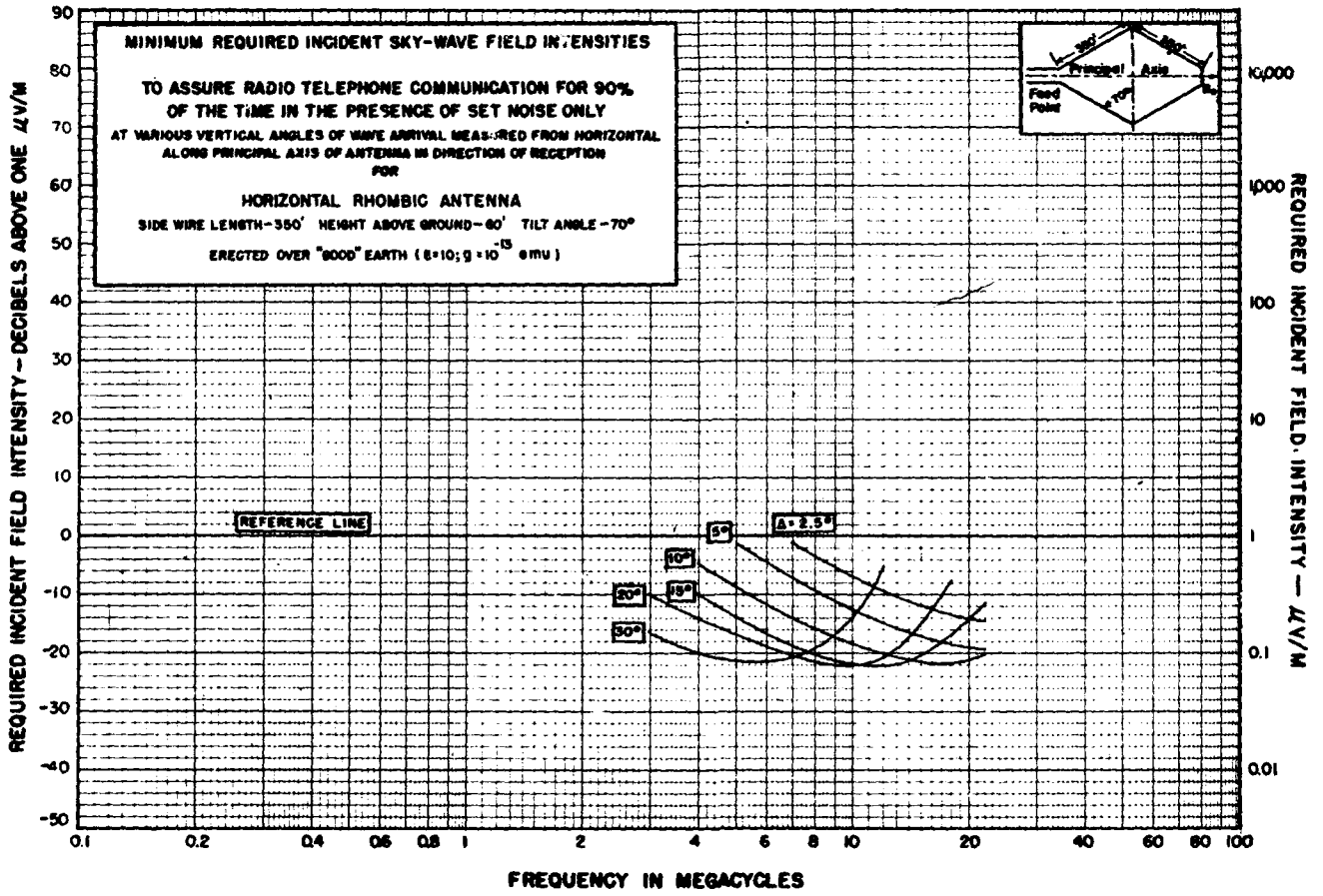
FIGURES 8.27 and 8.28.





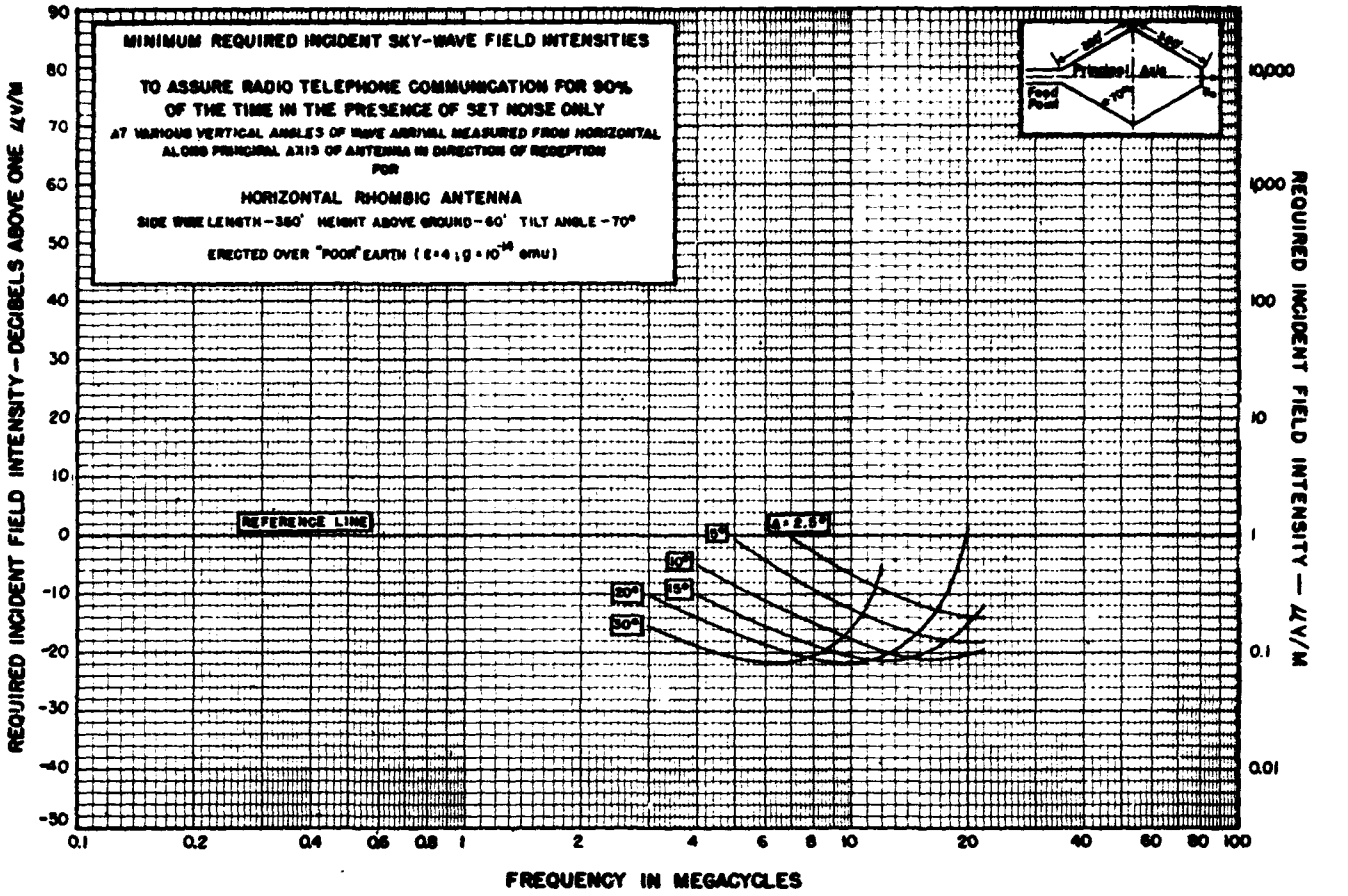
FIGURES 8.29 and 8.30.

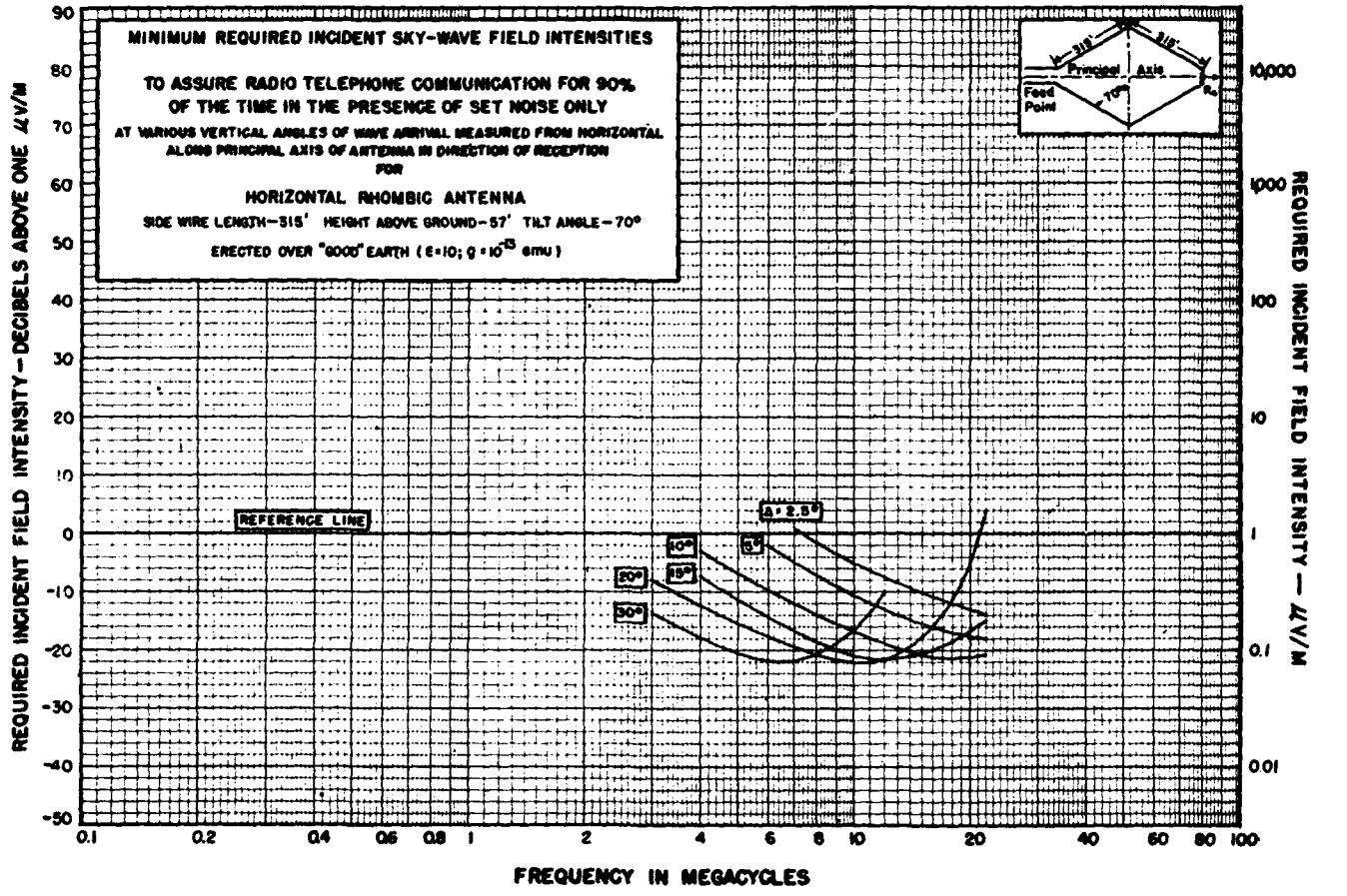




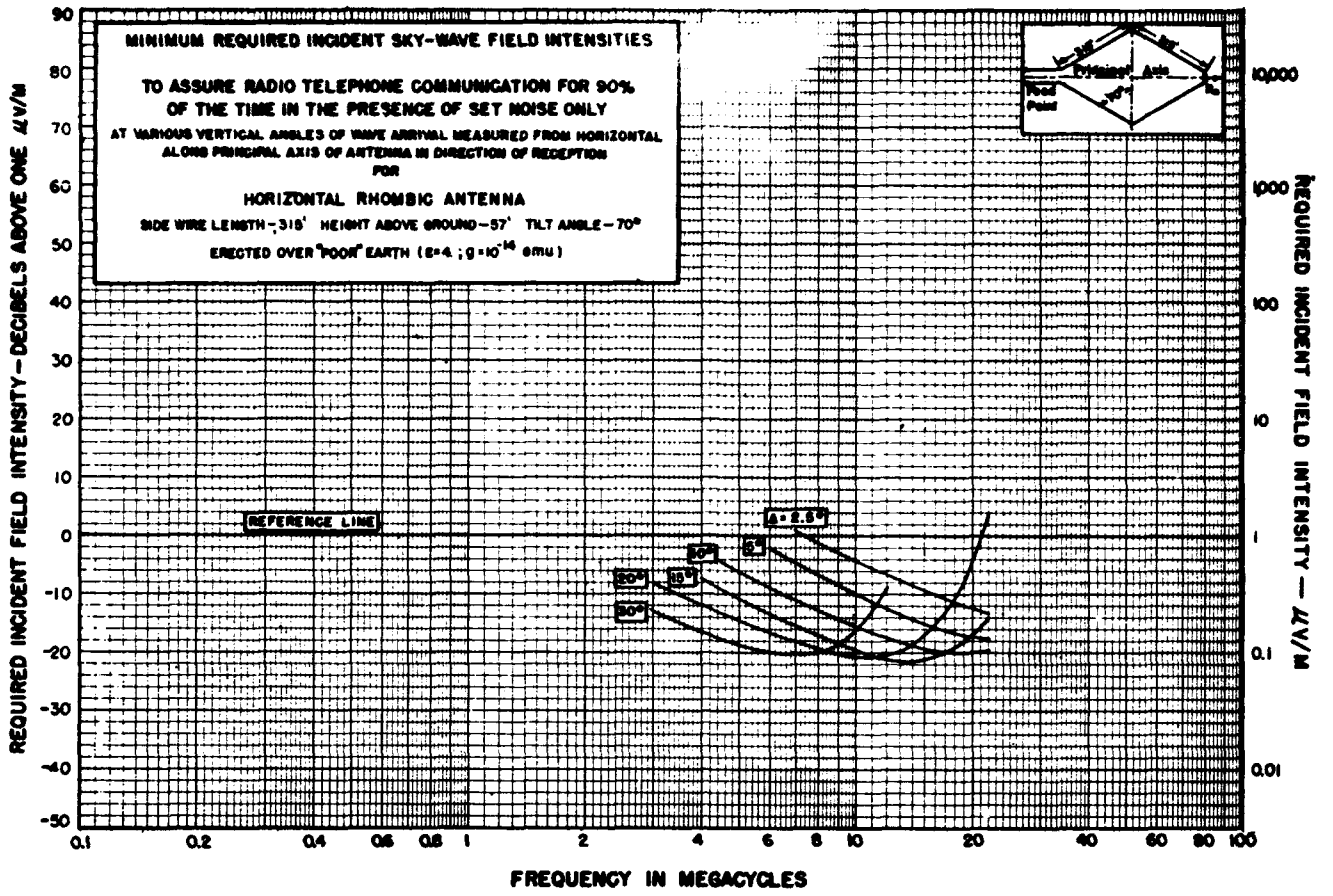
FIGURES 8.31 and 8.32.

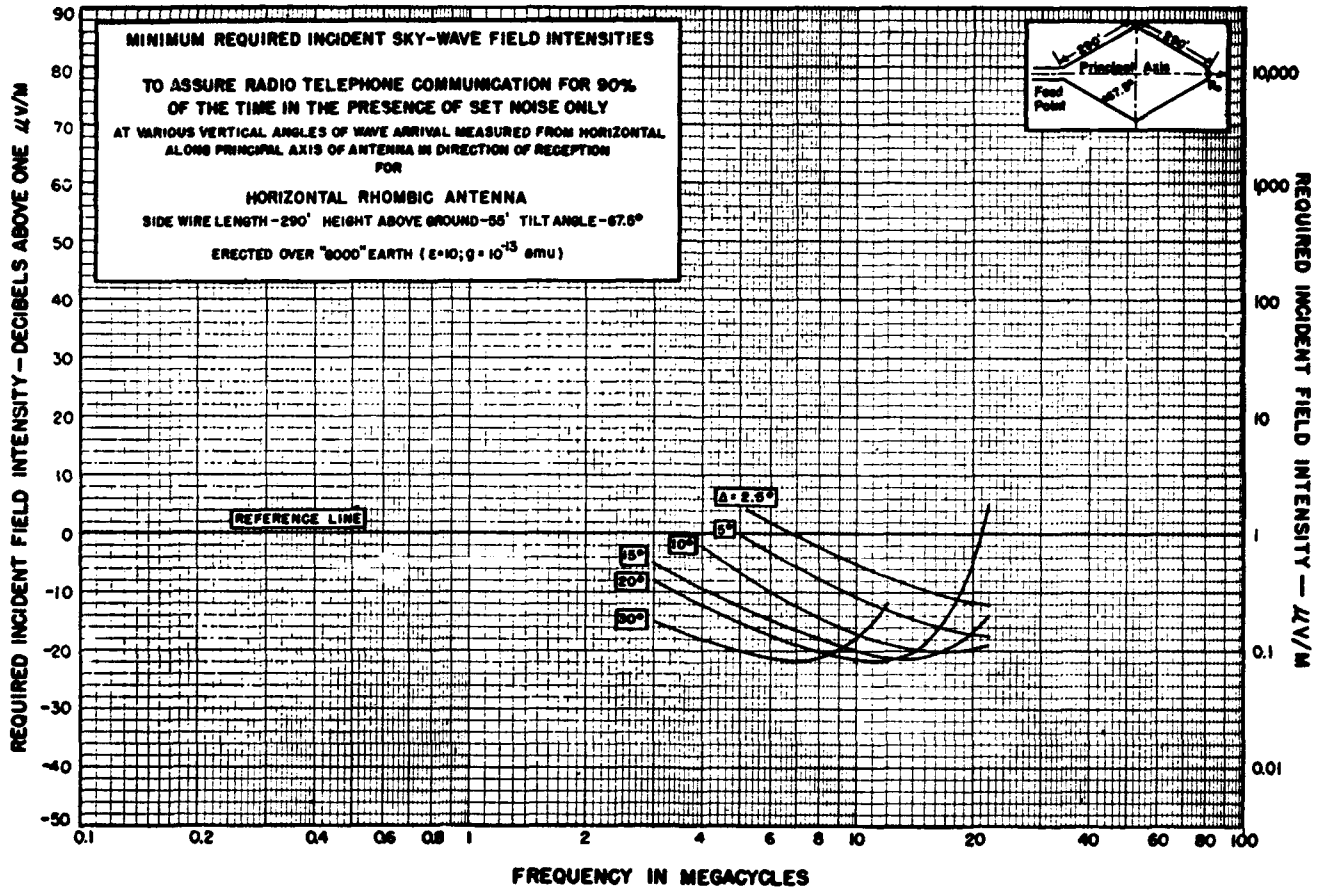
935333 O - 51 - 12



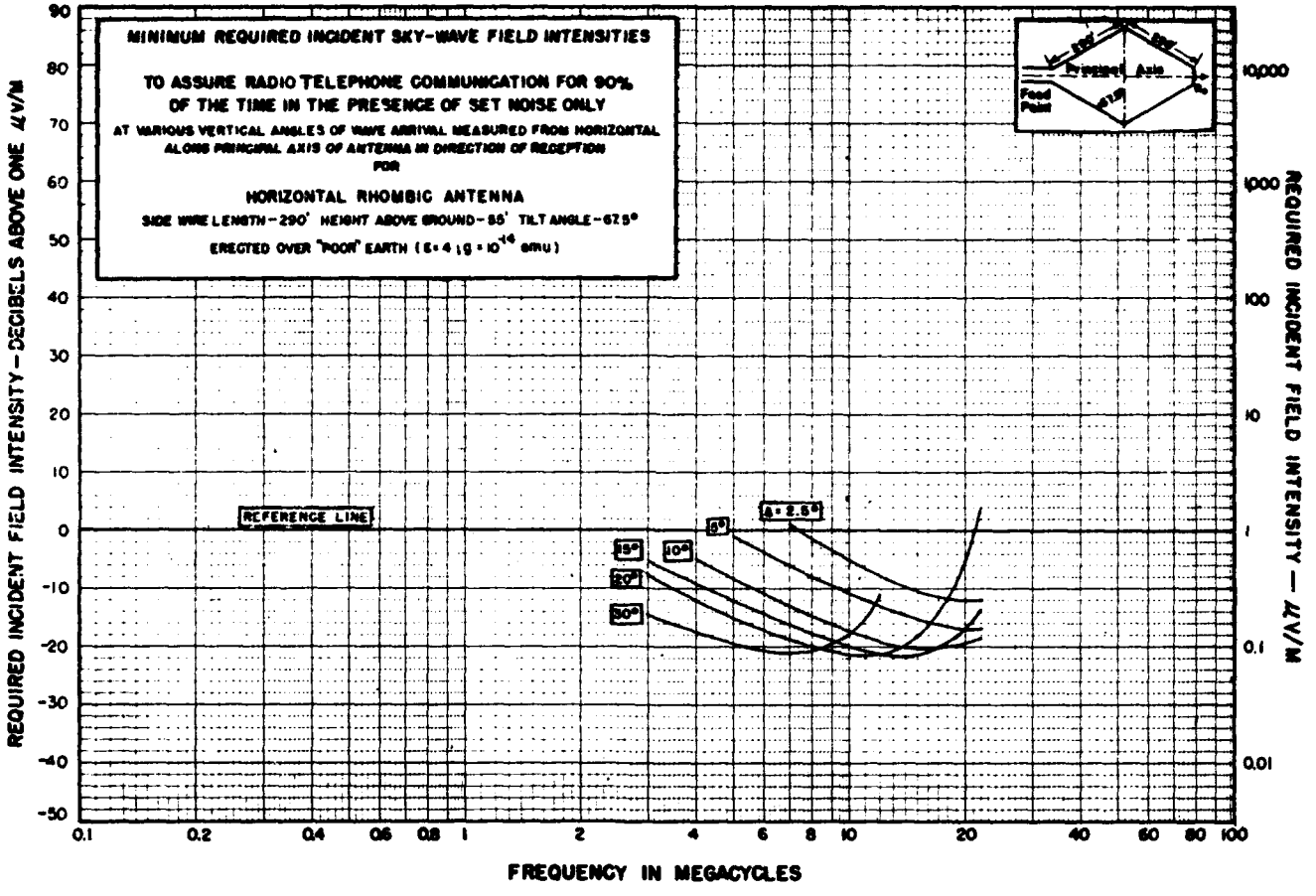


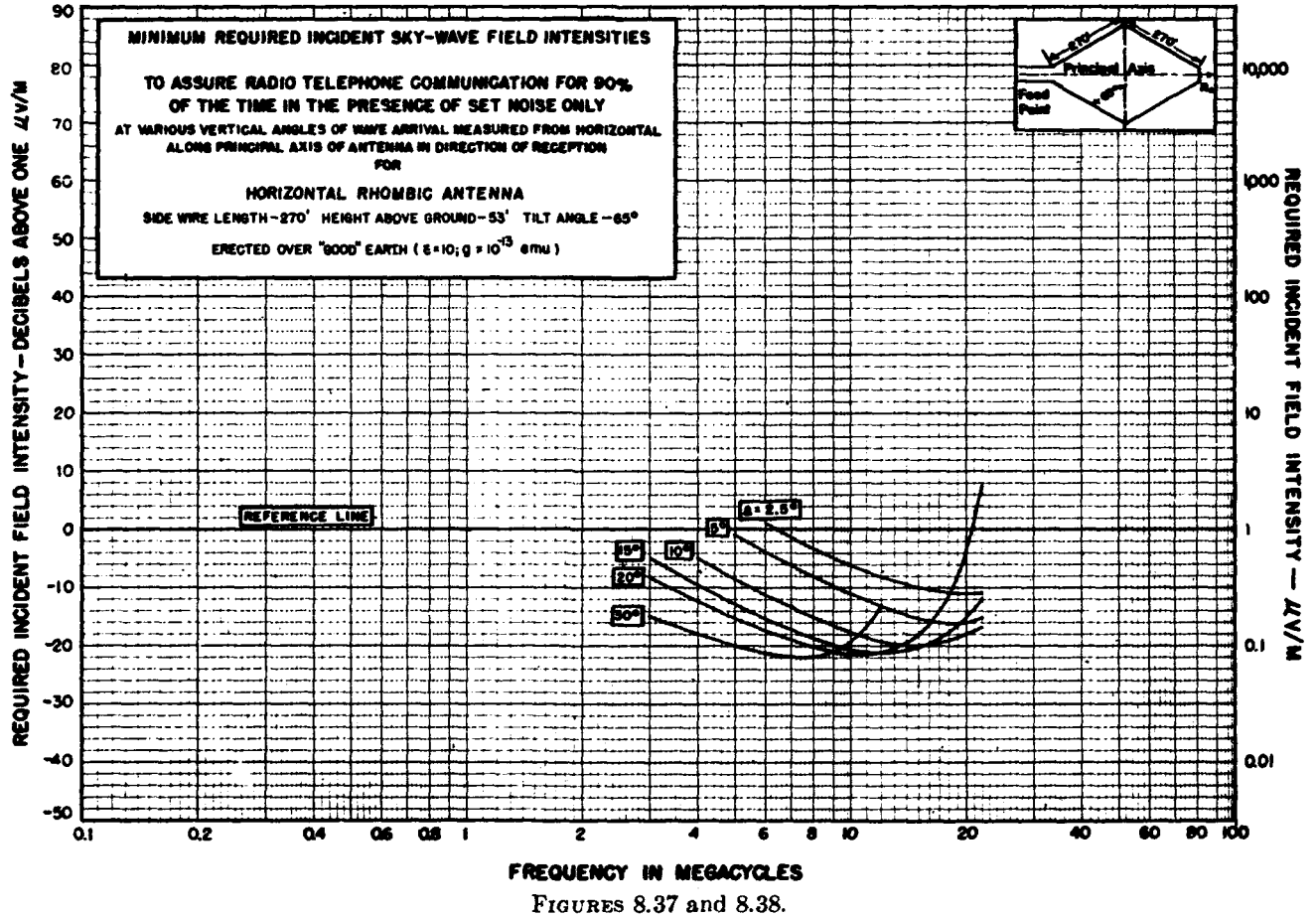
FIGURES 8.33 and 8.34.



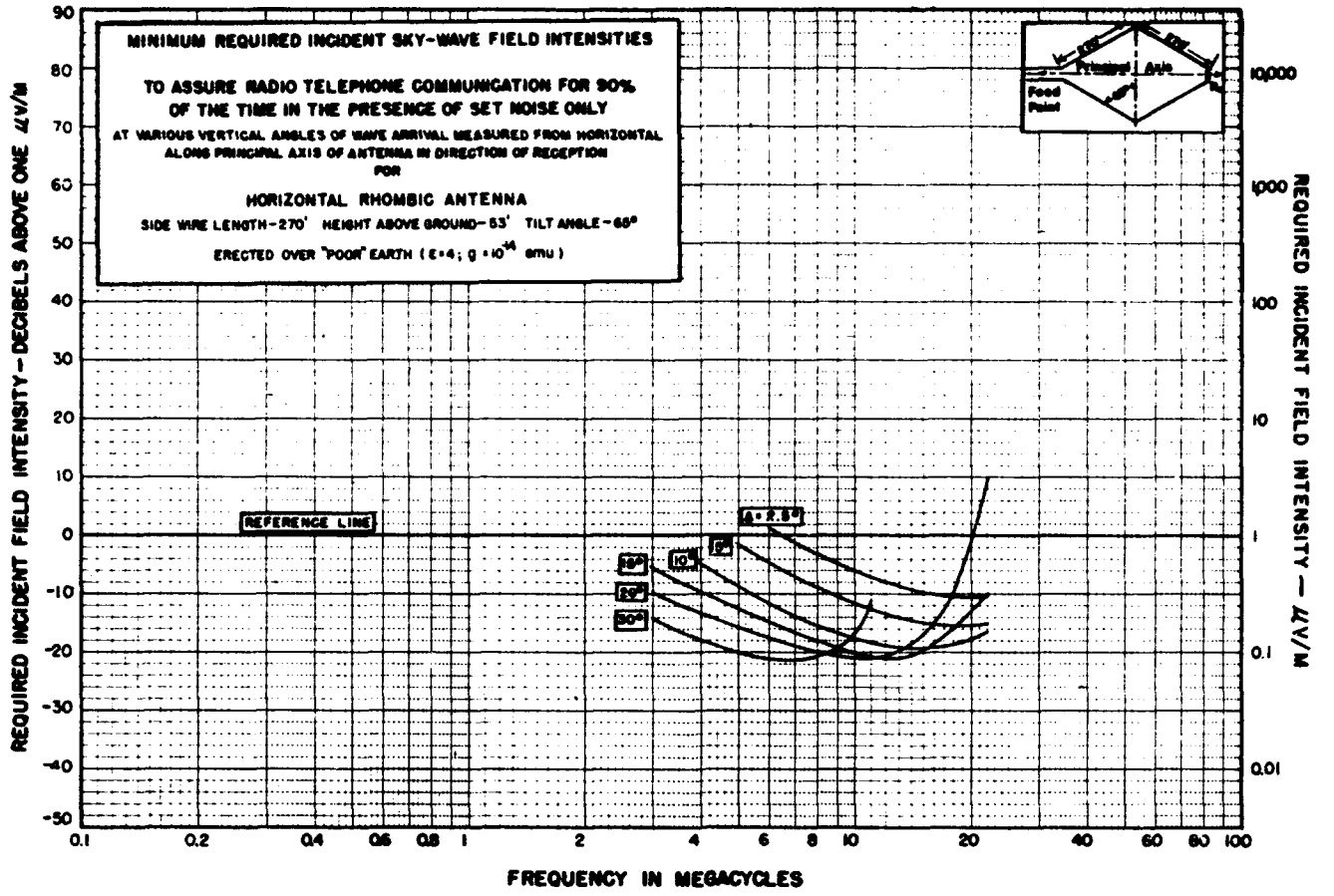


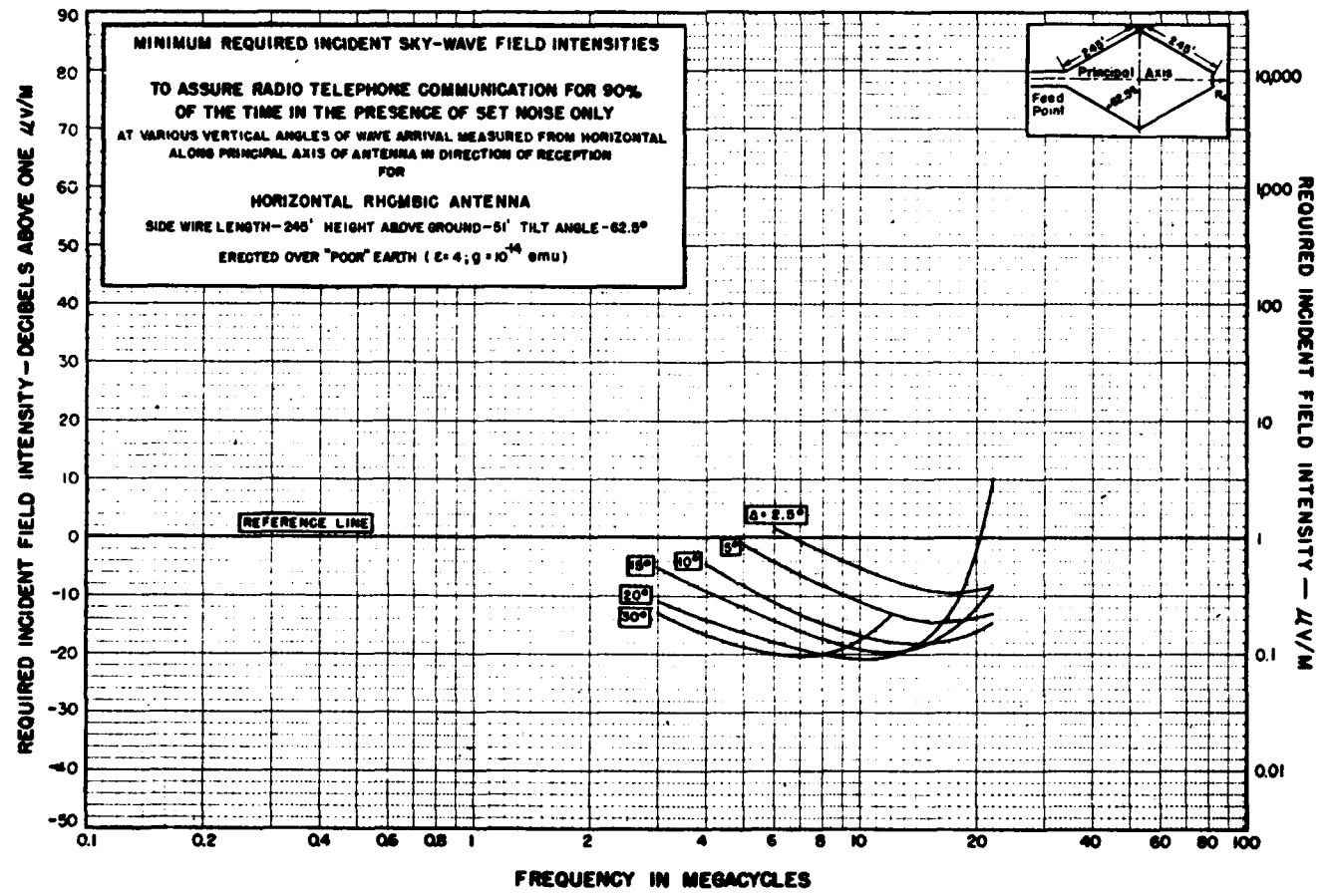
FIGURES 8.35 and 8.36.



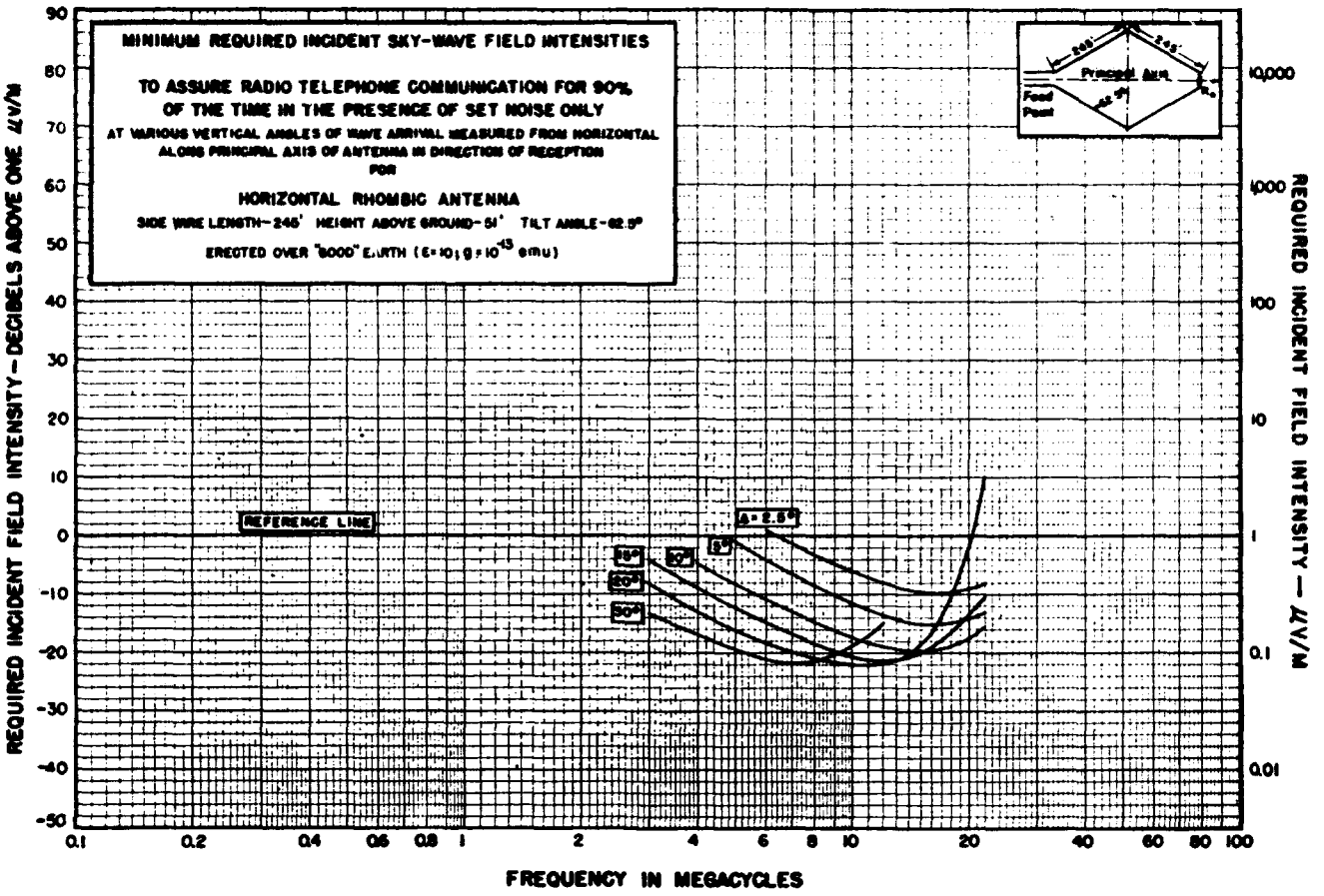


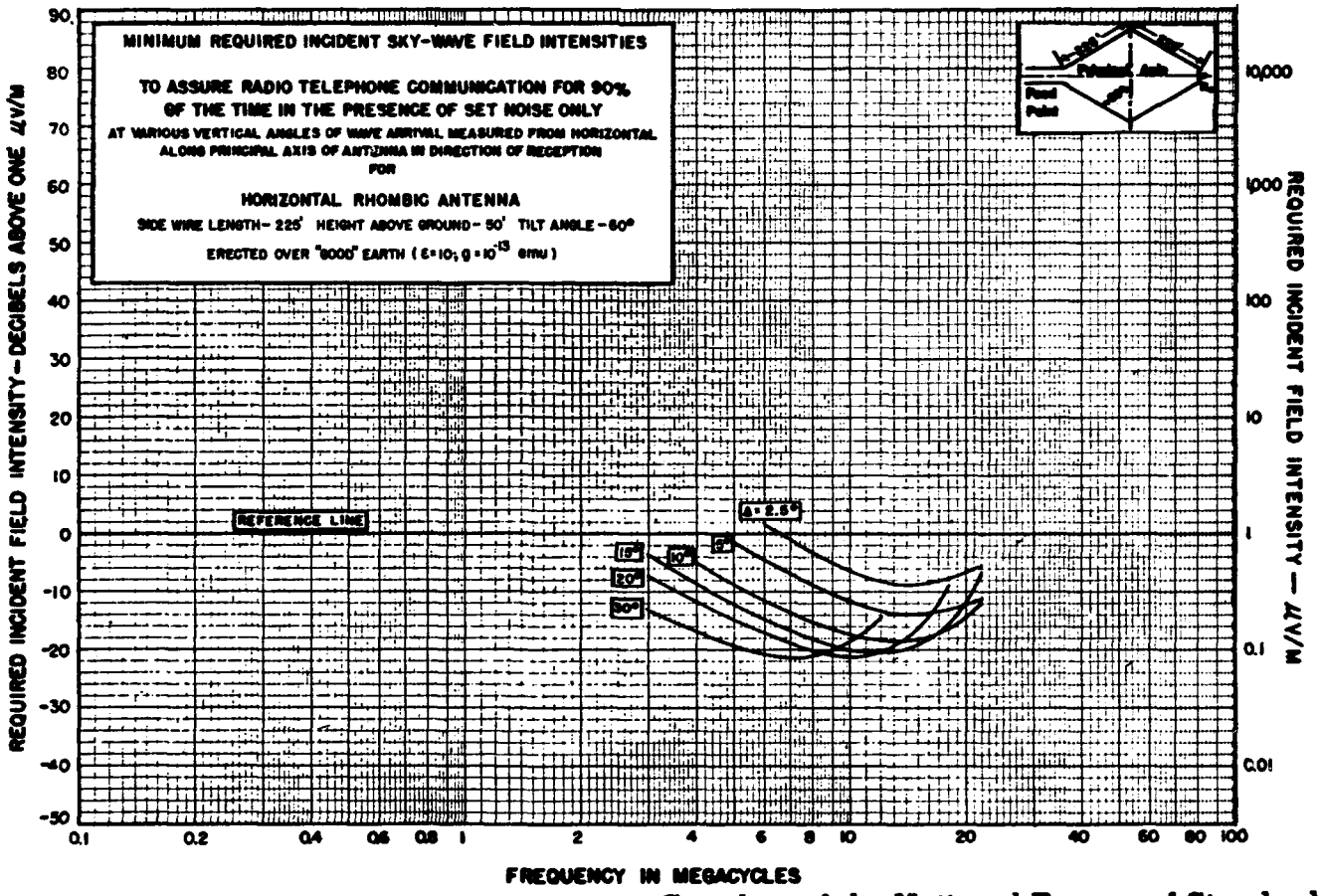
FIGURES 8.37 and 8.38.





FIGURES 8.39 and 8.40.


MINIMUM REQUIRED INCIDENT SKY-WAVE FIELD INTENSITIES
TO ASSURE RADIO TELEPHONE COMMUNICATION FOR 90% OF THE TIME IN THE PRESENCE OF SET NOISE ONLY
AT VARIOUS VERTICAL ANGLES OF WAVE ARRIVAL MEASURED FROM HORIZONTAL ALONG PRINCIPAL AXIS OF ANTENNA IN DIRECTION OF RECEPTION
FOR
HORIZONTAL RHOMBIC ANTENNA
SIDE WIRE LENGTH-245' HEIGHT ABOVE GROUND-51' TILT ANGLE-62.5°
ERECTED OVER "GOOD" EARTH (ε=10; g=10^-13 emu)
REFERENCE LINE
Δ=2.5°
5°
10°
15°
20°
30°
REQUIRED INCIDENT FIELD INTENSITY-DECIBELS ABOVE ONE μV/M
REQUIRED INCIDENT FIELD INTENSITY — μV/M
FREQUENCY IN MEGACYCLES


MINIMUM REQUIRED INCIDENT SKY-WAVE FIELD INTENSITIES
TO ASSURE RADIO TELEPHONE COMMUNICATION FOR 90% OF THE TIME IN THE PRESENCE OF SET NOISE ONLY
AT VARIOUS VERTICAL ANGLES OF WAVE ARRIVAL MEASURED FROM HORIZONTAL ALONG PRINCIPAL AXIS OF ANTENNA IN DIRECTION OF RECEPTION
FOR
HORIZONTAL RHOMBIC ANTENNA
SIDE WIRE LENGTH-225' HEIGHT ABOVE GROUND-50' TILT ANGLE-60°
ERECTED OVER "GOOD" EARTH (ε=10; g=10^-13 emu)
REFERENCE LINE
Δ=2.5°
5°
10°
15°
20°
30°
REQUIRED INCIDENT FIELD INTENSITY-DECIBELS ABOVE ONE μV/M
REQUIRED INCIDENT FIELD INTENSITY — μV/M
FREQUENCY IN MEGACYCLES

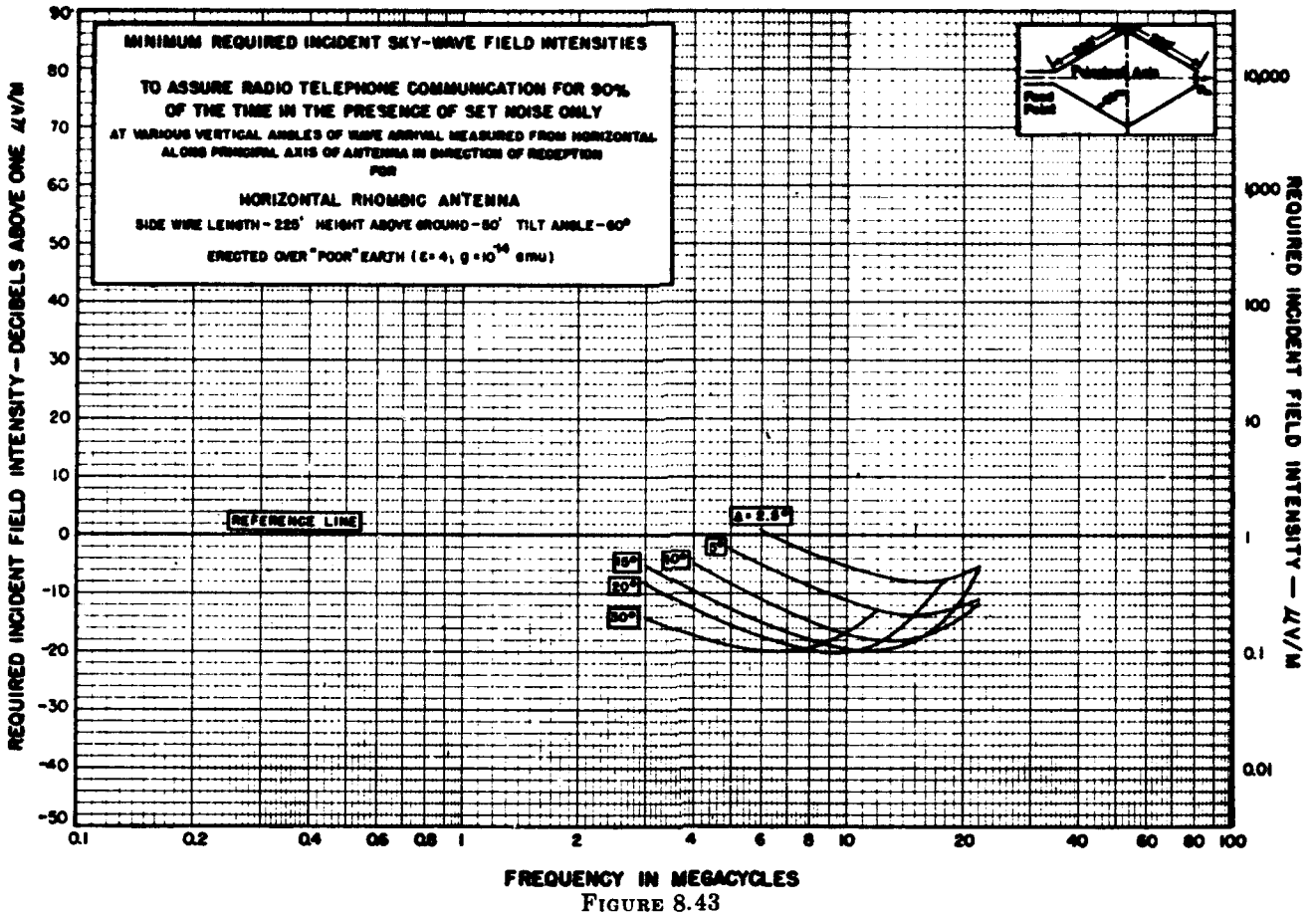


FIGURE 8.43

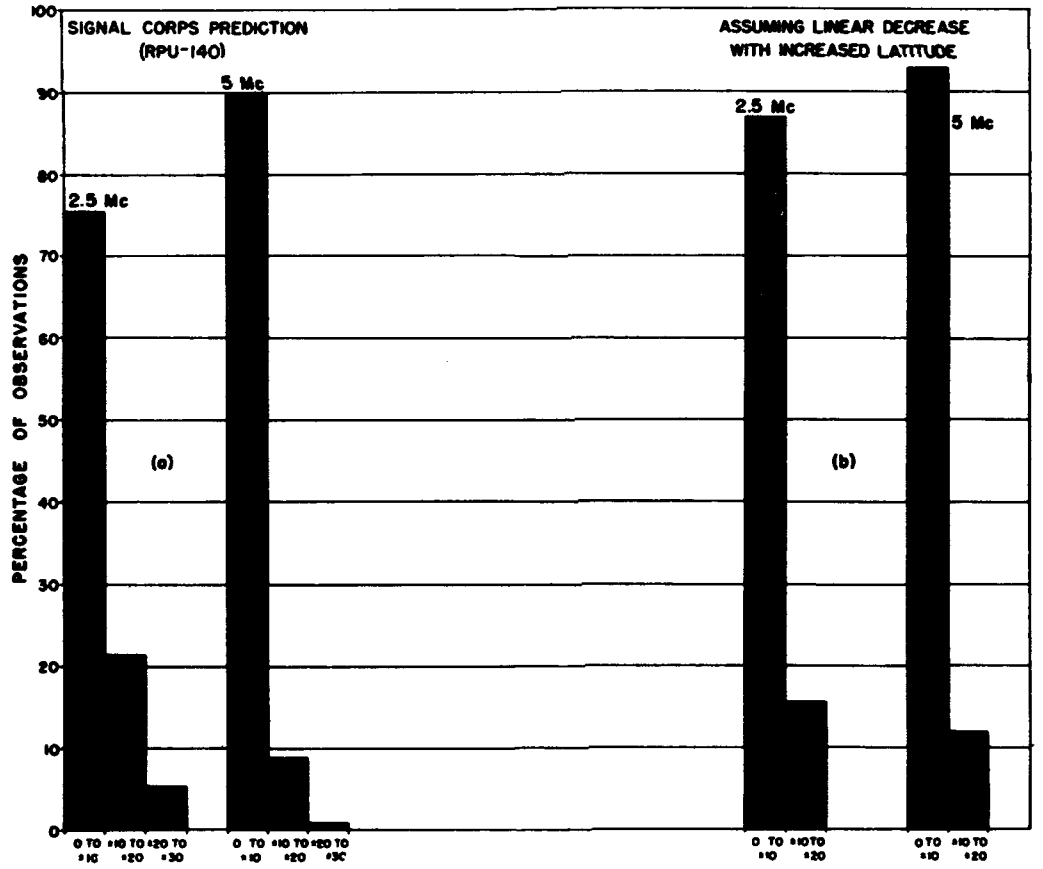


FIGURE 8.44. *Percentage of observed values deviating by a given amount from predicted values of required field intensity.*
Ratio of observed to predicted values given in 10 db class intervals. (Observations made on 2.5 and 5 mc at 17 stations for periods of 2 months to 2 years during period of sunspot minimum.)

CHAPTER 9

LOWEST REQUIRED RADIATED POWER AND LOWEST USEFUL HIGH FREQUENCY

9.1. General

The lowest required radiated power (lrrp) is the lowest radiated power that produces an incident field intensity greater than or equal to the minimum required incident field intensity discussed in the previous chapter. Determination of the lrrp is in some respects the reverse of the procedure described in chapter 7 for calculating the incident field intensity.

The lowest useful high frequency (luhf) is the lowest useful frequency for which the incident field intensity is greater than or equal to the minimum required incident field intensity.

Separate procedures for calculating the lrrp and luhf for short, intermediate, and long distances are given in sections 9.3 through 9.8. The procedures for intermediate distances make use of incident field-intensity curves (figs. 7.11 through 7.30) presented in chapter 7, and the required incident field-intensity curves (figs. 8.5 through 8.43) presented in chapter 8. The procedures for short and long distances involve the use of nomograms (figs. 9.2 to 9.7, and 9.8 to 9.19) of the type described in chapter 7 (figs. 7.10, 7.32, and 7.33) for calculation of short- and long-distance incident field intensities on which the minimum incident field intensities required to overcome atmospheric and cosmic noise presented in figures 8.5 through 8.10 have been plotted. It is to be noted that values for half-integral noise grades lie approximately halfway between curves for adjacent integral grades. Curves for grades 4 and 5 are the same for all seasons. For equinox months in grades 1, 2, and 3, use winter curves for 20, 00, and 04 local time, and summer curves for 08, 12, and 16 local time. Equinox values are to be used for all months of the year for stations between 30° N and 30° S latitude. These nomograms do not take complete account of set-noise required field intensities. The one set-noise curve drawn on each nomogram is for a half-wave dipole at medium angles of arrival. If desired, other set-noise curves may be plotted on field-intensity nomograms like those of figures 7.10, 7.32, and 7.33 and used in conjunction with the nomograms presented in this chapter.

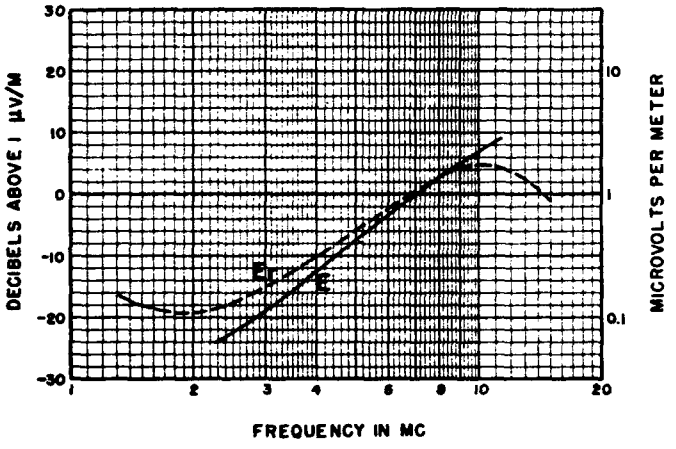


FIGURE 9.1. *Example in which the luhf is not well defined.*

E_r is the required field intensity for radiotelephone reception in noise grade 2 at 0800 local time in summer. E is the incident field intensity for 1-hop-E,1,200-km, $A=0.6$, and 3.2 w (25 db below 1 kw) equivalent radiated lower. The nominal luhf is about 8 Mc, but frequencies down to 3 or 4 MC should be nearly as serviceable.

The lrrp or luhf obtained following one of the procedures below may nominally be expected to work on 90 percent of the days of the month. This figure, however, is subject to considerable variation due to uncertainties in the values of incident and required field intensity involved, in the day-to-day variations of these quantities, and in the degree of correlation of these variations, as well as to possible errors in the user's estimate of such quantities as radiated power and noise discrimination. It is pointed out, that, as a rule, the user's errors are the least important.

The percentage of days is usually sensitive to variations in the operating frequency, but relatively insensitive to variations in quantities associated with equivalent radiated power. Thus a user who is getting only 50 percent usage with a certain power and frequency, say 500 w at 10 Mc, will find that in order to increase the percentage of use to 90 percent it is necessary to increase the power by a factor of 4 or more; whereas, the same increase in usage might be obtained by a relatively small increase in frequency, say to 12 or 13 Mc. A situation, discussed below, in which the percentage of use would be insensitive to frequency variation over a considerable range is shown in figure 9.1.

Whereas the lrrp is a quantity that always exists unambiguously, the luhf may not exist at all in the frequency range considered here, or may exist ambiguously in the sense of the above definition.

In figure 9.21(b) is illustrated the ideal situation. The curve marked noise grade 3.5 is the required field intensity, and the curve marked $A=1.56$ represents the incident field intensity plotted as a function of frequency. The curves intersect at approximately 9 Mc. Above this frequency the incident field intensity is greater than the required field intensity, while below this frequency the incident field intensity is less than the required field intensity.

In figure 9.21(d) the luhf does not exist. The incident field intensity never exceeds the required field intensity because of the muf limit.

In figure 9.21(a) also, the curves do not intersect. However, the luhf exists and is equal to the E-layer cut-off frequency at about 10 Mc.

Another situation that may arise is illustrated in figure 9.1. Here the curves intersect at about 8.5 Mc, but remain so close together below that frequency that frequencies down to 4 or 5 Mc might actually be usable.

9.2. Quantities Involved in lrrp and luhf Calculations

a. *Noise grade.* The noise grade for the receiving station is obtained from the charts presented in chapter 8, figures 8.1 to 8.4.

b. *Absorption factor* $A=JQ\overline{K}$ *or* $Ad=JQ\overline{K}d$. Refer to sections 7.5, f; 7.7 e; and 7.7, f, for details of these quantities.

c. *Modes of propagation, active modes of propagation, radiation angle, angle of arrival.* See section 7.1.

d. *Type of service factor T.* Since the required field intensities used in the calculations below are for radiotelephone service, a type of service factor T, defined as the ratio of the minimum radiated power required for radiotelephone service to that required for the desired type of service, is introduced. T factors for various types of service are listed in the last column of table 8.1. Expressed in decibels, T is equal to the negative of the type of service gain in decibels. For example, the type of service factor for manual CW telegraphy given in this table is 50 (+17 db).

e. *Noise discrimination factor Z.* When the receiving antenna discriminates either in favor of or against the signal wave with respect to the noise waves, the required field intensity for a particular type of service varies accordingly. The discrimination depends upon the vertical and horizontal angles of arrival of the signal and noise waves and upon the receiving-antenna pattern.

Very few data are available on the angles of arrival of atmospheric noise. It is thought that the bulk of such noise arrives from areas where thunderstorms are numerous, especially in equatorial regions. The noise-grade maps and the minimum-required-field-intensity curves presented in chapter 8 are in fact based on this assumption.

The horizontal angles of arrival of atmospheric noise are probably distributed more or less unevenly over a wide range of azimuths at most locations, and the distribution may vary considerably from season to season and possibly also from day to day, depending upon the prevalence and direction of nearby thunderstorms. In the absence of nearby thunderstorms, the vertical angles of arrival of most atmospheric noise are probably fairly low in the daytime (less than 20°), whereas at night a range up to 30° or 40° may not be unlikely. An average vertical angle of arrival of 20° has frequently been assumed for intermediate and low noise grades. The average angle should be somewhat higher than this at night and somewhat lower during the daytime. The average angle of arrival in the higher noise grades tends to be greater than in intermediate and low-noise grades.

In view of these uncertainties, general rules for obtaining the discrimination factor Z for atmospheric noise cannot be formulated. However, it is possible to give rules for estimating the factor in the case of receiving antennas which have no pronounced lobes, e. g., a single vertical wire, a horizontal half-wave dipole, or an inverted-L, the horizontal portion of which is not over a half wavelength long. For such antennas:

(a) If the angle of arrival of the signal wave is between 10° and 30°, take $Z=1$.

(b) If the angle of arrival of the signal wave is greater than 30° or less than 10°, take Z equal to the ratio of the antenna gain factor (as defined in chapter 7) for the angle of arrival of the signal wave to the average gain factor for vertical angles between 10° and 30° at all horizontal angles.

If cosmic noise is the limiting factor, which may be the case at the higher frequencies, take Z equal to the ratio of the antenna gain factor for the angle of arrival of the signal wave to the average gain factor for all angles of arrival.

In the case of antennas having pronounced lobes, if the signal wave is received on the major lobe, Z cannot be less than 1, but may take greater values, in some cases greater than the maximum gain factor of the antenna, depending upon the orientation of the lobes with respect to the angles of arrival of the radio waves.

f. *Gain factor G of transmitting antenna.* Discussed in section 7.2, c.

g. *Equivalent gain G′ and equivalent radiated power P′.* Defined by the formulas

$$P'=G'P_0,\ G'=TZG,$$

where T, Z, and G are the type of service factor, noise-discrimination factor, and the transmitting-antenna gain factor, respectively, and P_0 is the total radiated power (section 7.2, a). Note that both Z and G may vary with the mode of propagation.

h. *Equivalent distance d′ for long paths.* The distances labeled on the left-hand scale of the nomograms for long paths are valid only if the equivalent radiated power P' is 1 kw. For other values of P' the equivalent distance d' is located on this scale by means of the auxiliary P' scale printed alongside each nomogram as follows:

1. Mark the point corresponding to the path length d on the distance scale.

2. Aline the edge of a piece of paper with the auxiliary P' scale. Make one mark opposite 1.0 kw and another mark opposite the calculated value of P'.

3. Without rotating the paper, move it over to the distance scale and aline the 1.0-kw mark with the point d located in step 1.

4. Put a mark on the distance scale opposite the second mark on the paper. This locates the point d' to be used with the nomogram for calculations of luhf.

To obtain P', having found d', as in calculating the lrrp, follow the above procedure in reverse.

9.3. Calculation of lrrp for Short Paths, 0 to 400 km

1. Determine the atmospheric-noise grade for the location of the receiving station.

2. Calculate $A=JQ\overline{K}$ for the desired local time at the receiving station. Note that only certain local times are represented in the nomograms (figs. 9.2 to 9.7). If the desired local time coincides with one of these, the lrrp may be obtained directly from the nomogram. Otherwise, calculate A for several of the local times represented in the nomograms, obtain the lrrp for each of these times, and read off the lrrp for the desired time from a smooth curve drawn through the calculated values.

3. Plot A on the right-hand scale of the nomogram corresponding to the local time for which A was calculated.

4. At the desired frequency mark the appropriate noise-grade curve, or the set-noise curve, whichever is uppermost. Remember that the set noise curve is correct only for a half-wave dipole receiving antenna at medium angles.

5. Lay a straightedge between the points marked in steps 3 and 4.

6. Read P' at the intersection of the straightedge with the left-hand scale of the nomogram.

7. Determine the noise discrimination factor Z and the transmitting antenna gain factor G for each active mode of propagation. Only 1-hop-E and 1-hop-$F2$ need be considered, and usually only one of these is active.

8. Calculate P_0 for each active mode by substituting P', T, Z, and G in

$$P_0=P'/TZG.$$

The lower value so obtained is the desired lrrp. Note that P_0 is in watts if P' is read in watts.

Example: Find the lrrp for sky-wave reception of manual CW signals transmitted on 2.2 Mc over a 100-km path in the vicinity of Manila, P. I., at 1200 local time in June 1947, assuming $Z=1$ (0 db) and $G=0.5$ (−3 db). The atmospheric-noise grade obtained from figure 8.3 is 4. $A=1.56$ (see example worked out in section 7.7, b. The appropriate nomogram is that of figure 9.5. Assuming set noise can be neglected, atmospheric noise (curve labeled 4) controls the required field intensity at 2.2 Mc. Make a mark at 2.2 Mc on this curve, mark $A=1.56$ on the right-hand scale, and lay a straightedge between these two marks. Extended to the left, the straightedge intersects the P' scale at 400 w, which is 4 db below 1 kw. This is the required equivalent radiated power. For this problem, $T=50$ (+17 db). Thus the required lrrp is $P_0=P'/TZG=16$ w. (Using decibels, the lrrp $=-4-17-0+3=-18$ db referred to 1 kw.)

9.4. Calculation of luhf for Short Paths, 0 to 400 km

1. Determine the atmospheric noise grade for the receiving station.

2. Calculate P' for the radiation angle corresponding to one of the modes of propagation. (Only one-hop-$F2$ and one-hop-E need be considered). If P' is a function of frequency, calculate P' for several frequencies in the range where the luhf is expected to fall, and draw a smooth curve of P' as a function of frequency in this range.

3. Calculate $A=JQ\overline{K}$ for the desired local time at the receiving station. Note that only certain local times are represented in the nomograms (figs. 9.2 to 9.7). If the desired local time coincides with one of these, the luhf may be obtained directly from the nomogram. Otherwise, calculate A for several of the local times represented in the nomograms, obtain the luhf for each of these times, and read off the luhf for the desired time from a smooth curve drawn through the calculated values.

4. Select the nomogram corresponding to the local time for which A was calculated.

5. Plot A on the right-hand scale, and P' on the left-hand scale of the nomogram. If P' varies with frequency, plot the value corresponding to a frequency near the middle of the range selected in step 2.

6. Lay a straightedge between A and P'. Note the frequency corresponding to the intersection of the straightedge with the appropriate atmospheric-noise grade curve, or the set-noise curve, whichever is greater. Remember that the set noise curve is correct only for a half-wave dipole receiving antenna at medium angles.

7. If the frequency found in step 6 is not the same as the frequency to which P' corresponds, repeat step 6 by using the value of P' corresponding to the new frequency. Continue this process until the luhf is the same as the frequency for which P' was calculated. This frequency is the

correct luhf for the mode of propagation considered, provided the mode is active at this frequency.

8. Repeat steps 2 through 7 for the other mode of propagation. The lower of the two luhf thus found is the desired luhf for the path.

Example: Find the luhf for sky-wave reception of radiotelephone signals transmitted over a 100-km path in the vicinity of Manila, P. I., at 1200 local time in June 1947, assuming $P_0=25$ w, $G=0.5$ (−3 db), and $Z=1$ (0 db) irrespective of the mode of propagation involved. The noise grade is 4. The nomogram used is figure 9.5. For radiotelephone, $T=1$ (0 db). Substituting in $P'=P_0TZG$. we have $P'=12.5$ w. A straightedge is laid between $P'=12.5$ w on the left-hand scale of the nomogram, and $A=1.56$ on the right-hand scale (A was calculated in section 7.7, b.). The luhf is read at the intersection of the straightedge with the atmospheric-noise curve for grade 4, assuming set noise may be neglected. The luhf thus obtained is 3.4 Mc. Reference to figures 7.2 and 7.3 shows that this frequency is below the E-layer cut-off for F2-layer propagation and is therefore propagated by the E layer only.

9.5. Calculation of lrrp for Intermediate Paths, 400 to 3,200 km

1. Determine the noise grade for the receiving location.

2. Select the required field-intensity chart corresponding to the desired local time at the receiving station, figures 8.5 through 8.10. Also consult the appropriate set noise chart, figures 8.12 through 8.43 or a similar chart prepared for the antenna used if it does not correspond to one of those represented in these figures. At the desired frequency, read the field intensities for cosmic noise, for set noise, and for atmospheric noise of the grade found in step 1. The required incident field intensity is the highest of these. If the desired local time is not close to one of the local times for which the charts of figures 8.5 through 8.10 were constructed, read the required field intensity for several local times on the charts before and after the desired local time, and read off the required field intensity for the desired local time from a smooth curve drawn through these values.

3. Calculate $A=JQ\overline{K}$ for the path and desired local time at the receiving station.

4. Determine which modes of propagation are active at the time and frequency selected.

5. Select the incident field-intensity chart (figs. 7.11 through 7.30) corresponding to the path length and one of the active modes of propagation, and read the value of the field intensity corresponding to A and the desired frequency. If the path length is not close to one of the distances for which the charts were constructed, read the field intensity for the nearest chart distance less than and greater than the path length and interpolate linearly to find the field intensity for the desired distance.

6. As the field intensity determined above is for 1-kw effective radiated power, the equivalent radiated power $P'=TZGP_0$ required is equal to the square of the ratio of the required field intensity E_r found in step 2 to the incident field intensity E determined in step 5. The total radiated power P_0 in kilowatts required for this mode is therefore given by

$$P_0=(E_r/E)^2/TZG.$$

7. Find P_0 for each active mode of propagation. The least of these is the desired lrrp.

Frequently the lrrp is desired for several local times or a period of time. The lrrp should then be calculated for each of the local times represented in the charts, plotted versus local time, and a smooth curve drawn through the points. The lrrp for any local time may than be read from the curve.

Example: Find the lrrp for reception of manual CW signals transmitted on 13 Mc from Washington, D. C., to Miami, Fla., at 1200 local time in Miami in June 1947, if $Z=0.3$ (−5 db) for 1-hop-E, $Z=1$ (0 db) for other modes, $G=0.2$ (−7 db) for 1-hop-E, and $G=0.5$ (−3 db) for other modes. Both Z and G are less for 1-hop-E because of the low radiation and arrival angles involved (approximately 5°). It is further assumed that set noise may be neglected. For manual CW, $T=50$ (17 db). Referring to figure 8.3, the noise grade at Miami is approximately 3.5. Referring to figure 8.7, the required field intensity E_r for radiotelephone reception is 5.6μv/m (15 db above 1 μv/m), assuming set noise may be neglected. In the example worked out in section 7.7, c, only 1-hop-$F2$, and 1-hop-E were found to be active at 13 Mc. Incident field intensities for 1-kw effective radiated power for these modes were found to be 22.4 and 6.3 μv/m (27 db and 16 db above 1 μv/m), respectively. The lrrp for 1-hop-E need not be calculated, since it obviously is higher than that for 1-hop-$F2$. The desired lrrp is therefore that for 1-hop-$F2$ and is $P_0=(5.6/22.4)^2/50\times1\times0.5=0.0025$ kw$=2.5$ w. (Using decibels, the lrrp$=15-27-17-0+3=-26$ db, referred to 1 kw).

9.6. Calculation of luhf for Intermediate Paths, 400 to 3,200 km

1. Determine the noise grade for the receiving station.

2. Calculate $A=JQ\overline{K}$ for the path and desired local time at the receiving station. Note that only certain local times are represented in the required-field-intensity charts. If the desired local

time does not coincide with one of these, read step 12.

3. Calculate P' for one mode of propagation. If P' varies with frequency, calculate values at several frequencies covering the range in which the luhf should fall.

4. Figure 9.20 is a chart for plotting P' as a function of frequency. Lay a piece of transparent paper over this grid. Draw a horizontal line corresponding to $P'=1$ kw (0 db) and a vertical line at 1 Mc. Then plot P' on the overlay as a function of frequency for one mode of propagation, using the scales on the grid seen through the overlay. The values of P' calculated in step 3 are to be used. Draw a smooth curve through the plotted points. It is not necessary to copy the scales on the overlay. Note that P' increases downward on this chart.

5. Transfer the overlay to the required field-intensity chart (figs. 8.5 through 8.10), which has the curve for the noise grade found in step 1 and the local time selected in step 2. Adjust the overlay so that the 1-Mc lines coincide. Now slide the overlay vertically up or down until the P' curve intersects the reference line on the chart at one or more frequencies. Note the points on the appropriate required-field-intensity curve corresponding to these frequencies, i. e., the atmospheric noise curve or the cosmic noise curve, whichever is higher. Mark these points on the overlay. Slide the overlay up or down until the P' curve intersects the reference line at one or more other frequencies. Mark the points on the required-field-intensity curve corresponding to these frequencies. Repeat until enough points are plotted on the overlay to define a curve. Draw a smooth curve through the points.

6. Next, transfer the overlay to the appropriate set noise chart, figures 8.12 through 8.43, or a similar chart prepared for the antenna used, if it does not correspond to one of those represented in these figures. Adjust the P' curve to the reference line at the same frequencies as in step 5. Wherever the set noise curve exceeds that drawn in step 5, mark the set noise points and redraw the curve so that it passes through these points. The resulting curve is the required field-intensity curve for reception in the presence of atmospheric, cosmic, and set noise. It was necessary to adjust for variation in P' because the overlay will now be used with incident field-intensity curves drawn for 1 kw.

7. Repeat steps 5 and 6 for each desired local time. Curves for different local times may be plotted on the same overlay and labeled appropriately.

8. Transfer the overlay to the incident field-intensity chart (figs. 7.11 through 7.30), appropriate to the path length and mode of propagation. Adjust the overlay so that the 1-Mc lines coincide and the $P'=1$ kw (0 db) line on the overlay coincides with the reference line on the chart. The frequency read at the intersection of the required-field-intensity curve for the desired local time with the incident-field-intensity curve corresponding to the value of A calculated for this local time in step 2 is the luhf, provided the mode is active at this frequency. If the calculated value of A does not equal one of the values for which curves are plotted, an interpolated curve should of course be drawn or imagined. In the case of $F2$-layer modes, if the frequency determined above is less than the E-layer cut-off frequency, the cut-off frequency is the luhf.

9. Repeat steps 4 through 6 for each mode of propagation, preparing a separate overlay for each mode having a different P' curve to avoid confusion of curves.

10. The lowest of the luhf thus determined is the desired luhf for the path.

11. If the path length is not equal or nearly equal to one of the distances for which the incident field-intensity charts were constructed, follow steps 8, 9, and 10 for the nearest chart distance less and greater than the path length, and interpolate linearly between the two values of luhf thus obtained to obtain the luhf for the desired path length.

12. If the desired local time is not close to one of the local times for which the required field-intensity curves are drawn, obtain the luhf for the nearest local time before and after the desired local and interpolate. A better method is to calculate the luhf for several of the given local times. The luhf for any desired local time may then be read off a smooth curve drawn through the calculated luhf's.

Example: Find the luhf for reception of manual CW signals transmitted from Washington, D. C., to Miami, Fla., at 1200 local time in Miami in June 1947, assuming $P_o=400$ w (4 db less than 1 kw), and the same values of Z and G assumed in the problem worked out in the previous section. For simplicity, Z and G are further assumed to be independent of frequency. Also set noise is assumed negligible, so that step 6 in the above procedure may be omitted. For manual CW, $T=+17$ db. Thus for 1-hop-E, P' is found to be 1.26 kw (1 db above 1 kw). For other modes, $P'=10$ kw (10 db above 1 kw). For this path, $A=1.56$ (see example worked out in section 7.7, c). The path length is 1,500 km. Therefore, incident field-intensity charts for 1,200 km and 1,600 km, figures 7.18 through 7.25, chapter 7, are selected. The appearance of the two overlays constructed in step 5 of the above procedure when superposed on the incident field-intensity charts for 1,600 km is illustrated in figure 9.21.

In figure 9.21 (a) the curves do not intersect because of the E-layer cut-off at approximately

10 Mc. The cut-off frequency is therefore the 1,600-km luhf for 1-hop-$F2$. In figure 9.21 (b) the curves intersect at approximately 9 Mc, which is the 1,600-km luhf for 2-hop-$F2$. In figure 9.21 (c) the curves intersect at approximately 13 Mc, the 1,600-km luhf for 1-hop-E. In figure 9.21 (d) the curves do not intersect because of the muf limit. Consequently, the luhf does not exist for 2-hop-E under these conditions.

Results for 1,200 km (not illustrated) are 1-hop-$F2$ luhf is still the cut-off frequency; 2-hop-$F2$ luhf=8.0 Mc; 1-hop-E luhf=11.5 Mc; 2-hop-E luhf does not exist.

The 1,500-km luhf for 1-hop-$F2$ is the 1,500-km cut-off frequency=9.5 Mc. Interpolating linearly between the 1,200- and 1,600-km luhfs, the 1,500 luhfs for 2-hop-$F2$ and 1-hop-E are approximately 8.7 Mc and 12.5 Mc. The desired luhf for 1,500 km is the lowest of these, namely, 8.7 Mc.

9.7. Calculation of lrrp for Distances Over 3,200 km

The procedure is the same as for short distances, except that $Ad=JQ\overline{K}d$ is calculated and plotted on the right-hand scale of the nomogram (figs. 9.8 through 9.19) appropriate to the magnitude of Ad, d' is read on the left-hand scale, and P' is found by using the auxiliary scale on the left as instructed in section 9.2, h.

Example: Find the lrrp for reception of manual CW signals transmitted on 15 Mc from Washington, D. C., to Trieste, Italy, at 1200 local time in Trieste in June 1947, assuming $Z=1$ (0 db) and $G=4$ (6 db). Referring to figure 8.3, the noise grade at Trieste is 3. Since $Ad=7.6$ (see example worked out in section 7.7, d), the appropriate nomogram is figure 9.17. Assuming set noise may be neglected, the required field intensity at 15 Mc is controlled by atmospheric noise (curve labeled 3). Make a mark on this curve at 15 Mc, mark $Ad=7.6$ on the right-hand scale, and lay a straightedge between these two marks. Now mark the intersection of the straightedge with the d' scale on the left. The distance d for this path is 7,110 km. Following the procedure of section 9.2, h, P' is approximately 6.4 kw. Now $T=50$ (+17 db). Therefore, $P_0=P'/TZG=6.4/50\times1\times4=0.032$ kw=32 w. (Using decibels, the lrrp=8−17−0−6=−15 db referred to 1 kw).

9.8. Calculation of luhf for Distances Over 3,200 km

The procedure is the same as for short distances, except that $Ad=JQ\overline{K}d$ is calculated and plotted on the right-hand scale of the nomogram (figs. 9.8 through 9.19) appropriate to the magnitude of Ad, d' is located on the left-hand scale as instructed in section 9.2, h, and the straightedge is laid between d' and Ad.

Example: Find the luhf for reception of manual CW signals transmitted from Washington, D. C., to Trieste, Italy, at 1200 local time in Trieste in June 1947, assuming $P_0=250$ w, $G=4$ (6 db), and $Z=1$ (0 db). The noise grade is 3 (fig. 8.3). Since $Ad=7.6$ (worked out for example in section 7.7, d), the appropriate nomogram is figure 9.17. For manual CW operation $T=50$ (17 db). Substituting in $P'=P_0TZG$, obtain $P'=250\times50\times1\times4=50$ kw or 17 db above 1 kw. For this path $d=7{,}100$ km. Following the procedure of Section 9.2, h, locate d' on the left-hand scale, $A=7.6$ on the right-hand scale, and lay a straightedge between these points. The luhf is read at the intersection of the straightedge with the atmospheric noise curve labeled 3, assuming set noise may be neglected. The luhf so obtained is 12.8 Mc. Reference to CRPL-D31, or the example worked out in section 6.6, e, shows that the $F2$-owf is above 12.8 Mc for the path. Were the owf below 12.8 Mc, the expected utility of 12.8 Mc would be less than the usual 90 percent.

9.9. Paths Passing Through the Auroral Zone

If the transmission path lies wholly or in part in one of the auroral belts shown in figure 7.48, refer to section 7.7, g, for discussion of propagation effects and modification of $\overline{K}$ and $\overline{K}d$ used in the above calculations.

9.10. References

Calculation of sky-wave field intensities, maximum usable frequencies, and lowest useful high frequencies. RPU-144, Radio Propagation Unit Technical Report No. 6, March 1947. Prepared under the direction of the Chief Signal Officer by the Radio Propagation Unit, (9463d TSU), Helabird Signal Depot, Baltimore, Md. (Available at Office of Technical Services, Department of Commerce, Washington 25, D. C.)

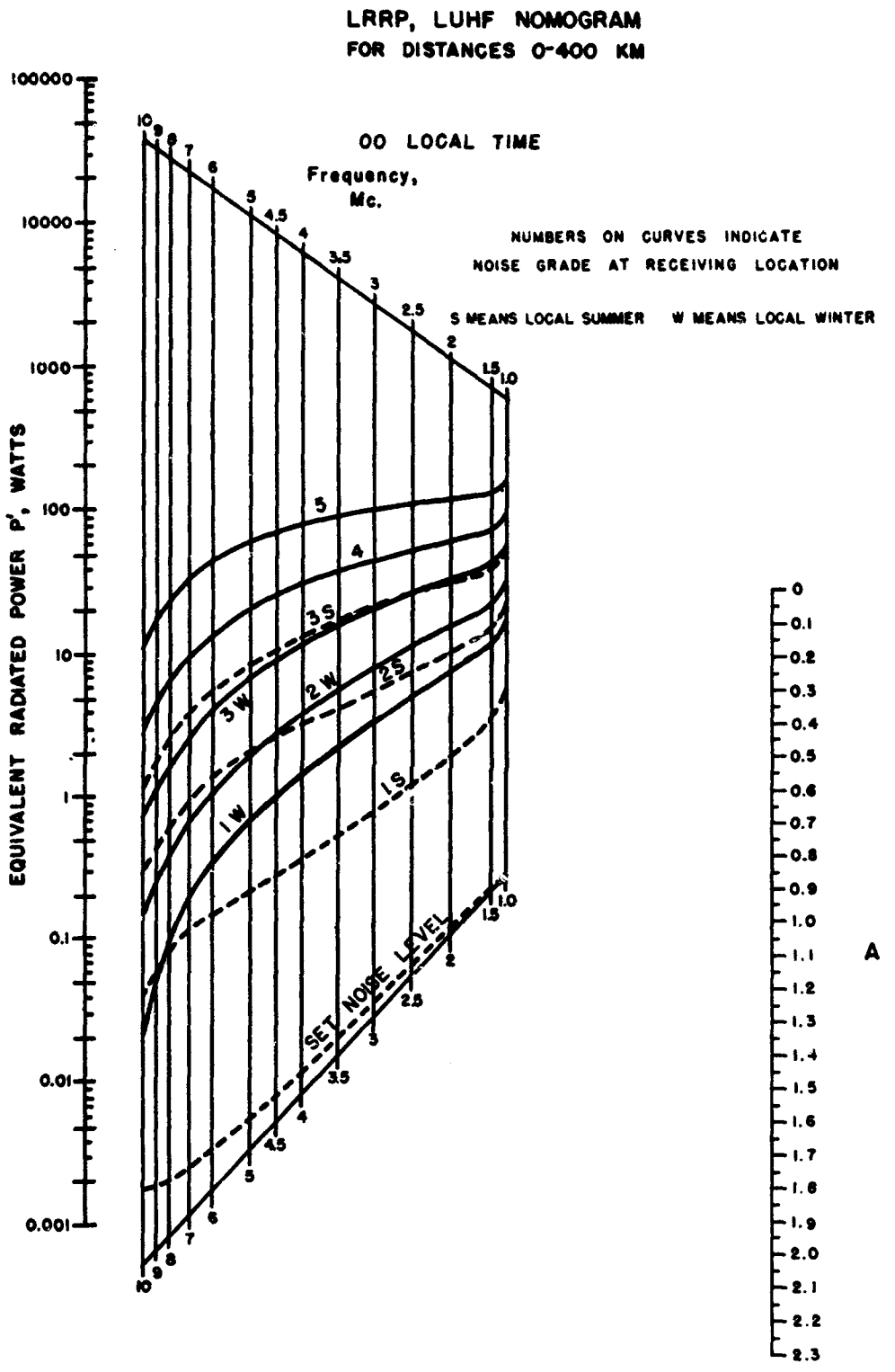

LRRP, LUHF NOMOGRAM
FOR DISTANCES 0-400 KM
00 LOCAL TIME
Frequency, Mc.
NUMBERS ON CURVES INDICATE
NOISE GRADE AT RECEIVING LOCATION
S MEANS LOCAL SUMMER W MEANS LOCAL WINTER
EQUIVALENT RADIATED POWER P', WATTS
100000
10000
1000
100
10
1
0.1
0.01
0.001
10 9 8 7 6 5 4.5 4 3.5 3 2.5 2 1.5 1.0
5
4
3S
3W
2W
2S
1W
1S
SET NOISE LEVEL
A
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2.0 2.1 2.2 2.3

FIGURE 9.2.

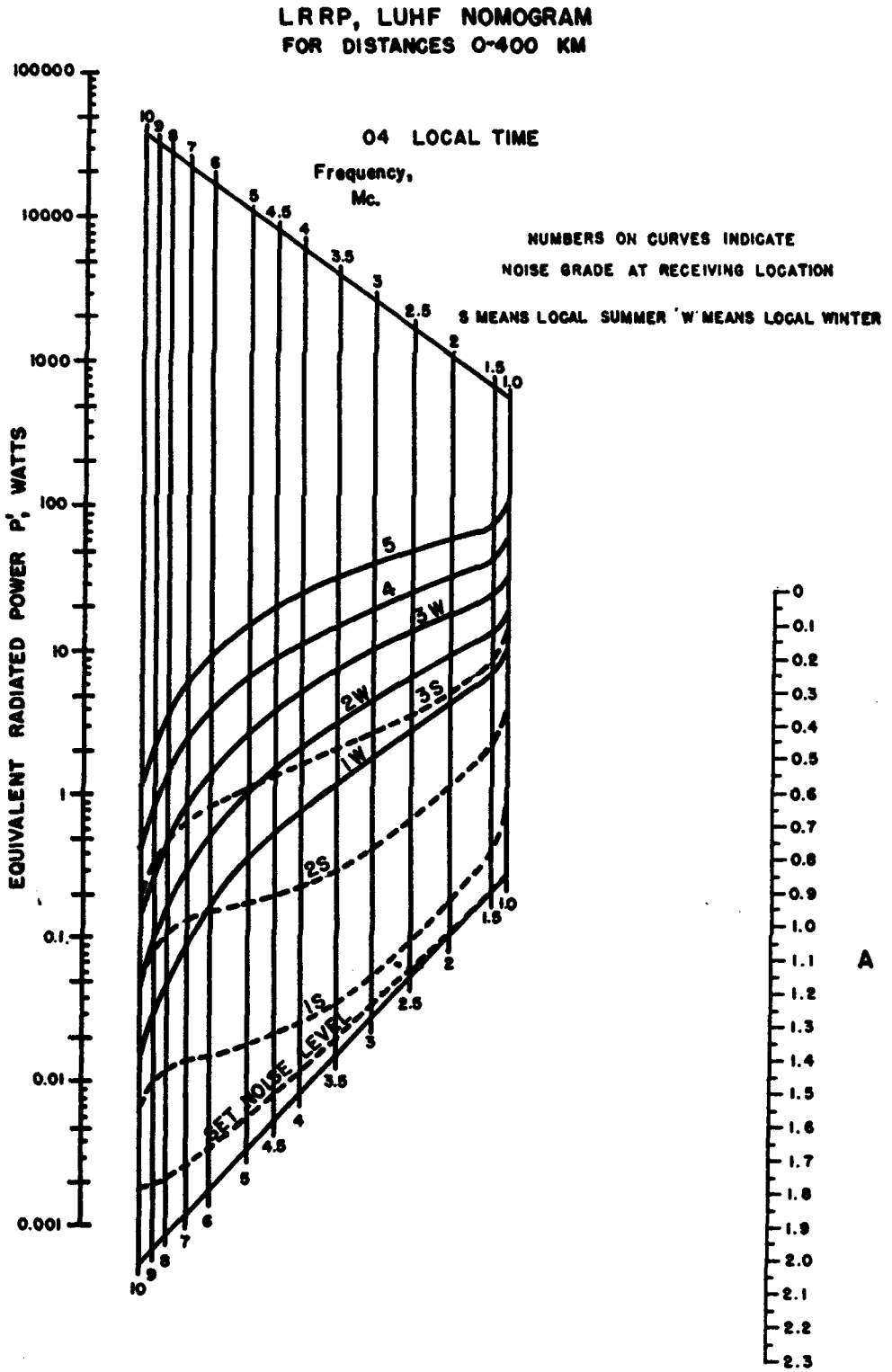

LRRP, LUHF NOMOGRAM
FOR DISTANCES 0-400 KM
04 LOCAL TIME
Frequency, Mc.
NUMBERS ON CURVES INDICATE
NOISE GRADE AT RECEIVING LOCATION
S MEANS LOCAL SUMMER 'W' MEANS LOCAL WINTER
EQUIVALENT RADIATED POWER P', WATTS
100000
10000
1000
100
10
1
0.1
0.01
0.001
5
4
3W
2W
1W
3S
2S
1S
REF. NOISE LEVEL
A
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
1.3
1.4
1.5
1.6
1.7
1.8
1.9
2.0
2.1
2.2
2.3

FIGURE 9.3.

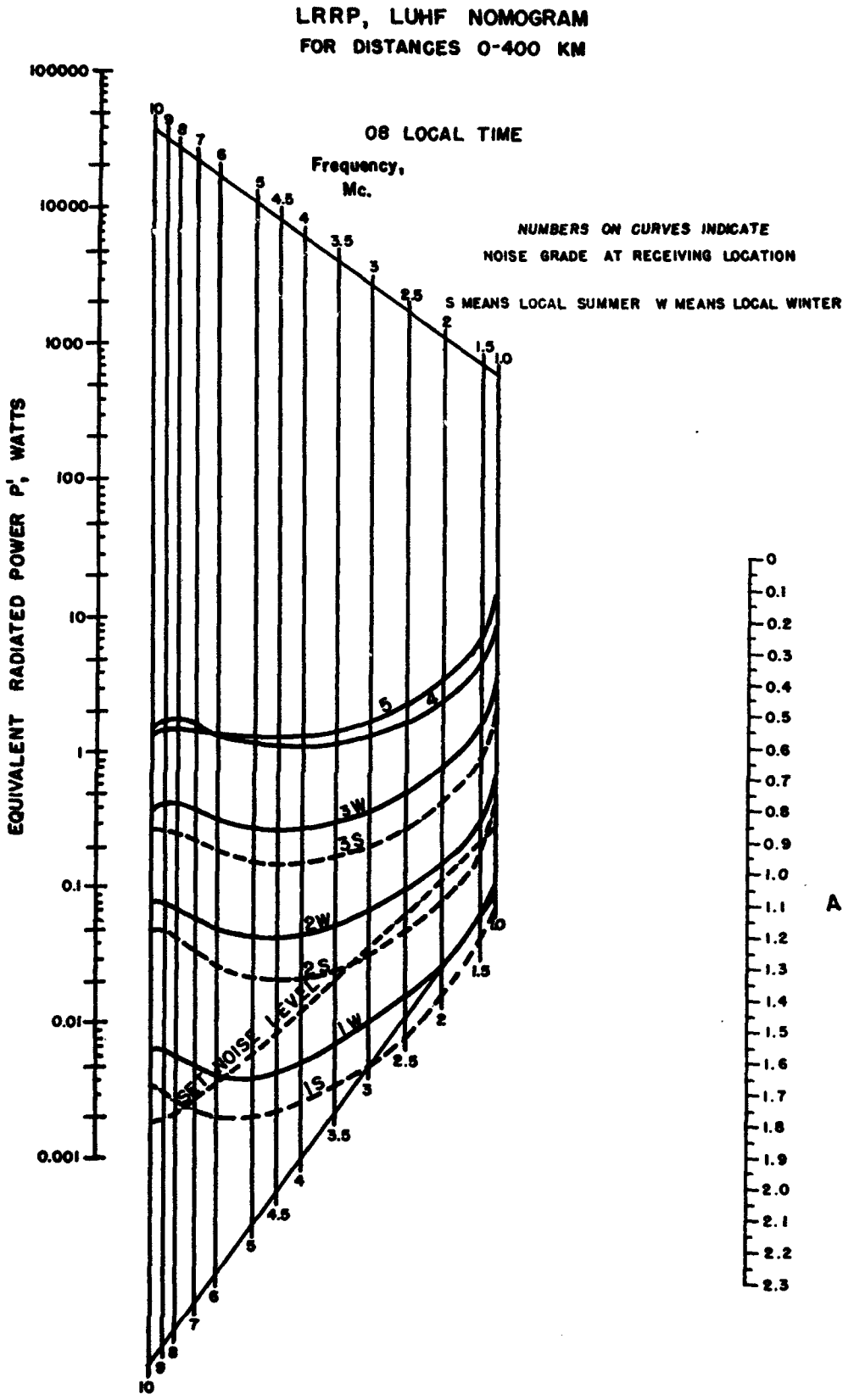

LRRP, LUHF NOMOGRAM
FOR DISTANCES 0-400 KM
08 LOCAL TIME
Frequency, Mc.
NUMBERS ON CURVES INDICATE
NOISE GRADE AT RECEIVING LOCATION
S MEANS LOCAL SUMMER W MEANS LOCAL WINTER
EQUIVALENT RADIATED POWER P', WATTS
100000
10000
1000
100
10
1
0.1
0.01
0.001
10
9
8
7
6
5
4.5
4
3.5
3
2.5
2
1.5
1.0
5
4
3W
3S
2W
2S
1W
1S
SET NOISE LEVEL
A
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
1.3
1.4
1.5
1.6
1.7
1.8
1.9
2.0
2.1
2.2
2.3

FIGURE 9.4.

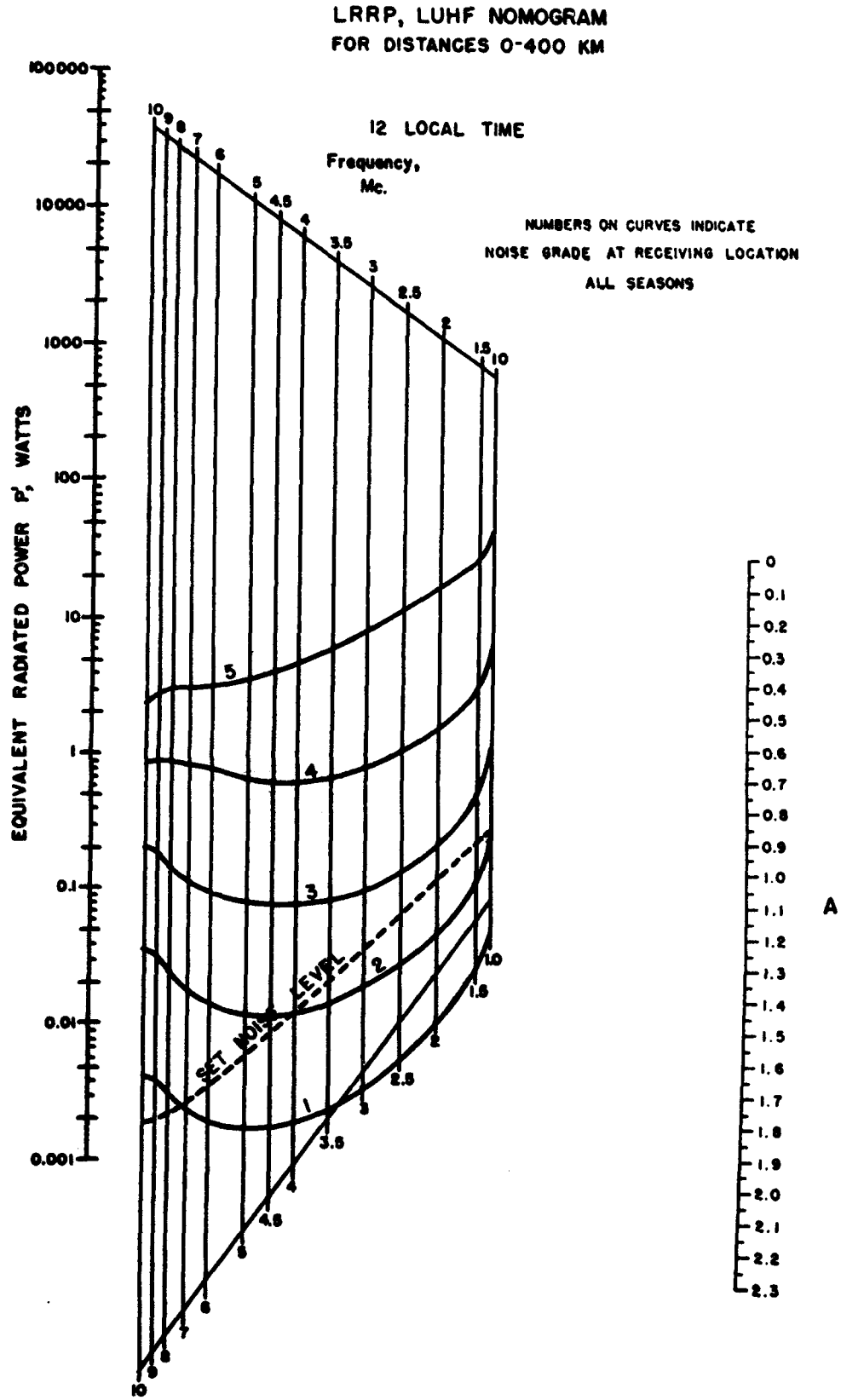


FIGURE 9.5.

935333 O - 51 - 13

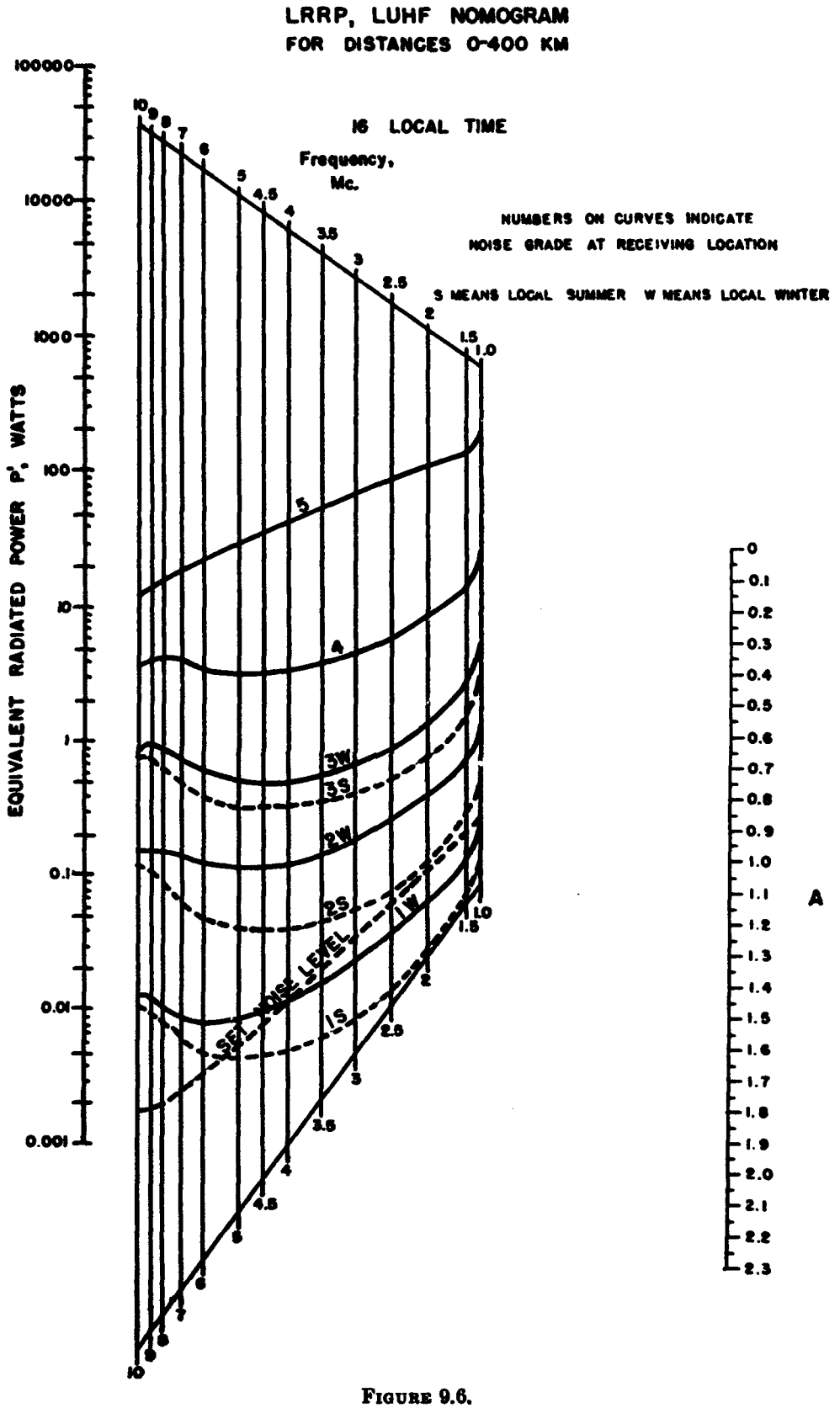

LRRP, LUHF NOMOGRAM
FOR DISTANCES 0-400 KM
16 LOCAL TIME
Frequency, Mc.
NUMBERS ON CURVES INDICATE
NOISE GRADE AT RECEIVING LOCATION
S MEANS LOCAL SUMMER W MEANS LOCAL WINTER
EQUIVALENT RADIATED POWER P', WATTS
100000
10000
1000
100
10
1
0.1
0.01
0.001
3W
3S
2W
2S
1W
1S
SEE NOISE LEVEL
A
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
1.3
1.4
1.5
1.6
1.7
1.8
1.9
2.0
2.1
2.2
2.3

FIGURE 9.6.

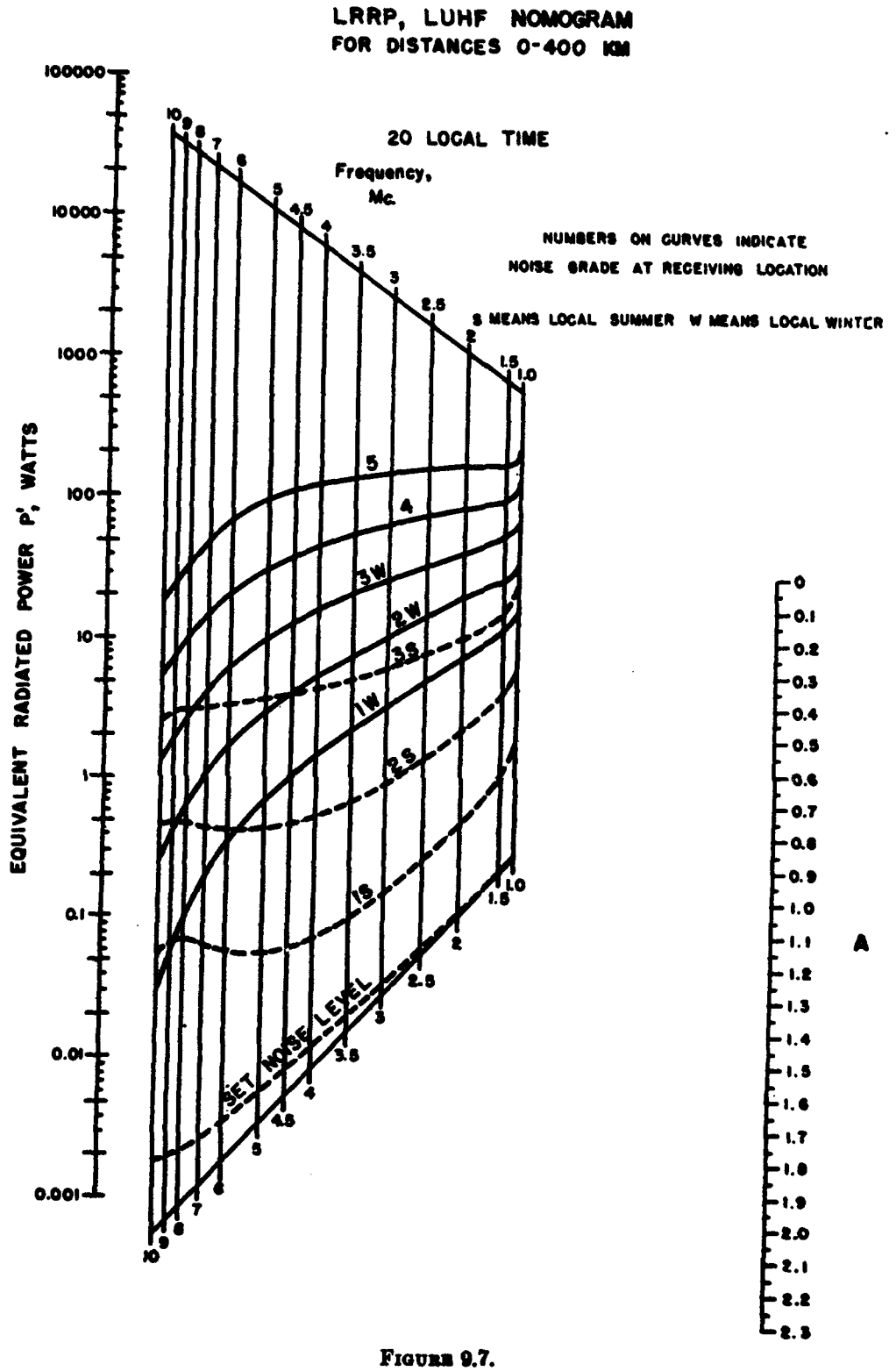

LRRP, LUHF NOMOGRAM
FOR DISTANCES 0-400 KM
20 LOCAL TIME
Frequency, Mc.
NUMBERS ON CURVES INDICATE
NOISE GRADE AT RECEIVING LOCATION
S MEANS LOCAL SUMMER W MEANS LOCAL WINTER
EQUIVALENT RADIATED POWER P', WATTS
100000
10000
1000
100
10
1
0.1
0.01
0.001
10
9
8
7
6
5
4.5
4
3.5
3
2.5
2
1.5
1.0
5
4
3W
2W
3S
1W
2S
1S
SET NOISE LEVEL
A
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
1.3
1.4
1.5
1.6
1.7
1.8
1.9
2.0
2.1
2.2
2.3

FIGURE 9.7.

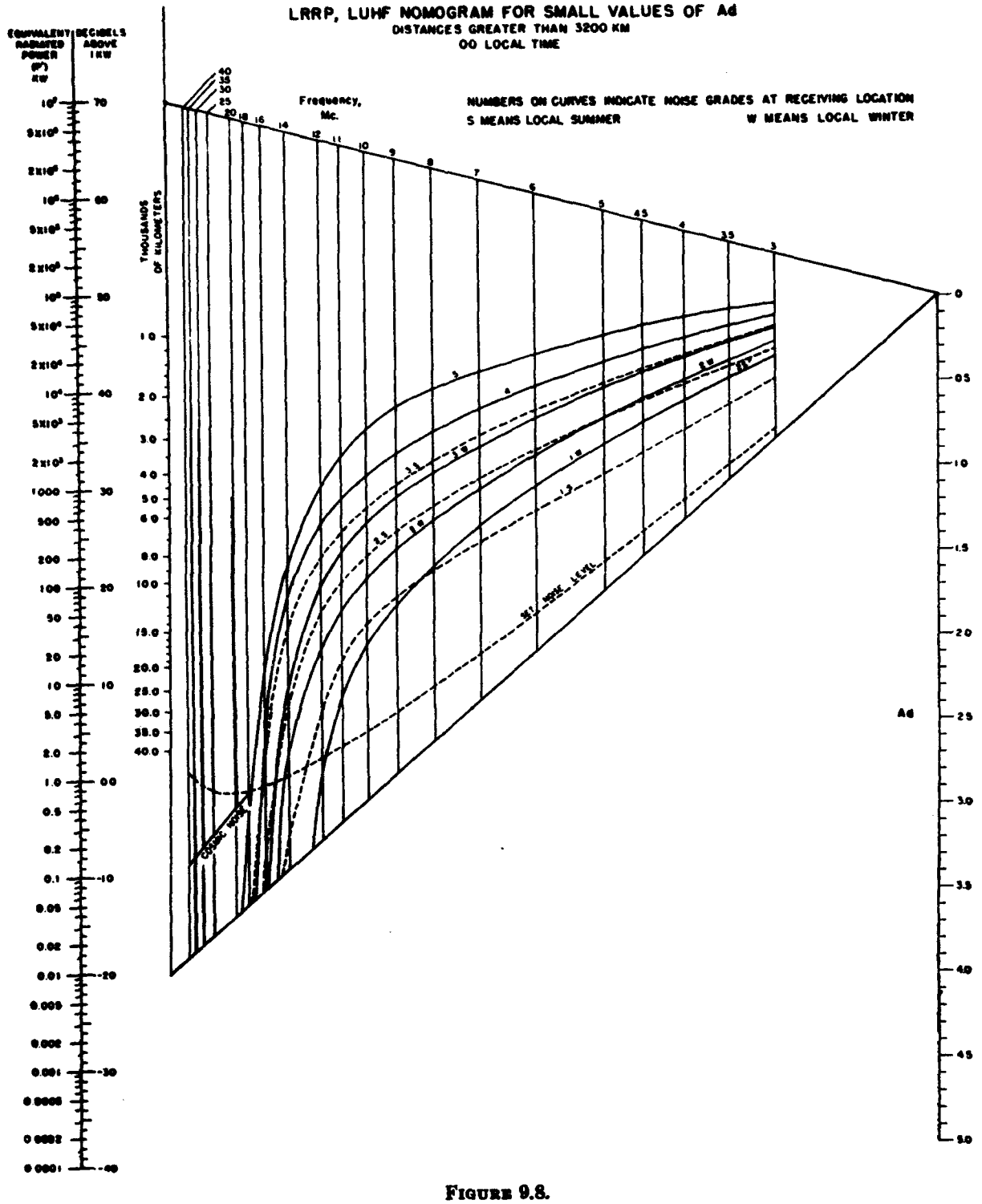

LRRP, LUHF NOMOGRAM FOR SMALL VALUES OF Ad
DISTANCES GREATER THAN 3200 KM
00 LOCAL TIME
EQUIVALENT RADIATED POWER (P') KW
DECIBELS ABOVE 1 KW
Frequency, Mc.
NUMBERS ON CURVES INDICATE NOISE GRADES AT RECEIVING LOCATION
S MEANS LOCAL SUMMER
W MEANS LOCAL WINTER
THOUSANDS OF KILOMETERS
Ad

FIGURE 9.8.

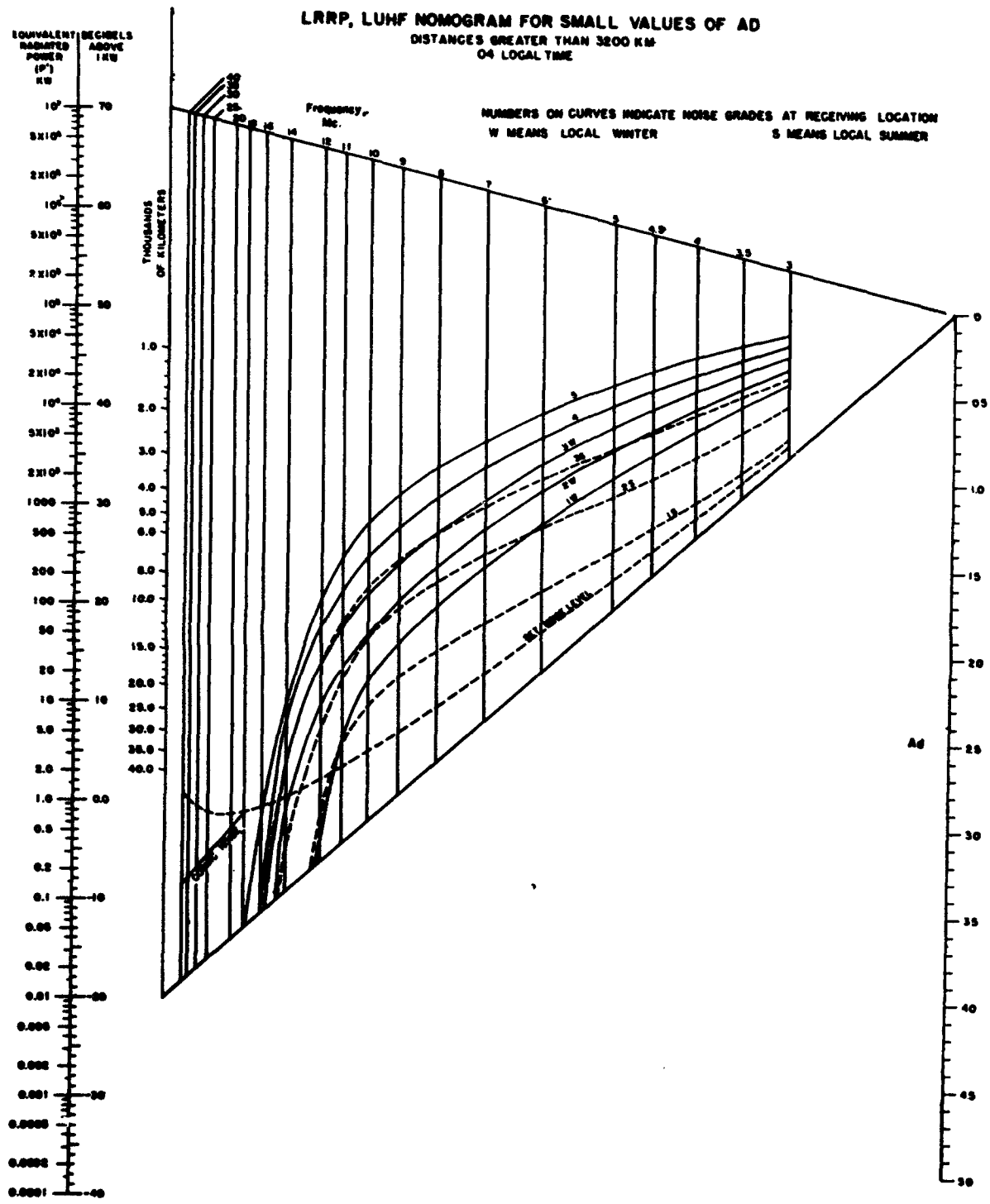

LRRP, LUHF NOMOGRAM FOR SMALL VALUES OF AD
DISTANCES GREATER THAN 3200 KM
04 LOCAL TIME
NUMBERS ON CURVES INDICATE NOISE GRADES AT RECEIVING LOCATION
W MEANS LOCAL WINTER
S MEANS LOCAL SUMMER
EQUIVALENT RADIATED POWER (P') KW
DECIBELS ABOVE 1 KW
THOUSANDS OF KILOMETERS
Frequency, Mc.
Ad

FIGURE 9.9.

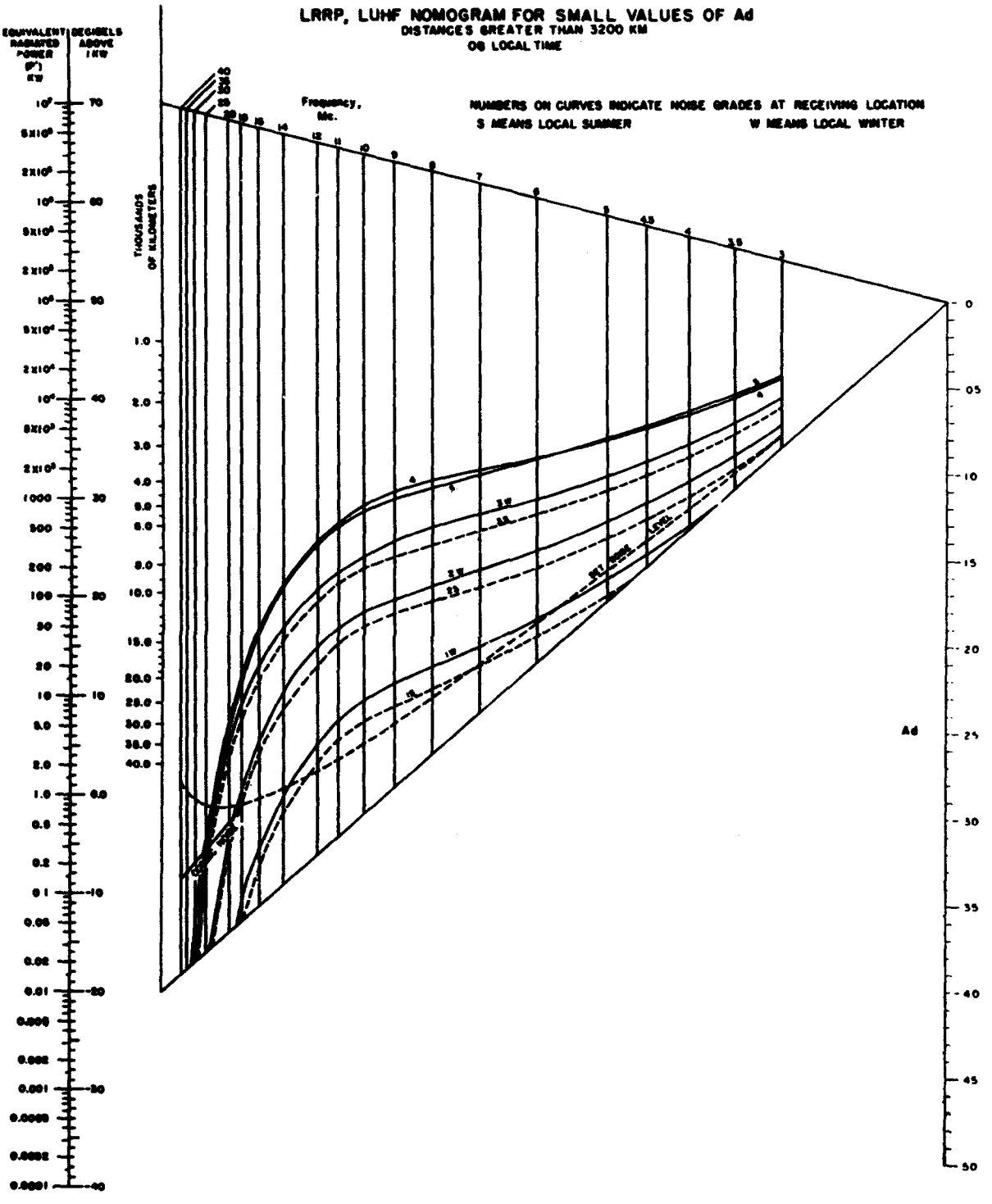

LRRP, LUHF NOMOGRAM FOR SMALL VALUES OF Ad
DISTANCES GREATER THAN 3200 KM
OB LOCAL TIME
EQUIVALENT RADIATED POWER (P') KW
DECIBELS ABOVE 1 KW
THOUSANDS OF KILOMETERS
Frequency, Mc.
NUMBERS ON CURVES INDICATE NOISE GRADES AT RECEIVING LOCATION
S MEANS LOCAL SUMMER
W MEANS LOCAL WINTER
Ad

FIGURE 9.10.

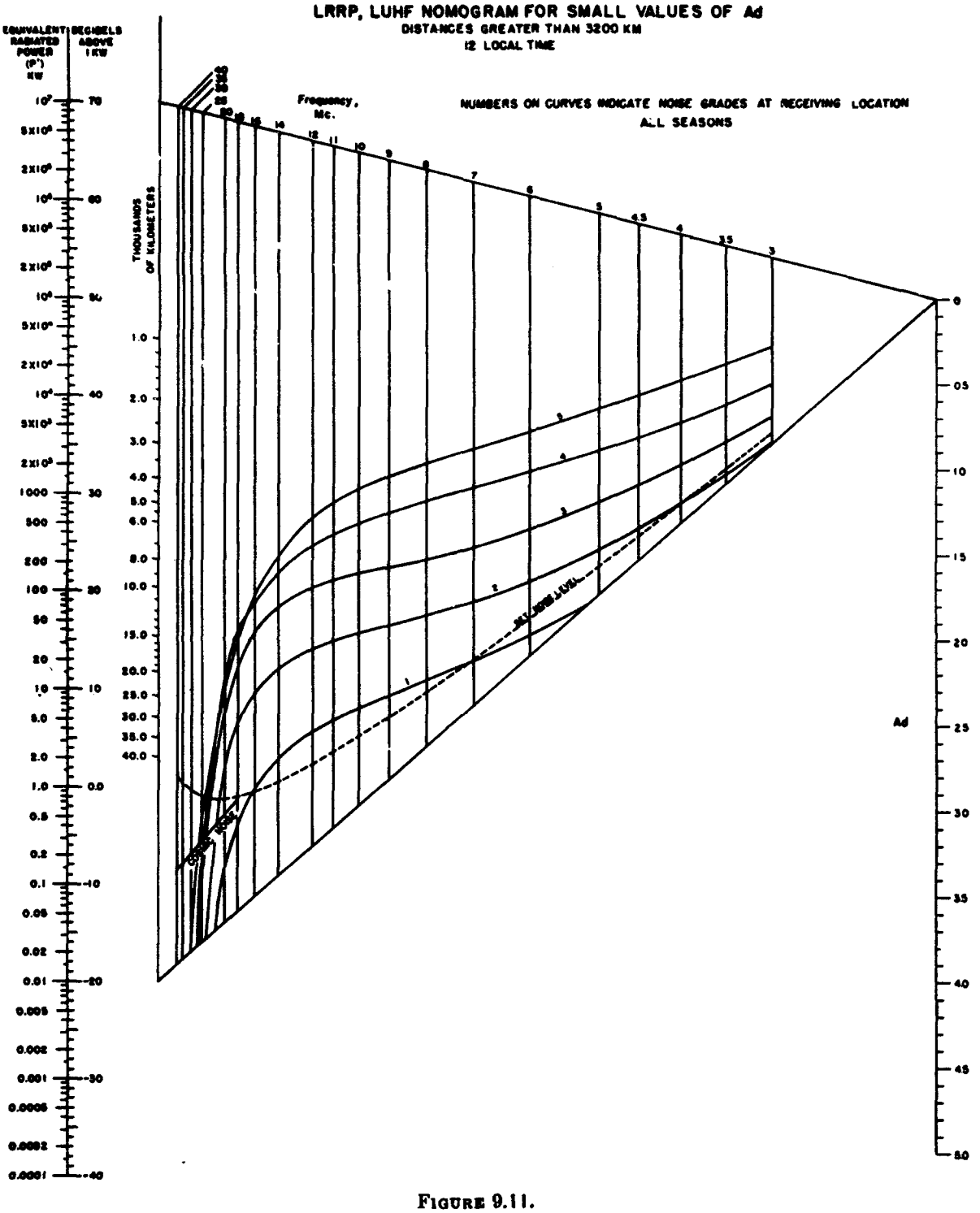

LRRP, LUHF NOMOGRAM FOR SMALL VALUES OF Ad
DISTANCES GREATER THAN 3200 KM
12 LOCAL TIME
NUMBERS ON CURVES INDICATE NOISE GRADES AT RECEIVING LOCATION
ALL SEASONS
EQUIVALENT RADIATED POWER (P') KW
DECIBELS ABOVE 1 KW
Frequency, Mc.
THOUSANDS OF KILOMETERS
Ad

FIGURE 9.11.

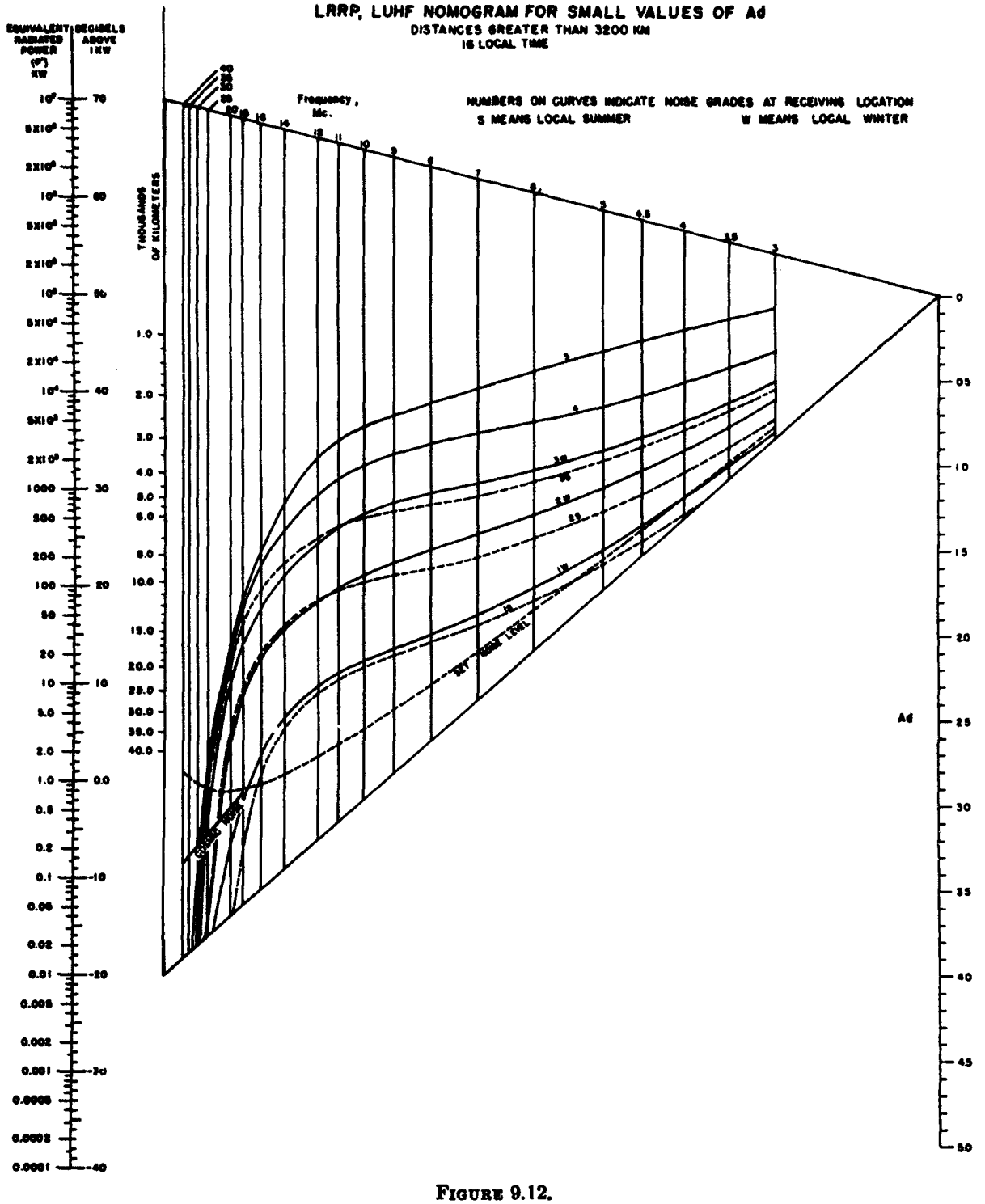

LRRP, LUHF NOMOGRAM FOR SMALL VALUES OF Ad
DISTANCES GREATER THAN 3200 KM
16 LOCAL TIME
EQUIVALENT RADIATED POWER (P') KW
DECIBELS ABOVE 1 KW
THOUSANDS OF KILOMETERS
Frequency, Mc.
NUMBERS ON CURVES INDICATE NOISE GRADES AT RECEIVING LOCATION
S MEANS LOCAL SUMMER
W MEANS LOGAL WINTER
SKY WAVE LEVEL
Ad

FIGURE 9.12.

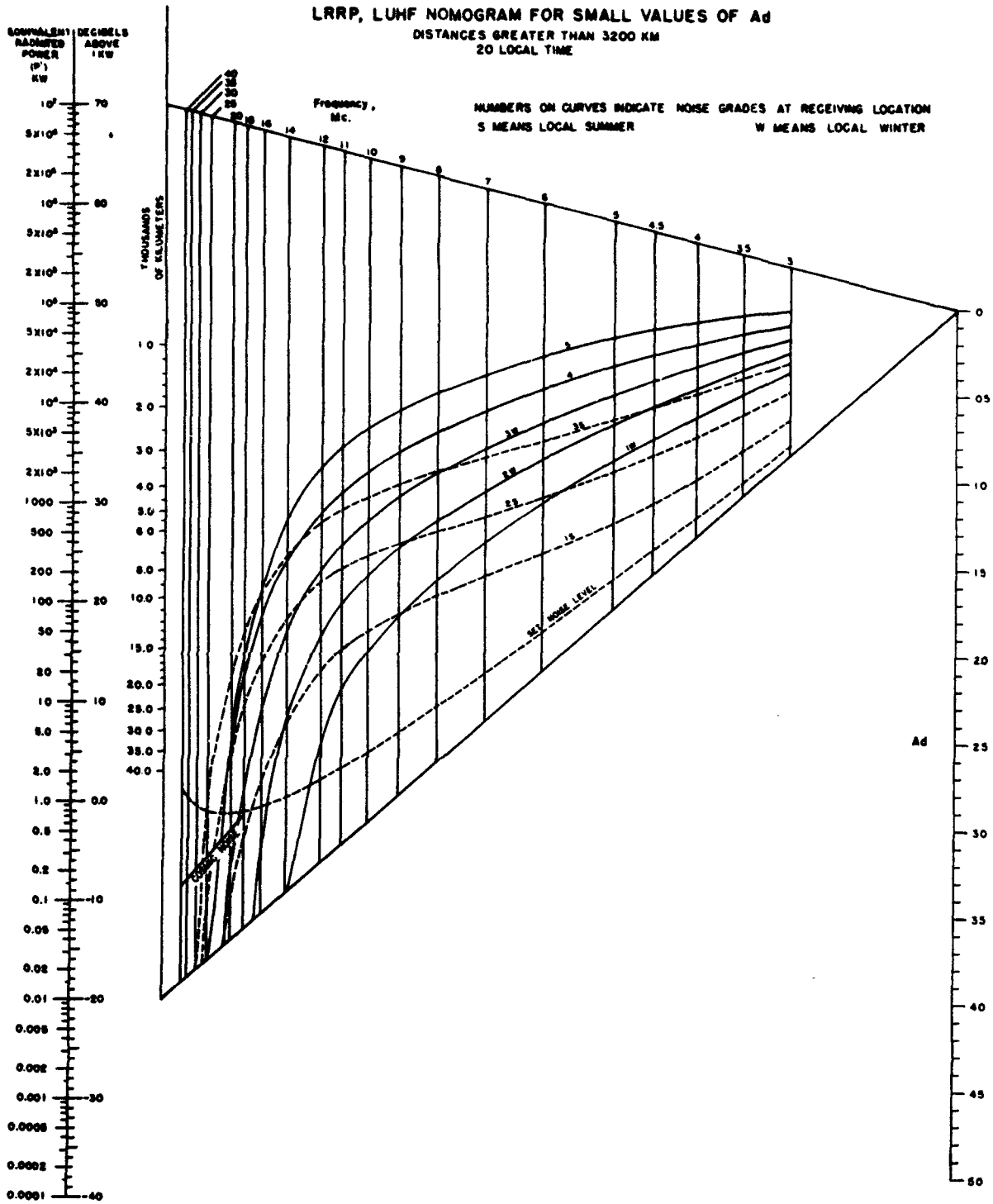

LRRP, LUHF NOMOGRAM FOR SMALL VALUES OF Ad
DISTANCES GREATER THAN 3200 KM
20 LOCAL TIME
Frequency, Mc.
NUMBERS ON CURVES INDICATE NOISE GRADES AT RECEIVING LOCATION
S MEANS LOCAL SUMMER
W MEANS LOCAL WINTER
THOUSANDS OF KILOMETERS
DECIBELS ABOVE 1 KW
Ad

FIGURE 9.13

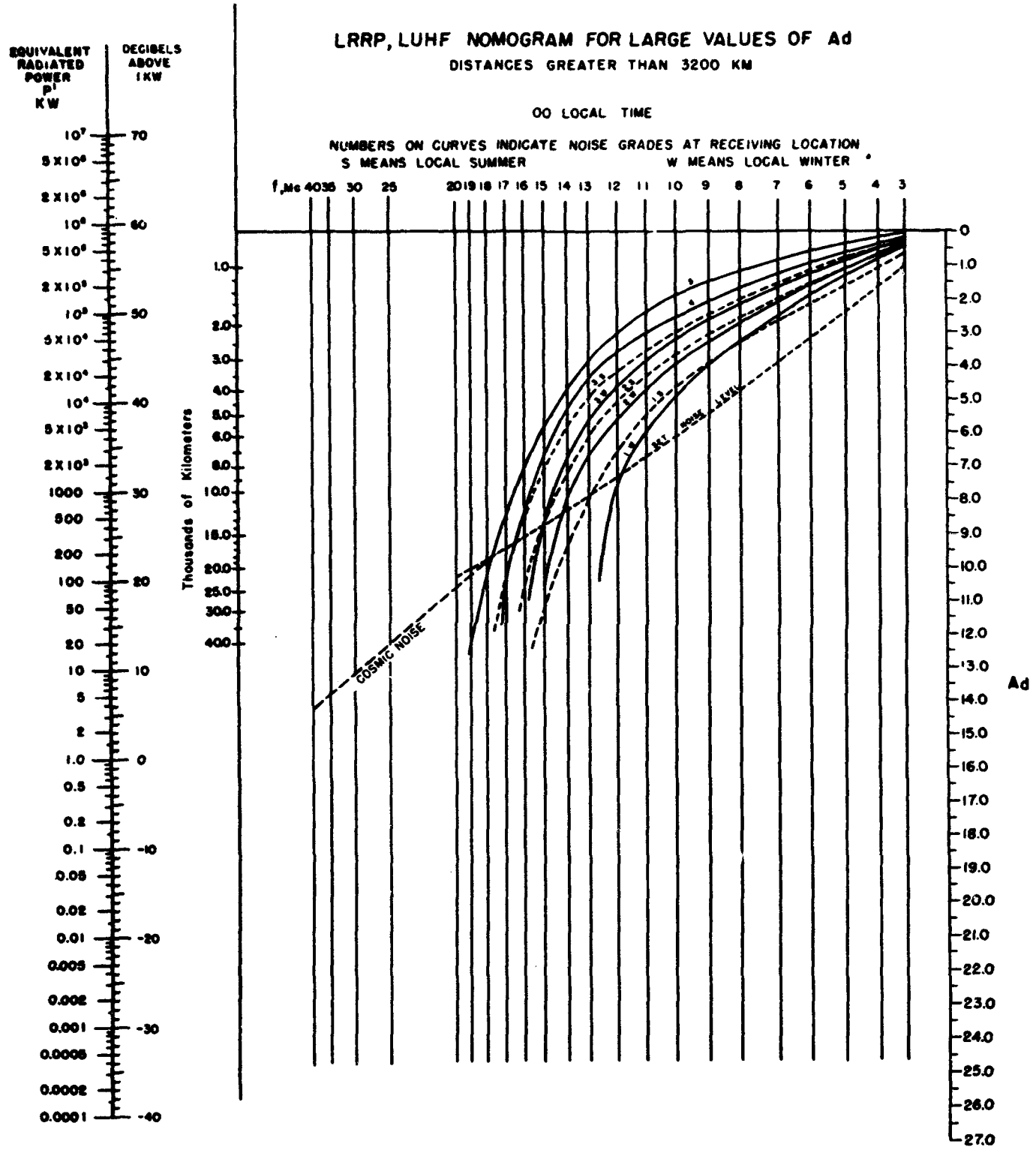

LRRP, LUHF NOMOGRAM FOR LARGE VALUES OF Ad
DISTANCES GREATER THAN 3200 KM
00 LOCAL TIME
NUMBERS ON CURVES INDICATE NOISE GRADES AT RECEIVING LOCATION
S MEANS LOCAL SUMMER
W MEANS LOCAL WINTER
EQUIVALENT RADIATED POWER p' KW
DECIBELS ABOVE 1 KW
f, Mc 40 35 30 25 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3
Thousands of Kilometers
COSMIC NOISE
Ad

FIGURE 9.14.

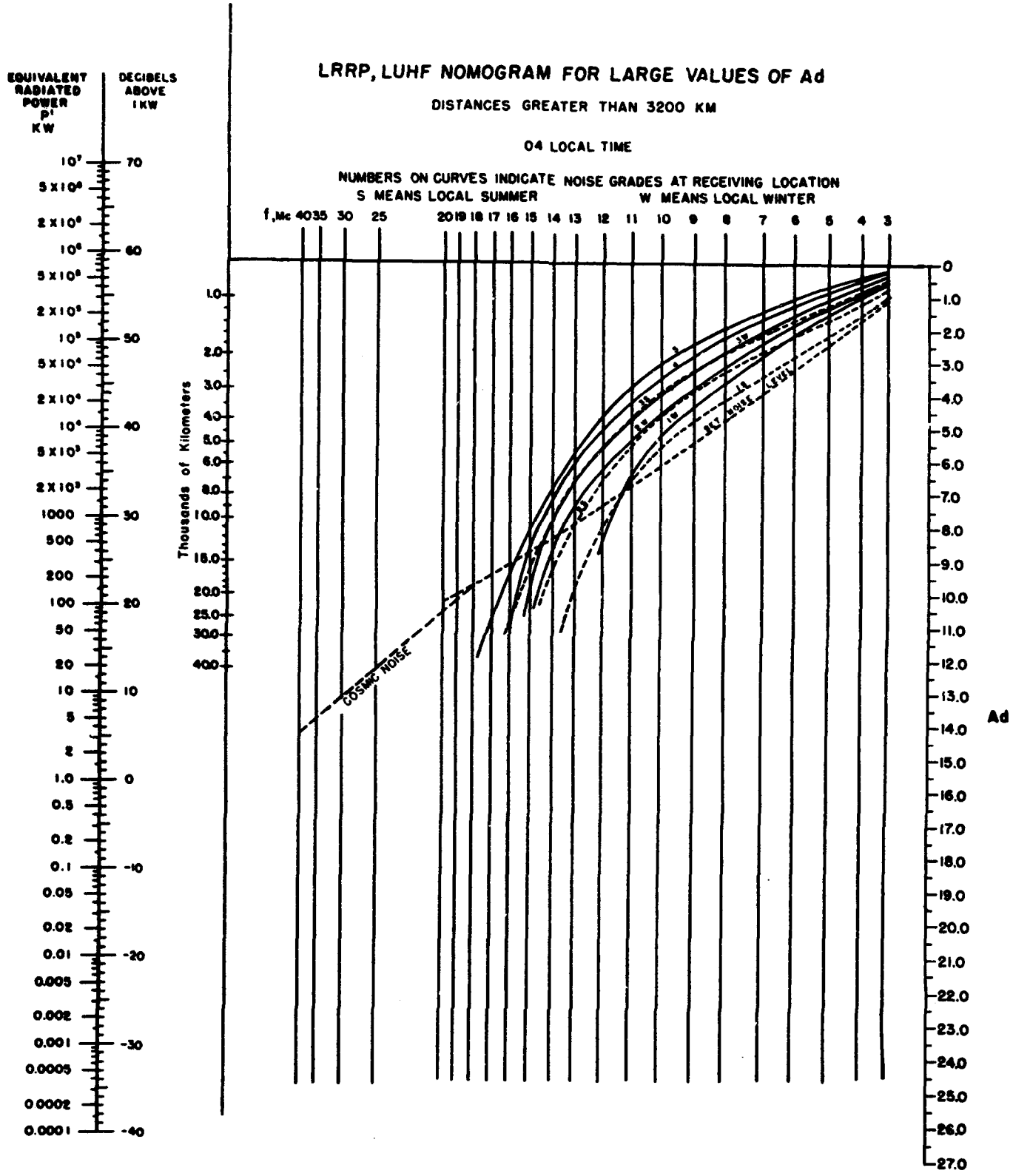

LRRP, LUHF NOMOGRAM FOR LARGE VALUES OF Ad
DISTANCES GREATER THAN 3200 KM
04 LOCAL TIME
NUMBERS ON CURVES INDICATE NOISE GRADES AT RECEIVING LOCATION
S MEANS LOCAL SUMMER
W MEANS LOCAL WINTER
EQUIVALENT RADIATED POWER P' KW
DECIBELS ABOVE 1 KW
f, Mc 40 35 30 25 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3
Thousands of Kilometers
COSMIC NOISE
Ad

FIGURE 9.15.

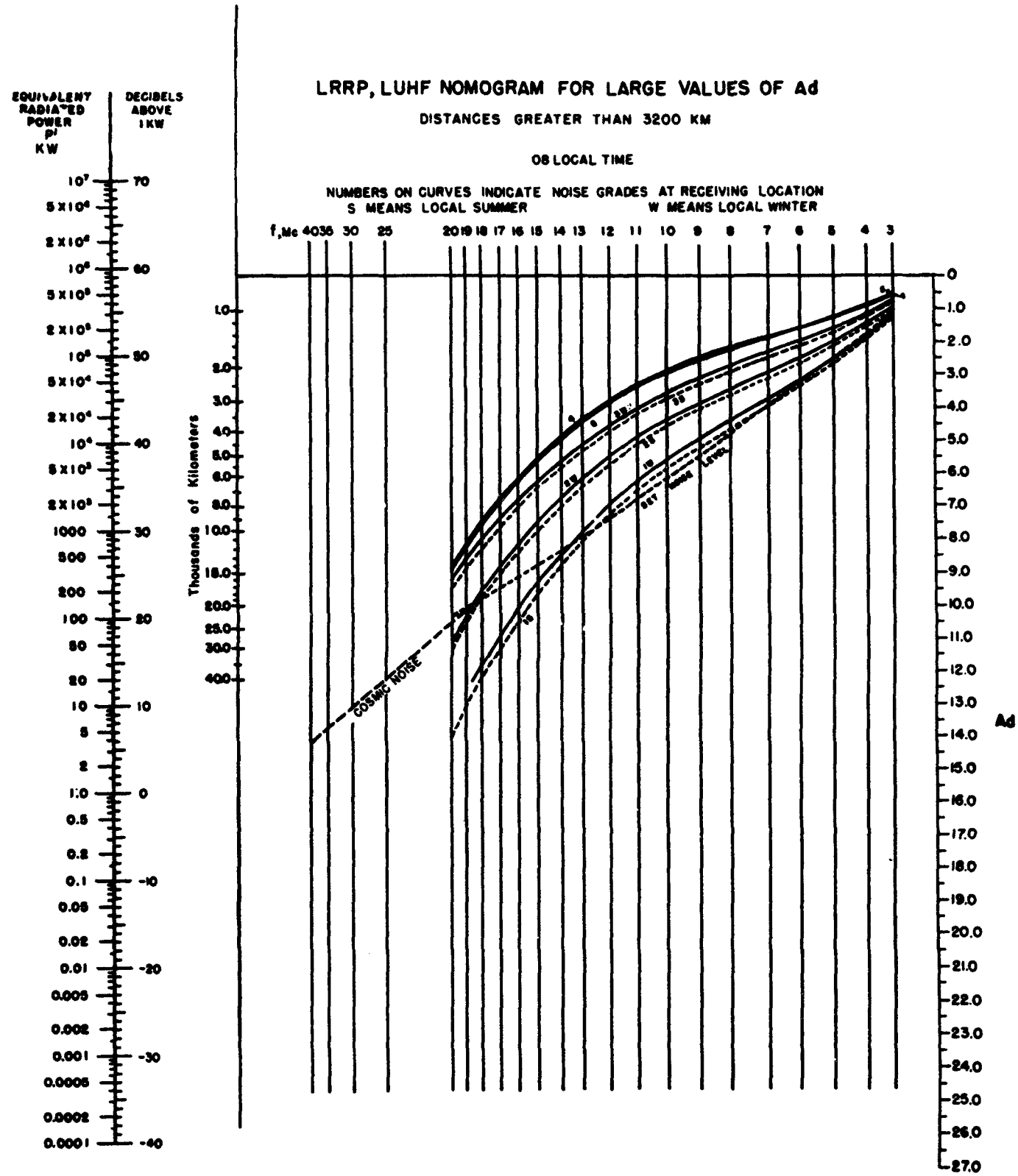

LRRP, LUHF NOMOGRAM FOR LARGE VALUES OF Ad
DISTANCES GREATER THAN 3200 KM
0S LOCAL TIME
NUMBERS ON CURVES INDICATE NOISE GRADES AT RECEIVING LOCATION
S MEANS LOCAL SUMMER
W MEANS LOCAL WINTER
EQUIVALENT RADIATED POWER P' KW
DECIBELS ABOVE 1 KW
f, Mc 40 35 30 25 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3
Thousands of Kilometers
COSMIC NOISE
Ad

FIGURE 9.16.

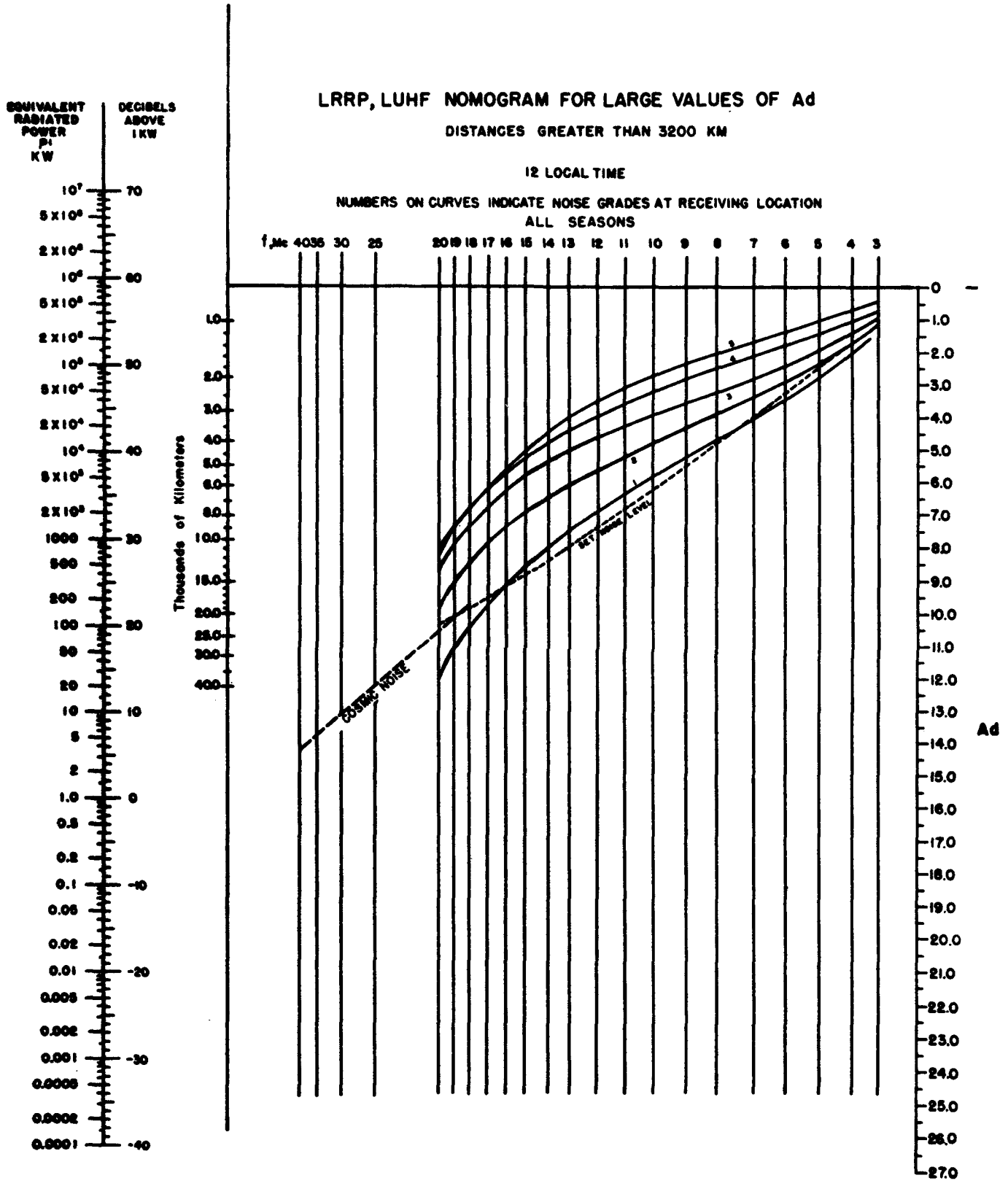

LRRP, LUHF NOMOGRAM FOR LARGE VALUES OF Ad
DISTANCES GREATER THAN 3200 KM
12 LOCAL TIME
NUMBERS ON CURVES INDICATE NOISE GRADES AT RECEIVING LOCATION
ALL SEASONS
EQUIVALENT RADIATED POWER Pt KW
DECIBELS ABOVE 1 KW
Thousands of Kilometers
f, Mc 40 35 30 25 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3
COSMIC NOISE
SET NOISE LEVEL
Ad

FIGURE 9.17.

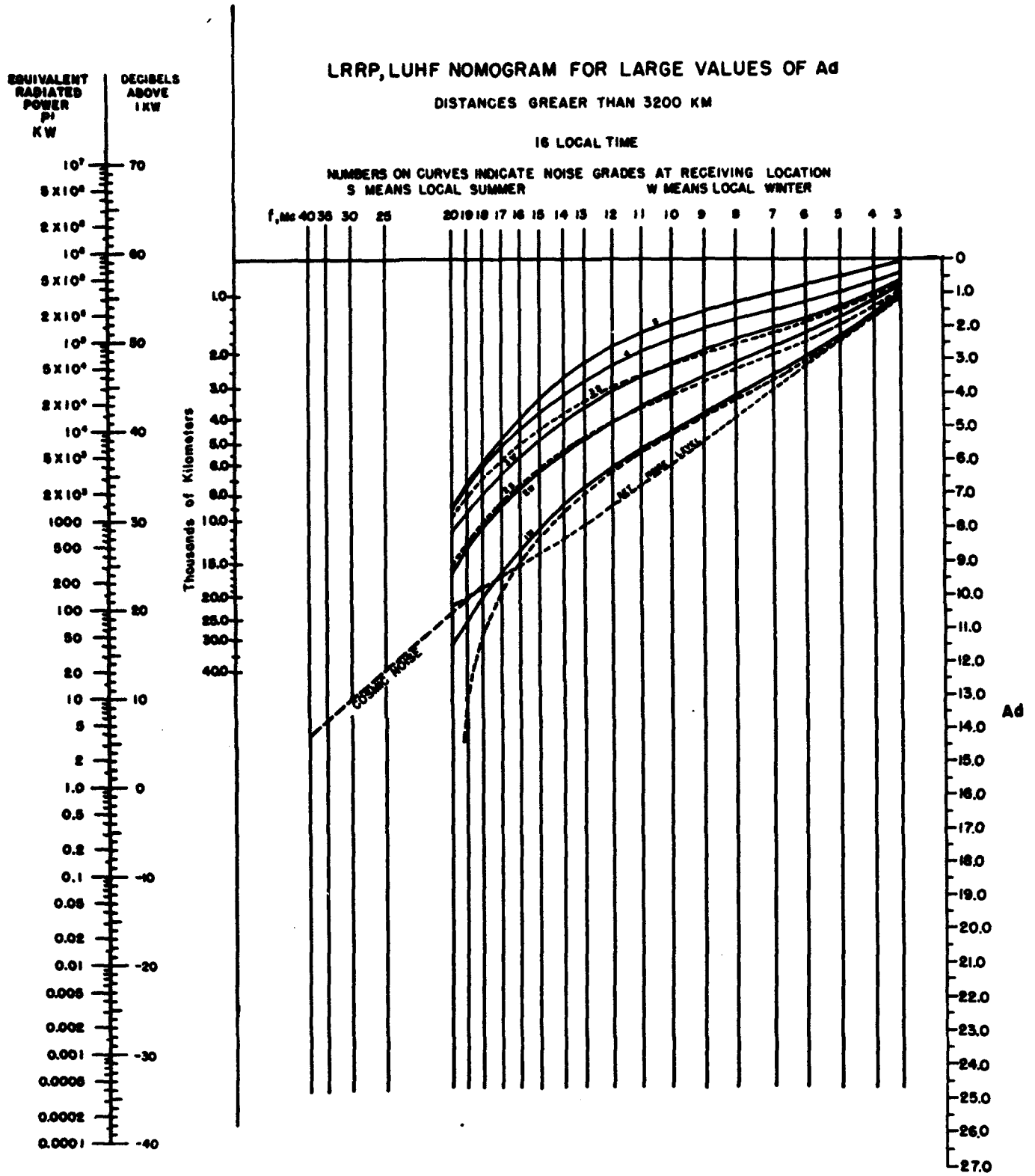

LRRP, LUHF NOMOGRAM FOR LARGE VALUES OF Ad
DISTANCES GREAER THAN 3200 KM
16 LOCAL TIME
NUMBERS ON CURVES INDICATE NOISE GRADES AT RECEIVING LOCATION
S MEANS LOCAL SUMMER
W MEANS LOCAL WINTER
EQUIVALENT RADIATED POWER P' KW
DECIBELS ABOVE 1 KW
f, Mc 40 36 30 25 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3
Thousands of Kilometers
COSMIC NOISE
Ad

FIGURE 9.18.

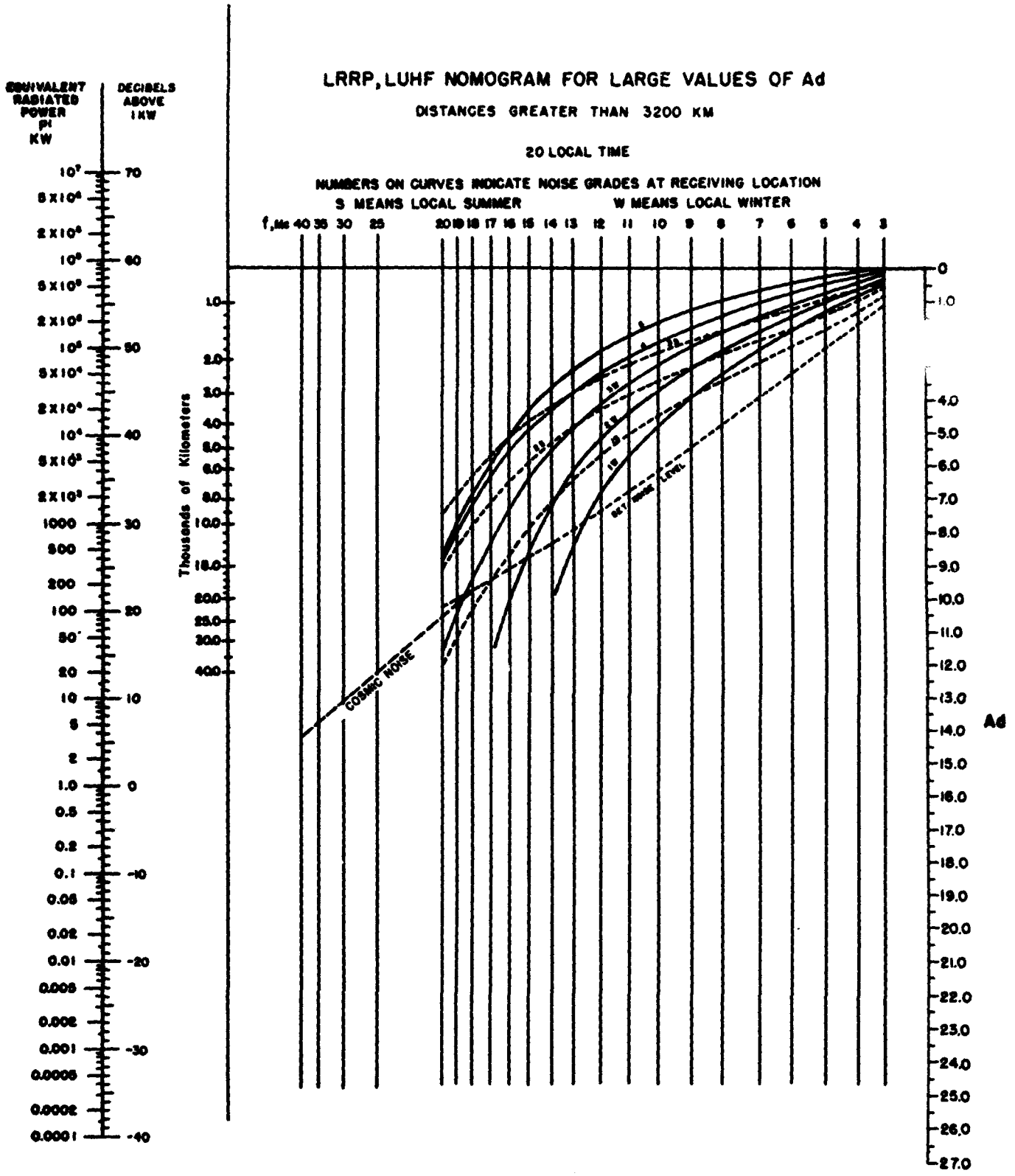

LRRP, LUHF NOMOGRAM FOR LARGE VALUES OF Ad
DISTANCES GREATER THAN 3200 KM
20 LOCAL TIME
NUMBERS ON CURVES INDICATE NOISE GRADES AT RECEIVING LOCATION
S MEANS LOCAL SUMMER
W MEANS LOCAL WINTER
EQUIVALENT RADIATED POWER pi KW
DECIBELS ABOVE 1 KW
f, Mc 40 35 30 25 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3
Thousands of Kilometers
COSMIC NOISE
Ad

FIGURE 9.19.

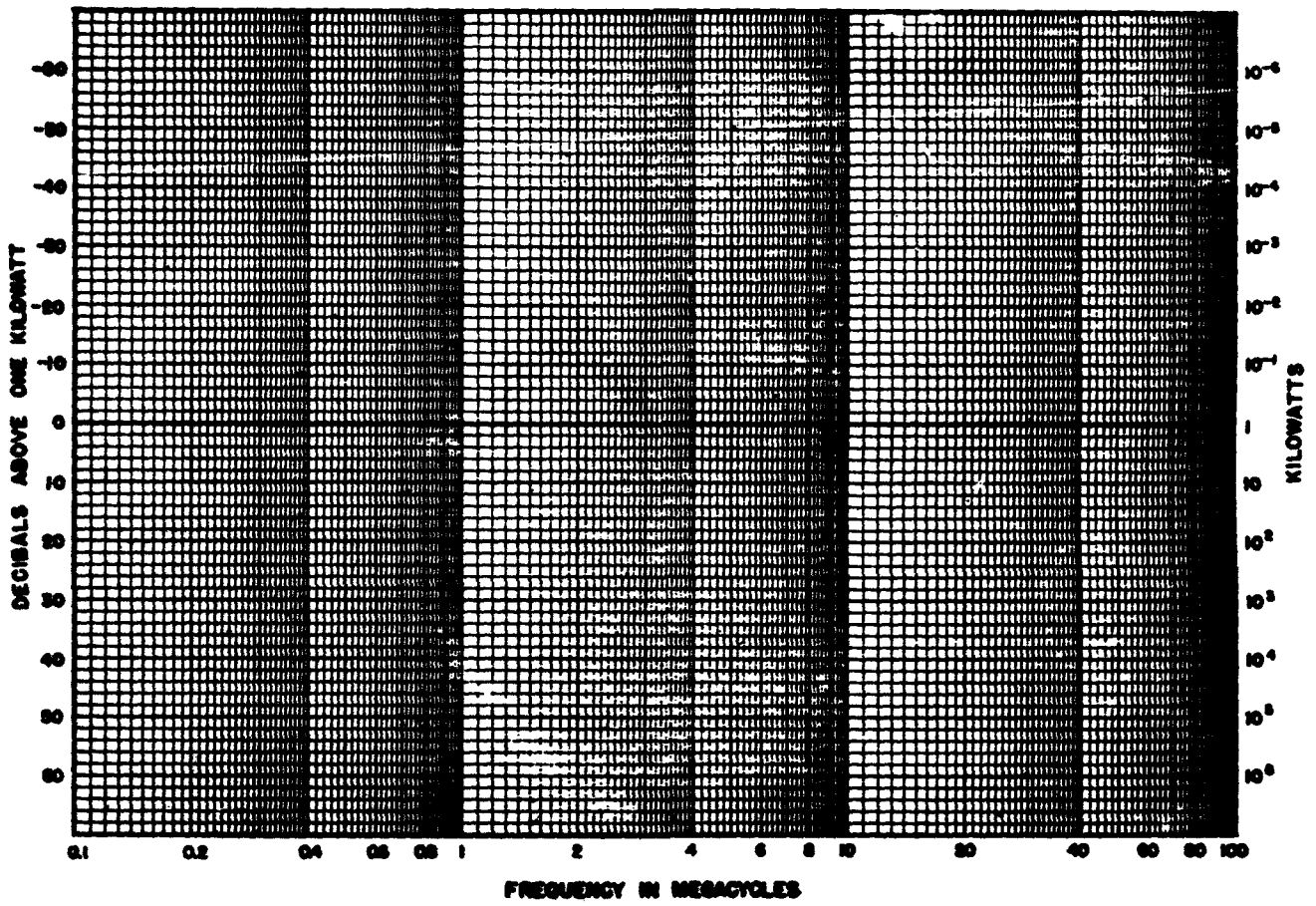


FIGURE 9.20. *Grid for plotting equivalent radiated power (P') as a function of frequency for a given mode of propagation in the solution of intermediate distance luhf problems (see section 9.6).*

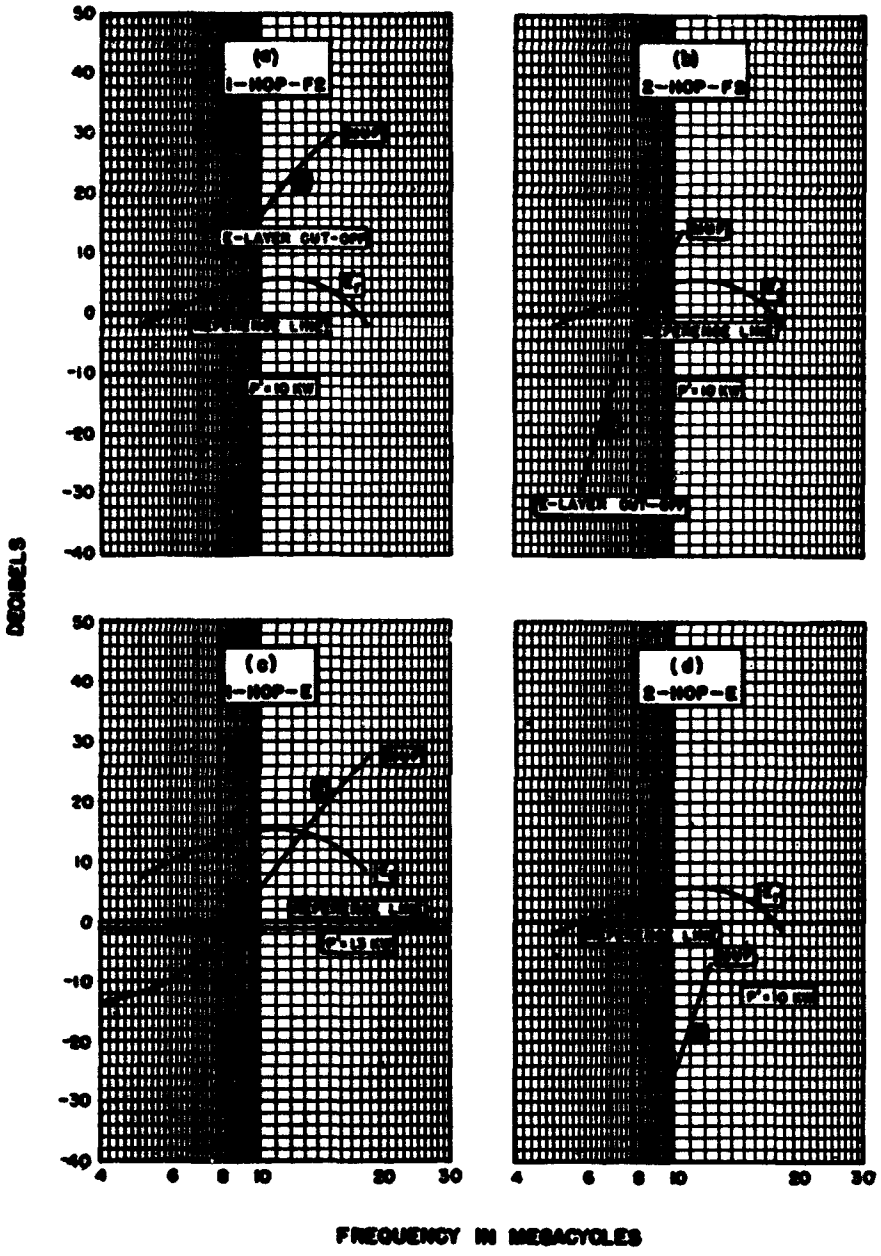


FIGURE 9.21. *Illustration for luhf example worked out in section 9.6.*

E=incident field intensity. E_r=required field intensity.

996833 O - 51 - 14

10. APPENDIX

THE mks SYSTEM OF UNITS

Electrostatics and magnetostatics grew up independently, and each devised its own system of units. When the relation between the two subjects was recognized, it was necessary to reconcile the two systems of units, and this was done by inserting the conversion factor $c=3\times10^{10}$ (the number of esu of charge in 1 emu) in various places. Furthermore, a troublesome factor of 4π was usually prominent, caused by the fact that the area of a unit sphere is 4π square radians, and thus there were 4π lines of electric or magnetic force conceived of as emanating from a unit charge or pole, respectively. Also, the units of electrical quantities in the two systems were either too large or too small for convenience in practical work.

In addition to the basic quantities of length, mass, and time, one other basic quantity is necessary to provide a system of units in electrodynamics. A convenient choice of such a fourth quantity is electric charge. With the coulomb as the unit of charge, the meter as the unit of length, the kilogram as the unit of mass, and the second as the unit of time, the mks system of units may be derived. Having selected these basic units, the units of potential, current, resistance, charge, and power in this system are the practical units, namely, volt, ampere, ohm, coulomb, and watt, respectively. The mks system of units for various quantities and their equivalents in the cgs electrostatic and electromagnetic systems are given in the following table:

Quantity	Symbol	Dimensions	mks	cgs (esu)	cgs (emu)	Defining equation
Length	L	L	1 meter	$=10^2$ cm	$=10^2$ cm	
Mass	M	M	1 kilogram	$=10^3$ g	$=10^3$ g	
Time	T	T	1 second	$=1$ sec	$=1$ sec	
Force	F	ML/T^2	1 newton	$=10^5$ dynes	$=10^5$ dynes	$F=ma$
Energy	W	ML^2/T^2	1 joule	$=10^7$ ergs	$=10^7$ ergs	$W=\frac{1}{2}mv^2$
Power	P	ML^2/T^3	1 watt	$=10^7$ ergs/sec	$=10^7$ ergs/sec	$P=dW/dt$
Charge	Q	Q	1 coulomb	$=3\times10^9$ esu	$=10^{-1}$ emu	
Current	I	Q/T	1 ampere	$=3\times10^9$ statamp	$=10^{-1}$ abamp	$I=dQ/dt$
Resistance	R	ML^2/Q^2T	1 ohm	$=\frac{1}{9}\times10^{-11}$ statohm	$=10^9$ abohms	$R=P/I^2$
Potential	V	ML^2/QT^2	1 volt	$=\frac{1}{3}\times10^{-2}$ statvolt	$=10^8$ abvolts	$V=IR$
Electric field	E	ML/QT^2	1 volt/meter	$=\frac{1}{3}\times10^{-4}$ statvolt/cm	$=10^6$ abvolts/cm	$E=-\nabla V$
Electric flux density*	D	Q/L^2	1 coulomb/square meter.	$=12\pi\times10^5$ esu	$=4\pi\times10^{-5}$ emu	$D=\frac{Q}{4\pi r^2}$
Dielectric constant	κ	Q^2T^2/ML^3	1 farad/meter	$=36\pi\times10^9$ esu	$=4\pi\times10^{-11}$ emu	$\kappa=D/E$
for free space	κ_0	----------	$\frac{1}{36\pi}\times10^{-9}$ farad/meter.	$=1$ esu	$=\frac{1}{9}\times10^{-20}$ emu	
Capacitance	C	Q^2T^2/ML^2	1 farad	$=9\times10^{11}$ statfarad	$=10^{-9}$ abfarad	$W=\frac{1}{2}CV^2$
Magnetic flux	ϕ	ML^2/QT	1 weber	$=\frac{1}{3}\times10^{-2}$ esu	$=10^8$ maxwells	$V=d\phi/dt$
Magnetic flux density	B	M/QT	1 weber/square meter	$=\frac{1}{3}\times10^{-4}$ esu	$=10^4$ gauss	$B=\phi/A$
Inductance	L	ML^2/Q^2	1 henry	$=\frac{1}{9}\times10^{-11}$ esu	$=10^9$ abhenry	$W=\frac{1}{2}LI^2$
Magnetic intensity*	H	Q/LT	1 ampere turn/meter	$=12\pi\times10^7$ esu	$=4\pi\times10^{-3}$oersted	$\int Hds=I$
Permeability	μ	ML/Q^2	1 henry/meter	$=\frac{1}{36\pi}\times10^{-13}$ esu	$=\frac{1}{4}\pi\times10^7$ emu	$\mu=B/H$
for free space	μ_0	----------	$4\pi\times10^{-7}$ henry/meter	$=\frac{1}{9}\times10^{-20}$ esu	$=1$ emu	

e (electronic charge) $=1.59\times10^{-19}$ coulomb.
e/m (specific electronic charge) $=1.77\times10^{11}$ coulombs per kilogram.

*Rationalizing factor 4π introduced in these quantities.

WASHINGTON, *June 2, 1947.*

11. INDEX

S

T

U, V

W, Z

U. S. GOVERNMENT PRINTING OFFICE: O—1954