ESD-TR-67-631

A GENERAL ANALYSIS OF NATURAL ENVIRONMENTAL EFFECTS ON ELECTROMAGNETIC RADIATION UTILIZED FOR COMMUNICATIONS

Dr. Norman E. Gaut
Major Kenneth E. German

January 1968

TECHNICAL REQUIREMENTS AND STANDARDS OFFICE
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

LEGAL NOTICE

OTHER NOTICES

Change 1

DEPARTMENT OF THE AIR FORCE
HEADQUARTERS ELECTRONIC SYSTEMS DIVISION (AFSC)
HANSCOM AIR FORCE BASE, MASSACHUSETTS 01731

REPLY TO
ATTN OF: WE (Capt Breslin/3237) 9 April 1979

SUBJECT: Suggested Change to ESD-TR-67-631

TO: ESD/TOSC

1. Reference 6 Apr 79 telecon between Capt Breslin (ESD/WE) and Mr. Courtoglous (ESD/TOSC).

2. At the Staff Meteorology Office we have found ESD-TR-67-631, A General Analysis of Natural Environmental Effects on Electromagnetic Radiation Utilized for Communications, a very good reference document. It addresses environmental effects clearly and concisely.

3. However, we recommend the following change be made. Reference page 119, second paragraph. Change "(i) it is <u>directly</u> proportional to the square root of the electron density" to read "(i) it is <u>inversely</u> related to the electron density."

<u>REASON</u>: From eqn (15) $n=(1-X)^{\frac{1}{2}}$ Squaring this yields $n^2=1-X$. Through mathematical manipulation it follows that $n^2=(1-X)\frac{(1+X)}{(1+X)}=\frac{1-X^2}{1+X}\doteq\frac{1}{1+X}$ for $X \ll 1$. Taking the square root shows $n \doteq \frac{1}{(1+X)^{\frac{1}{2}}}$, and given that the plasma frequency is directly proportional to the electron density"; it follows that the index of refraction is inversely related to the electron density.

4. If you have any questions contact Capt Breslin.

CLARENCE B. GIVENS, Lt Col, USAF
Chief Staff Meteorologist

Cy to: AFSC/WER (Mr. Durham)
AWS/DNT

ESD-TR-67-631

A GENERAL ANALYSIS OF NATURAL ENVIRONMENTAL EFFECTS ON ELECTROMAGNETIC RADIATION UTILIZED FOR COMMUNICATIONS

Dr. Norman E. Gaut
Major Kenneth E. German

January 1968

TECHNICAL REQUIREMENTS AND STANDARDS OFFICE
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

FOREWORD

The invisible electromagnetic waves which carry information between sender and receiver in any electronic communication system are always affected by the environment through which they propagate. The interaction may alter either the direction, speed of propagation, phase, amplitude, frequency or polarization of the signal, or any combination of these features, from their free space values. When environmental effects are predictable, they can be minimized or, in many cases, utilized to provide more reliable signals. When environmental effects are unpredictable, signal degradation usually occurs.

This Technical Report analyzes, in a general manner, the interaction between the natural aerospace environment and propagating electromagnetic energy used for communications. Its purpose is to give the reader a combined but simplified view of this very complex subject.

PUBLICATION REVIEW

This technical report has been reviewed and is approved.

FOR THE COMMANDER

DELOS C. JENSEN, LtCol, USAF
Chief, Meteorology Division
Tech Rqmts & Stds Office

ABSTRACT

The earth and its atmosphere provide a highly complex environment for the propagation of electromagnetic waves. A convenient method to study the interplay between this environment and propagating electromagnetic signals is to isolate various quasi-homogeneous regions of the earth-atmosphere system, and study the effects they have on different frequency signals. The troposphere, the ionosphere, the oceans and the solid crust are such quasi-homogeneous subdivisions. This report presents a short review of the physical make-up of each of the subdivisions important to radio propagation, together with the most important effects each has upon the propagation of electromagnetic signals.

TABLE OF CONTENTS

SECTION 1
INTRODUCTION

SECTION 2
GENERAL PROPERTIES OF
ELECTROMAGNETIC RADIATION

TABLE OF CONTENTS (Cont'd)

TABLE OF CONTENTS (Cont'd)

TABLE OF CONTENTS (Cont'd)

TABLE OF CONTENTS (Cont'd)

TABLE OF CONTENTS (Cont'd)

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (Cont'd)

LIST OF ILLUSTRATIONS (Cont'd)

LIST OF ILLUSTRATIONS (Cont'd)

LIST OF ILLUSTRATIONS (Cont'd)

LIST OF ILLUSTRATIONS (Cont'd)

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

A	Constant in waveguide equations determined by exciting voltage; symbol standing for an atom in ionospheric processes; modified refractivity
Å	Angstrom = 10^{-8} cm
a	Particle diameter
ac	Alternating current
C	Degrees centigrade $\cong {}^{\circ}K-273$; constant depending upon the strength of the geomagnetic field

CO_2	Carbon dioxide
c	Velocity of light in a vacuum $\cong 3 \times 10^{10}$ cm/s; cycles
cm	Centimeters = 10^{-2} meters = $\frac{1}{2.54}$ inches
D	An ionospheric layer which is found between 50 and 90 km
d	Distance between walls of a waveguide; duct thickness
db	Decibels = $10 \log_{10} \frac{Pin}{Pout}$, where P is power
dc	Direct current
E	An ionospheric layer found between 100 and 110 km
$\vec{E}$	Electric field strength vector with components E_x, E_y, E_z along the x, y, z axes of a cartesian coordinate system
E_o	Maximum amplitude of electric field strength
E_s	The sporadic E ionospheric layer found at about 105 km
E_t	Translational energy of a molecule
E_e,E_r E_v, E_f E_i	Quantized energies of a molecule, where e, r, v, f, and i refer to electronic, rotational, vibrational, final, and initial, respectively
E_1,E_2 E_y,E_z	Electric field strength in medium 1, medium 2, and along the y and z axes respectively
EHF	Extremely high frequency
e	The base of the Naperian logarithm system = 2.718...; the partial pressure of water vapor in the atmosphere
$\vec{F}$	Force vector
F_1,F_2	Ionospheric layers appearing roughly between 175-250 km, and 250-400 km, respectively
FOT	Frequency of optimum transmission
f	Frequency of an electromagnetic wave in cycles per second (c/s) or Hertz (Hz); $f = \omega/2\pi$
ft	Feet

LIST OF ABBREVIATIONS AND SYMBOLS (Cont'd)

G	G Prefix giga = 10^9
GHz	Gigahertz = 10^9 Hz
Gc/s	Gigacycles/second = 10^9c/s
gm	Gram
gm/kgm	Grams per kilogram
H	Scale height = RT/mg where R is the universal gas constant, T is temperature in degrees Kelvin, m is molecular weight of gas, and g is the acceleration of gravity. H is the depth of an atmosphere of constant density or, in an isothermal atmosphere, it is the distance one must ascend to have the density decrease by the factor $\frac{1}{e}$
H	The magnetic field strength vector with components H_x, H_y, H_z, along the x, y, and z axes of a cartésian coordinate system
H_N	Scale height for refractivity N
Hz	Hertz (one cycle per second)
H_1,H_2 H_y,H_z	Magnetic field strength in medium 1, medium 2, and along the y and z axes respectively
HF	High frequency
H_2O	Water
h	Planck's constant = 6.624 x 10^{-27} erg-seconds; height
hr	Hour
I	Intensity of an electromagnetic wave $\sim$ (amplitude)^2
j	Imaginary number = $\sqrt{-1}$
K	Degrees Kelvin $\cong$ $^\circ$C + 273
K	Kilo = 10^3
Kc/s	Kilocycles/second = 10^3 c/s
KHz	Kilohertz = 10^3 Hz
km	Kilometers
kW	Kilowatts = 10^3 watts

LIST OF ABBREVIATIONS AND SYMBOLS (Cont'd)

LF	Low frequency
M	Mega = 10^6; water content in grams per cubic meter
MF	Medium frequency
MHz	10^6 Hertz
MUF	Maximum usable frequency
m	Meters; waveguide mode number; molecular weight of a gas; milli = 10^3
mb	Millibars of pressure = 10^3 dynes/cm^2, 1013.25 mb = 1 atmosphere = 760 mm of Mercury
mc/s	Megacycles/second = 10^6 c/s
N	Refractivity = $(n - 1) \times 10^6$; atomic nitrogen
N_1	Refractivity at an altitude of 1 km
N_2	Molecular nitrogen
N_m	Mean sea level refractivity
NO	Nitric oxide
N_s	Ground or surface value of refractivity
n	Index of refraction for a dielectric medium, in this case the troposphere; waveguide mode number
n	Unit vector normal to a surface
n_1, n_2	Index of refraction in medium 1 and medium 2
O	Atomic oxygen
O_2	Molecular oxygen
O_3	Ozone
P	Total pressure
PCA	Polar cap absorption
PCE	Polar cap event

q	Specific humidity = grams of water/kilograms of moist air; rate of ion production in the Chapman layer equation; charge on a particle
R	Universal gas constant; receiver; a number which measures sunspot activity
R_1	Power reflection coefficient
r_m	Duct height
S	Distance
S	Poynting's vector, a measurement of magnitude and direction of energy flow of an electromagnetic wave
S_1,S_2	Distance in first and second medium
SCNA	Sudden cosmic noise absorption
SEA	Sudden enhancement of atmospheric
SHF	Super high frequency
SID	Sudden ionospheric disturbance
SPA	Sudden phase anomaly
SWF	Short wave fade-out
T	Period, temperature; transmitter
t	Time coordinate
UHF	Ultra high frequency
$\vec{V}$ & V	Velocity of an electromagnetic wave in a dielectric
V_1,V_2 V_y,V_z	Velocity in medium 1, medium 2, and along the y and z axes respectively
VHF	Very high frequency
VLF	Very low frequency
W	Watt

LIST OF ABBREVIATIONS AND SYMBOLS (Cont'd)

X — Square of the ratio between the plasma frequency of an electron concentration and the frequency of the penetrating electromagnetic wave = $(\omega_n/\omega)^2$

x — One distance coordinate of a cartesian coordinate system

Y — A variable in the equation for the ionospheric index of refraction depending on the geomagnetic field strength

y — Distance coordinate in a cartesian coordinate system

yd — Yard

Z — A variable in the equation for the ionospheric index of refraction depending on the geomagnetic field strength, height

Z_1, Z_2 — Reference heights (arbitrary levels)

Z_m — Reflection height corresponding to mode m, height of maximum electron concentration

z — Distance coordinate (vertical) in a cartesian coordinate system

α — Radiation at 1215.7 Å (Lyman α); angle from the horizontal for a transmitted electromagnetic wave; absorption coefficient per unit mass for the atmosphere

α_m — Launch angle for mode m

β — Angular deviation from a straight line path of an electromagnetic ray passing through a planar ionosphere; number of ions produced by unit quantity of absorbed radiation

γ — Angular deviation from a straight line path of an electromagnetic ray passing through a spherical ionosphere; propagation constant for the equation governing the electric and magnetic fields in a waveguide

Δ — Indicates a small increment of a quantity

$\Delta\beta$ — A small change in look angle for a radar beam

$\Delta S_1, \Delta S_2$ — Increment of distance in medium 1 and in medium 2

ΔT_1 — Change of temperature

$(\Delta T_2)_D$ — Change of temperature after dry adiabatic ascent

LIST OF ABBREVIATIONS AND SYMBOLS (Con'td)

$(\Delta T_2)_W$	Change of temperature after wet adiabatic ascent
Δt	Small interval of time
ϵ	Permittivity of a dielectric medium
η	Electron concentration or density
θ	Angle entry into or exit from a planar ionosphere by an electromagnetic ray; phase angle
θ_1, θ_2	Angle of incidence in first and second medium
λ	Wavelength of an electromagnetic wave
λ_c	Wavelength associated with cut-off frequency in a waveguide
μ	Permeability of a dielectric medium, micron = 10^{-4} cm = 10^{-6} m
$\mu v/m$	Microvolts/meter (1 microvolt = 10^{-6} volts), a measurement of electric field strength
ν	Collisions per second of electrons in the ionosphere, i.e. collision frequency
π	3.1416
ρ	Density of a gas
σ	Conductivity
ϕ	Angle of reflection or incidence
ψ	Tilt angle of a guided surface wave from the local vertical
ω	Angular frequency of a signal = (2π) x (frequency)
ω_p	Plasma frequency $\sim \sqrt{\eta}$
∇	A vector operator = $(\frac{\partial}{\partial x}\hat{i} + \frac{\partial}{\partial y}\hat{j} + \frac{\partial}{\partial z}\hat{k})$
$\equiv$	(Is defined as)

SECTION 1

INTRODUCTION

The U. S. Air Force is a large, complex organization with installations and commitments all over the world. It is therefore strongly dependent upon efficient communications to carry out its mission. Several fundamentally different techniques are employed to move information but all operate within the natural environment and are more or less affected by it. For those people who use communication systems and for those who design them it is important to have a general understanding of these effects.

By far, the most important effects of the natural environment are felt by electromagnetic radiation propagating directly through the earth or its atmosphere. Such wireless techniques have been in use since the beginning of the 20th Century, but the advent of the space age has placed new priority on improving their performance. To no small extent, improvement will be closely coupled to better understanding of the limitations imposed by the natural environment on electromagnetic signals.

For optimum results, then, the design of communication systems requires a great deal of knowledge about the earth-atmosphere environment, both to avoid the limitations imposed by it and to take advantage of its peculiarities. Much has already been learned, but even after decades of theoretical and experimental research, many important questions remain to be fully answered. Without all of the details, however, the general outline of the interactions are well understood, and it is exactly this general outlined of the effects of the earth-atmosphere system on electromagnetic radiation which is the subject of this report.

Its purpose is to survey, in a basic way, the fundamental restrictions which the earth-atmosphere system imposes upon electromagnetic radiation. It is decidedly not meant to be a comprehensive technical review from either the communications or the natural environmental standpoint; it is a broad brushstroke examination of the interface between these two complex fields of knowledge.

This report is organized into two major parts, and so written that little background is required to appreciate its contents. One part is devoted to a short review of the fundamentals of electromagnetic radiation. The second part then examines various quasi-homogeneous regions of the earth-atmosphere system to determine the major effects which they have upon electromagnetic radiation.

Section 2, which is devoted to basic electromagnetic theory, might be considered beyond the primary subject matter. The reason for inclusion of this material is that a comprehension of the information presented is mandatory for an understanding of the interactions between electromagnetic radiation and the natural environment covered in Section 3. The electromagnetic theory is admittedly basic. Those readers having a complete command in this area are invited to skip Section 2 and to begin Section 3.

Graphs, illustrations, and diagrams are freely used to facilitate understanding of the discussion. A list of abbreviations and symbols used in the text is provided near the beginning of the paper. Altogether, the report is self-contained as much as is practicable.

Justification for a report such as this lies in its combined views of the subject matter. Countless hundreds of articles have been written which mathematically analyze minute segments of the radiation-environment problem. They are essential to quantitatively understand the phenomenon involved, but they are far too technical for general understanding. A painless general understanding in a single package is the goal of this report.

SECTION 2

GENERAL PROPERTIES OF ELECTROMAGNETIC RADIATION

I. FIELDS AND WAVES

The ultimate nature of electromagnetic energy is a metaphysical question. Everything we know about it is in the form of its effects on material objects, effects which we can measure and from which we can infer the fundamental characteristics of electromagnetism.

A. DETECTION OF ELECTRIC AND MAGNETIC FIELDS

A method to measure the presence of either electric or magnetic fields utilizes their effects on an electrically charged particle. An electric field may be detected by observing the force it exerts on a stationary charged particle: the strength of the electric field is directly proportional to the measured force; and its direction is in the direction of the force on the particle. If, with such a charge, we could take an unlimited number of measurements of the strength and direction of the field, arbitrarily close in time, we could define the time dependence of the field. Knowledge of the time dependence of the strength and direction of the field at all points in space would define the field completely. Figure 1 illustrates the detection of the electric field.

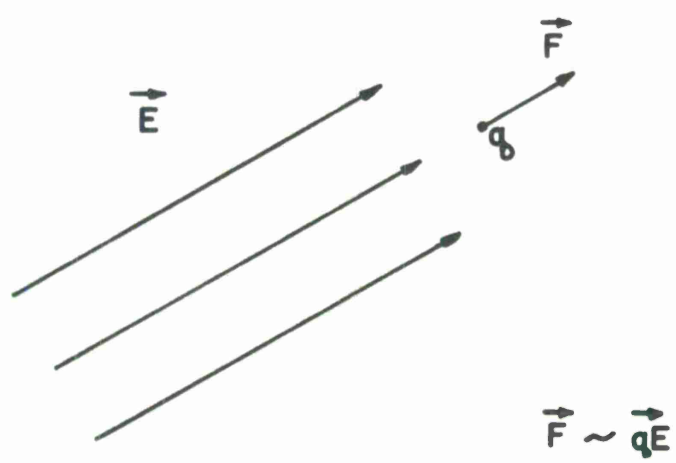
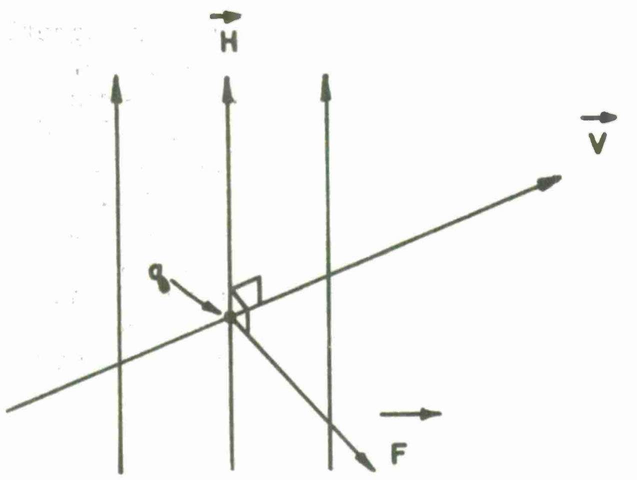


Figure 1. Detection of an electric field using a charged particle. An electric field ($\vec{E}$) may be detected by a charged particle (q). The force ($\vec{F}$) on the particle is proportional to the strength of the electric field, and the force is in the direction of the field. (The arrow over a quantity indicates a vector with both magnitude and direction associated with it.)

If there existed a stationary magnetic field in the vicinity of our probing charge, it could be detected by moving the charge at some speed and measuring the force on it. If it were moving in any direction except parallel to the magnetic field, a force would be felt perpendicular to the motion. If the direction of travel of the point charge were varied but not its speed, then, when the force measured was a maximum, the investigator would know that the direction of travel, the force, and the magnetic field would all be pointed at right angles to each other. Figure 2 will clarify this concept.



$$\vec{F} \sim q(\vec{V} \times \vec{H})$$

Figure 2. Detection of a magnetic field using a moving charged particle. The force ($\vec{F}$) on the particle is proportional to its charge (q) and its velocity ($\vec{V}$), and to the magnetic field ($\vec{H}$). When the force is a maximum, given a constant speed, the magnetic field, the direction of motion of the particle, and the force on the particle are mutually perpendicular. When $\vec{V}$ is rotated into $\vec{H}$, through their smallest angular separation, the rotation advances a right hand screw in the direction of the force on a positively charged particle.

B. COMBINATION OF ELECTRIC AND MAGNETIC FIELDS

A characteristic of electric and magnetic fields which greatly aids in their study is their additive nature. If two fields, either electric or magnetic, are measured separately by the forces they exert on a charged particle, then when both fields are applied simultaneously, their combined effect is the vector sum of the individual fields. This is illustrated in Figure 3 for electric fields. The same illustration holds true for magnetic fields, if E is replaced by H.

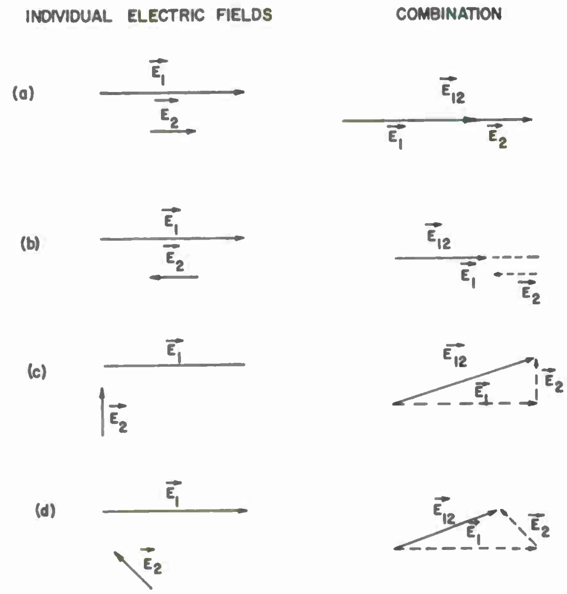


Figure 3. The vector combination rule for electric fields. The combined effect of two individual electric fields, E_1 and E_2, may be found by the rules of vector combinations.

Fields whose direction of action are parallel add. Fields whose direction of action are anti-parallel tend to cancel. Fields acting at angles to each other combine by the rules of vector combination. The individual fields do not interact. If all the independently produced electric and magnetic fields are known, then the resultant fields are known. The convenience of this characteristic will be seen to be even more important when time varying fields are considered below.

C. PRODUCTION OF SIMPLE ELECTRIC AND MAGNETIC FIELDS

Up to now we have examined the electric and magnetic fields as separate entities. Is this justified? In a classical sense, for static fields, the answer is yes. But the moment that either the electric or magnetic fields change their magnitude or direction over a period of time, or electrical charges move, a field of the opposite kind is produced. Michael Faraday showed that a changing magnetic field would produce an electric field. Embodied in Ampere's Law is the relation between the movement of charges and the induced magnetic field. Although a detailed review of these laws is outside the scope of this report, we can review qualitatively how electric and magnetic fields may be produced. We shall also consider the very important concept of dipoles.

Electric fields are produced in two ways: (i) by a collection of electrical charges; or (ii) by magnetic fields which change in time. A single positive charge in space will have a simple radial electrical field around itself pointing outward (i.e., a positive test charge in this field will feel a force directed radially away from the positive charge we are examining). A negative charge will have the same field only directed radially inward (see Figure 4a and 4b).

If a positive and negative charge are brought close to one another their fields combine in such a way as to give what is called the dipole field shown in Figure 4c. This simple combination plays an important role in the study of the interaction between fields and matter because many of the phenomena can be understood if it is assumed that matter is composed or can form equivalent dipoles of atomic dimensions.

Magnetic fields, on the other hand, apparently have no isolated sources or sinks. That is, there is no equivalent to positive or negative charge. All magnetic fields, as now known, are produced by either: (i) movement of charges either singly or in currents; or (ii) by electric fields changing in time. Even on the atomic scale, tiny currents of electrical charge are set up by the orbital motion of electrons or by their intrinsic spin to produce atomic magnetic fields. Nevertheless, the concept of a magnetic dipole is of equal use to that of the electric dipole. To produce the equivalent of a dipole field composed of closely spaced magnetic charges, a ring current of electricity may be used similar to that shown in Figure 4d. The strength of the magnetic field is proportional to the magnitude of the current in the ring.

In the case of the electric dipole, any impressed uniform electric field which is not parallel to the axis joining the charges will repel one of the charges and attract the other until it is parallel. The same will happen to the magnetic dipole in a uniform magnetic field, just as if magnetic charges had existed in the positions which the electric charges occupy in the electric dipole. This property is important and we will come back to this subject when we discuss the electric and magnetic properties of matter.

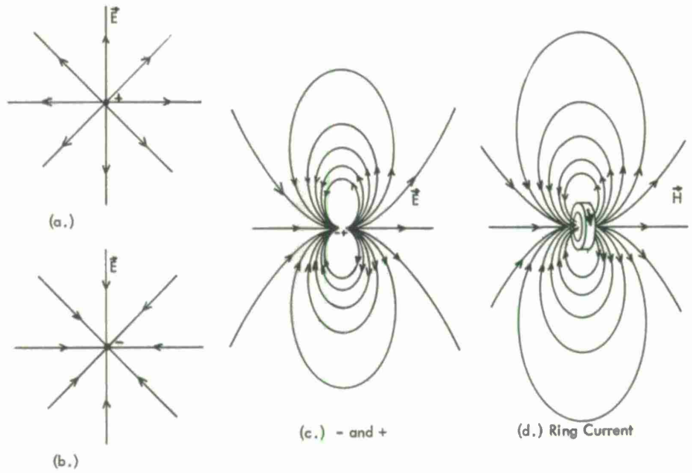


Figure 4. Monopole and dipole fields. The electric fields associated with a positive charge, a negative charge, and their combination in a dipole configuration are shown in figures 4a, 4b, and 4c respectively. Figure 4d shows the magnetic field produced by a ring current which is the equivalent to a magnetic dipole. The last result is a consequence of Ampere's Circuital Law.

D. PROPERTIES OF SIMPLE SINUSOIDAL WAVES

1. The Description of the Wave

Maxwell is really the originator of the concept of electromagnetic radiation. For it was he who carefully sifted through the experimental work which had been performed in electricity and magnetism up to his day (1865) and condensed out their mathematical essence. By adding several missing pieces, he was able to provide a neat, concise, theoretical basis to all classical electromagnetic phenomena. The equations which bear his name remain today the starting place for the study of the interaction between electromagnetic radiation and macroscopic matter.

Perhaps the most startling and unexpected new idea to be derived from Maxwell's work was that when disturbances occurred in the electric or magnetic field, these disturbances should propagate away from their origin as an electromagnetic wave; that is, the electric and magnetic fields should be coupled and be detectable at a point remote from the place of the original disturbance. His calculations showed that they should travel in a vacuum at the speed of light. Hertz, in 1887, confirmed that this radiation did indeed exist.

The properties of the radiation which emerged from Maxwell's equation can best be illustrated by assuming that the strength of the electric field being produced at some point in a vacuum varies sinusoidally. Figure 5 shows the details of this type of periodic function. The maximum strength of the electric field (E_o in Figure 5) is called the <u>amplitude</u>. One complete oscillation of the field strength, i.e., from any point on the variation to the point where the pattern begins to repeat itself is one <u>cycle</u>. The time required for one cycle (T in Figure 5) is the <u>period</u>. The number of cycles which can occur in one second is called the <u>frequency</u>. Note that the frequency is simply the inverse of the period. The argument of the sine function generating the disturbance is called the <u>phase</u>.

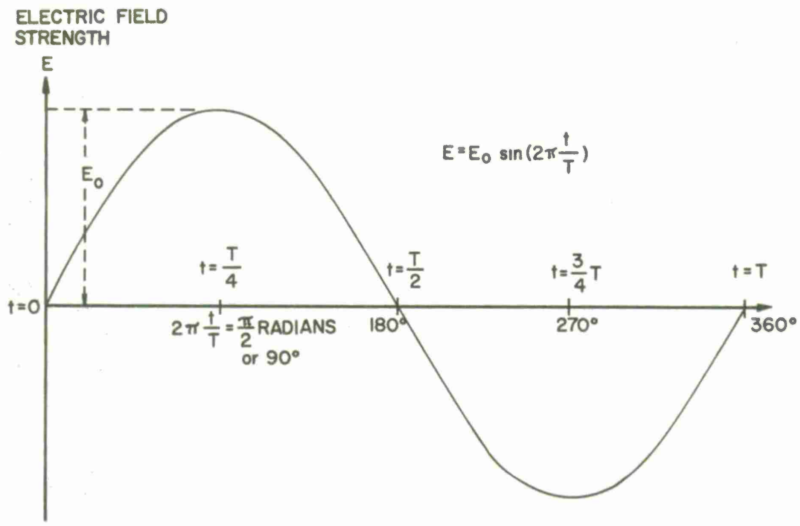


Figure 5. Example of a sinusoidally varying electric field. This example is one cycle of a simple, but very common, periodically varying electric field which describes a sine wave. The angular notation (radians and degrees) refers to the argument of the sine which gives the various amplitudes along the cycle when multiplied by E_o; it is referred to as the phase angle. E_o is the maximum amplitude; and T is the period (seconds per cycle).

It should be made absolutely clear that any disturbances in an electric or magnetic field will cause energy to be radiated. Disturbances which trace sine waves, however, are extremely important because many natural and artificial sources of radiation produce almost pure sinusoidal oscillations. And, as it will be shown later on, many real disturbances can be represented to any degree of accuracy by superimposing pure sine wave oscillations. Therefore, the concepts just introduced to describe such oscillations, i.e., frequency, amplitude, and phase, should be crystal clear to the reader. Further important definitions will be introduced shortly.

The spatial properties of the wave we have defined are illustrated in Figure 6. The variable electric field at the origin in Figure 5 is now assumed to be constant over the Y-Z plane which passes through the origin. The reason for this is to ensure that the amplitude of the fields which move down the X-axis will remain constant, simplifying the analysis. The wavecrests of the generating electric field propagate in a vacuum at the speed of light, and either faster or slower in material media. Their speed is called the phase speed of the disturbance. (The fact that phase speeds can travel faster than the speed of light should not be disturbing. Einstein's prediction that the speed of light should be the maximum attainable in the universe refers to energy carrying signals. It will be shown that in a dispersive medium, i.e., one in which the speed of propagation varies with the frequency, energy is carried by what is called the group velocity which always moves at a speed equal to or less than the speed of light.) The magnetic field produced will be rotated 90° to the direction of the electric field. The direction of propagation will be related to the direction of these two fields by what is called the right hand rule: if the electric field vector is rotated into the magnetic field vector in the direction of their smallest angular separation, then that rotation will advance a right hand screw in the direction of propagation.

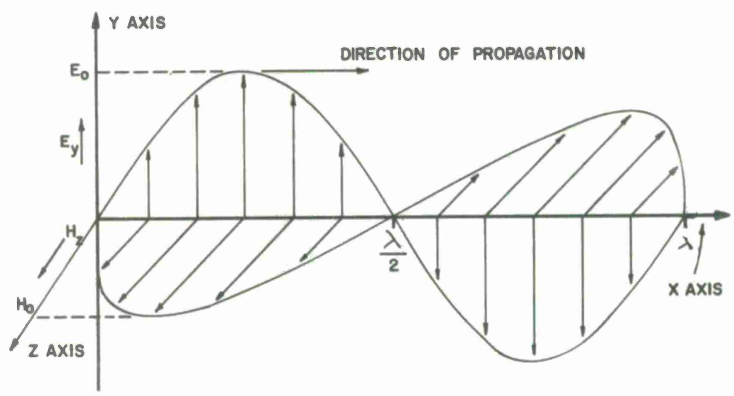


Figure 6. A propagating sinusoidal electromagnetic field. The relationship between the sinusoidally varying electric field E of Figure 5, the induced magnetic field H, and the direction of propagation of the electromagnetic radiation. In this case, the electric and the magnetic fields are of equal strength and direction everywhere in any Y-Z plane. The wave travels in the indicated direction at speed c. In time T, the wave travels a distance λ, which is equal to c times T. λ is called the wavelength.

One complete oscillation will occur in the time period defined as T in Figure 5. The radiation associated with the beginning of the oscillation will have traveled in a vacuum in that time a distance c times T or what is normally designated λ, the wavelength.

Later parts of the cycle will leave the origin at a later time and therefore not travel as far. The sinusoidal generating field will be reflected as a sinusoidal field along the X-axis.

A remark should be included here about the sinusoidal solution to the wave equation which Maxwell derived from his electromagnetic theory. It is of the form:

$$E = E_o \sin\left[2\pi\left(\frac{x}{\lambda} - \frac{t}{T}\right)\right] \tag{1}$$

This equation contains both the time dependence and spatial variations of a plane sinusoidal radiation field. If we examine the strength of the electric field up and down the X-axis at one instant in time (t = constant), then we get Figure 6. If we examine over a period of time the strength of the field at one point on the X-axis (x = constant), then the argument of the sine must be constant, i.e., $2\pi(x/\lambda - t/T)$ = constant. Choosing x to be zero when t is zero, the constant will be zero. Then we have $x/\lambda = t/T$ or rearranging $x/t = \lambda/T$. But λ/T is just the speed of propagation c in the vacuum case we are examining. Therefore a feature propagates down the X-axis at speed c.

2. Phase Differences Between Two Waves

In Figure 7, the concept of <u>phase difference</u> is illustrated. If two electric fields similar to that one which generated the radiation of Figure 5 are allowed to begin varying at two times separated by 1/4 of their oscillation period (T/4), then their separate wave forms would appear as in Figure 7. Wave 2 would "lead" wave 1 in phase angle by 1/4 period or $\pi/2$ radians in the argument of the generating sine function. Both would be identical in frequency, maximum amplitude, and speed of propagation, but they would be "out of phase" by 90^o with each other. Phase differences can occur because two sources of similar frequency radiations are out of phase;

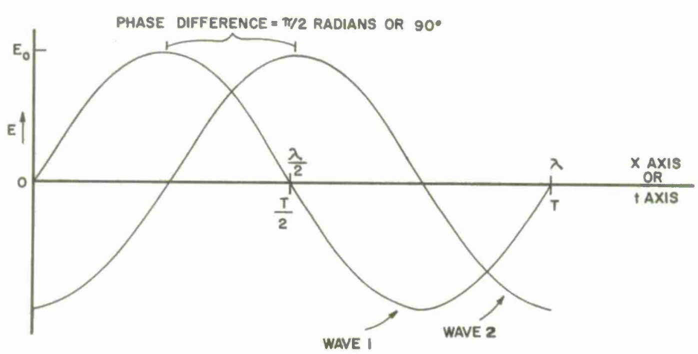


Figure 7. Phase difference for similar waves. The spatial and time dependence of two similar waves with a phase difference of $\pi/2$ radians or 90^o. The same phase difference would also exist for the two associated magnetic fields.

or they can occur at a distance from a single source of radiation when one part of the signal must travel a longer path to the point of detection than the other parts; or two fields oscillating in different directions (say along the Y and Z axes) may be propagated at different speeds through a medium, changing, therefore, their phase relationship with distance.

Two signals which show a steady or easily described relationship between their phases are said to be <u>coherent</u> with each other. Coherent signals will combine in the manner shown in Figure 8. This will be the case for almost all sources of electromagnetic waves used in communication systems. If several waves, on the other hand, have randomly changing phase differences, they are said to be incoherent and they will not combine as straightforwardly as in Figure 8. We will not discuss this latter possibility; most of the effects we are interested in are related to the coherent condition.

Figure 8 shows the consequences of phase differences between two signals. Utilizing the additive nature of the electric vector at each point along the X-axis, the resultant signal can have any value from twice the amplitude everywhere, to zero amplitude everywhere, when two identical waves are shifted in phase and combined.

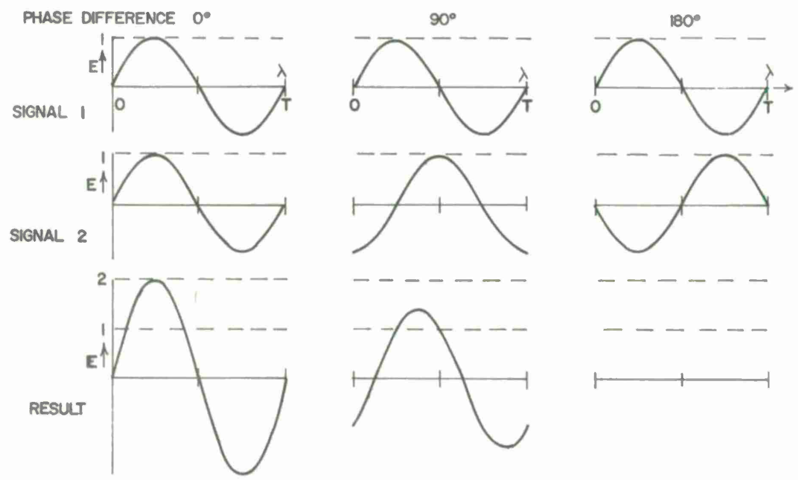


Figure 8. Some combinations of two coherent signals. The effect of phase difference on the combination of two otherwise identical plane signals can lead to a resultant signal which may be anything from twice as strong (phase difference of 0°) to zero (phase difference of 180°). In a vacuum the phase relation is preserved in time as well as space, hence the double labeling of the horizontal axis.

3. Polarization of the Wave

Another important property of electromagnetic radiation is its polarization. This term refers to the direction in which the electric or magnetic fields point as a function of time. If the direction of the field always remains along a line in space, it is called linearly polarized. If the vector magnitude (length) remains unchanged but the direction changes at a constant angular rate so that the tip of the field vector describes a circle, it is called circularly polarized. If the magnitude and direction change in such a manner that an ellipse is described, it is called an elliptically polarized field. If a field vector shows no periodic behavior at all, then it is said to be randomly polarized. Figure 9 illustrates the first three possibilities and how they can occur from the combination of two sinusoidal perpendicular electric fields.

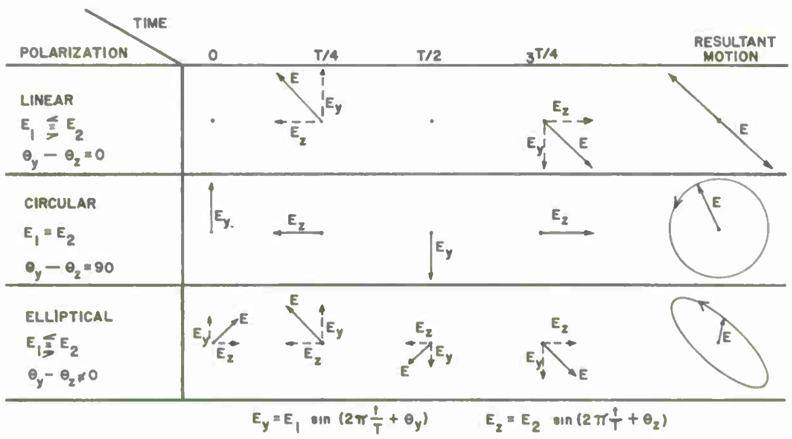


Figure 9. Possible polarizations of an electric field. Various polarizations can be generated by vector addition of sinusoidal components with varying phase differences $(\Theta_y - \Theta_z)$ and amplitudes (E_1, E_2).

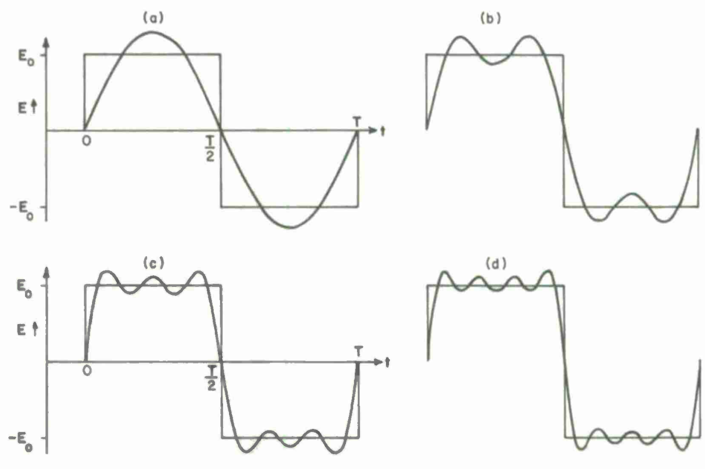
Many times in experimental work, it is helpful to study one linear component of a more general field; sometimes this is unavoidable because the antenna being used is only able to accept one direction of linear polarization. The state of polarization which the radiation is in can often be helpful in determining either its origin or through what physical media it has propagated before reaching the detection equipment.

With the discussion of polarization we have introduced the last necessary concept for fields whose amplitudes vary sinusoidally. Each of the properties are important because, in any interaction between a radiation field and the earth or atmosphere, one or more of them will be changed by that interaction.

E. THE SYNTHESIS OF ARBITRARY SIGNALS FROM SIMPLE SINUSOIDAL WAVES

The preceding paragraphs dealt with pure sine wave oscillations of the radiation producing fields. Are they applicable to real signals? Strictly speaking, there are no pure sinusoidal signals in nature. However, the fact remains that all concepts referenced to sinusoidal oscillations may be used with complete clarity on most signals for one of two reasons: (i) some signals differ from true sinusoidally produced signals by a negligible amount; (ii) real signals of arbitrary waveform and finite length may be synthesized to any degree of accuracy from pure sinusoidal components. An example of the latter is given in Figure 10.

Figure 10 shows the first 4 components of what must be an infinite series for a perfect synthesis. In reality, enough terms are used so that the synthesized waveform fits the true waveform well enough to understand the frequency dependence of the signal. In the case of the square wave of Figure 10, if higher frequencies were attenuated more strongly, say as it propagated through some medium, then the signal would gradually change its shape toward the approximation given by fewer and fewer of the sine series, i.e., towards (a) in the figure.



$$F(t) = \frac{4}{\pi}\left[\sin\left(2\pi\frac{t}{T}\right) + \frac{1}{3}\sin\left\{3\left(2\pi\frac{t}{T}\right)\right\} + \frac{1}{5}\sin\left\{5\left(2\pi\frac{t}{T}\right)\right\} + \frac{1}{7}\sin\left\{7\left(2\pi\frac{t}{T}\right)\right\}\right]$$

Figure 10. Synthesis of a square wave by combination of simple sinusoidal components. The square wave is approximated by the function F(t) above: (a) first term of F(t); (b) first two terms; (c) first three terms; (d) all four terms. A perfect fit would only occur if an infinite number of sine curves were combined. As can be seen, however, only a few terms can give a remarkably close fit.

The possibility for synthesis of arbitrary signals allows us to talk about the effects on a signal in terms of frequency, with the understanding that those effects apply only to the components of the signal oscillating at or near that frequency. It is quite possible and indeed expected that there will be alteration of the waveform of a non-sinusoidal signal which passes through a medium whose effects depend on frequency (a dispersive medium).

F. PHASE SPEED, GROUP SPEED, AND SIGNAL SPEED

The speed of a pure sinusoidal wave in a vacuum or a material is its phase speed. However, if two waves of slightly different frequency are combined, the phase speeds of the individual waves would now have less importance because of the structure of the new signal produced. Figure 11 shows what occurs when two signals, slightly different in frequency, are combined. The waves are in phase (zero phase difference) in some places and show as much as a 180° phase difference in other places. The resultant waveforms shows, therefore, in the former case, an amplitude which is the sum of the amplitudes of the combining signals and at other places it shows their difference. The envelope of the combined wave is now a periodic function broken into "beats" or "groups".

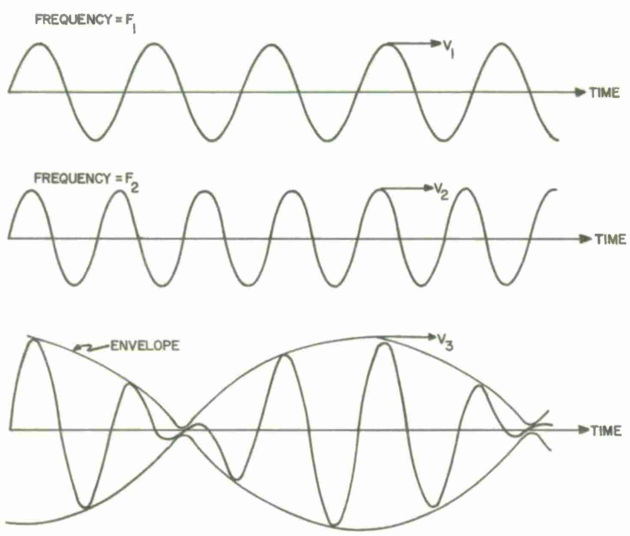


Figure 11. The formation of groups from two signals of slightly different frequency. If the phase speeds, V_1 and V_2, are equal, then the groups move with this same speed. If V_1 does not equal V_2, then V_3, the group speed, is different from both.

If the medium in which these groups were formed shows a constant phase speed for every frequency, these groups travel at this speed. If, however, the phase speeds of the combining waves depend upon frequency, the group speed shows a distinct speed of its own. In all cases where a group speed is meaningful, energy is propagated at the group speed. This concept will be found to be important for a number of geophysical media we will be concerned with.

Although the phase speed and group speed are easily understood concepts, it is not always completely clear what the relationship is between these ideas and the speed of the real signal. A fundamental consideration must be what the criterion is for recognizing the beginning of the signal. Is it the first time the average amplitude of the radiation exceeds a certain level? Or is it the passage of the first recognizable waveform? Clearly, the signal speed for both criteria depends upon how we answer the questions posed. The definition of such criteria, and their application, go well beyond the scope of this report.

It is sufficient now to emphasize that when dealing with reflection and refraction, we are mostly concerned with the phase speed. For transmission of energy and momentum, the group speed is more applicable. How both of these relate in detail to signal speed we will not take up.

II. ELECTRICAL AND MAGNETIC PROPERTIES OF MATTER

A. ATOMIC STRUCTURE

Although we will be interested almost exclusively in the gross effects which matter has upon electromagnetic radiation, it will be of interest and use to outline the physical make up of matter on an atomic scale, for the gross effects are but the result of innumerable atomic interactions.

The smallest particle which can still be associated with a macroscopic material is the atom. Its scale of diameter size is in the vicinity of 10^{-8} cm. It has a positively charged nucleus composed of numerous varieties of sub-atomic particles. The nucleus retains a net positive charge equal to its atomic number multiplied by the fundamental unit of electrical charge. Its diameter is of the order of 10^{-12} cm. Electrons with one unit of negative electrical charge circulate about the nucleus. In an electrically neutral atom, the number of electrons surrounding the nucleus is the same as the atomic number of the atom, balancing exactly the positive charge of the nucleus.

The atomic scale of things differs from the world which we will be concerned with in this report in one overwhelmingly important way. It is found from theory and experiment that all quantities which describe the atomic world come in multiples of discrete units. In the macroscopic world in which we live, we intuitively feel that there is a continuous choice of quantity for electrical charge, mass energy, etc.. This is true only because the smallest directly measurable amounts of any of these properties represent countless numbers of their fundamental units. Even though the quantum discreteness of the atomic scale is mostly non-applicable to our discussion (with the notable exception of spectral lines) it will be helpful and illuminating to refer on occasion to the atomic origin of certain macroscopic effects.

B. THE ELECTRICAL PROPERTIES OF MATTER

To be more specific, the effects of matter on radiation are in large measure due to the fact that it is possible for electrons to become separated from their nuclei, or in some way reorient themselves with respect to those nuclei.

The first condition applies to the general class of matter known as conductors. In fluids and solids, atomic electrons farthest from the nucleus are sometimes easily detachable. In particular, if many atoms of a material which exhibits this property are packed together, these outer electrons circulate throughout the entire assemblage. It is the average drift of these free electrons in some direction which is the electrical current.

All substances exhibit conductivity to some extent, but the range is enormous. Copper will conduct electricity 10^7 times more easily than sea water. And sea water will conduct 10^{12} times more easily than ordinary glass. The division between good and bad conductors, therefore, is a matter of choice. Sea water, as a passing note, is generally considered to be a good conductor.

A perfect dielectric, on the other hand, has no free charges. It will be able to interact with an impressed field, however, by showing molecular polarization. That is, an electric field, in the case of symmetrical molecules, will cause the negative and positive charges to shift away from their average position, or in the case of unsymmetrical molecules, to align themselves partially with the field. Figure 12 illustrates these points.

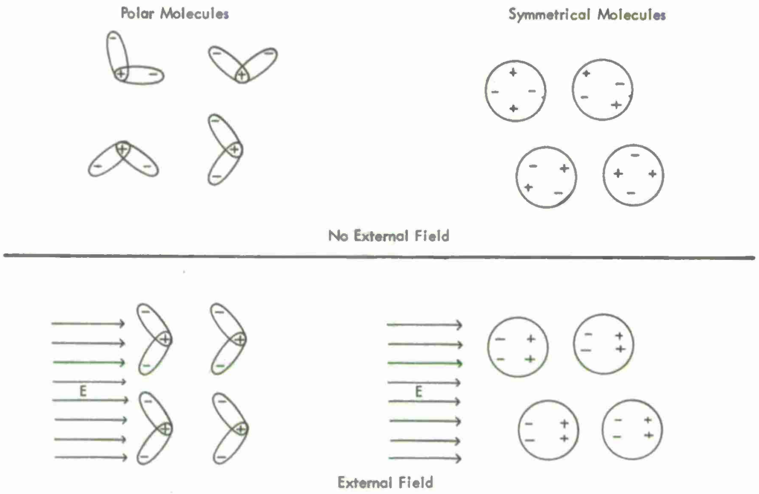


Figure 12. Polarization of electrically symmetric and asymmetric molecules. In the presence of an external electric field, polar molecules show alignment while symmetrical molecules show internal polarization.

For electrically unsymmetric or polar molecules, the average number of particles aligned with the field will be less the higher the ambient temperature of the material, since thermal agitation of the molecules will tend to randomize their alignment. There will be no temperature effect for the induced polarization on symmetrical molecules.

Substances are called dielectrics when the approximation to the ideal conditions above is close. No perfect dielectrics exist, since all substances conduct electricity to some extent. The division is usually made at the point where the conduction and the polarization phenomena are of equal importance.

It is known from experiment and theory that the conductivity of a material and the ease with which it may be electrically polarized play a very large role in determining how well and in what manner an electromagnetic wave will propagate through it. The magnetic properties can also be important, but at many frequencies, for many media, the magnetic effects are so small that the electrical properties determine their propagation characteristics.

C. THE MAGNETIC PROPERTIES OF MATTER

Three types of magnetic materials exist. They are called diamagnetic, paramagnetic, and ferromagnetic. A diamagnetic material contains no permanent magnetic dipoles (see Figure 13) but they may be induced by an external magnetic field. The magnetic dipoles are the result of changes in the motion of electrons around the core of an atom caused by the impressed field. The changes have the same effect as setting up small ring currents within the atom similar to the magnetic dipole ring current illustrated in Figure 4. Just as in the case of the induced electrical polarization, the orientation of the induced magnetic dipoles or current rings is such as to oppose the impressed field. Diamagnetism is usually a very weak phenomenon although it occurs in all matter.

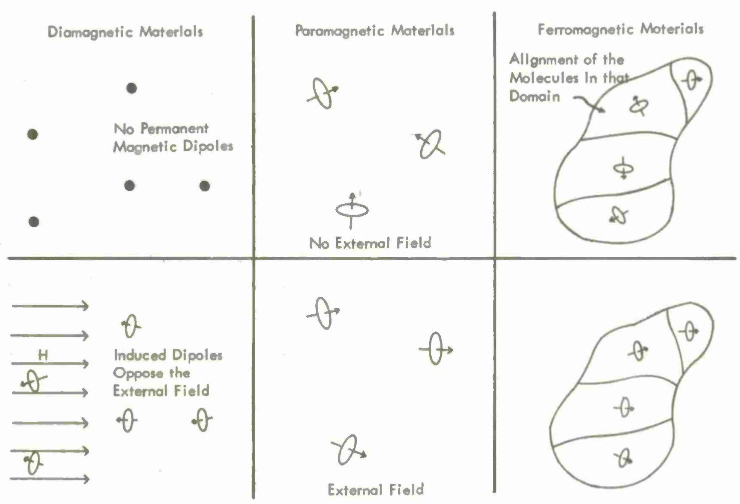


Figure 13. Diamagnetic, paramagnetic, and ferromagnetic materials. Diamagnetic, paramagnetic, and ferromagnetic materials all react to the influence of an external magnetic field.

Paramagnetic materials do contain permanent dipoles. The dipoles result from the movement of electrons about the nucleus and from the spin of the electrons themselves (both tiny current rings). Some current rings cancel others by being oriented oppositely. This is exactly true in diamagnetic substances, but in paramagnetic material there remains a residual magnetic dipole effect. When an external magnetic field is applied to a paramagnetic substance, the permanent dipoles tend to align themselves with the field. Thermal agitation, as in the case of electric dipoles, opposes the aligning process. Paramagnetism in most instances is a larger effect than diamagnetism. Both are usually proportional to the applied field.

Ferromagnetism is an extreme case of paramagnetism. Over small macroscopic regions, quantum mechanical effects cause the atomic magnetic dipoles to align themselves without the influence of an external field. Such a region is called a domain. Domains can be lined up by an external field very easily, resulting in large total magnetic fields.

Ferromagnetic substances make up very little of the earth-ocean-atmosphere environment and are therefore a negligible factor in the gross effects of the environment upon electromagnetic radiation. Diamagnetic substances, though universal, have such a small effect on radiation that, for practical investigations, their magnetic properties are almost always neglected. Only paramagnetic substances will be of interest to us. One such substance is oxygen, and it will be considered in our discussion of absorption of electromagnetic radiations by the atmosphere.

D. IONS AND ELECTRONS

In the highest parts of the atmosphere, the sun's radiation is so powerful that electrons are actually totally separated from their nuclei when this radiation interacts with atoms. The result is a plasma of free, light, negative electrons and heavy, positive nuclei called ions. The free electrons are able to interact very strongly with a radiation field propagating through such a region. The result is very special and must be considered in some detail to understand all of the consequences. In the highest regions of the atmosphere, the results are complicated because of the importance of the magnetic field of the earth on the movements of charged particles. These special conditions are treated in the discussion of the ionosphere.

III. IMPORTANT MACROSCOPIC EFFECTS OF MATTER ON ELECTROMAGNETIC RADIATION

Before we enter into a full discussion of the effects of the earth and atmosphere on radiation, it will be valuable to discuss in general terms some of the most important of these expected effects. Accordingly, several sections follow in which the physical ideas of propagation speed, reflection, refraction, diffraction and scattering, absorption, dispersion, and changes of polarization are outlined and in some cases schematically represented.

A. CHANGES IN THE SPEED OF PROPAGATION

Electromagnetic radiation in a vacuum propagates at what is called the speed of light. This speed seems to be one of the fundamental constants of the universe, for, as Einstein showed, no matter how a beam of light is launched or received in a vacuum it will always appear to the sender and receiver to be traveling at the constant speed c ($c = 3 \times 10^{10}$ cm/sec). The consequences of this fact have led us into a new realm of physics in which our normal three dimensional space must be expanded to accept a new dimension in which time itself becomes a variable. Furthermore, the speed of light appears to be the upper limit at which energy or momentum can be transferred in the material world. Therefore, we expect and find that electromagnetic radiation which carries energy and momentum in a material medium does so invariably at speeds equal to or less than c.

The exact speed of propagation of a signal in a material medium is a complex question to answer because it is difficult to define exactly what the signal is. But, in general, four things may be said about the speed of propagation: (i) it depends upon the ease with which the electric and magnetic dipoles can be polarized by an external electric or magnetic field; (ii) it depends upon the conductivity of the medium; (iii) it depends upon how these parameters change as the frequency of the radiation changes; and (iv) it will always be equal to or less than c.

It will be recalled that in the discussion of simple sinusoidal waves the phase speed in a material medium could be less than, equal to, or greater than the speed of light in a vacuum. This is true because the phase speed, in most instances, is different from the signal speed. The signal is always made up of many simple sinusoidal waves, and when it is found that these simple waves travel faster than light, it is always the case that their combination leads to signal speeds less than c.

In many low conductivity media, especially at radio frequencies and below, the complexities of determining the signal speed greatly diminish. The propagation speed in these cases is the square root of the parameter which describes the ease with which the atomic electric dipoles are formed or aligned.

B. REFLECTION

A commonly observed effect of matter on electromagnetic radiation is reflection. We see electromagnetic energy in the form of light waves being reflected every time we look into a mirror or observe a placid lake surface. In general, there is some energy reflected from any material body upon which electromagnetic energy is impinging.

The physical process of reflection involves the oscillatory motion of both free charges (free electrons or ions) and bound charges (atomic electrons, nuclei, etc.) in the surface of the reflecting material; these charges are set in motion by the electric and magnetic fields of the impinging radiation. The oscillatory motion of the charges causes a secondary field to be produced which radiates both back into the medium from which the original field came and also forward into the reflecting medium. Exactly how much energy is reflected and how much passes from one medium into the other is determined by the boundary conditions which must be satisfied at the interface. These conditions are a consequence of Maxwell's equations.

Figure 14 shows a particularly simple case of reflection at the interface between two media. Medium 1 on the left has no free charges. Its electrical and magnetic qualities are such that the speed of propagation is reduced to V_1, below the velocity of propagation in a vacuum, c. The ratio c/V_1 we will designate as n_1. It is called the index of refraction for medium 1.

Similarly, on the right, is medium 2 with electrical and magnetic qualities which slow the speed of propagation to V_2, less than V_1. Its index of refraction is therefore n_2, greater than n_1. It also does not contain any free charges.

Now, let E_o be the maximum amplitude of the electric field strength approaching the interface and H_o the maximum amplitude of the associated magnetic field strength. Let E_1 and H_1 and E_2 and H_2 be similar designations for the reflected and transmitted waves respectively.

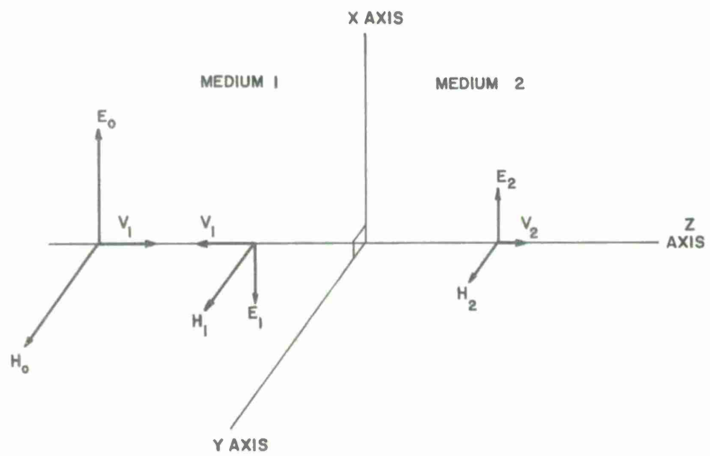


Figure 14. Reflection at the interface of two electrically different media. The propagation speed in Medium 1 (V_1) is greater than in Medium 2 (V_2); consequently the index of refraction in Medium 1 ($n_1 = c/V_1$) is less than in Medium 2 ($n_2 = c/V_2$).

Then, from the theory, it is found that the amplitude of the reflected wave is simply:

(2)
$$E_1 = - \frac{n_2 - n_1}{n_2 + n_1} E_o$$

The energy contained in such a wave is proportional to E^2. The fraction of energy reflected, therefore, is given by E_1^2/E_o^2 and is designated R_1 the <u>power reflection coefficient</u>. We can see that the larger the difference between the propagation speeds in the media, the greater the percentage of energy which will be reflected.

The transmitted amplitude is given by:

(3)
$$E_2 = \frac{2n_1}{n_2 + n_1} E_o$$

If $n_1 = n_2$ (propagation speeds equal), it is clear from equations (2) and (3) that all of the incident wave would be transmitted.

When the amplitude of the reflected and transmitted waves are added up, the boundary condition which was originally applied becomes evident:

$$E_o + E_1 = E_2$$

The reversal of the direction of E_2 from E_o in the diagram, making it possible (by the right hand rule) for the reflected wave to move back into medium 1, is caused by a 180^o phase shift of the reflected electric field at the boundary. If n_1 were greater than n_2, the magnetic field would have undergone the phase shift instead of the electric field.

The simple example of Figure 14 does not include either the effects of oblique incidence or the effects of free charges on the amplitude of the reflected wave. Oblique incidence leads to the fact that if a plane is defined which includes the direction of the incident wave and the normal to the surface of reflection, the electric field and magnetic field components in this plane are reflected according to a different law from the polarization component perpendicular to this plane. This will be shown to be important in the case of radio wave reflections from bodies of water and land. The introduction of free charges into the reflecting medium increases the reflection coefficient at all angles of incidence.

C. REFRACTION

Upon crossing the boundary between two media with different speeds of propagation, the transmitted part of the radiation suffers a change in direction at all angles of incidence except perpendicular to the boundary. This phenomenon is called refraction. For simple media, it is easy to show why the difference in propagation speed causes such a directional change. Refer to Figure 15. A portion of an incoming plane wave front arrives at an oblique angle θ_1 and therefore enters the second media over a short period of time, say Δt. The wave edge still in medium 1 travels over this time a distance $S_1 = V_1 \Delta t$. The wave edge which had entered medium 2 at the beginning of Δt has traveled only $S_2 = V_2 \Delta t$, however. The wave front, consequently, has tilted toward the normal to the interface if V_2 is less than V_1 or away from it if V_2 is greater than V_1. The relationship follows that the index of refraction in medium 1 times the sine of the angle of incidence is equal to the index of refraction in medium 2 times the sine of the angle of refraction, i.e. $n_1 \sin \theta_1 = n_2 \sin \theta_2$. This relationship is known as Snell's Law.

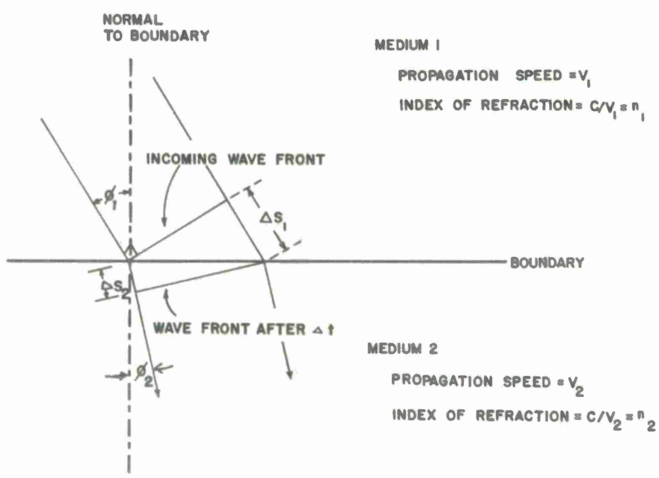


Figure 15. A simple physical illustration of refraction. The illustration shows that if the speed of propagation decreases as the wave front passes into an electrically different medium then the direction of propagation of the wave front turns toward the normal to the boundary between the mediums.

D. DIFFRACTION AND SCATTERING

The definitions of diffraction and scattering are best made in terms of word pictures. Consider a thin opaque screen upon which falls a plane electromagnetic wave at perpendicular incidence. Let us say that the screen is infinite in its height and width but a small rectangular slit exists through it somewhere. If the electromagnetic field behind this screen is measured, it will be found that the portion of the wave which passed through the slit has spread out, so that if the slit width is small enough in comparison to the wavelength of the radiation, radiation can be detected behind the screen at angles as great as 90° away from the direction of the original plane wave. The initially plane wave has been diffracted as it passed the slit.

An approximate but straightforward means of calculating the effects of this type of diffraction and others wherein the initial wave front encounters a totally absorbing obstacle was originally given by Huygens and later modified by Fresnel. The principal idea of the former was to assume that every point on an advancing wave front was the source of a spherical wavelet. Then at any point ahead of the advancing wave a new field could be calculated from these new sources by adding up the contributions of each wavelet. Fresnel modified the procedure by taking into account both amplitude and phase of the wavelet. This has been shown to be a good approximation to the real diffraction pattern when the wavelength of the radiation is small compared to the dimensions of the obstacle. An example of the intensity pattern formed from a small slit is shown in Figure 16a.

The general diffraction problem, of which the Huygens-Fresnel problem is a special case, occurs when an electromagnetic wave encounters an obstacle of given size, form, and composition. The diffracted energy may then be defined by removing the object and substituting for it a completely black screen (i.e., one which absorbs all the electromagnetic energy falling on it) with exactly the same cross-sectional dimensions as the original object. An initially plane wave falling on this screen will be modified because of it. The electromagnetic fields which are needed to explain the deviations from the original plane wave represent the diffraction fields.

Another aspect of the general problem is associated with the energy intercepted by the object which, for the diffraction problem, was assumed to have been totally absorbed. In reality the incident wave forces the bound and free charges within the body to oscillate synchronously with the applied field. These oscillating charges radiate energy internally and externally in a manner which depends upon the composition and geometry of the body. The incident energy is said therefore to be scattered by the object.

The final electromagnetic fields which result are then a combination of the incident electromagnetic fields, the diffracted fields, and the scattered fields. Of course, if no energy is converted to other forms within the body, the total energy in the new fields must be just equal to the energy in the original plane wave when the object did not interfere. All real material objects, however, do absorb some electromagnetic energy (normally transforming it into heat) so that the new fields will not carry as much energy as the original undiffracted, unscattered field.

Two special cases of scattering arise when the ratio of wavelength of the propagating signal to the dimensions of the scattering objects is either very small or very large.

In the former case, one may employ the techniques of geometrical optics. Rays of the radiation can be traced as they are reflected and refracted at the internal and external boundaries of the diffracting object to establish the resulting intensity pattern. Light passing through raindrops can be analyzed successfully by this method.

In the latter case, the wavelength of the radiation is very much larger than the objects encountered. Lord Rayleigh developed the applicable theory of scattering and it is called Rayleigh scattering in his honor. The result showed that the energy scattered was dependent upon the inverse of the wavelength to the fourth power, i.e., $1/\lambda^4$.

This explains why we see a blue sky away from the sun. The shorter wavelengths of blue light are scattered out of the direct sunlight much more effectively than red light by the atoms and molecules of air. This results in a blue cast to the sky and a reddish tint to the sun. As the sun descends towards the horizon and its rays must penetrate more and more of the scattering molecules, it becomes more and more deeply red. The blue is almost totally removed. Figure 16b may be of help to visualize how small particles may scatter electromagnetic energy.

In between the two extremes discussed above there is a transition region where the analysis becomes very difficult. One effect in particular, however, should be mentioned. Objects whose dimensions are similar to the wavelength of impinging radiation can show scattering effects of very large amplitude. These anomalies occur at quite sharply defined wavelengths and are due to the excitation of natural modes of oscillation within the object. Essentially, what happens is that a wave crest from the external field arrives at a surface just as the wave from an internal reflection reaches the same surface. The internal electromagnetic field becomes larger at every reflection until internal conduction and other losses just equal the external supply of energy. The scattered fields for these natural resonant oscillations can be very large. Such a phenomenon is usually called resonance scattering.

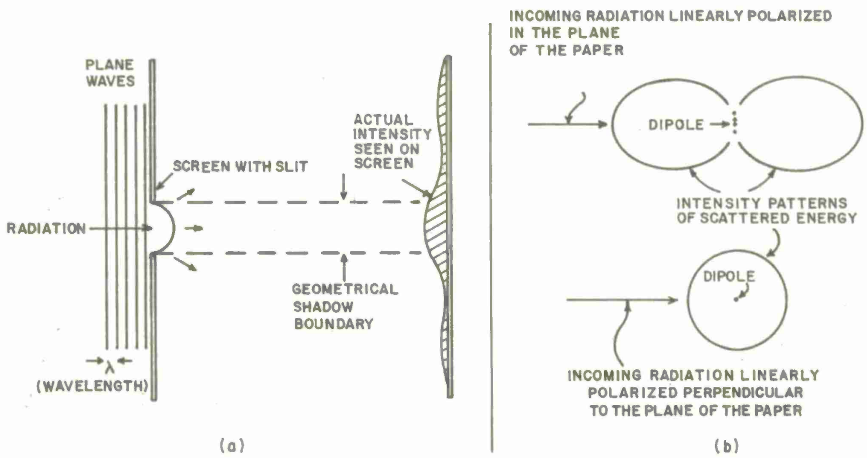


Figure 16. Aspects of the diffraction and scattering problem. (a) A plane wave initially traveling in one direction passes a narrow slit; the direction of the emerging waves is quite complex. (b) An incoming linearly polarized electromagnetic wave causes a dipole to oscillate. The dipole radiates energy in all directions except parallel to the polarization of the exciting field. Energy is redistributed or scattered away from the original direction of propagation.

E. ABSORPTION

It has been pointed out that the electrical charges of which matter consists are forced to move under the influence of the changing electric and magnetic fields of a propagating electromagnetic signal. Much of the energy required for these motions is either returned to the signal or radiated away in new directions (scattered). However, some of this energy disappears in other ways. Free electrons can collide with heavier neutral particles and the energy they would normally reradiate contributes instead to the random thermal energy of the gas. Electrons bound to atoms can be forced, by the radiation field, to take up different modes of circulation around the nucleus or they can be removed completely (ionized). Both processes require energy from the signal, and this energy is not immediately returned. In the quantum world of molecules, the vibrations and rotations of the particles can be changed by an external electromagnetic field, again at the expense of the signal energy. Compounds can increase their chemical energy, liquids and solids their internal energies, all from the energy originally part of the radiation field. Any process, including those above, which removes energy from the electromagnetic wave is included under the general heading of absorption.

For atmospheric gases, some detail can be put down to clarify the basic physical ideas which govern absorption. Let us examine a single molecule to expose the numerous interactions which it may have with the ambient electromagnetic wave.

If, at first, one neglects the interaction between modes of energy within an isolated molecule and the molecules rest mass energy we would find its total energy to be:

(4) $$E = E_e + E_v + E_r + E_t$$

where E_e is the electronic energy, E_v the vibrational energy, E_r the rotational energy, and E_t the translational energy of the molecule. Quantum mechanics shows us that E_e, E_v, and E_r can only take on discrete values. The energy emitted or absorbed by these modes is therefore also quantized and follows the Planck relation given in Equation 5:

(5) $$E_f - E_i = hf$$

where E_f is the final energy of the molecule and E_i is the initial energy of the molecule for that mode; h is Planck's constant; and f is the frequency of the radiation emitted or absorbed.

Changes in the molecular energy states of these three modes account for spectral line absorption. These "lines" are narrow frequency regions where a particular gas under given physical conditions absorbs very strongly compared with frequencies nearby. For the atmospheric gases, the lines of lowest frequencies lie in the microwave region. (For a definition of frequency regions, see Figure 21.) These lines proliferate at wavelengths shorter than 1 mm so that they overlap and absorb very strongly over broad bands of frequencies lying in the infrared.

It is important to remember that absorption implies some interaction between the electromagnetic wave and the electric or magnetic properties of the gas. The electromagnetic wave must cause some movement, reorientation, stretching, vibration, rotation, or internal reorganization in the molecule in order to lose energy. In the case of free electrons, the electrical field attempts to move them physically, setting them into vibrational motion at the frequency of the wave. Absorption in this case occurs when the moving electrons collide with other particles in the gas and lose kinetic energy derived from the wave. Absorption of this type shows a continuous spectrum, without the sharp lines seen in those cases where quanta of energy are emitted or absorbed. Continuous absorption occurs in the ionosphere where free electrons are abundant.

Where electromagnetic waves come into contact with fluids and solids, many processes similar to those which act on individual molecules in a gas act to extract energy. Conduction electrons are set in motion, electron orbits can change, crystals can vibrate, photoionization can occur; in short, all the electric and magnetic properties of the liquids and solids can interact with the wave, and in so doing absorb energy.

F. DISPERSION

A dispersive medium is simply one in which the phase speed of electromagnetic waves is dependent upon frequency. This is due to the fact that the electric and magnetic properties of the medium change with frequency. Technically speaking, all media are dispersive. But just like there are approximately perfect dielectrics, there are some geophysical media which are very close to being non-dispersive over certain frequency bands.

The importance of dispersion may be illustrated by a simple example. Refer to Figure 10. If the electric and magnetic properties change as the frequency changes in a medium transmitting the square wave shown synthesized in the figure, then the speed of propagation will most likely be different for each harmonic component. If this be the case, then, after traveling some distance through the medium, the individual harmonic components will be separated in space resulting in severe distortion of the square wave shape.

Dispersion can be explained satisfactorily by the mechanical properties of the medium in question. The orientation of the dipoles and the polarization of charges have natural frequencies of oscillation; that is, they will be very easily set in motion at certain frequencies. At lower and higher frequencies they will follow the oscillatory motion of the radiation field, but at reduced amplitude. Far away from the regions of natural oscillation, most media show little change in electric and magnetic polarizability with frequency. However, as the natural frequency of oscillation is approched, the properties change significantly and dispersion results.

G. CHANGES IN POLARIZATION

Since any vector may be decomposed into perpendicular components, any interaction with matter which acts on one of these components in a manner different from how it acts on the other will, as a necessary consequence, change the polarization of the original signal.

This will happen, in general, for reflections and refractions. Figure 17 shows two other ways in which polarization changes are induced by a propagating medium. If one of the perpendicular components of an originally linear polarized field is attenuated more than the other component, then the polarization vector will rotate toward the less attenuated member. If, however, no differential attenuation occurs, but one perpendicular component travels slightly slower than the other member, the phase difference between them is altered. Linearly polarized radiation becomes elliptical, circular, elliptical, linear, etc., as the phase difference moves through each 360° change.

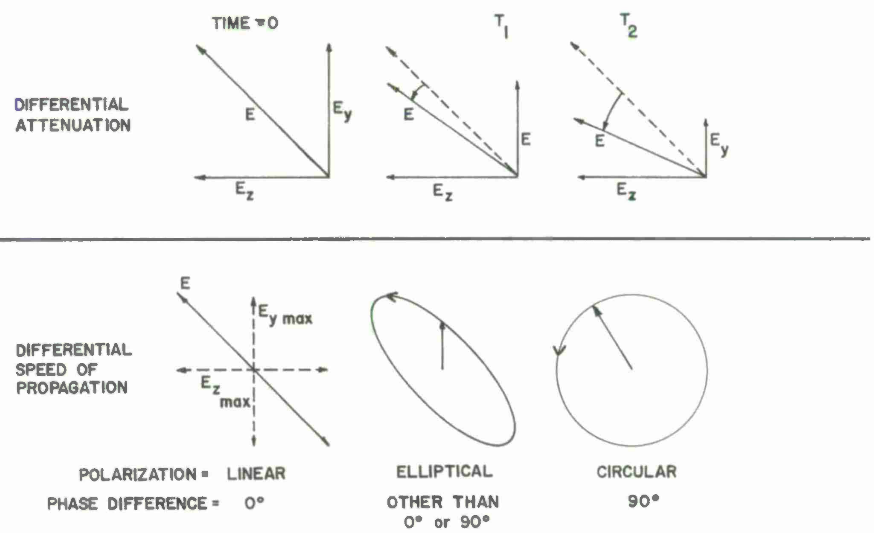


Figure 17. Two ways in which polarization changes might be induced by a propagating medium. The upper half shows the polarization changes due to differential attenuation of one component of a signal. E_y decreases with time, and the resultant electric vector rotates toward E_z. The lower half shows the polarization changes due to a medium in which the speed of propagation (index of refraction) depends upon the orientation (polarization) of the electric vector.

IV. PROPAGATION OF ELECTROMAGNETIC WAVES IN CERTAIN IDEALIZED CIRCUMSTANCES

A. ELECTROMAGNETIC WAVES PROPAGATING IN A VACUUM

When we refer to effects which the earth and atmosphere have upon communication signals, it will be very useful to already understand clearly the general characteristics of electromagnetic wave propagation in certain ideal environments. The first of these we have referred to often before. We now collect and repeat the properties of an electromagnetic wave moving through a vacuum:

Since there are no charges, no dielectrics, no magnetic materials, the speed of the wave is the maximum attainable in the universe, the speed of light c. All frequencies travel at this speed so that no distortion from dispersive effects can occur. The propagation is rectilinear and no energy is lost to absorption. These characteristics will be contrasted with those found in matter.

B. ELECTROMAGNETIC WAVES PROPAGATING IN A DIELECTRIC

1. Non-dispersive Dielectric

The phase speeds of all frequencies in a non-dispersive medium are equal. In general, this condition is met only in a vacuum and a perfect dielectric. In fact, in a perfect dielectric, which has the same electric and magnetic properties everywhere (homogeneous), and in all directions (isotropic), the only difference from free space propagation is the speed of the wave. It is always less than c. As a consequence, the wavelength in the dielectric is also less than in free space for a wave of the same frequency.

If the dielectric is non-uniform in space, i.e., its electric and magnetic properties change along the wave path, then refraction can occur, and the direction of propagation can be altered. Reflection and refraction can occur, as already pointed out, at the boundary of two perfect dielectrics of differing properties. However, in all these cases, no loss of electromagnetic energy is experienced.

2. Dispersive Dielectric

The phase velocity of a medium can change with frequency because: (i) the ease with which the atomic and molecular dipoles can be aligned changes with frequency; or (ii) because there is finite conductivity. It is therefore evident that all media are dispersive to some degree because all have some conductivity. And, generally, the polarizability of the atomic dipoles does change with frequency. These effects may be so slight, however, that some media approach a perfect dielectric very closely over some frequency regions.

Finite conductivity introduces another factor not found in propagation in a vacuum. The conducting electrons, in their movement throughout the material, can be thought of as colliding with the fixed heavy atoms making up the material and imparting some of their energy to the vibrations of these atoms. This energy loss is detected as heat and is called ohmic or resistive loss. The attenuation of the penetrating wave is exponential with distance in a conducting medium.

Figure 18 shows some of the major effects of a wave entering media with different properties.

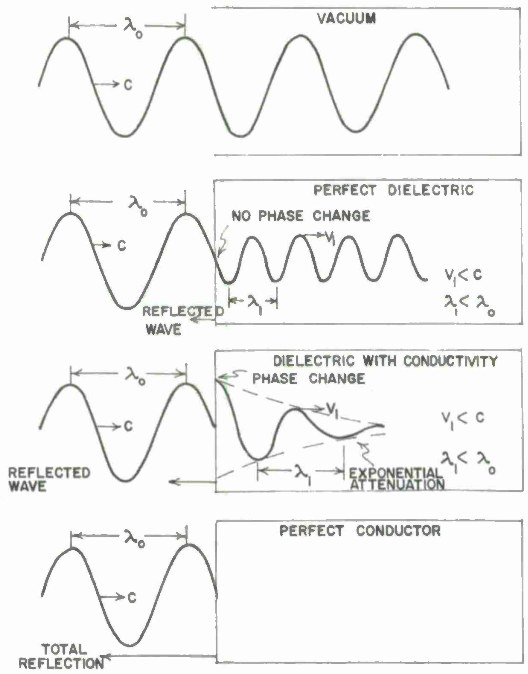


Figure 18. The effects of dielectrics and conductors on a propagating electromagnetic wave. The effects in all cases are restricted to perpendicular incidence and propagation from vacuum into a dielectric whose index of refraction is greater than one. The perfect dielectric may or may not be dispersive; the dielectric with conductivity will always be.

C. ELECTROMAGNETIC WAVES PROPAGATING IN A PLASMA

The atmosphere above about 60 km shows a rapid increase in ionized atoms and molecules and free electrons. An electromagnetic wave which propagates through such a dilute ionized gas, or plasma, experiences several effects which differ markedly from propagation in a dielectric. We will investigate first a plasma which has no permanent external magnetic field affecting it.

In order to understand more easily what occurs in a plasma, we will define a common quantity which will be very useful in the discussion. It is the so-called plasma frequency. It depends solely on the square root of the number of electrons per unit volume, η, when electrons dominate the plasma, as they do in the ionosphere. We will designate the plasma frequency ω_p and, therefore, we can write

(6) $$\omega_p \sim \sqrt{\eta}$$

Then the index of refraction for the non-magnetic case (i.e., the speed of light divided by the phase speed in the medium) is given simply by

(7) $$n = \sqrt{1 - \omega_p^2/\omega^2}$$

where ω equals 2π times the propagating wave frequency f.

If ω is greater than ω_p, n is real and less than 1. That means that the phase speed in a plasma is faster than the speed of light. But also, n depends upon frequency, and a plasma is therefore dispersive. Groups are formed (see Figure 11) from the combination of the constituent frequencies in a signal, and these groups always travel at a speed less than c.

The opposite condition, that is when ω is less than ω_p, leads to the square root of a negative number for the index of refraction. In effect, the wave is attenuated exponentially as it enters the plasma. Therefore propagation cannot be supported, and total reflection takes place. Since ω_p increases as the electron content of the plasma increases, it can be seen that the greater the electron density, the higher the frequency must be to penetrate such a plasma.

The conditions become more complex when the magnetic field is introduced. When a charged particle is set in motion in the presence of a magnetic field, it will be recalled from Figure 2 that the particle will experience a force perpendicular to its direction of motion. In a uniform field, a free electron will constantly be forced to one side of a rectilinear path because of its motion perpendicular to the lines of magnetic force. Its components of motion along the field lines will be unaffected by the field. The result is the electron spirals along the magnetic field lines. Figure 19 will clarify these ideas.

More important to the propagation of electromagnetic waves is this preference for motion along the magnetic field lines. If the oscillating electric field is parallel to the magnetic field, the electrons feel no constraints and pass the wave as if no magnetic field existed. However, if the electric vector of the electromagnetic field has a component perpendicular to the magnetic field lines, the electrons are constrained and the wave is propagated in an entirely different manner. Phase speed is different between the two components, and therefore refraction is different (see Figure 15). Phase differences (see Figure 7) occur between the components of the wave perpendicular and parallel to the magnetic field, introducing phase changes between them as they propagate (as depicted in the lower half of Figure 17). The total effect of the magnetic field will be more completely brought out in the discussion of the ionosphere which is a plasma subjected to an external magnetic field.

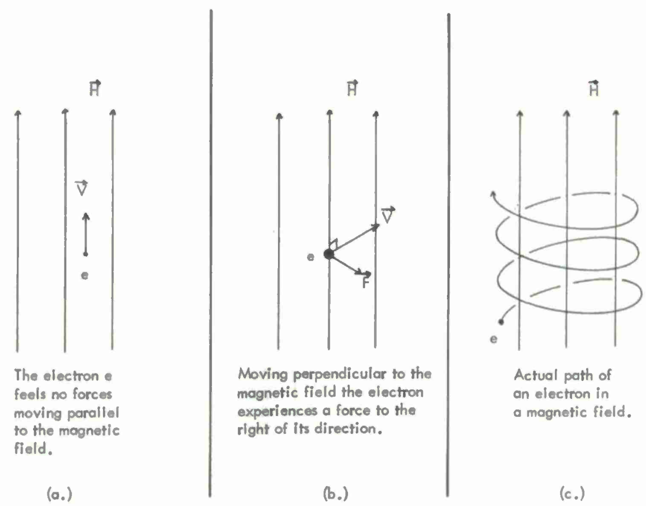


Figure 19. The forces on a moving electron in a uniform magnetic field and the resulting motion. The spiral, of course, can be a closed circle or can stretch to the case of (a.).

D. GUIDED ELECTROMAGNETIC WAVES

In certain instances, conductors and dielectrics can act to guide electromagnetic radiation. That is, most of the energy in the wave can be confined to move within some geometrical shape, or along a boundary. The natural environment provides frequent examples of such behavior. This section will, therefore, identify and discuss several of the most common forms of guided waves.

1. Surface Waves

If a conductor and a dielectric meet at a plane boundary, it can be shown that an electromagnetic wave can be excited which travels along the interface, radiating no power into either medium. This wave is known as a surface wave and is very important for lower frequency propagation on the surface of the earth.

The wave shows certain identifiable characteristics. For a perfect dielectric and perfect conductor, and a perfectly plane surface, the above properties hold. The amplitude of the fields will decrease exponentially into each medium. However, if a finite conductivity exists in the conducting medium, then the waves propagated along the interface are attenuated. The attenuation increases for higher frequencies and lower conductivities. The attenuation may be attributed to resistive losses in the motions of the charges which are set in motion by the wave as it passes over a region of the conducting surface. Because of the movement of the charges, currents within the conductors do exist. That is why this mode of propagation in radio terminology is sometimes classified as under earth current propagation. Figure 20a shows the configuration of the electric field around the interface of a perfect conductor and dielectric.

Irregularities in the surface, curvature of the surface and variations in the electric and magnetic properties of either surface will cause energy to be radiated and thus lost to the wave. Such considerations will be very important when the earth-air interface is examined.

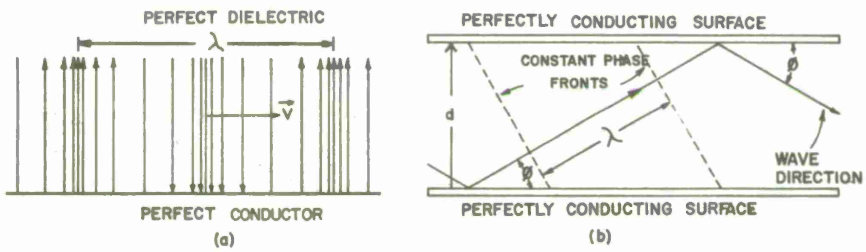


Figure 20. Two examples of electromagnetic propagation via guided wave modes. (a) One type of guided wave is the surface wave, which propagates without loss along the plane interface between a perfect dielectric and a perfect conductor. In the drawing the strength of the electric field is proportional to the spacings and the directions are as indicated. V represents the direction of propagation. (b) Only certain frequencies can propagate between two perfectly conducting parallel planes because the electric field parallel to the plates at their surfaces must be zero. In this drawing the electric field is into and out of the paper.

2. Waves Guided Between Conducting Surface

If two perfectly reflecting plane surfaces were situated parallel to each other, it is intuitively clear that radiation introduced at an angle to one of them would be reflected back and forth without losing energy. They would be "guided" by the surfaces, so that their energy would be propagated in a direction parallel to the plates.

However, upon closer inspection, the fact that the plates are conducting imposes certain restrictions on the wavelength of the waves propagating between the two surfaces. Refer to Figure 20b. The electric vector of the wave launched between the two surfaces is parallel to the plates, i.e. , into and out of the paper. The wavelength is the distance between the dashed lines in the figure perpendicular to the direction of propagation. The planes are infinitely conducting so that if the electric field is anything other than zero at the two surfaces, charges within them instantly move to cancel the applied field. Therefore, the electric field is zero at these surfaces at all times. For that to occur, the wavelength of the radiation must be such that there is an integral number of half wavelengths from one reflection to the next. If the separation of the plates is a distance d, then the path length from one reflection to the next is simply $d/\sin \phi$, where ϕ is the angle of incidence or reflection of the ray. That means that the following equation must hold for zero field strength at both surfaces,

$$\frac{d}{\sin \phi} = m\frac{\lambda}{2} \tag{8}$$

or rearranging,

$$m\frac{\lambda \sin \phi}{2} = d \tag{9}$$

$\lambda \sin \phi$ is just the effective wavelength in the direction perpendicular to the plates. An integral number of one half of one of these effective wavelengths must be able to fit into the distance between the plates to have zero electric field at both surfaces.

The number m, in this case, is called the waveguide mode number. It is clear that lengths greater than $\lambda_c = 2d/\sin \phi$ cannot be propagated by the surfaces we have postulated because there is no way that an integral number of effective half wavelengths can fit between them. The frequency associated with λ_c is called the cutoff frequency. Any frequency below cutoff is therefore rapidly attenuated and not propagated.

Other geometrical configurations of conductors produce other configurations of guided waves. For the earth-atmosphere, we will meet guided waves between concentric spherical surfaces in several places. The general physical process of wave guidance, however, remains the same as that given above.

SECTION 3

GENERAL EFFECTS OF THE NATURAL ENVIRONMENT ON THE PROPAGATION OF ELECTROMAGNETIC RADIATION USED IN COMMUNICATION

The electromagnetic spectrum is theoretically unlimited. However, only a portion of the spectrum can be used for communications. The upper limit of usable frequencies occurs where materials and techniques can no longer produce or control the radiation. At frequencies below this upper limit, the restrictions are primarily imposed by interference from the natural geophysical environment. Figure 21 gives an overview of this spectrum indicating important regions for communications.

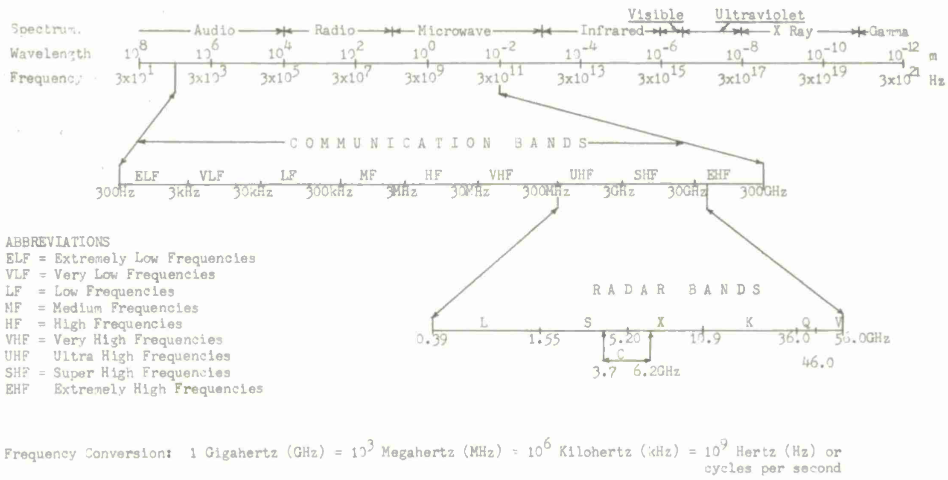


Figure 21. The electromagnetic spectrum from one Hertz to 10^{20} Hertz. The limits of each frequency region are somewhat arbitrary on the upper scale. The designations Very Low, Low, etc, are still in use today but they no longer cover all usable communication frequencies.

Communication systems in widespread use today are generally confined to the centimeter region and below (frequencies less than about 3×10^{10} Hz). However, the invention and perfection of lasers and masers has produced a surge of activity directed toward utilizing the higher frequencies produced by these devices for communication purposes. This will mean that coherent electromagnetic signals like those now produced by transmitters at lower frequencies will be available at many frequencies throughout the infrared and visible portion of the spectrum.

The effect of the natural environment on usable frequencies depends very much on the particular region of the natural environment of which one speaks. For that reason this section of the paper dealing with natural environmental effects attempts to analyze quasi-homogeneous regions of the earth-atmosphere system within which environmental effects on radiation remain approximately constant.

The regions which eventually emerge which are of interest to our discussion may be conveniently thought of as spherical shells roughly situated at greater and greater distances from the center of the earth. The innermost of these regions is composed of the crust and oceanic waters. Next is the boundary between this region and the atmosphere. The lower atmosphere itself is a region of many important and interesting effects and will be studied next. High in the atmosphere of the earth, solar radiation separates electrons from their atomic nuclei producing what is called the ionosphere. The study of this final region includes the effects upon signals which do not penetrate through the ionized gases but are reflected back towards the surface of the earth, and the effects encountered when signals pass entirely through the ionosphere.

By organizing Section 3 in this manner, it is hoped that a picture emphasizing the role of the natural environment in propagation will emerge.

I. THE SOLID EARTH

A. PROPAGATION THROUGH THE CRUST

A great deal of speculation exists on the electrical properties of the earth below its surface. The feasibility of transmitting signals over relatively long distances within the earth's crust depends upon what those properties are and how they are distributed with depth. Numerous laboratory measurements have been made, and on the basis of these, several theories about propagation within the crust of the earth have been offered. We will investigate the basic ideas behind these theories and show the necessary conditions which must exist within the earth for them to operate.

A fundamental limitation on all earth propagation is introduced by the finite conductivity of solid earth; all signals are attenuated exponentially. The conducting ions and electrons collide with the solidly bound atoms of which the rocks and soil are composed and thereby lose energy. Therefore it is essential that a region of very low conductivity exist to minimize attenuation.

There is the possibility that a such a region exists, i.e., there may be extensive regions of the earth which have a highly conducting surface layer overlying basement granitic rock of much lower conductivity. And it is generally believed that if one penetrates to sufficient depth, the low conductivity region will give way to another region of higher conductivity again, the result of thermal agitation of the crystalline material at the elevated temperatures found in earth's interior. Such a situation is illustrated in Figure 22.

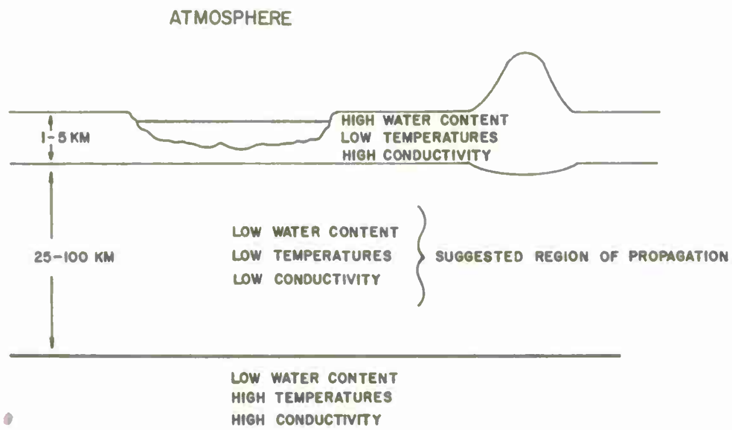


Figure 22. A hypothetical illustration of the distribution of conductivity with penetration distance into the earth.

If this be truly the case, then several possibilities for propagation exist. They include: (i) guided waves between the two regions of high conductivity; (ii) surface waves at the boundary of the low and high conductivity strata; and (iii) direct and reflected propagation between points within the low conductivity layer.

Figure 23 will illustrate the physical picture we have described and the different propagation modes which could conceivably be supported. Those shown are not fundamentally different from propagation modes which will be found to exist in the atmosphere, excepting that strong exponential attenuation, always present in earth propagation, is of much less importance in most atmospheric transmission. The intrusion of highly conducting material into the stratum of low conductivity would have an effect similar to that of mountains and buildings on atmospheric propagation. The radiation fields would be reflected, refracted, and diffracted around the high conductivity regions.

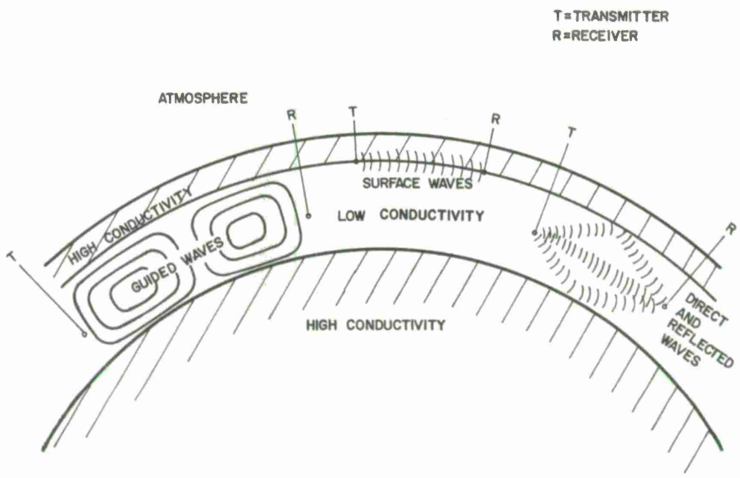


Figure 23. Three hypothetical modes of propagation within the earth's crust.

Attenuation of the radiation would be highly frequency dependent; that is, attenuation would be much more pronounced at higher frequencies. Presumably, very low frequencies would be utilized for any operational communication system. It has even been suggested that "d-c signaling" might be the best (or only) suitable technique to use. Dots and dashes of direct current would be the information carriers. The pulse rate, however, would be restricted in such a system by the dispersive nature of the medium and the intended distance to be spanned by the signal. Overlapping of the pure sine waves of one pulse with those of another (see Figure 10) would occur if the pulses were spaced too closely or traveled too far.

A great many facts and opinions have been marshaled which attempt to show the weaknesses in the earth propagation hypotheses. Perhaps the most convincing of these arguments centers around the unproven existence of the low conductivity layer upon which the theories so far discussed have been based.

It is argued that although laboratory measurements of conductivity on the types of rocks thought to make up the basement strata show very low values, low enough, indeed, for long range transmission through them, these measurements are not valid for several reasons. One of these is that water, which is known to fill the pores of surface rocks and to substantially raise their conductivity, will seep down and saturate the lower rock strata also. It is pointed out that even in the deepest wells, saturation does not seem to lessen.

A second important factor, it is believed, would be the large scale folding and fracturing which would greatly alter and confuse the propagation paths. No detailed knowledge of deep faulting and fracturing is known, but it is believed that the complex surface features will not be appreciably simpler and more stratified beneath the surface.

Although the controversy is not resolved, there are interesting reasons to learn if such propagation can be accomplished. One is the invulnerability of such a communication system to events on the surface of the earth. Thunderstorms, ionospheric disturbances, and all other natural or man made noise would be unable to penetrate the overlying conducting region. This fact would allow much lower signal strengths to be detected, and would greatly increase reliability.

The measurements of the conductivity which have been performed to date over the United States and in Europe indicate that, in general, the conductivities in most of the deep strata granitic rock through which communication was hoped to be possible, was too high and therefore attenuation was too great to be of practical use. Some localized high resistivity samples were found, but were minor in areal extent. The maximum distance over which communication were successfully completed at a frequency of 10 KHz was 1300 meters.

Tests were also performed using very low frequency conduction currents through earth strata. Again the bulk resistivity between transmitter and receiver resulting from the resistivity of the rock and its fracturing was such that the maximum distance over which signals were received was 32 km.

B. PROPAGATION THROUGH THE OCEANS

Ocean water is a relatively good conductor of electricity because of the ion forming salts and minerals dissolved in its waters. A propagating electromagnetic wave is therefore highly attenuated in sea water; free ions and electrons set in motion by the field collide with other constituents, passing energy from the wave into the random motions of the fluid. The attenuation of electromagnetic radiation in sea water is depicted in Figure 24.

It may be seen that at 100 Hz, the attenuation is nearly 1 decibel per yard (db/yd), i.e., for every yard of penetration, the radiation is attenuated to about 79% of its value at the beginning of that yard. (For the definition of decibel see db in List of Abbreviations.) After 10 yards, the signal strength would be reduced to one-tenth of its starting value, and so on.

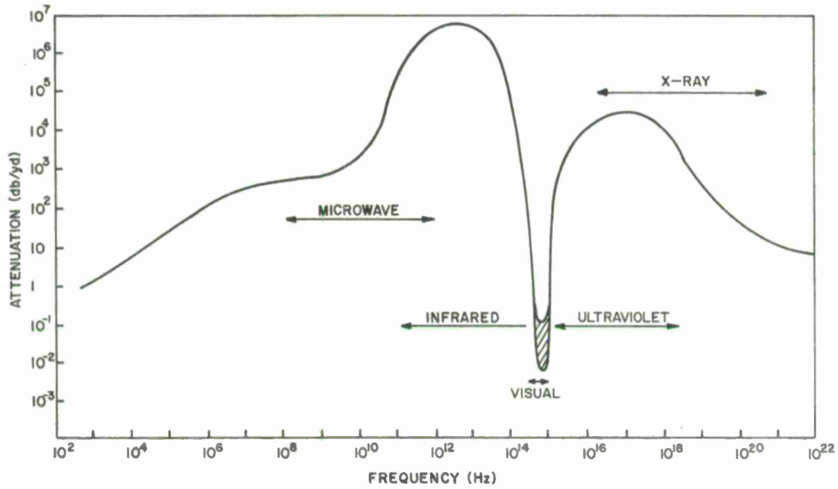


Figure 24. Attenuation as a function of frequency in ocean water. The hatched area shows the effects of suspended matter in the propagation path.

The attenuation builds to a maximum in the infrared region where it becomes over one million times stronger than at 100 Hz. A well documented hole in the attenuation scheme appears in the visual region where attenuation dips at least to the 100 Hz value if not considerably lower. Into the ultraviolet portion of the spectrum the attenuation is again severe but trails off towards the X-ray region.

The velocity of propagation is also a strong function of frequency. Some representative values for propagation speed together with penetration distance of an electromagnetic wave are given in Table 1. The penetration distance is based upon a 40 db reduction in the signal strength from its initial value, i.e., reduced to one ten thousandth of its initial value.

Table 1. Propagation of electromagnetic waves through sea water. (After Hill)

Frequency	Velocity	Penetration Distance
1 Hz	1.8×10^3 m/s	1160 m
100 Hz	18.0×10^3 m/s	116 m
1 MHz	2.0×10^6 m/s	1 m
10 GHz	2.7×10^8 m/s	4 mm

Direct propagation through sea water, because of the properties outlined above, is not practical over long distance. However, it is found that submerged transmitters can communicate with submerged receivers over distances far beyond those thought possible based upon the theoretical attenuation coefficients. Upon closer examination, it is found that the energy received has not traveled directly through the body of ocean water but has usually traveled to the boundary between the ocean and overlying air, and there excited a surface wave. The surface wave then propagates at a greatly reduced attenuation in all directions, leaking energy into the ocean as it passes. This leakage is what is detected. This particular surface wave will be discussed more completely in the next section.

There is the possibility, although no proof exists, that conditions might be such for a surface wave to propagate along the ocean-floor to ocean interface under some ocean basins. For this to occur, the very low conductivity basement rock would have to be directly beneath the body of water. It would act as the dielectric half-space, while the ocean water would act as the conductor (see Figure 22). Exponential attenuation would occur with distance because of the finite conductivity of sea water but lower signal levels could be detected because atmospheric and man made noise would be effectively screened out by the overlying conducting fluid.

II. THE SOLID EARTH - ATMOSPHERE INTERFACE

The section which we are now beginning starts the discussion of the region where most radio communication occurs, i.e., at and above the surface of the earth. Less speculation is necessary for this section than for the previous one; many years of regular usage have firmly established the facts about propagation over the earth and how it is affected by the natural environment.

A transmitter above the surface of the earth which radiates energy in all directions can have some of this energy reach a receiver in a number of ways. They are schematically represented in Figure 25. Radiation can be sent skyward and refracted back to the receiver by the ionosphere; radiation propagating in the troposphere can have its direction of travel altered sufficiently so that it reaches the receiver; radiation can propagate almost directly to the receiver; it can be reflected by the earth to the receiver; or it can travel as a surface wave at the earth-atmosphere interface. The characteristics of the surface wave and the reflected waves are the topics of this section of the report.

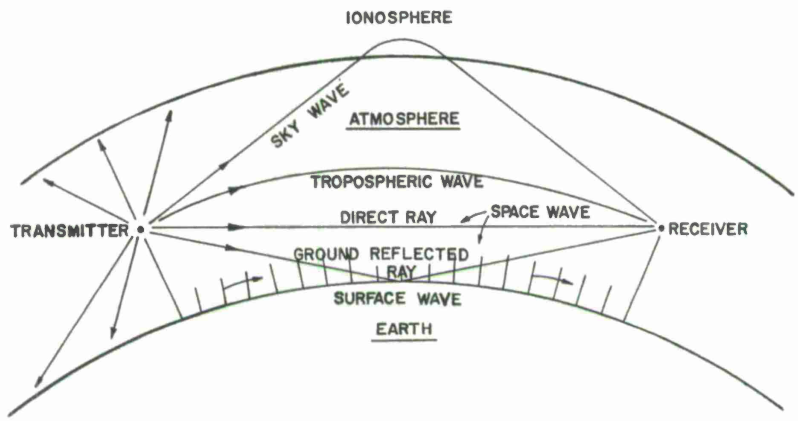


Figure 25. A schematic view of the various propagation paths which a signal might utilize between a transmitter and a receiver which are near the surface of the earth. Only the surface wave and the ground reflected part of the space wave depend upon the earth to reach the receiver.

A. THE SURFACE WAVE

The surface of the earth represents a large change in electrical properties from the atmosphere overlying it. Consequently, radio waves which are directed against the surface may be substantially altered in three important respects: (i) some of the energy can excite a surface wave; (ii) some of the energy can be reflected; and (iii) some of the energy will be absorbed and show up ultimately as heat. We will discuss the first two phenomena beginning with the surface wave.

A vertically polarized antenna (i.e., one in which the electric vector is oriented vertically) can launch an electromagnetic wave whose lower edge is supported by the earth's surface. It represents a guided surface wave whose general properties were discussed in Section 2, page 33. The atmosphere acts as the dielectric half space and the earth as the conducting half space. The wave glides over the surface of the earth in a manner similar to Figure 26. Surface charges are carried along with the wave in a manner which constitutes a current flow.

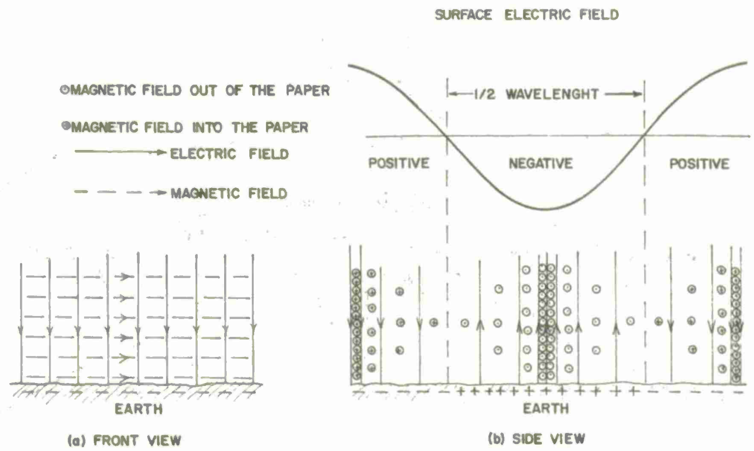


Figure 26. Front and side view of a surface wave showing the polarization of the electric and magnetic fields. Solid lines represent the electric field; dashed lines and small circles represent the magnetic field. (After Terman)

In a perfect conductor, any horizontal electric field would be instantaneously short circuited, and therefore only the vertical electric field would exist. The surface currents under such idealized conditions would not dissipate any energy and the only loss to the wave would be caused by spreading horizontally over the surface. The region in which the surface currents flowed would be limited to an infinitesimal thickness near the interface.

For a surface of finite conductivity, like the earth, the idealized circumstances are altered. The ground offers resistance to the flow of any current, so that the energy of the surface wave below the surface is absorbed and dissipated as it progresses. Energy for the earth currents must therefore be supplied by that part of the wave above the surface. This can only happen when the wavefront itself tilts slightly downward allowing energy to propagate into the earth. Figure 27 illustrates the tilt of a wavefront associated with a surface wave.

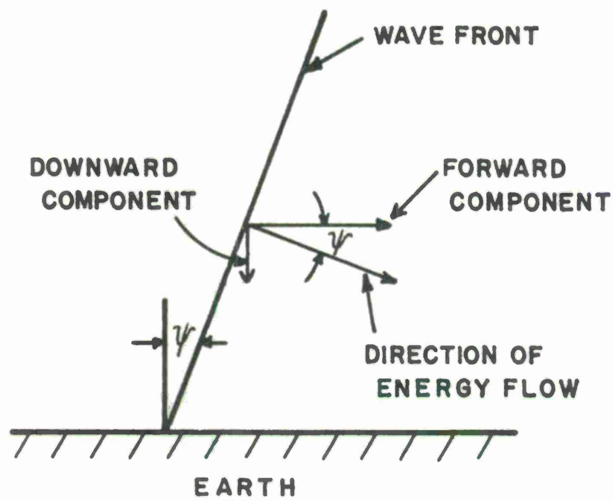


Figure 27. The tilt of a surface wave propagating over a surface with finite conductivity. Over sea water the angle of forward tilt (Ψ) will be approximately 10 seconds of arc. Over land, it will vary but will always exceed 10 seconds. (After Terman)

The attenuation per unit distance of a traveling surface wave is increased by lowering the conductivity of the conducting surface and by increasing the frequency. This can be readily seen by studying Figure 28. The diagram shows the strength of a surface wave versus distance from a one kilowatt transmitter whose antenna is vertically polarized and radiates over a smooth earth. Three frequencies are plotted. Conditions for good soil are used for the set of solid curves, and conditions found in sea water are used for the dashed curves. The points to notice are the following: (i) for a given frequency and distance from the transmitter, all intensities are higher over the water path; (ii) for the given surface conductivity, the lower frequencies are attenuated less; and (iii) the deviations of the plots from straight lines, indicating earth curvature effects, always start closer to the transmitter site for higher frequencies.

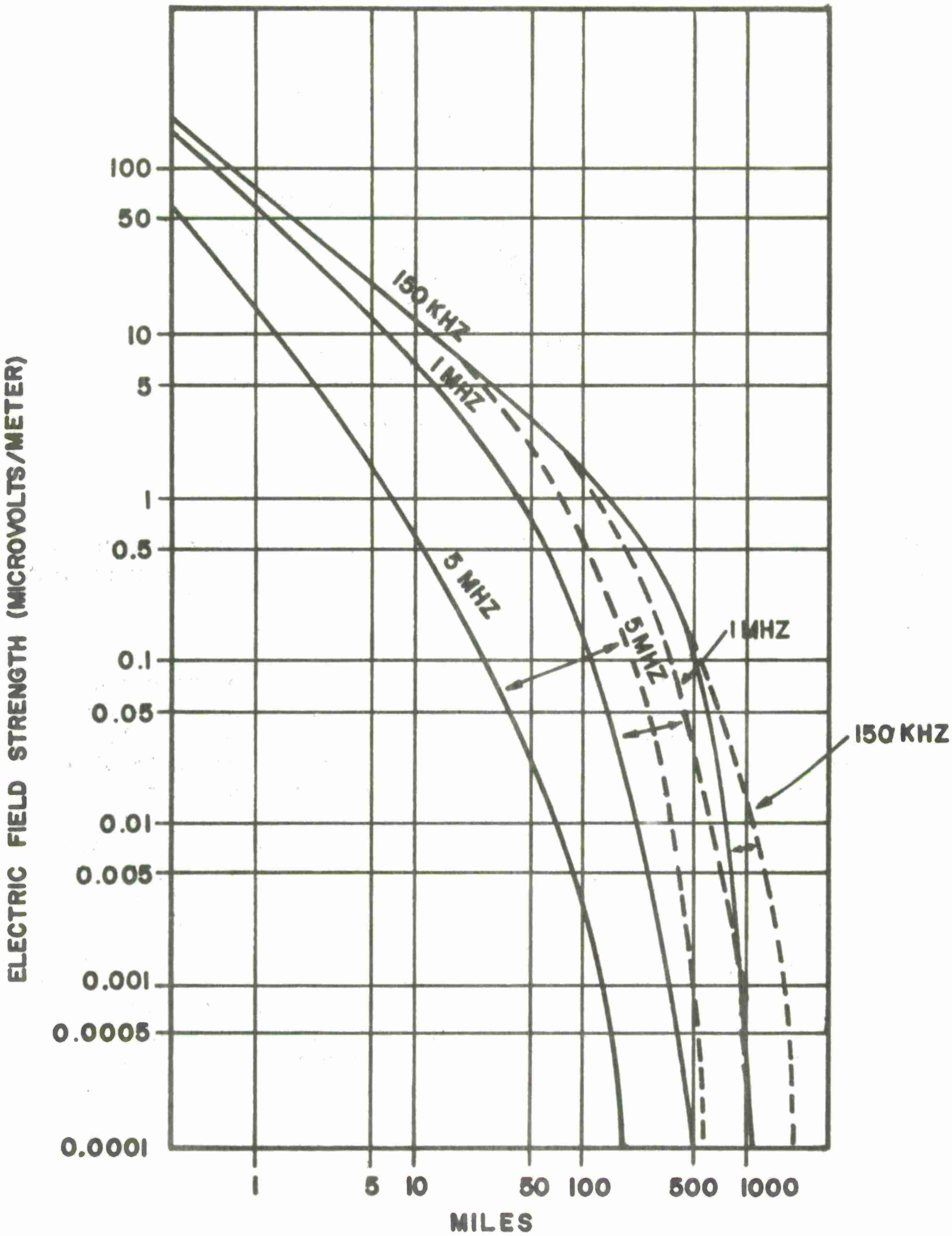


Figure 28. Strength of the surface wave as a function of distance, frequency, and surface conductivity. Solid lines are for conditions found in good soil and the dashed lines are for sea water. Transmitter power is 1 kilowatt and the radiated field strength varies as the cosine of the elevation angle.

Besides the above mechanisms of attenuation, other physical realities of the earth's surface contribute to losses in the surface wave. Deviations from a flat surface cause energy to be radiated away. The surface irregularities must be of an order of magnitude similar to the wavelength of the surface wave in order to be effective, so that, once again, higher frequencies (shorter wavelengths) are more affected. The spherical shape of the earth's surface may be considered an irregularity that greatly favors longer wavelengths.

The values for conductivity and other electrical properties of the earth which are effective in determining the attenuation of the surface wave are averages for those properties over vertical distances into the ground within which important earth currents flow. These distances are perhaps 5 to 10 ft at frequencies used in short wave communication (2-30 MHz) and over 50 ft at commercial broadcast frequencies (550 KHz to 2 MHz). Therefore, the attenuation is not strongly dependent upon conditions at the actual surface of the earth caused by such natural phenomena as rain and snow.

Practical propagation distances for the surface wave are most importantly dictated by frequency. In the frequency region between 20 KHz and 100 KHz, surface waves account for the greatest part of the energy received, at all times, up to about 1000 km from the transmitter. Under ideal conditions, much greater distances can be covered. Because the average electrical properties in the top 50 to 500 ft of such long transmissions paths change little with time, the received signal will not vary much diurnally, seasonally, or even yearly. Fading, caused by the vector combination of signals passing over different and changing path lengths, is not a problem at these frequencies because of the dominance and stability of the surface wave. Of course, at distances where the strengths of the surface and sky waves are comparable, fading as well as diurnal and seasonal changes are observed. Figure 29 shows the July average for diurnal variations at two frequencies illustrating the stability of lower frequency propagation.

Between 100 KHz and 2 MHz, which includes the commercial broadcast bands (550 KHz to 2 MHz), the practical ground wave propagation distance diminishes rapidly. In this frequency region, for a high power transmitter (100 KW), the maximum distance of effective communication even over sea water is limited to approximately 1200 km, while over land, it is reduced to perhaps 200 km.

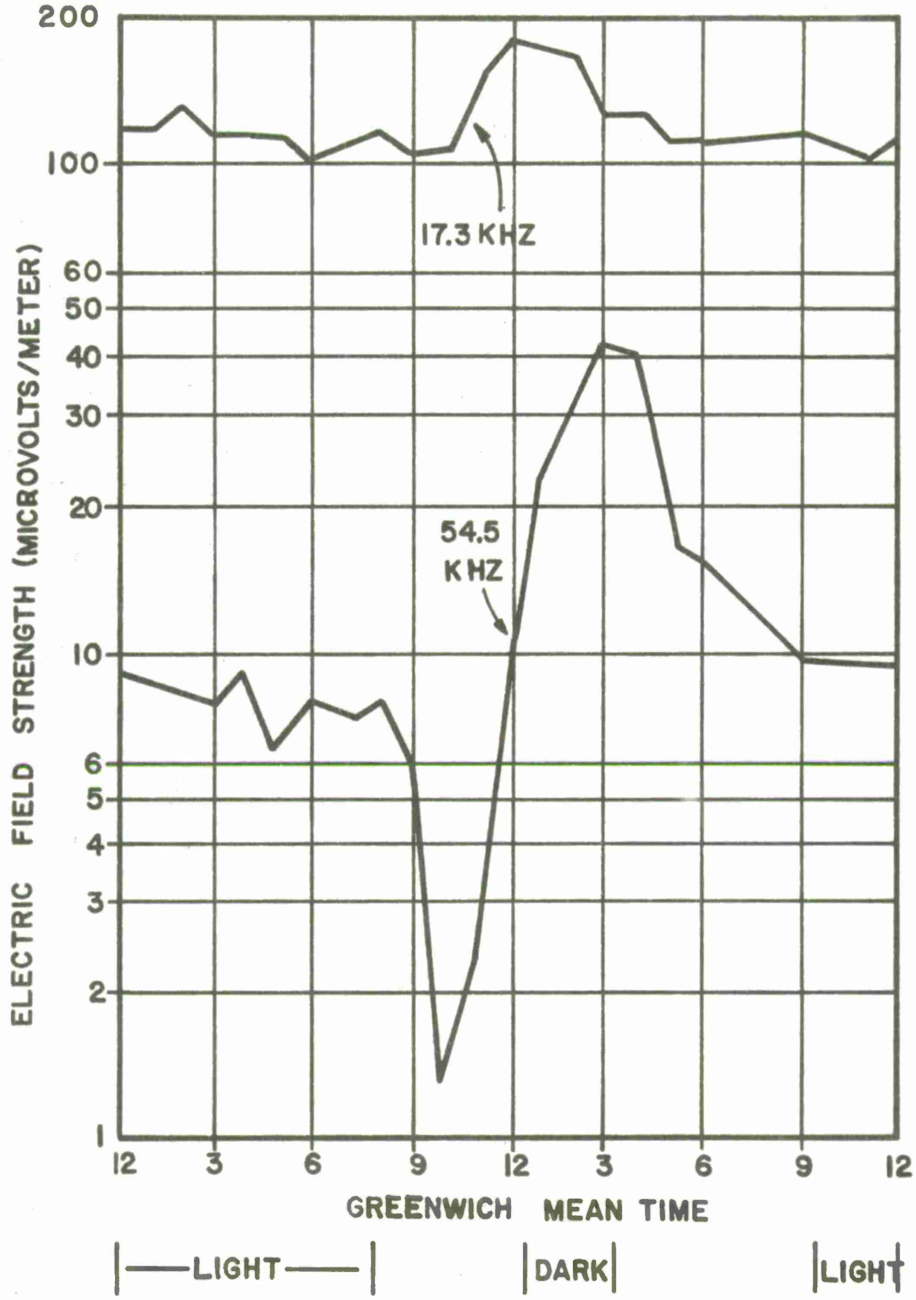


Figure 29. The average diurnal variation of long wave signals propagated across the North Atlantic in July. Only the variation is meaningful since the two signals are of unequal strength. The intervals at the bottom indicate when the propagation paths are entirely in light and entirely in darkness.

At frequencies above 2 MHz, the surface wave attenuates so rapidly with distance that long range communication is completely unfeasible. The ground wave is relied upon at these frequencies to provide coverage, at most, over city wide regions.

B. THE SPACE WAVE

The second important effect which occurs to radio waves at the surface of the earth is the reflection of signals. In Section 2, part III, reflection was shown to be the result of matching, at the interface between two media of differing electrical properties, certain conservative quantities of the wave. And the only way in which this could be done was to postulate that some of the energy, if not all, was reflected rather than transmitted into the second medium.

Propagation over the earth results, in most cases, in a part of the energy being reflected by the earth to the receiving antenna. Together with the direct ray, this contribution makes up the so-called space wave (depicted in Figure 25). The direct ray is that part of the signal which reaches the receiver over a path which is effected only by the normal gross atmospheric refraction. Both of these parts of the received signal are little attenuated by their passage through the atmosphere; they are weakened mostly by spreading. The most important change occurs to the reflected vertically polarized ray. At grazing incidence, as is almost always the case, the direction of its electric vector is reversed, i.e., it undergoes a phase change of 180° or, more simply, a phase reversal as shown in Figure 14. At the receiving antenna the two signals are of about equal strength but almost opposite in phase. They do not cancel completely because the reflected ray has traveled a somewhat longer path than the direct ray so that they are not directly out of phase. The final result of their vector combination is that the intensity of the received signal shows an overall inverse proportionality to the square of the distance rather than the distance to the first power.

A graph of the intensity of the received space wave signal is shown in Figure 30 for a frequencies of 300 MHz and 50 MHz. The free space intensity is inversely proportional to distance and shows up as the dashed straight line. The actual intensity shows large oscillations for the 300 MHz signal at distances not far from the transmitter, ranging from very small intensities to intensities twice as great as would be found in free space. This is explained by realizing that the phase difference between the direct and reflected rays changes with distance away from the transmitter.

When the difference is a full wavelength, the two signals will tend to cancel (because of the 180° phase shift in the reflected ray). At differences of one half wavelength, the two signals will add and be nearly twice the free space magnitude. At distances beyond about 15 miles away from the transmitter, the signal is almost inversely proportional to the square of the distance and appears as a gently curved line on the logarithmic plot. Finally, as the radio horizon is reached, the signal strength decreases very rapidly; only small amounts of energy are diffracted into the shadow zone.

The non-oscillating curve is for a frequency of 50 MHz. The wavelength is now so large with respect to the antenna height (6m : 330m), that the phase cannot be totally reversed by differences in path length.

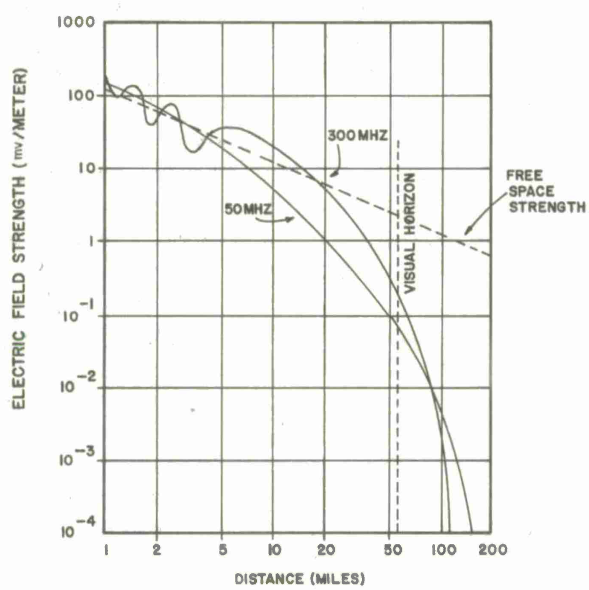


Figure 30. The intensity of the space wave as a function of distance from the transmitter for two frequencies. Values shown are for a 1000 foot, 1000 watt antenna broadcasting a horizontally polarized signal. The receiving antenna is 30 feet high. The oscillations in the 300 MHz signal are due to interference between the direct ray and the ground reflected ray. (After Terman)

The previous discussion on the surface wave was based upon vertical polarization of the transmitting and receiving antennas. The requirement is necessary in order to launch a surface wave.

The general characteristics of the received space wave, on the other hand, would not be changed by substituting a horizontally polarized signal; the path length difference argument would still apply. However, the general characteristics of reflection for horizontally versus vertically polarized radiation at a material surface is quite different and will be clarified below.

If, as in Figure 31, a plane wave strikes a plane surface, it is reflected with an angle of reflection equal to the angle of incidence. If the plane surface is a perfect conductor, the intensity of the reflected beam will be equal to the intensity of the incident beam. If the polarization of the incident electric field is horizontal, i.e, is perpendicular to the plane of incidence (the plane containing the normal to the surface as well as the vector describing the propagation direction), no phase change occurs. If the incident radiation is polarized so that its electric vector lies in the plane of incidence, a phase reversal occurs.

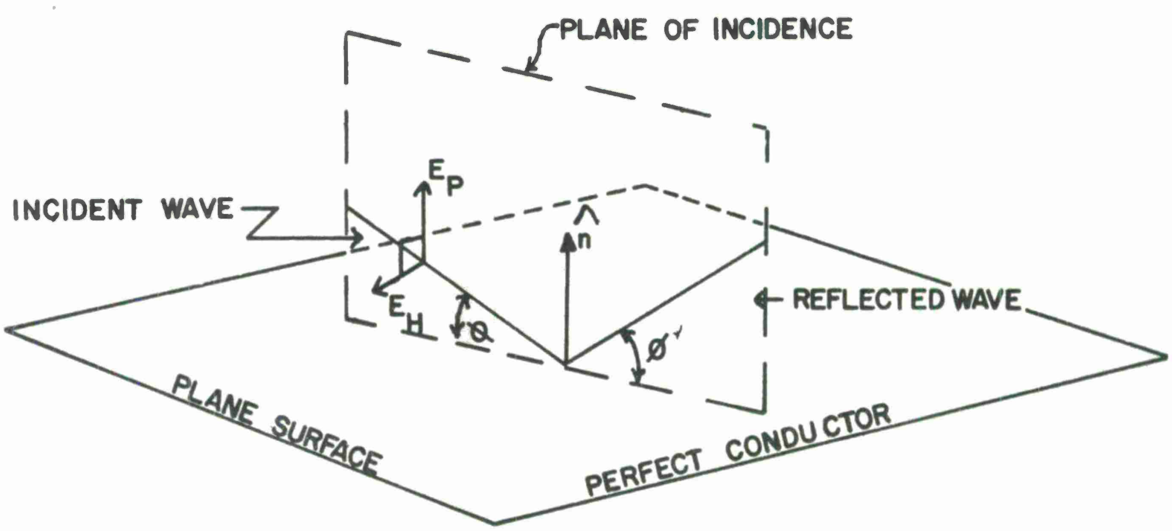


Figure 31. The geometry of a signal reflected at a plane surface. A component part of the incident electric field lies in the plane made up of the direction of propagation and the normal to the reflecting surface and is labeled E_p (polarization parallel). A horizontally polarized part of the field is parallel to the reflecting surface and is labeled E_h (polarization horizontal).

For a real surface, like the earth or sea water, a much more complicated situation exists. The reflection coefficient will depend upon many factors: the polarizability of the atomic and magnetic dipoles; the conductivity of the surface; the frequency of the radiation; and the angle of incidence of the radiation on the surface.

Figure 32 shows typical values of the reflection coefficients for both polarizations for reflection by the earth. There are also traces depicting the phase shifts which occur to each polarization component on reflection at different angles of incidence. The appearance of the graphs is typical of conducting surfaces in general and points out that the phase of a horizontally polarized signal is little affected by reflection, whereas a vertically polarized signal shows an almost complete phase reversal from angles near grazing incidence to angles near vertical incidence.

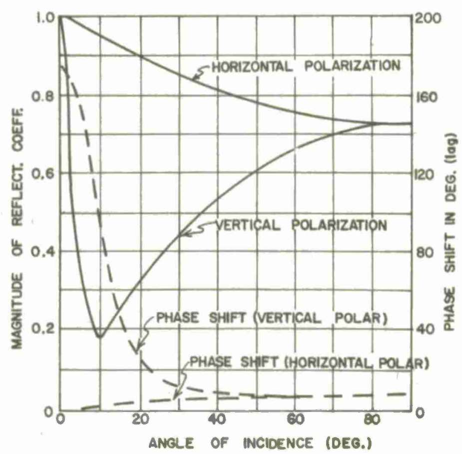


Figure 32. Typical magnitudes of the reflection coefficient and phase shift for horizontally and vertically polarized signals reflected by the earth. The reflection coefficients are depicted by solid lines; the phase shifts by dashed lines.

III. THE LOWER ATMOSPHERE

A. DEFINITION AND GENERAL CONDITIONS

The region which we will arbitrarily refer to as the lower atmosphere will be confined to that part of the envelope of gases which surrounds the earth which does not contain significant numbers of electrons or ions. The ionosphere, which is distinguished by having significant numbers of such particles, is best treated separately; the effects on radio wave propagation in the two regions are quite different. This will mean that in the daytime the lower atmosphere will extend from the surface to about 55 km; while at night one might push this upper level, under normal ionospheric conditions, up to about 75 km.

The lower atmosphere of the earth shows a great deal of variation in both the horizontal and vertical directions. Figure 33 will give some idea of how the pressure, density, and temperature vary with altitude in a model atmosphere thought to resemble average conditions found above the earth at mid-latitudes. It may be seen that both pressure and density decrease exponentially with the altitude, at least up to the 100 km level. The temperature profile, on the other hand, is not nearly as smooth. In fact, the sharp changes in the vertical temperature gradient are the basis of a convenient classification of various vertical zones.

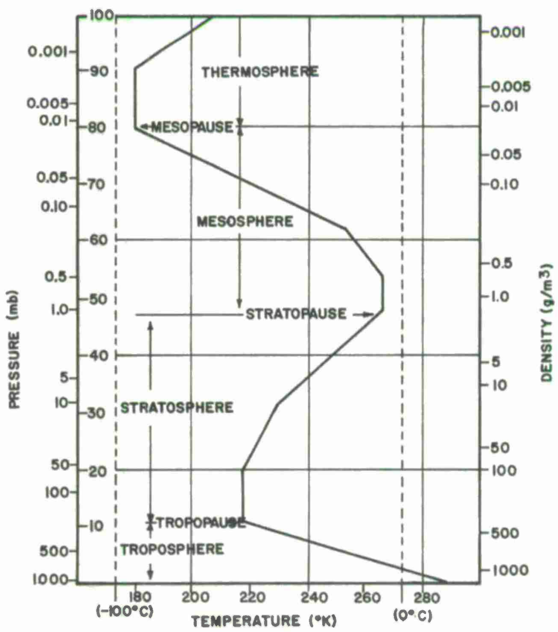


Figure 33. The 1962 U. S. Standard Atmosphere. The nomenclature and boundaries for the atmospheric shells shown above are those adopted by the World Meteorological Organization. (Handbook of Geophysics and Space Environments)

From the surface to about 11 or 12 km in mid-latitudes the temperature decreases quite rapidly (approximately 6.5° C/km). Above this level and below about 20 km, the temperature falls slowly or not at all. The region of constant temperature decline has been labeled the <u>troposphere</u> from the Greek word tropos, meaning turning and sphere. In this region the air is constantly mixed and overturned producing the majority of weather phenomena seen and felt at the surface. The boundary dividing this region from the constant temperature region above is known as the <u>tropopause</u>.

The stratosphere is a word derived from Latin stratum plus sphere meaning layered sphere. This name was applied to the region above the troposphere when atmospheric soundings which first penetrated beyond the tropopause found, on the average, a stable horizontally stratified atmosphere. The name has been extended today so that it includes the region up to the stratopause at about 48 km. Above the isothermal (constant temperature) region extending from the tropopause to around 20 km, the temperature in the stratosphere rises steadily. The rise continues to the stratopause where a new isothermal region is encountered.

Above the stratopause, lies the mesosphere, the word originating from Greek "mesos" meaning middle or intermediate. In this region the temperature decreases steadily as one ascends up to a level of about 80 km. This level is designated the mesopause and appears to be the last region in the atmosphere which provides the necessary conditions for clouds to form which are visible from the surface of the earth. These are the famous noctilucent clouds seen in high middle latitudes during the long summer twilight.

Above the mesosphere, the temperature rises steadily for hundreds of kilometers. The predominance of the high temperatures in most of this region has led to its designation as the thermosphere, from thermos the Greek word for heat.

Finally, above about 550 km, atoms and molecules can escape into space at thermal velocities, and this region, as a consequence, has been called the exosphere.

The physical reasons leading to this vertical profile of temperature are intimately related to the absorption, by consitituents in the atmosphere, of radiation originating in the sun and at the surface of the earth and to the general planetary circulations of the atmosphere. In addition, the tropospheric profile is profoundly influenced by its intimate association with the surface of the earth.

The air of the troposphere is quite transparent to the bulk of radiation which streams into its upper boundary from the sun. The sea and ground both are effective absorbers of the sunlight falling upon them, and it turns out that the troposphere is heated by radiation from and contact with the surface of the earth. The earth reradiates much of the energy it absorbs from the sun but at much lower frequencies than those found in the solar spectrum. The air is not so transparent to this long wave radiation, and the radiation is effectively absorbed by the troposphere, especially by water vapor and by liquid water in the form of clouds. The troposphere, as a result of these energy sources, is warmest at its base and shows a temperature decline with increasing altitude.

The stratosphere breaks the steady decline of temperature which marks the troposphere by responding to the energy absorbed by ozone. Ozone is a triatomic molecule of oxygen and is formed by photochemical reactions in the stratosphere. It is particularly effective in absorbing energy from certain frequency regions of sunlight (2100 Å to 3200 Å); in fact, ozone is so effective that these frequencies are not seen at all at the surface of the earth. Although the maximum concentration of ozone occurs at heights of perhaps 25 km above sea level, the temperature maximum related to its absorption occurs at the stratopause, in the vicinity of 50 km; the reason for this discrepancy is that the radiation which is absorbed most efficiently by ozone is mostly depleted by the 50 km level.

The decline of the temperature in the mesosphere is a result of the efficient radiation at infrared frequencies by the neutral atmospheric gases. No strongly absorbing constituents, like for example ozone, are present to compensate for this loss. The temperature declines steadily until important absorption reoccurs. The absorbing particles are various atomic fragments of gases and the neutral gases themselves which make up the atmosphere below the mesopause. The difference in the region above 80 km from that below may be attributed to the presence of ultraviolet solar radiation. Ultraviolet radiation can break up molecules into their constituent atoms and in so doing increases the kinetic energy of the gas. However, it only penetrates down to the 80 km level. By this height, ultraviolet radiation is mostly absorbed, its energy having gone to maintain the high kinetic temperatures found in the thermosphere.

It should be pointed out here that the model atmosphere of Figure 33 is just that. The real atmosphere, at any particular moment, will show strong deviations from this model both in longitude and latitude. However, averaged over a sufficiently long period of time, it is believed that the mid-latitude atmosphere approaches the model quite closely.

Both experiment and theory dealing with radio wave propagation in the lower atmosphere show that the most important effects are found near the surface of the earth, up to and including the tropopause. Therefore, it is quite sufficient to study the effects of the troposphere alone to understand what the role of the lower atmosphere is in radio wave propagation.

B. THE TROPOSPHERE

To obtain a broader and more complete picture of the interaction between the troposphere and radio waves, it is necessary to present a somewhat more detailed view of this part of the atmosphere than can be obtained from Figure 33. Therefore, in the next few pages, conditions in the troposphere which might be found at various times and places and which are important to radio propagation will be discussed; the part that solar radiation plays in the troposphere will be pointed out; and a general description of the most important dynamical influences in the lower atmosphere will be presented. With this background, a considerably better understanding of the problems which exist in radio propagation through the atmosphere will be possible.

1. Physical Details of the Troposphere

a. Composition

Disregarding water vapor, the composition of the atmosphere remains quite constant up to about 100 km. By volume, air consists of approximately 78 percent nitrogen, 21 percent oxygen and about 1 percent argon. At least 12 to 14 other gases are found in trace amounts.

Water vapor is the only major gas which has an appreciable variation in time and space, going from as little as a trace amount to as much as 4 percent by volume of a given air sample. It is noteworthy that water vapor is the only gas in the atmosphere which exists in other than the gaseous phase. Its uniqueness in this respect is compounded by the fact that it actually can exist in all three phases in significant quantities throughout most of the troposphere. The liquid and solid phases of water vapor constitute the major non-gaseous parts of the atmosphere. All three phases, by their intrinsic physical nature and by their spatial distribution play an extremely important part in the dynamical and radiational processes which are constantly at work in the troposphere.

b. Temperature

The temperature below the tropopause, in the mean, decreases from equator to pole and from surface to tropopause. The height of the tropopause itself ranges from about 15 to 19 km in height at the equator down to about 6 to 10 km in the polar region and varies daily and seasonally. A vertical cross-section in height and latitude showing the gross temperature structure of the troposphere is shown in Figure 34. The tropopause is marked by a heavy dashed line.

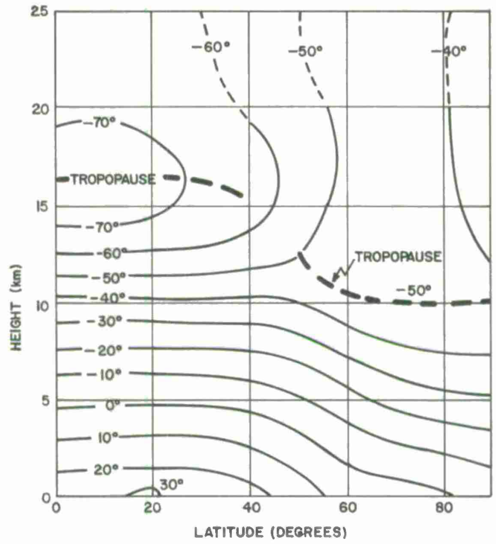


Figure 34. A vertical cross-section in height and latitude showing the gross temperature structure of the troposphere. (After Meyer)

At an individual geographical location, the vertical temperature profile might be depicted as in Figure 35. The standard atmosphere is shown as a dotted line whereas the hypothetical vertical temperature profile might follow the solid curve. Two shallow regions occur where the temperature increases with height. These are called <u>temperature inversions</u> and, if they exist over an extensive horizontal region, can be quite important for radio propagation at high frequencies. The origin and distribution of such inversions will be discussed in following sections which describe radiational effects and dynamical conditions in the troposphere. It is enough now to know that such inversions commonly exist.

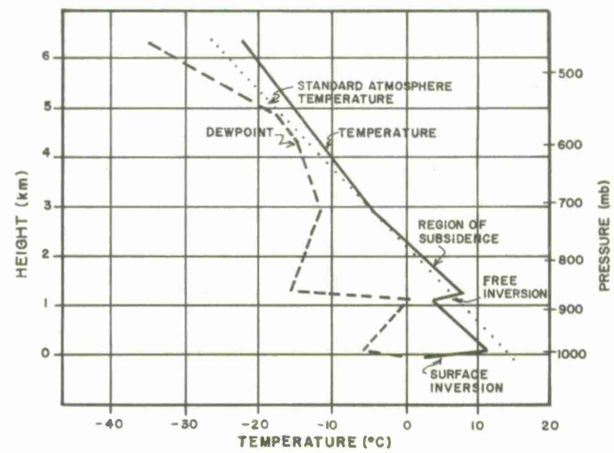


Figure 35. A hypothetical vertical sounding in mid-latitudes. Temperature is represented by the solid curve, and the dew point is represented by the dashed curve. The dotted line represents the temperature of the standard atmosphere. Inversions are regions in which the temperature increases with altitude. (After Meyer)

c. Water Vapor

The distribution of water vapor in the troposphere is constantly changing and varies widely in space and time. As an example only, Figure 35 includes a vertical profile of the dew point which might exist simultaneously with the temperature profile discussed above. The dew point is the temperature to which the atmosphere must be cooled in order for it to be saturated. A dew point equal to the free air temperature would mean that a particular parcel of air is saturated with water vapor. Any more vapor would only condense and fall out. The saturation or dryness of a parcel of air depends upon the proximity and extent of free surfaces of liquid water, the past dynamical history of the parcel, and upon its past history of temperature.

The maximum density of water vapor is normally found at the surface of the earth and diminishes roughly exponentially with height. This fact is true for two reasons: (i) the source of water vapor is at the surface of the earth; and (ii) the absolute maximum amount of water vapor which can be held by a parcel of air depends only upon its temperature (the higher the temperature of a parcel the greater its capacity for water vapor), and the temperature of the troposphere normally decreases with height. Deviations from this behavior, particularly as shown in Figure 35, will be an important aspect of the discussion of anomalous radio propagation through the troposphere.

d. Pressure

The pressure one would measure at any point in the atmosphere is very closely the weight of a column of air of unit cross-section extending from the measurement level vertically outward to the edge of the atmosphere. Because of the rough, uneven nature of the earth's surface, the uneven solar radiation reaching it at different latitudes and because of the somewhat random and changing distribution of clouds and surface conditions over the earth reflecting sunlight upward, the pressure changes in space and time. However, the fact that it changes is not so important as what weather is normally associated with those changes. Therefore, we will not discuss further the pressure relationships in the troposphere, but move on to specific weather conditions.

2. Solar Radiational Effects in the Troposphere

a. Heating

The ultimate source of energy for the motions and warmth of the troposphere is the sun. The bulk of its radiation passes through the atmosphere to fall upon the earth. The percentage of this radiation which is absorbed, and the percentage reflected, depends upon the physical nature of the surface. Snow reflects most of the radiation falling on it, whereas dark ploughed ground absorbs most of the radiation which strikes it. The troposphere receives most of its energy in three forms from the earth: (i) direct conduction of heat at the interface; (ii) absorption of long wave radiation emanating from the earth; and (iii) heat of condensation from water vapor which condenses in the atmosphere.

The geographical unevenness of the radiational input leads to the great weather systems which cross the earth and to the local variability of that weather. Also, it leads to the formation of distinct air masses and to widespread temperature inversions, both of which are important to radio propagation.

b. Air Masses

Air can remain long enough over large homogeneous areas of the globe to take on a vertical temperature distribution, a vertical moisture profile, and a vertical density distribution which will be approximately similar for all the air in that locale. These air masses are eventually constrained to move to other regions of the earth, but because of their size and the dynamic influence on the winds which they produce, their identifying characteristics can be retained for days and sometimes weeks. It will be illustrative to examine two extreme types of air masses; where they are formed, their characteristic temperature distribution, and the moisture which they hold.

(1) Structure of arctic air masses

In the high latitudes of the earth, arctic air masses are evolved. They are produced in regions of almost continuous snow and ice coverage. The air mass will evolve with characteristics which are illustrated in Figure 36 and generally described by the following statements: (i) overall temperatures are very low; (ii) a strong temperature inversion usually exists at or near the surface; (iii) moisture content is very low; and (iv) normally, the specific humidity (grams of water vapor per kilogram of moist air) increases from the ground to around 850 mb and decreases above.

Of the features of the arctic air mass shown in Figure 36, the temperature inversion at the surface is the most important one for radio propagation. It arises due to very strong radiational cooling at the snow and ice surfaces, and may be explained as follows: the crystalline nature of ice and snow reflects solar radiation very efficiently. It also radiates long wave radiation. Hence the surface tends to cool continuously. The air is cooled because the atmosphere transports heat to the surface via conduction where it is promptly radiated toward space. Some of this radiation is absorbed by the air. The remainder escapes to space. The amount absorbed depends mostly upon the amount of water vapor present, and under the existing conditions very little water vapor is present. Since the air gains less through absorption than it loses through conduction, it cools forming an inversion. To break the inversion the surface itself must be heated, and this is very difficult with highly reflective snow or ice cover.

The water vapor capacity of the atmosphere increases with height up to the base of the inversion because the temperature increases. Moisture in the very lowest layers sublimates out, and the profile of the percentage of water vapor to total air shows the maximum at the peak of the inversion.

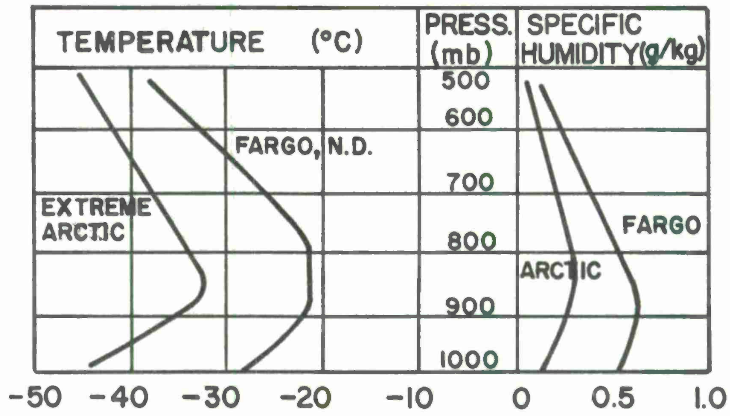


Figure 36. Characteristics of two arctic air masses. (After Pettersen, Vol. II)

(2) Structure of maritime tropical air masses

The other extreme type of air mass, as opposed to the arctic type above, is observed at the eastern end of the semi-permanent high pressure zones in subtropical latitudes. The features which identify the air above the oceans in tropical regions are, at least partially, dynamically produced; and to that extent, air which moves out of this source region will not show the extremes of the vertical temperature distribution and humidity which we will illustrate below. But for such air masses in their source regions experiencing large scale subsidence (descending air from aloft), Figure 37 is typical. The characteristics of air masses in the trade wind inversion regions can be described as follows: (i) the overall temperature is high; (ii) the temperature decreases for a distance above the ocean surface until the bottom of the inversion is encountered at 500 to perhaps 2500 m; (iii) an extremely well defined boundary region (inversion) between the lower layer and upper layer normally exists (The inversion can be anywhere from a few meters to as much as 2500 m thick.); and (iv) mixing in the lower layer is complete making it uniformly moist while the upper layer is very dry.

The very strong free subsidence inversion in the maritime tropical air mass is a result of adiabatic (meaning no external heat sources) compression of the descending air. At the periphery of these regions air is lost at the surface and gained at high levels. There is, therefore, a slow sinking over the interior of such a region. And the air which subsides, even if it were initially saturated, becomes capable of holding more and more water vapor as its temperature increases. However, its only source, the ocean surface, in insulated from it by the very strong inversion and it remains dry.

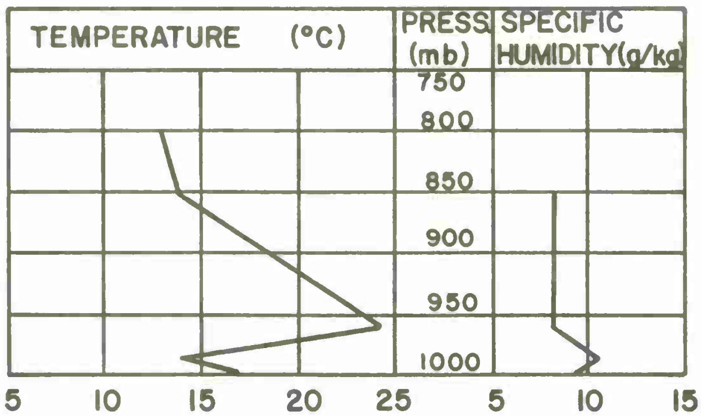


Figure 37. Characteristics of a maritime tropical air mass. (Modified from Rhiel)

The static properties of air masses are important to radio propagation, but their movements and interactions create meteorological phenomena which are probably of even greater importance. In the next section, dealing with inversions, the movement of air masses will be seen to be an important source of those features. In the remainder of this section, some of the basic considerations of the interaction between air masses will be presented.

c. Interactions Between Air Masses (Fronts)

Air masses differ in their mean temperatures, mean water vapor content, and therefore their mean densities. In order to satisfy the energy budget of the earth, and to balance the angular momentum budget of the atmosphere, the general circulation of the air is constrained in such a way as to move the colder, drier, denser air masses from polar latitudes toward the equator, and warmer, moister, lighter air masses from subtropical regions toward the poles. In mid-latitudes these air masses meet to form the familiar weather patterns which we see from day to day.

The boundaries between well defined air masses are called <u>fronts</u>, and for dynamical reasons they remain narrow distinct transitions between the air masses. Along these fronts, a great deal of meteorological activity takes place, and it is that activity which we will investigate now.

Colder, denser air, as might be expected, will normally replace warmer, lighter air at the earth's surface. This occurs along a so-called cold front. Also, however, warmer air can be dynamically forced to replace more dense, cooler air. Such a condition occurs along a warm front. The conditions which exist through a cold front are illustrated in Figure 38 and through a warm front in Figure 39.

In Figure 38, the salient features to observe are the following: (i) the frontal slope is steep with colder air pushing under warmer air; (ii) the band of weather is narrow but intense; (iii) the speed of the front is rapid, perhaps 30 to 50 km/hr; (iv) the likelihood of large cumulus clouds and therefore severe thunderstorms is high; (v) strong, gusty winds are normal; (vi) a strong temperature inversion will occur through the front but both air masses near the front will most likely be saturated with moisture because of rain showers; and (vii) rain will be general along the front but with great local variation in intensity. The large amount of moisture, the form of the moisture, the existence of electromagnetic noise energy (lightning), all of which are found in cold frontal activity, are quite important to most radio transmission.

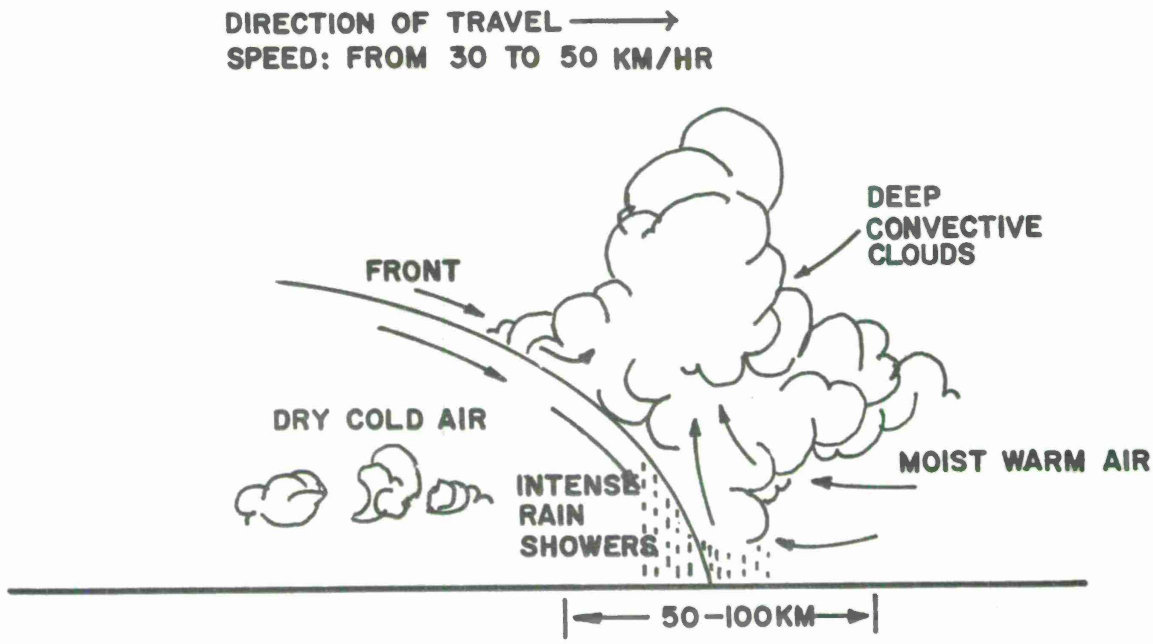


Figure 38. Cross-section through a cold front.

A warm front, on the other hand, has other unique features some of which are illustrated in Figure 39 and enumerated below. Of importance to notice are: (i) the very shallow slope of the frontal surface; (ii) the extended region of weather; (iii) the layered stratiform nature of the clouds; (iv) the less rapid movement with respect to a cold front; (v) the excellent chance that the overrunning warm air will create a widespread but weak inversion through the frontal surface with moisture above and below the inversion; (vi) rain will be widespread, steady, and composed of a smoothly varying raindrop size distribution; and (vii) there will be a good chance for fog to occur near the intersection between the earth and the front.

The less violent nature of the warm front and its discontinuities usually makes the warm front phenomena less important to radio propagation than cold frontal weather, although this will depend upon the frequency and the spatial relationships between the front and the transmission path.

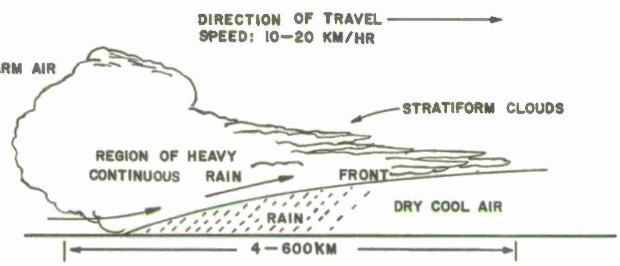


Figure 39. Cross-section through a warm front.

Several more tropospheric phenomena which are very important to propagation are covered in the next two sections. The first elaborates and draws together information on temperature inversions. The final section of our discussion of the troposphere will then attempt to outline some of the most important characteristics of stability and turbulence.

3. Tropospheric Temperature Inversions

Already we have seen, in two types of air masses, how two types of temperature inversions are produced. One was a result of strong radiational cooling at the surface of the earth; the second was formed aloft because of subsidence of the air. These and several other possible temperature inversions are elaborated upon below.

a. Radiational Inversions

Inversions due to radiational cooling are very common over land areas, forming particularly at night in clear dry air. The land acts as a very poor reservoir of heat and, after the sun sets, quickly cools. The air in contact with the surface is also cooled forming a shallow layer of cold air which intensifies until dawn. Usually, soon after the sun rises, this form of inversion disappears. Moderate surface winds prevent such a shallow surface inversion from forming by mixing the lowest layers of the air.

b. Subsidence Inversions

The most extreme version of a subsidence inversion occurs in Maritime Tropical air masses as shown in Figure 37. However, these types of inversions can form wherever subsidence occurs. This is a general condition in the eastern regions of migrating high pressure cells over land or water. The subsidence inversions over land show somewhat the same moisture trends as over water, merely less accentuated since the sources of moisture are less extensive.

c. Inversions Caused by the Movement of Air Masses

Finally, surface and free inversions (free meaning above the surface) can be caused by the movement of air masses. Consider the following three cases: (i) warm dry air moving over cooler water (The lowest layers of the air mass are cooled and water vapor is evaporated into them. A very strong inversion can result with a very marked contrast in water vapor through the inversion.); (ii) warm air moving over cold land (the same process as (i) above takes place but without the moisture contrast); and (iii) warm, usually moist, air, overrunning cold air. (Free, usually weak inversions can be caused by warmer, lighter, air riding up and over colder denser air. Many times the colder air is trapped in valleys or depressions.).

The most typical regions for all of these inversions to form can be pointed out with respect to a typical weather chart which plots lines of equal surface pressure and shows thereby moving air masses. Figure 40 is such a synoptic (meaning instantaneous) pressure chart over the United States during winter. Seven regions are shown which can typically form various types of inversions. They are listed and explained below:

Region I: A strong free inversion is formed by the subsiding air at the eastern end of the anti-cyclone.

The surface layers above the ocean but below the inversion are quite moist while the air above the inversion is rather dry. Normally no surface inversion can exist here because the ocean water will be warmer than the air above it.

Region II: A region of snow and ice cover and weak pressure gradient. A surface inversion is the expected feature of this area forming for the same reasons as in an arctic air mass. It will be somewhat weakened in the daytime but probably not entirely wiped out.

Region III: Another region of weak gradients but no snow cover. Shallow surface inversions form in the clear dry air, intensifying until dawn, then disappearing soon after the sun rises.

Region IV: A warm frontal system is approaching; warm air is replacing cooler air. The density differences caused by the temperature change across the front leads to the warm air riding up and over the cooler air. Usually a shallow free inversion occurs at the boundary between the two air masses. The warm air is usually moist and the cooler air considerably drier.

Region V: Conditionally unstable air between the warm and cold front is moist to high levels and generally well mixed. Winds are moderate and help to keep inversions from forming.

Region VI: Cold, dry, unstable air leads to a great deal of turbulence and mixing. Gusty winds, uneven solar insolation, and wet patches contribute to the turbulence eliminating all possibility of stable inversions.

Region VII: Similar conditions exist here as are portrayed in Figure 37. A subtropical high over an oceanic region creates a very strong, clearly defined subsidence inversion with great moisture contrast between the air above and below the inversion.

The strength and even the existence of the inversions discussed above will vary widely depending upon the intensity of the weather systems and the past history of the regions over which they are moving.

We will leave the subject of inversions for the moment and move on to a study of atmospheric stability and turbulence.

In recent years, the turbulence of the troposphere has been found to produce small random reflections and refractions of radio signals at high frequencies so that beyond the horizon communication has become routine using this effect. Our knowledge of the atmosphere must include, for completeness then, some of the facts which apply to turbulence. That will be our last topic in the meteorological study of the troposphere.

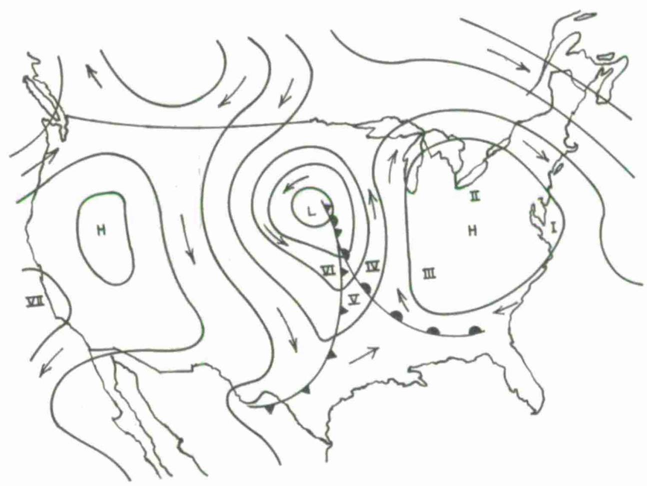


Figure 40. The occurrence of inversions can be related to specific patterns on weather maps. Details can be found in the text. (After Meyer)

4. Turbulence in the Troposphere

All the motions of the atmosphere can be considered turbulent as opposed to the periodic or laminar motions which can occur in certain idealized fluid systems. Even the largest motions of the atmosphere, the continental sized waves with which we associate the movement of air masses, are only quasi-periodic, and they take on semi-random shapes, intensities, and movements. They are merely the most convenient size for the atmosphere to transport the excess of heat energy, which would otherwise accumulate in the tropics, to the heat deficit regions near the poles.

On any scale, the horizontal and vertical movements of the atmosphere are random functions of time and space, even down to the smallest measurable movements. The average amplitudes of the impulsive movements which we see as gusty winds, or as weather systems, or feel as minute puffs or air against our faces, however, are limited. The limit is imposed by the dynamic stability of the fluid itself. By way of an introduction to turbulence, therefore, it is most valuable to describe the basic physical concepts of atmospheric stability. Many of the properties of atmospheric turbulence can be qualitatively predicted if the stability of the atmosphere with respect to small vertical displacements is known.

a. Stability of the Troposphere

Refer to Figure 41. The figure shows three lines that have a common origin on the temperature axis. The solid line represents the standard atmosphere; the other two lines represent the temperature of a parcel of air as it ascends in the atmosphere under two sets of conditions. Both sets include the stipulation that no heat is added to this parcel from its exterior. Any process occurring under such conditions is said to be an adiabatic process. The more sharply sloped line to the left is for an unsaturated (or dry) parcel of air. As it rises, the pressure around it lessens; the parcel expands, doing work against the local environment, thus decreasing its internal energy. Since its internal energy is only dependent upon its temperature, the temperature of this parcel must decrease. And it will do so, if dry, so that its temperature follows the dashed line labeled dry adiabat. In the real atmosphere, in most situations, this behavior is followed quite closely for dry air moving vertically.

What if the parcel is saturated with water vapor? Then its temperature will follow the dot-dashed saturation adiabat to the right of the dry adiabat in Figure 41. As the parcel cools, some of the vapor must condense out because, as was pointed out previously, the maximum amount of water vapor any parcel can hold depends only upon its temperature. Condensation of water vapor releases heat in proportion to the difference between the internal energy of water vapor and liquid water at the same temperature. This heat (called heat of condensation) causes the temperature of the saturated air parcel to decrease more slowly with height than that of the dry parcel; therefore, the more vertical slope of this temperature curve. As the parcel rises higher and higher, the water vapor is depleted and the slope of the saturation adiabat approaches asympotically to that of the dry adiabat. For comparison purposes only, the standard mid-latitude mean temperature profile was included. It is interesting and important to note that the slope of this standard atmosphere in the lower troposphere falls in between the slopes of the dry and wet adiabats.

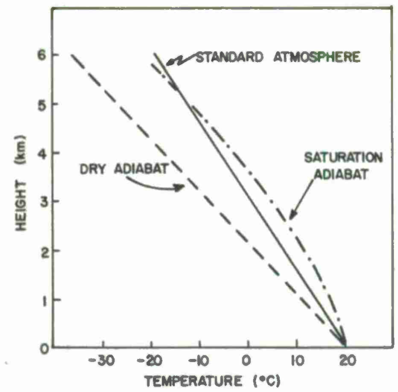


Figure 41. Temperatures of ascending dry and moist air parcels.

Figure 42 shows the implication which dry and wet adiabats have for stability in the atmosphere. The three solid lines schematically represent three possible temperature profiles versus height in the real atmosphere with some exaggeration, however, of the slopes of the outside solid curves for clarity. The dashed lines starting at height Z_1 from the same temperature are the dry and wet adiabats of Figure 41. The lower horizontal line is a reference height Z_1, from which we will launch several air parcels and investigate their behavior.

Refer to profile (1). If we take a parcel of dry air at the reference height Z_1, which has the temperature of the surrounding air, and cause it to ascend over a short distance to Z_2, we will find its temperature has followed the slope of the dry adiabat. Its temperature will be higher than the new surroundings by ΔT_1. Because it is warmer, it will be less dense. Because it is less dense it will be buoyant and continue to rise. The parcel, therefore, is unstable to vertical movements. If the parcel were initially saturated at Z_1, the difference T_1 would have been the difference between the temperature of profile (1) and the wet adiabat at Z_2 and it would have ascended even more rapidly. Profile (1) describes an <u>absolutely unstable atmosphere</u>.

Profile (2) shows a different situation. If the parcel is initially dry when it is raised from Z_1 to Z_2 its temperature will follow the slope of the dry adiabat. At Z_2 the temperature of the parcel will be less than its surroundings by an amount $(\Delta T_2)_D$. Its density will be greater than its surroundings causing it to fall back to its starting level.

If the parcel were initially saturated at Z_1, it would have followed the saturated or wet adiabat to Z_2 and found itself there warmer than its surroundings by $(\Delta T_2)_W$, less dense, and therefore unstable. Profile (2) describes what is called a <u>conditionally unstable atmosphere</u>.

Profile (3) shows an absolutely stable atmosphere. Whether the parcel is dry or saturated at Z_1, it arrives at Z_2 cooler, hence more dense than its surroundings, and falls back to Z_1. Even though the illustrations of Figure 42 are designed to show the consequences of air parcel movements upward, exactly the same arguments will show that unstable atmospheres are unstable for downward displacements also. If a parcel in an unstable atmosphere is forced downward, it will continue downward, whereas a parcel forced downward in a stable atmosphere will return to its starting level. With this knowledge, it will now be rewarding to investigate the characteristics of turbulence in the lower atmosphere.

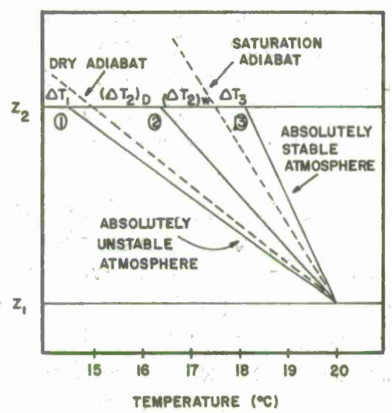


Figure 42. The reaction of a parcel of air to vertical displacement for several temperature profiles. (After Meyer)

b. The Origin of Turbulence in the Atmosphere

Even though the dominant feature of atmospheric motions on all size scales in turbulence, it is an unfortunate fact that it remains one of the least well understood of meteorological phenomena. However, we may still describe qualitatively its origins and the most important controlling conditions in the troposphere. We will limit our discussion to the turbulence most important to radio propagation; it ranges in linear size from a few meters to perhaps a few kilometers.

The ultimate energy source for all turbulence in the atmosphere is the differential heating of the earth by the sun. Local turbulent eddies, of the size which interest us, receive their energy, however, from either of the two sources: (i) convection, the rise of warm air parcels in a generally cooler surrounding; or (ii) from friction. The latter can be internal friction of the fluid or it can be friction with the earth itself.

We can begin our survey of atmospheric turbulence by referring to Figure 43. It shows an idealized mean profile of the wind speed as a function of altitude. Three distinct regions can be identified: (i) a surface influenced layer, sometimes identified as the planetary boundary layer; (ii) a well mixed layer of convective activity; and (iii) a region above which the wind increases slowly with height.

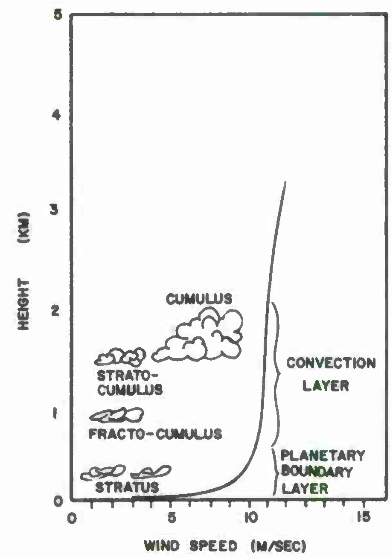


Figure 43. An idealized profile of the average wind speed with height. (After Meyer)

One would correctly expect that right at the surface of the earth the wind speed would be nearly zero due to friction. The height above the surface which is influenced by this surface friction is determined by the stability of the atmosphere. Remembering our discussion on stability, it will be recalled that a very stable atmosphere retards the vertical motions of an air parcel. In such a stable atmosphere, higher velocity parcels cannot descend to inject energy into the boundary layer flow; the influence of friction may then reach as high as 500 m. An unstable atmosphere mixes more easily and the vertical influence of surface friction is reduced. The boundary layer in this case is below 500 m.

The second zone of Figure 43 identifies the region of most active cumulus cloud production. It is the region in which a great deal of convective activity takes place and is consequently well mixed. The mean wind speed which is measured, is relatively constant over the layer since any differential wind speeds are soon diffused and dissipated throughout the entire region.

The role of stability in damping turbulence can be illustrated further by observing the diurnal variation of the wind speed over sunny days. At night the surface radiation stabilizes the lowest layers of the atmosphere and the wind speed near the surface decreases to its minimum value. Several hundred meters above the surface, however, the wind speed may remain strong. During the daytime, as convection sets in because of solar heating at the ground, the surface winds pick up at expense of the winds several hundred meters above the surface. In this latter case convective mixing brings down impulses of higher speed air and lifts impulses of frictionally slowed air, causing, in the mean, a more uniform vertical wind speed profile.

The friction between the moving air and stationary earth is easy to visualize. But within the moving air streams themselves, what is the mechanism for the generation of turbulence? The answer is parcel exchanges between adjacent air masses. Such exchanges between adjacent air masses take place constantly. Friction between two parcels may be viewed as existing when exchanged parcels are moving at different speeds. The exchange tends to slow down one parcel and speed up the other. One look at Figure 43 shows that this condition is certainly fulfilled in the boundary layer near the surface, i.e., there are moving parcels adjacent to each other which have different speeds. Turbulence then is to be expected near the surface whenever the wind blows.

The horizontal size of the elements of tropospheric turbulence seems to be an important parameter in electromagnetic propagation, at least in tropospheric scatter communication. The clouds and air bubbles associated with rising thermals give some idea of the horizontal scale of a few of these elements. They range from a few meters in the smallest thermals to perhaps 30 km in the largest cumulo-nimbus clouds. The maximum differential velocity with respect to the surrounding medium which these vertically ascending bubbles can have is in the neighborhood of 40 m/s in the largest thunderstorms.

Internal friction plays the major role in the severe turbulence which is associated with the jet stream, the narrow river of air which occurs near the tropopause and which can reach speeds in excess of 400 km/hr. The rapid transition between slower moving air and this jet creates tremendously strong turbulent eddies; strong enough to break apart large aircraft. This turbulence has been called clear air turbulence because it is not normally associated with visible features in the atmosphere like clouds are with low level convection. It is believed that the turbulence around the jet stream can lead to pressure differentials in the atmosphere as large as 3 percent between nearby regions. This could be a significant mechanism for the scattering of electromagnetic waves propagating through the jet stream.

One final word on turbulence will be concerned with the cascade of energy which is believed to occur in turbulent conditions. That is, if energy is injected in some manner into large scale turbulent eddies, the energy will eventually cascade toward smaller and smaller sized disturbances until it is dissipated at the molecular level. Therefore, large scale turbulence guarantees the existence of the smaller scales.

c. Turbulence and Weather Phenomena

It must be obvious by now that any weather conditions that promote stability in the atmosphere, promote a decrease in turbulence. We may return to Figure 40 and associate different parts of weather systems with different degrees of turbulence.

Region I: Cold dry air is flowing out over a warmer ocean. In the zone below the strong subsidence inversion, the atmosphere will be quite unstable, and a great deal of mixing and overturning will occur distributing moisture uniformly through this marine layer. Turbulence should be strong in this region, and considerably weaker above the inversion.

Regions II and III: Both of these areas share weak pressure gradients (therefore weak winds) and stabilizing nocturnal surface inversions. At night, calm conditions should prevail near the surface with stronger steady winds aloft. During the daytime, particularly in Region III, the inversion breaks, mild convection occurs, and turbulence is at its diurnal maximum.

Region IV: An approaching warm front stabilizes the atmosphere above the inversion. Clouds are predominately horizontal in their distribution and distinctly layered. Internal friction can cause turbulence in this warm air aloft if the winds are strong as is often the case. Normally, the winds speed, and turbulence, will increase with height. Below the inversion of the front, cool stable air remains. It is being dynamically removed, however, and as the front approaches the depth of the cool air diminishes, turbulence increases until it reaches a maximum at the front itself. In the transition zone, a sudden shift in the direction of the wind is normal and can lead to considerable turbulence.

Region V: The warm sector is subject to instabilities if sufficiently strong impulses are provided to set them off. One sees widely scattered showers and cumulus clouds indicating convection. Winds are usually moderate. Turbulence, again, is likely to increase with height as do the winds.

Region VI: Cold, dry, very unstable air marks the region behind the cold front. A great deal of mixing occurs at all levels, accompanied by strong winds. Surface turbulence is normally a maximum in this region and because of vertical shear in the wind field turbulence remains strong completely up through the tropopause.

Region VII: Similar to Region I except the air below the inversion has been over the ocean for a long period and will be less unstable. Nevertheless, small convective cells will be prominent in the lower layer indicating some steady turbulence at least partly induced by mixing caused from surface friction.

C. THE EFFECTS OF THE TROPOSPHERE ON ELECTROMAGNETIC PROPAGATION

With the information concerning the troposphere which we now possess, a much more physically realistic picture can be drawn of the changes which electromagnetic radiation undergoes as it propagates through the lower atmosphere. The effects of importance can be grouped under three main headings: (i) refraction; (ii) scattering and diffraction; and (iii) absorption. We will begin our discussion with the refraction effects of the lower atmosphere.

1. Refraction Effects

On page 23, the refraction of an electromagnetic wavefront was shown to occur when the wave passed obliquely from one medium into another in which the speed of propagation was different. In order to predict refraction effects, one must know the variation in time and space of the speed of propagation.

Many important aspects of tropospheric refraction depend upon transient atmospheric effects. But also, the mean conditions in the atmosphere cause refraction and are in themselves important to know. Therefore, it will be most convenient to discuss tropospheric refraction in two parts: tropospheric refraction from mean atmospheric conditions; and anomalous tropospheric refraction.

a. Mean Tropospheric Refraction

It has been theoretically determined and experimentally verified that the index of refraction, n (speed of light in a vacuum, c, divided by the speed of propagation, V) may be closely approximated below 50 GHz (50×10^9 Hertz) by the following formula:

$$(n - 1) \times 10^6 = N = \frac{77.6}{T} \left(P + \frac{4810e}{T}\right) \tag{10}$$

where n is the index of refraction (c/V), N is the refractivity and is used for convenience since n differs so little from 1, P is the total pressure in millibars, T is the temperature in degrees Kelvin, and e is the partial pressure of water vapor in millibars. The speed of propagation for these wavelengths is very closely independent of frequency.

The total pressure, the temperature, and the partial pressure of water vapor all decrease with height but in such a manner that for a mean atmosphere equation 10 leads to an almost exponential decrease of N with height. In terms of propagation speed, this means that a wave travels more rapidly at higher elevations. As a consequence, a wavefront proceeding non-vertically through the atmosphere will tend to be refracted downward, toward the surface of the earth (see Figure 15 for the reason behind this). Figure 44 shows the smoothed refractivity as a function of height for a hypothetical atmospheric sounding which has two temperature inversions and a moisture profile indicated by the dew point temperature. The difference in speed between a wavefront and the speed of light in a vacuum, is indicated at several heights.

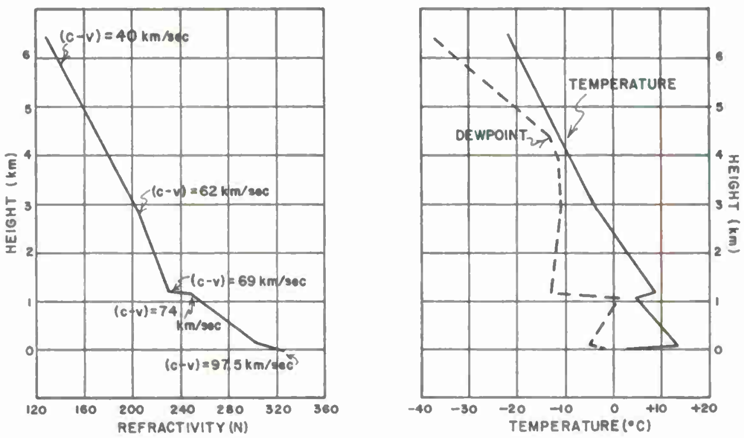


Figure 44. A possible relationship between atmospheric temperature and moisture and the refractivity. V is the velocity of propagation at a given height in the atmosphere; c is the speed of electromagnetic signals in a vacuum ($\sim 3 \times 10^5$ km/sec).

For radio propagation it is often necessary to know the refraction which will occur between two widely separated points. Details of atmospheric structure sufficient to compute the refractivity are seldom available. However close approximations can be made based upon the known mean characteristics of the atmosphere as a function of geographical location and height. Three models are currently in use, and it will be instructive to review each and point out their shortcomings when applied to the real atmosphere.

The first model has been used for many years in propagation problems and is called the Effective Earth's Radius Model. It assumes an earth which is larger than it really is by an amount which corrects for the normal curvature of radio waves propagating in the atmosphere near the earth's surface. This hypothetical earth has a radius 4/3 as large as the real one. Such a model appears to be accurate, however, only for the first, or at best, the first two kilometers in the atmosphere. Figure 45 shows the average refractivity above Portland, Maine, in February and for Miami, Florida, in August and the 4/3 radius approximation to each. The two averages represent extremes of the N profile. The plot of the real refractivities extend into an envelope which includes a wide sample of N profiles in the United States. It can be seen that the 4/3 radius approximation to N in each case and for the broad sample decreases far too rapidly with height above one kilometer.

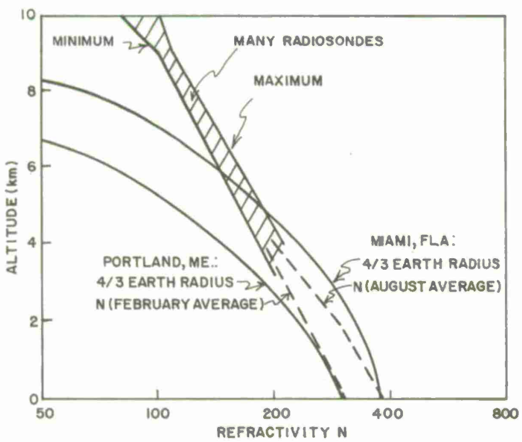


Figure 45. A comparison of refractivity profiles with the Effective Earth's Radius Model for two locations. Envelope for many observations is also shown. (After Bean and Dutton)

The plot of Figure 45 is logarithmic in N and the envelope of many N profiles shows up almost straight. This indicates that the average N profile follows an exponential decrease with height.

This is the basic assumption of a second model for atmospheric refractivity, the Exponential Model. This model simply assumes that the refractivity declines exponentially with height from the surface value, N_s, at a rate determined by the value of N_1, the refractivity at one kilometer in height. Long term statistics have shown that a relationship exists between N_s and N_1 so that one needs only to know N_s to have a complete model for N in the vertical.

A third model for the refractivity in the atmosphere is really a three part model and is known as the Modified Effective Earth's Radius Model. From the surface to one kilometer, the $\sqrt{4/3}$ earth radius model is preserved. From one kilometer to 9 km, an exponential decay is used between the values at one kilometer and a constant value at 9 km derived from long term statistics of the atmosphere. Above 9 km, another exponential decay is used with a scale height (distance required to reduce the value by 1/e) of approximately 7 km.

Of the three models, the Exponential Model turns out to be probably most accurate in the crucial first three kilometers of the atmosphere. Climatological values of the sea level refractivity derived from the average refraction between any two points may be computed using either the Exponential or Modified Effective Earth's radius model assumptions.

The actual bending of a ray under the mean conditions which have been discussed and modeled above can be exemplified by using the Effective Earth's Radius model for a few rough calculations. If an atmosphere exactly duplicated the linear decrease of N units which this model calls for (39.2 N units per kilometer), then a ray launched horizontally would bend approximately 0.67 degrees per one hundred km (0.0067°/km); whereas the earth's surface curves about 0.89 degrees in 100 km (0.0089°/km). After 250 km, this horizontally launched wavefront would be about 4.2 km above the surface of the earth, and after 1000 km its height would be calculated to be about 63 km by this model, low by perhaps 8 to 12 km, however, from the real ray path.

b. Anomalous Tropospheric Refraction

Under this heading will be included all effects which are not explained by the refraction characteristics of the mean conditions in the troposphere. Some of these effects we can only gain a qualitative understanding of, mostly by studying the tropospheric conditions which give rise to them. Such conditions occur for instance in frontal systems. Other effects, most notably trapping of radio energy in ducts, may be studied in more detail because they are somewhat more regular atmospheric features.

(1) Anomalous refraction in the vicinity of fronts

Fronts present generally chaotic, highly variable meteorological conditions over small linear dimensions. Temperature, moisture and winds change abruptly from one air mass to the other. It is little wonder that the refractivity also shows rapid changes and extreme behavior in the vicinity of frontal systems. Figures 46 and 47 illustrate in a particularly revealing manner how an ideal cold front and an ideal warm front appear in terms of what are called A units. A units are simply the refractivity, N, corrected for the normal exponential decrease of N with height. The conversion from N to A is accomplished using equation (11) below:

$$A = N + N_m (1 - e^{-h/H_N}) \tag{11}$$

where N is the observed refractivity, h is the height above sea level, N_m the mean sea level value of N (313), and H_N is the averaged scale height for N (7km). In the figures, the variation of N caused by the peculiarities of the fronts themselves are clearly brought out.

The most interesting features of Figure 46 and 47 are the clear indication of air mass features. In the cold front of Figure 46, the stratified cold air mass may be easily distinguished from the warm, moist air being lifted in front of its approach. In Figure 47 the warm, moist tongue of air sliding up and over the cold air interface is easily identified. Also, the region of precipitation below the front is obvious from the area of high surface values of A.

Such chaotic conditions as found in frontal systems lead to propagation effects which we can only qualitatively explore. It can be said that the effects will be frequency dependent, the higher frequencies being more affected because of the increased inhomogeneities and scattering bodies present on smaller size scales. One would expect many possible and rapiply changing propagation paths through frontal weather. Consequently, one would expect fading due to multipath interference to be more prevalent in such regions than in less turbulent, more homogeneous air masses. Through frontal systems, the increased number of refractions out of the main beam of a focused signal might lead to a greater loss of energy for such a signal being propagated over a well defined line of sight path than would normally be experienced. On the other hand, a signal which depended upon scattered and reflected energy out of a sharply focused beam, like that which occurs in troposcatter communication, might possibly be aided by the increased turbulence associated with frontal activity.

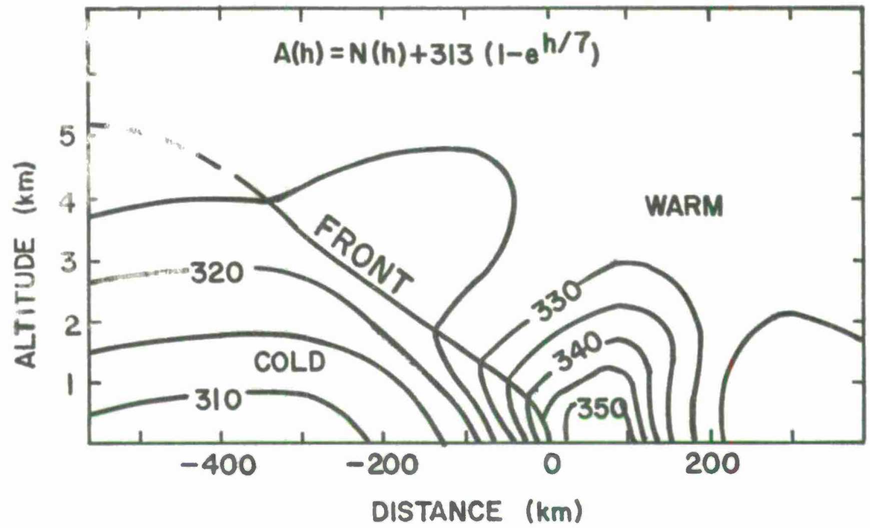


Figure 46. The pattern of A units near a cold front. Movement of front is to the right. The presence of strong gradients of refractivity (A units) is clearly shown. (Bean and Dutton)

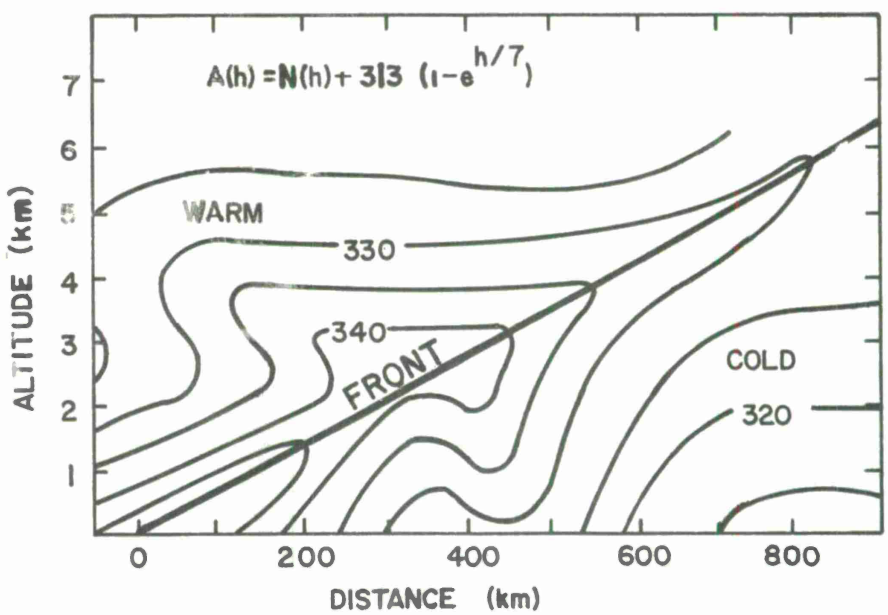


Figure 47. The pattern of A units near a warm front. Movement of front is to the right. Considerably less extreme refraction changes occur near warm fronts than near cold fronts. (Bean and Dutton)

In truth, all of the above effects are observed when propagation paths lead through frontal systems. Signals below 30 MHz are seldom affected because of their reliance for long distance communication on ionospheric reflections and because the wavelength (10 m at 30 MHz) is too large to appreciably interact with the turbulent medium. Signals whose wavelengths are shorter than 10 cm (frequencies greater than 3 GHz) are appreciably affected by frontal activity for just the opposite reasons to signals below 30 MHz. They cannot rely on ionospheric propagation and therefore energy received from a transmitter must pass through the lower regions of the troposphere where frontal activity is greatest. The wavelength of such signals allows them to interact not only with the turbulent medium but also with the cloud particles and raindrops present. The highest frequency signals are both scattered and, more importantly, can be strongly absorbed. Intermediate frequencies between 30 MHz and 3 GHz (wavelengths of 10 m to 10 cm) can be either enhanced or degraded depending upon their usage. More details about this region will be brought out in the discussion covering troposcatter communication.

(2) Anomalous refraction from horizontally stratified layers

On page 64, an examination was made of temperature inversions in the troposphere and how they are regularly formed. It was pointed out that many times significant changes in the water vapor which the atmosphere holds, also occurs across these temperature inversions. It is now possible to show why such stable layers and other horizontally stratified regions are important to tropospheric radio communication.

If one reviews equation (10) for the effects of a temperature rise and a moisture decrease over a thin vertical region, it will be quickly apparent that both processes will tend to cause the value of N to decrease. Some temperature inversions are strong enough and the moisture contrast high enough to cause N to decrease so rapidly with height that an electromagnetic signal can be refracted so that its curvature is equal to or greater than the curvature of the earth's surface; hence the signal will parallel or return to the earth. Such a situation was illustrated in Figure 44 where the surface inversion and elevated inversion both show decreases in N greater than the 157 N units per kilometer, the vertical rate of decrease needed to refract a signal with curvature equal to that of the earth's surface. It is these so called <u>ducts</u>, where the gradient of N is greater than 157 N units per kilometer in the vertical direction, and the associated but weaker superrefracting layers, which we wish to examine in some detail in this subsection. We will begin by defining various anomalous layers, talk about their distribution in space and time, and analyze, as best is known, their individual effects on tropospheric signals.

Four categories of refractive index profiles may be identified and related to propagation: (i) subrefractive layers; (ii) unstratified regions; (iii) superrefractive layers; and (iv) radio ducts. Subrefractive layers are those which show an increase of N with height. Unstratified regions are defined as those layers having the vertical gradient of N near to the normal expected for that particular height. Superrefractive layers are those in which the gradient of N is at least twice as great as the average gradient for that level in the atmosphere. And a duct, as already defined, is a layer whose N gradient is large enough that radio waves are trapped, or better, guided along the geometry of the stable layer.

In order to better understand physically what an average N gradient is it may be recalled that the best fit to the mean N profile in the atmosphere was exponential with height, hence we may write

(12) $$N(h) = N_o \exp(-h/H)$$

where N (h) is the refractivity at a height h, N_o is the sea level refractivity determined from the local surface value, and H is the scale height of the exponential decrease of N. The mean vertical gradient of N is therefore given by

(13) $$\frac{dN(h)}{dh} = \frac{-N_o}{H} \exp(-h/H)$$

another exponential. At sea level, dN(h)/dh = - (313/7) = -44.7 N units per vertical kilometer, an increase in the speed of propagation per kilometer in height of about 14.9 km/sec. At 7 km, dN(h)/dh = -16.4 N units/km, an increase of the speed of propagation of only 5.8 km/sec for each kilometer in height.

The significance of the anomalous N gradients we have defined lies with the propagation of signals above 30 MHz. These signals are not able to be propagated using normal (reflection-refraction) ionospheric techniques, and therefore depend upon the refraction found in the troposphere plus diffraction and scattering to reach over the geometrical horizon from the transmitter site. We already know that mean conditions in the troposphere lead to some bending of rays downward causing the normal radio horizon to be beyond the geometrical horizon. Each condition above will modify this normal trans-horizon propagation.

Refer to Figure 48. This idealizes the effects of each of the four categories of N profiles. A subrefractive zone causes a wavefront moving obliquely to it to deviate from a straight line but in a direction away from the surface of the earth. The radio horizon for a signal propagating in a subrefractive region is quite obviously nearer than the geometrical one. The incidence over the earth of such layers seems to be low, and consequently no study of their regular appearance is readily available. This general effect, however, is clearly to cause anomalously short ranges for tropospheric propagation.

An unstratified or normal atmosphere can be thought of as one in which the N gradient is between zero and twice the average for that height. An unstratified region refracts radio waves toward the surface of the earth but in a way similar to that discussed in the section dealing with mean tropospheric refraction.

A superrefracting layer is arbitrarily identified as one having an N gradient at least twice that normal for the height of the layer. Most enhancement of trans-horizon tropospheric communication is caused by superrefractive layers. The horizontal extent the persistence, and the intensity of such layers has been the subject of widespread interest for several years but a great deal still remains to be done before they are well understood.

It appears that superrefractive layers above the earth's surface are more prevalent than once thought. Indeed, recent evidence has tended to show the troposphere to be generally more layered and horizontally stratified than distantly separated and slowly responding radiosondes report. Airborne refractometers, which measure the index of refraction of the free air directly and in most cases with very rapid response times, show layers as thin as 10 m extending over a few kilometers in horizontal distance. More typically layers of 100 m thickness are detected whose horizontal extend may be tens of kilometers. Many times the detected layers are found in the stable air of elevated temperature inversions. In the regions of the tropics where the oceanic trade wind inversion exists superrefracting layers exist normally and may be 300 to 400 meters thick and extend thousands of kilometers.

More effective reflection plus scattering and diffraction from these layers has been invoked as the physical reason that over-the-horizon propagation enhancement is highly correlated with the existence of such layers.

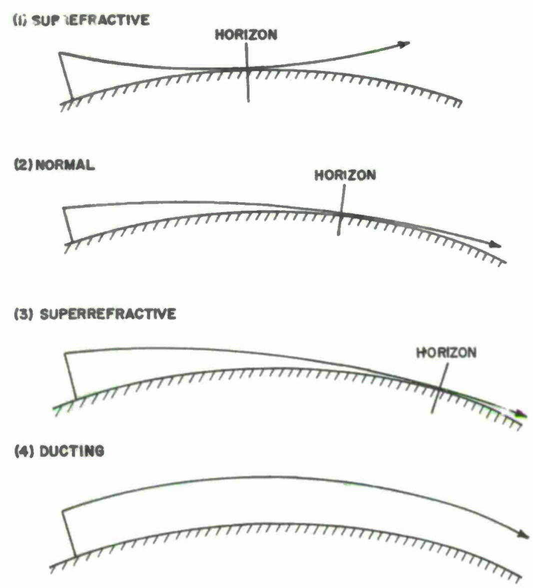


Figure 48. Four general categories of the vertical changes of the index of refraction and the resultant effects on propagation. The atmosphere is subrefractive when the index of refraction increases with height; superrefractive, when it decreases at twice the normal rate; and ducting, when it decreases at the rate of 157 N units/km or greater.

Besides elevated superrefracting layers, there are many conditions which cause superrefracting layers to form with their bases on the surface of the earth. These also enhance trans-horizon propagation by increased scattering and diffraction. These layers are closely associated with ground inversions, many times forming from radiational cooling at night.

The last category of horizontal layers important to radio propagation is really a special case of the superrefracting one. In the case of radio ducts, the magnitude of the vertical N gradient exceeds 157 N units per kilometer leading to the refraction of radio waves so that the curvature of their paths is greater than the curvature of the earth's surface. If the layer is thick enough with respect to the wavelength of the radiation being propagated, a high percentage of the energy will be trapped within the duct and be guided by it.

The complete physical picture of the operation of a radio duct is quite complex. However, we may qualitatively explain the basic idea by using Figure 49, and then pointing out how this picture must be modified for a more exact representation of the phenomenon.

In Figure 49, we assume that a superrefracting layer is based on the ground and has thickness d. A transmitter with an isotropic radiation pattern (i.e. radiates in all directions equally effectively) exists at height Z. Three paths which the radiation might take are depicted. Path 1 is launched at a shallow angle with respect to the horizontal. The refraction is great enough so that after a short distance the ray is bent downward and travels until it reaches the earth's surface. Reflection takes place causing the ray path to again move away from the earth. Further investigation, however, shows that the ray cannot leave the superrefracting layer and proceeds in a series of hops. The ray is trapped and the layer is regarded as a radio duct.

If we decrease the launch angle, similar results occur. Different results occur if we follow a part of the signal launched at a steeper elevation angle. Eventually a path will be found (Path 2) where the signal can just reach the edge of the duct. It will asymptotically approach the boundary, whereas all steeper launch angles away from the horizontal will actually penetrate the boundary and be lost (Path 3). All lesser launch angles will be trapped.

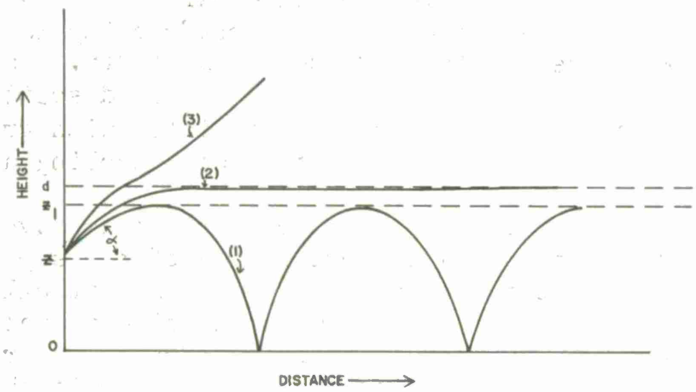


Figure 49. A radio duct and the effect of launch angle (α) on a hypothetical signal. Rays whose launch angles are greater than that for ray path 2 will not be trapped; those with launch angle less than for ray path 2 will be trapped. (After Freehafer and Kerr)

Upon closer examination, it is found that any signal may be broken down into two components: one which travels along the duct; and one which travels vertically. The vertical component is restricted, however, because it can be thought of as reflected from the surface of the earth and from a level above the earth determined by the vertical distribution of the index of refraction. Using the same type of reasoning which we used in Section 2, page 34, for the formation of special modes between two reflecting plates, we can postulate correctly that only reflections which reinforce each other can exist; all others will have random phase differences and cancel. This leads to the establishment of preferred launch angles $\alpha_1, \alpha_2, \ldots, \alpha_m$ where m describes different allowed modes. The analysis shows further that each m corresponds to a particular reflection height Z_m. This height, called the track height or duct thickness increases with the order of the mode (i.e. increasing m).

Modes may be thought of as existing at all times in propagation over the earth even when the troposphere does not refract. However, if less than the trapping gradient of the index of refraction exists in the vertical, a great deal of energy from each mode leaks above and beyond the reflection height Z_m as the wave is forced to bend around the curvature of the earth. The conditions for a duct appear when the effect of tropospheric refraction is greater than the curvature of the earth through a depth at least as great as the reflection height for the first mode Z_1. Then this mode leaks very little energy out of its track (between the surface and Z_1) and the wave is guided around the curvature of the earth with small loss.

The angle α_m and consequently the track heights Z_m, are functions of the wavelength λ. It is found that the first mode track height for a wavelength of 10 centimeters is about 11 meters; for a wavelength of 1 kilometer, the track height is about 5 kilometers. Since we know that ducting layers seldom exceed 300 to 400 meters in thickness and are usually much less, it is obvious why the shorter wavelengths are greatly favored for trapping.

2. Gaseous Absorption Effects

a. General Comments

The atmosphere, as perceived and treated in the study of weather, is taken to be a continuous fluid. And, for most studies of the interaction between electromagnetic waves and the atmosphere, it is sufficient to treat the atmosphere as an infinitely divisible fluid. However, when we talk about absorption, we must take into account, explicitly, the fact that individual molecules are the building blocks of this fluid and that particles of various kinds are suspended in it, because changes in the internal energies of the individual molecules and particles control the most important form of atmospheric absorption of electromagnetic energy.

As was pointed out in Section 2, page 26, molecules possess energies of translation, rotation, vibration, and electronic excitation, of which the last three are quantized. It was not only pointed out that attenuation from the interaction between molecules and the radiation field occurs throughout the microwave region, but that there also exist narrow frequency regions which are of particular interest because the molecule exhibits greatly enhanced absorption. These frequency regions were called spectral lines. It might also be recalled that the energy of translation of the molecule is not quantized, i.e., it can have any value below some upper limit (and therefore the molecule may have, theoretically, any speed below that of light in a vacuum). As a consequence, translational energy changes do not produce sharp maxima of absorption or emission.

The absorption lines of lowest frequency are in most cases due to discrete changes in the various rotational modes of individual molecules. Generally, the lowest of these lines occur at frequencies in the region from 10 to 100 GHz, but can be as high as several thousand GHz. Most lines of a given molecule resulting from rotational energy changes are usually found in the microwave or far infrared part of the electromagnetic spectrum.

Changes of the vibration modes rarely occur at wavelengths longer than 16 microns. The energies involved are so much larger than the minimum for rotational changes that vibrational changes almost never occur alone; there are almost always many simultaneous rotational changes. The result is a group of absorption lines called a vibration-rotation band. These usually occur in the intermediate infrared region.

Electron transitions usually involve energies of at least a few electron volts; hence they absorb and emit predominantly in the visible or ultraviolet spectral regions. Although atoms exhibit isolated electronic line spectra, molecules usually have complex absorption bands involving simultaneous changes in all three modes of energy.

b. Absorption by Atmospheric Gases

Although attenuation of radiation in the lower atmosphere occurs at all frequencies, it is normally too weak to be an important consideration until the wavelength of the radiation becomes less than 10 cm. For electromagnetic signals with a wavelength shorter than 10 cm, various physical constituents of the atmosphere become important as absorbers of energy from the radiating wave. We will discuss first the resonant absorption due to the important atmospheric gases.

(1) Nitrogen

Of the gases in the atmosphere of the earth, only oxygen, water vapor, carbon dioxide, and ozone are important absorbers of electromagnetic energy at infrared and longer wavelengths. The reason that nitrogen, the most abundant atmospheric gas, does not absorb in these wavelengths is due to its symmetric structure; that is, the electrons are distributed over the molecule so that no dipole effects like those illustrated in Figure 4 exist. Because no electrical or magnetic dipole moments are present, the nitrogen molecule can be oriented arbitrarily with respect to a surrounding electromagnetic field and feel no effect. Because of this, no spectral lines are produced when the molecule rotates in different configurations. Vibrations occur only along the axis connecting the two nitrogen atoms and consequently do not upset the symmetry of the molecule. Therefore, not even a transient dipole moment is produced which can interact with the ambient electromagnetic field, as does happen, for instance, with carbon dioxide.

(2) Oxygen

Oxygen is a diatomic molecule like nitrogen. It is shaped like a dumbbell with the two oxygen atoms separated by 1.2 Å as depicted in Figure 50a. For the same symmetry arguments used for nitrogen, no electrical dipole moment can exist. However, each electron spins on its own axis and acts therefore as a tiny magnetic dipole as illustrated in Figure 4d. There are an even number of electrons all of which, because of quantum physical laws, have there spin axes aligned. The resulting dipoles are either parallel or opposed in their orientation. Again, because of quantum laws, it is found that all but two of the dipoles are canceled by oppositely spinning electrons. These two are aligned and cause the oxygen molecule to have a residual magnetic dipole moment. If the spin axes of these two electrons were oppositely oriented, then oxygen would be as neutral as nitrogen. With the residual magnetic dipole moment, however, changes in the rotation or vibration rotation of the molecule can cause radiation to be absorbed or radiated. The rotation and vibration of the molecule is schematically illustrated in Figure 50c and d.

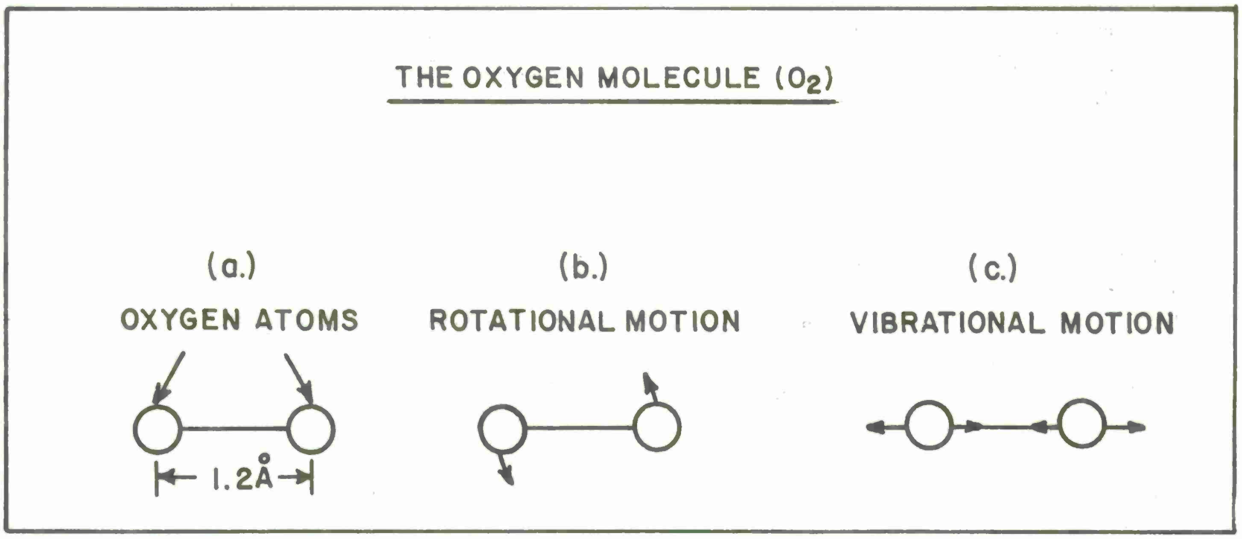


Figure 50. Diagram illustrating the size, and the rotational and vibrational motions of the oxygen molecule.

The oxygen spectrum at microwave frequencies is produced by changes in rotation only. It is plotted in Figure 51 for standard sea level atmospheric pressure and a temperature of 293°K (20°C). The absorption of radiation by molecular oxygen as a function of frequency between 3 GHz (10 cm) and 300 GHz (1 mm) is shown. For frequencies less than 30 GHz (1 cm) absorption is slight, never exceeding 0.01 db per km. The absorption in this lower frequency region is dominated by so called non-resonant absorption, as opposed to the resonant absorption evident in the region of the spectral lines. Below 15 GHz, non-resonant oxygen absorption is the dominant absorption mechanism in cloudless atmospheres.

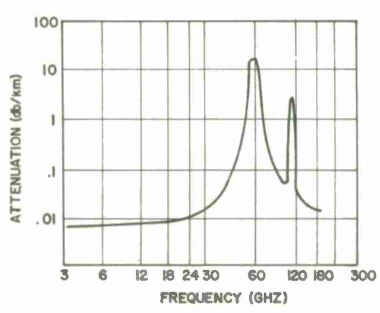


Figure 51. Absorption in db/km for oxygen as a function of frequency at sea level pressure and temperature of 293 °K. (After Van Vleck)

Twenty-five individual lines make up the 60 GHz absorption peak, while only one line exists at the higher frequency (118 GHz). The strongest absorption region has a peak of 15 db/km, while the peak attenuation in the single line approaches 3 db/km for the conditions stated. In either case, the absorption is strong enough to be of considerable importance over distances spanned by direct, line of sight, microwave links.

(3) Water vapor

Figure 52a illustrates the structure and properties of the water vapor molecule. It consists of two hydrogen atoms attached to an oxygen atom at an angle of approximately 104°. Because of the unique geometrical shape, the molecule shows a residual negative charge in the vicinity of the oxygen atom, and an equal residual positive change in the vicinity of the hydrogen atoms. The result is very like a tiny electric dipole as shown in Figure 4 of Section 2, page 7.

Water vapor molecules can rotate about any of the three axes shown in Figure 52b. They can vibrate in any of the three modes depicted in Figure 52c. The properties of absorption and emission of energy depend explicitly upon the physical properties, the rotational configuration, and the vibrational geometry shown in Figure 52.

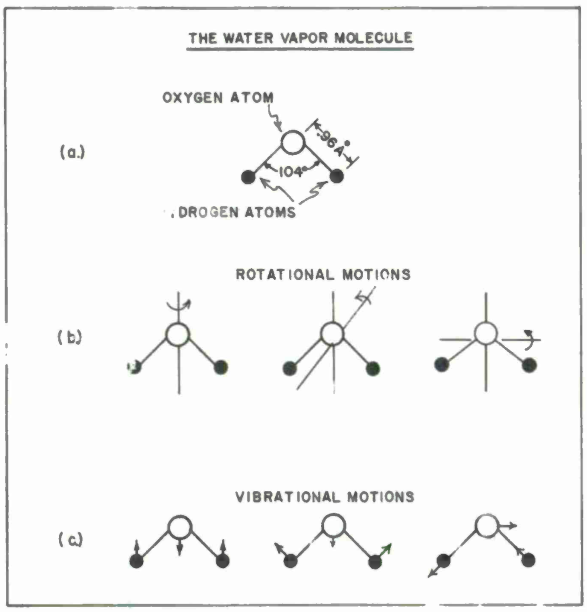


Figure 52. Schematic representation of the configuration of the water vapor molecule and its rotational and vibrational modes.

The absorption of water vapor in the microwave and infrared regions of the electromagnetic spectrum is extensive in frequency and very complex. In contrast to the oxygen molecule case, no non-resonant absorption occurs; all absorption is due to a particular resonance line plus the effects of the wings from resonance lines in the vicinity of the one of interest. The lowest two lines are shown in Figure 53. The lower of these two lines occurs at a frequency near 22.2 GHz. Both of these lines arise from changes in rotation only; they do not involve vibrational changes.

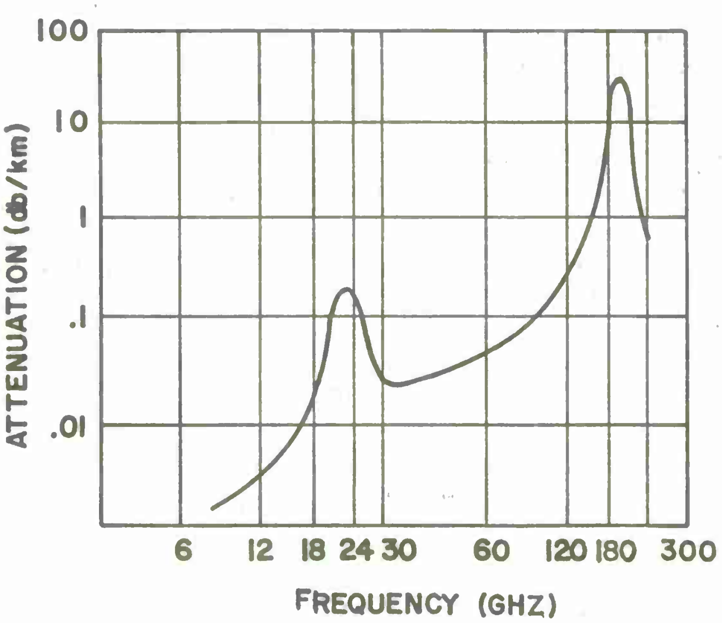


Figure 53. Absorption in db/km for water vapor as a function of frequency at sea level pressure, temperature of 293 °K, and 7.5 g/m^3 of water. (After Van Vleck)

The important features of Figure 53 to be noticed are: (i) under the conditions which this diagram is valid (sea level pressure, 273°K temperature, and an atmosphere consisting of one percent water molecules 7.5 gms/m^3) the peak attenuation below 30 GHz is roughly 0.2 db/km at 22.2 GHz; (ii) the attenuation drops off rapidly toward lower frequencies; (iii) the effect of the wings of resonant lines at higher frequencies is roughly inversely proportional to λ^2 in the region between the resonant peaks. At frequencies at and below about 15 GHz, the water vapor absorption is less than the normal level for oxygen absorption.

Above the 22 GHz line, the absorption increases until the second lowest frequency rotational resonance is encountered near 183 GHz. This line is over one hundred times as intense as the lower line and reaches attenuation rates above 220 db/km in tropical climates. In dry arctic regions the maximum attenuation can be as low as 20 db/km.

Above 183 GHz, the base attenuation continues to rise due to additional higher frequency water vapor lines. More and more very strong lines are encountered until they are so close and so strong that very little radiation can pass vertically through the atmosphere. (See Figure 54) The water vapor vibration and rotation spectrum remains very strong, keeping the atmosphere opaque up to 15 microns. And even above 15 microns, to the low frequency side of the visible part of the spectrum, numerous important water vapor bands occur.

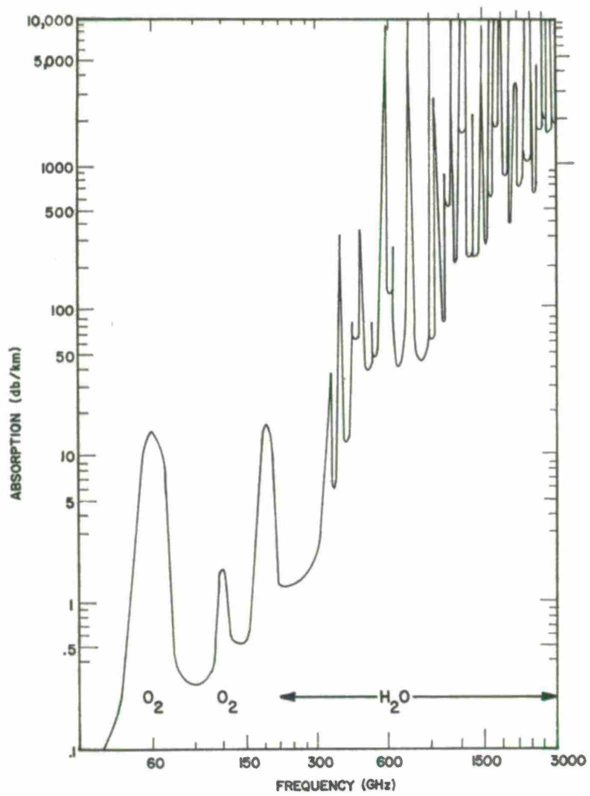


Figure 54. Oxygen lines and the proliferation of water vapor lines in the far infrared region. (After Deer)

(4) Carbon dioxide and other gases

Carbon dioxide is a linear symmetric molecule. Its configuration is shown in Figure 55. Because of its symmetry it has no residual dipole moment and therefore no pure rotational spectrum. Also, the first fundamental vibration mode, shown in Figure 55a, does not disturb this symmetry and therefore does not produce an absorption spectrum. Important infrared spectra are produced, however, by the vibrational mode near 15 microns as shown in Figure 55b and at shorter wavelengths by the mode shown in Figure 55c. Other gases contributing to atmospheric absorption in the infrared frequency region are: most notably ozone, near 9.6 microns; and to lesser extents, nitrous oxide, methane, carbon monoxide, and probably nitric oxide.

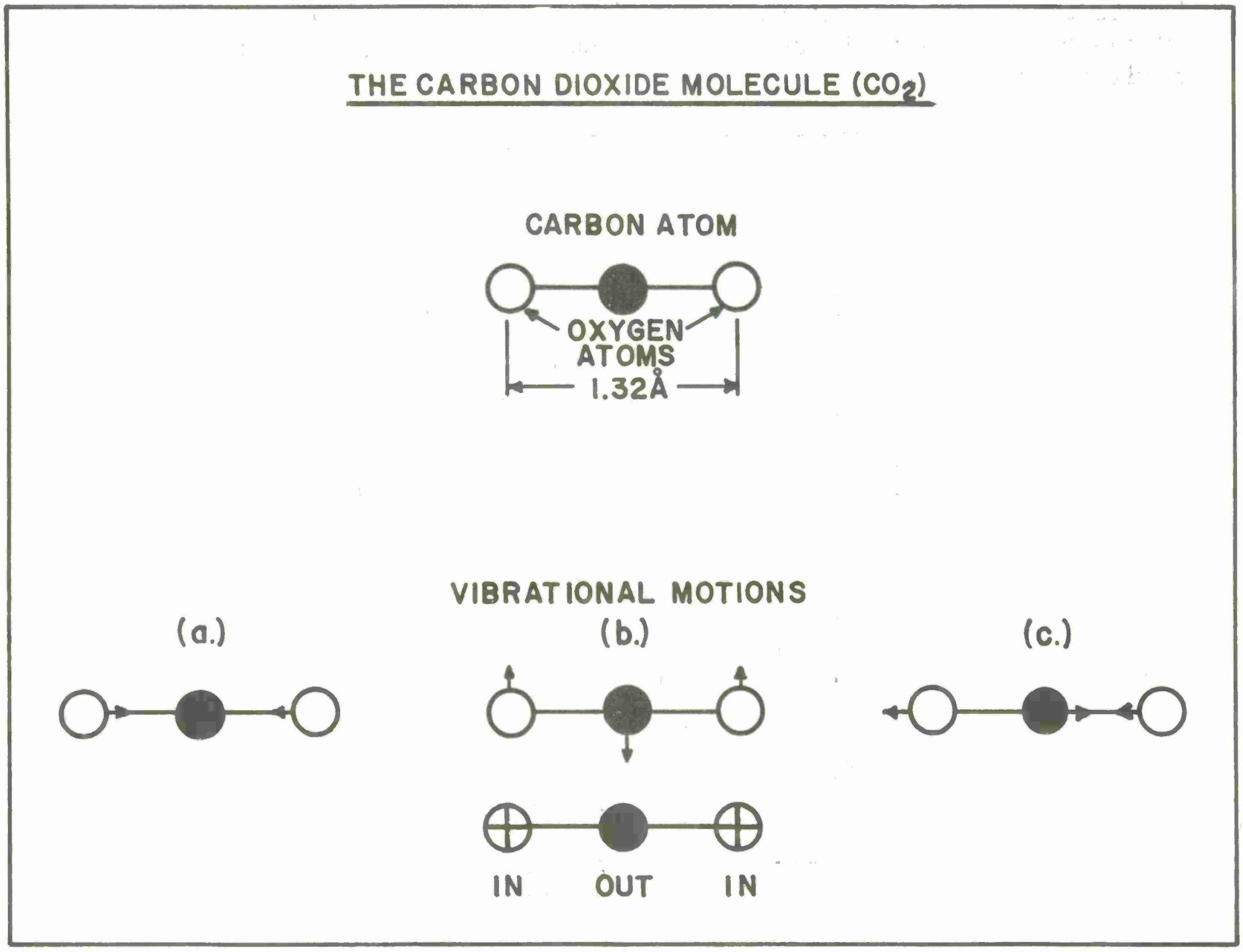


Figure 55. Diagram illustrating the size and the vibrational motions of the carbon dioxide molecule.

3. Attenuation by Hydrometeors

In the previous section absorption by atmospheric gases was discussed; the second aspect of attenuation, scattering, will be discussed later. This section describes the various mechanisms for the attenuation (including both absorption and scattering) of electromagnetic energy by hydrometeors, i.e, the liquid and solid particles which are found suspended in the atmosphere and are seen as clouds, rain, fog, snow, etc..

a. General Discussion

The first solution to the general problem of the interaction between electromagnetic radiation of any wavelength and spherical particles of arbitrary diameter and electrical properties was published in 1908 by Mie. This general solution contained all of the ordinarily separable interactions between electromagnetic signals and matter. The phenomena of scattering, reflection, absorption, and diffraction were not studied independently of one another but were included in the formulation of the general perturbation field which the particle produces as a result of the impinging electromagnetic signal. The results revealed that even for the simple geometry of a sphere the solution of the problem was highly complex except for certain limiting cases determined primarily by the ratio of the wavelength of the incoming signal to the diameter of the sphere. For the discussion which follows, the limiting cases will be investigated first because they have important applications in communications problems and there are several natural illustrations of each.

(1) Limiting case 1: Wavelength much longer than particle diameter.

When an electromagnetic wave strikes a material object, the electrons which are bound to atoms and molecules and those which are associated with the conduction of electrical currents orient themselves as best they can to cancel the electric fields of the incoming wave. This movement causes new electromagnetic fields to be generated, some of which propagate internally in the body and some of which are radiated away externally. These external fields plus the diffracted part of the original signal represent the total effect of the body on the propagating signal.

As pointed out, the general solution even for a simple spherical shape is complex and difficult to visualize. However, if the particle size with respect to the wavelength is quite small, then the external field generated by the particle closely approximates that of a dipole as illustrated in Figure 4.

The effectiveness of the particle to radiate (or scatter) energy out of the incoming signal is proportional to $(a/\lambda)^4$ (where a is the particle diameter and λ is the wavelength of the signal) exactly that of a Rayleigh scatterer as discussed in Section 2, page 23.

Energy can be absorbed from the signal by the particle through several mechanisms associated with the movement of the internal charges. The conduction electrons can be impeded in their movements by collisions and therefore convert electromagnetic energy into random thermal energy. Likewise, the oscillations of polar molecules caused by the oscillations of the external field can be opposed by the equivalent of an internal viscosity and again electromagnetic energy can be converted to internal thermal energy. For the limiting case where (a/λ) is very small, the energy absorbed by atmospheric particles is quite small and can normally be disregarded when compared with other losses which a signal will encounter.

(2) Limiting case 2: Wavelength much shorter than the particle diameter

If a particle's size is very much larger than the wavelength of the impinging radiation, then another limiting case is encountered. Geometrical optics can be applied to such conditions. Wavefronts will be reflected, transmitted, and refracted in a way which can be traced using the ideas of rays of radiation. The understanding of such natural phenomena as rainbows, haloes, and coronas is simplified by using ray tracing techniques. The droplets of water suspended in the atmosphere are so large in comparison to the wavelength of visible light that their surfaces represent essentially plane smooth boundaries between media of differing electrical characteristics (air and water).

Absorption occurs for the same reasons as in the limiting case 1: conduction electrons and polar molecules are set in motion and because of collisions and equivalent viscous effects convert some of the electromagnetic energy into random thermal motions of the particle's atoms and molecules.

(3) Intermediate case: Wavelength comparable to the particle diameter

Between the limiting cases discussed above, there is the region where the wavelength of the signal is comparable to the dimensions of individual particles. In this region, the full complexity of the solutions of the interaction between particle and signal can be discussed only qualitatively.

First, the representation of the induced field, as a dipole even in the case of a spherical particle, is no longer valid. As a/λ approaches unity, higher and higher multipoles are excited; that is, instead of a simple dipole field, the induced field has components of quadrupoles (a configuration of two anti-parallel dipoles), octupoles, etc..

Second, resonance effects appear. If one follows the amplitude of the scattered energy as a function of frequency, multiple maxima and minima appear. These represent the effects of internally reflected waves which reinforce or cancel the surface reflected components of the particle produced wave. The ratio of (a/λ), the electrical properties of the particle material, and of course the shape of the particle control where these maxima and minima occur. The total attenuation of a given particle (absorption plus scattering) will fluctuate rapidly with the wavelength in this region.

Because of the complexity of the problem, it is difficult to predict how signals of, let us say, millimeter wavelength will be attenuated through rainstorms. The ratio a/λ, for such wavelengths approaches unity for the largest raindrops, and further the dielectric constant is so large for water that ratios of a/λ considerably removed from unity begin to show important resonance effects.

b. Liquid Water in the Atmosphere

The scattering and absorption of electromagnetic energy by small water particles is observed almost daily in the scattering and absorption of sunlight by clouds. On this basis alone, some related effects on infrared and microwave radiation might be anticipated. In this and the following sub-section the general effects discussed above will be applied directly to liquid and solid atmospheric hydrometeors.

For a better physical grasp of the effects in which we are interested it is important to have some knowledge of the ranges of particle sizes which will be found in the atmosphere. Table 2 will be helpful in this respect. A further useful generalization is that radars and communication equipment seldom operate at wavelengths shorter than 4 mm; therefore, and because atmospheric attenuation is generally negligible for wavelengths longer than 10 cm, the discussion will be concentrated on the general wavelength range shorter than 10 cm and longer than a few millimeters.

Table 2. Size ranges and typical values of drop diameters of hydrometeors. (After Vaughn)

Type of Hydrometeor	Altitude (km) Range	Drop Diameter (microns) Range	Rep
Layer Clouds	sfc-1.5	1-40	11
Layer Clouds	1.5-7.5	1-50	12
Layer Clouds (ice crystals)	7.5-15.0	10-10,000	100
Convective Cloud			
Fair Weather Cumulus	0.5-8.0	1-75	12
Cumulus Congestus	0.5-13.0	1-200	15
Continuous Type Rain	sfc-6.0	500-3000	1000
Shower Type Rain	sfc-13.0	500-7000	2000
Coalescence (warm) Rain	sfc-5.0	100-1000	500
Hail	sfc-13.0	0.01-13cm	0.8cm
Ice and Snow Crystals	sfc-13.0	100-20,000	5000

(1) Clouds and fogs

From Table 2 it may be seen that the particles in clouds and fogs are generally small compared to the wavelengths of the incident radiation with which we are chiefly concerned. As pointed out above, there is both scattering and absorption in this region. By far the most important effect, however, is due to absorption of the incident radiation by the droplets of water. Water has a high dielectric constant and exhibits a finite conductivity. Since water drops often form on a salt nucleus, the conductivity of rain water is usually greater than pure distilled water. Investigations have shown that absorption by water droplets is directly proportional to the mass of water present, and fairly independent of the drop size distribution. Attenuation by scattering is negligible for all clouds and fogs for incident wavelengths of 1 mm or greater. Figure 56 shows how attenuation varies with wavelength. It can be seen that for wavelengths longer than 1 cm attenuation will generally be less than 1 db/km since liquid water content values as great as 1 gm/m^3 seldom occur except in heavy sea fogs or in low altitude clouds in tropical latitudes.

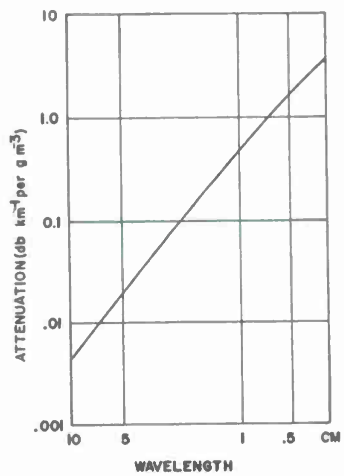


Figure 56. Attenuation by very small water drops at a temperature of 18°C. (After VandeHulst)

It should be noted that the dielectric constant of water is a function of temperature and wavelength. As a result, at lower temperatures and wavelengths the values shown in Figure 56 will be larger. Some representative corrections are shown in Table 3.

Typical attenuations in various fogs and clouds are shown in Figure 57 as a function of wavelength.

Table 3. Correction factors for attenuation at various temperatures and wavelengths. (After Goldstein)

λ, cm	0°C	10°C	18°C	20°C	30°C	40°C
0.5	1.59	1.20	1.0	0.95	0.73	0.59
1.25	1.93	1.29	1.0	0.95	0.73	0.57
3.2	1.98	1.30	1.0	0.95	0.70	0.56
10.0	2.0	1.25	1.0	0.95	0.67	0.59

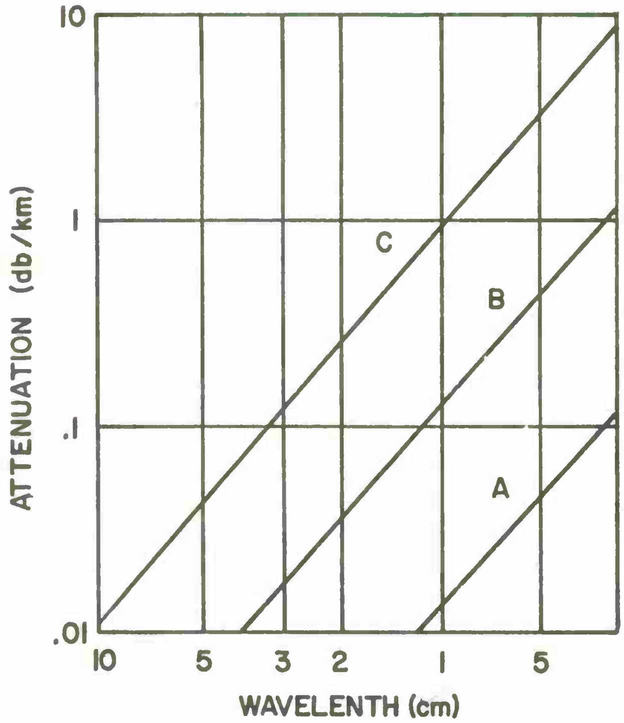


Figure 57. One way attenuation by fog or cloud. Curve A represents a liquid water content of 0.032 gm/m^3 (visibility about 2000 feet); curve B, 0.32 gm/m^3 (visibility about 400 feet); and curve C, 2.3 gm/m^3 (visibility about 100 feet). (After Goldstein)

(2) Rain

The size of raindrops ranges from 100 microns to 7 mm, and a representative size may be taken as 1 mm (see Table 2). The most predominant size for a particular situation varies with precipitation rate as shown in Figure 58. Many radars and some communications systems operate at centimeter and millimeter wavelengths. For all of these, the wavelength is comparable to the droplet size; therefore the effects as described for the intermediate case of the ratio (a/λ) and for liquid water in the atmosphere are applicable.

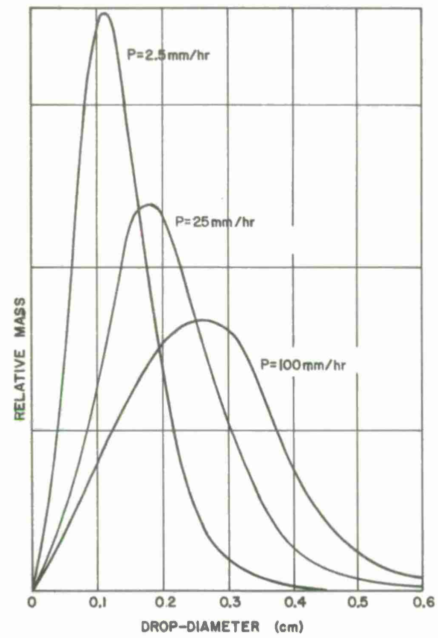


Figure 58. Relative total mass of liquid water at various precipitation rates as a function of drop diameter. (Prepared by Bean and Dutton)

In the drop size region in which raindrops are found, the dielectric absorption (i.e. the absorption associated with the equivalent viscous effects on oscillating polar molecules) increases over that of the smaller drops common to fogs and clouds. It varies with drop size in a complicated way, generally increasing fairly rapidly with the mean size. As the mean drop size and wavelength become comparable, scattering becomes more and more important until its contribution to the total attenuation is comparable to that from absorption. For several rain conditions Figure 59 shows the distance a signal must travel in order that absorption and scattering are comparable.

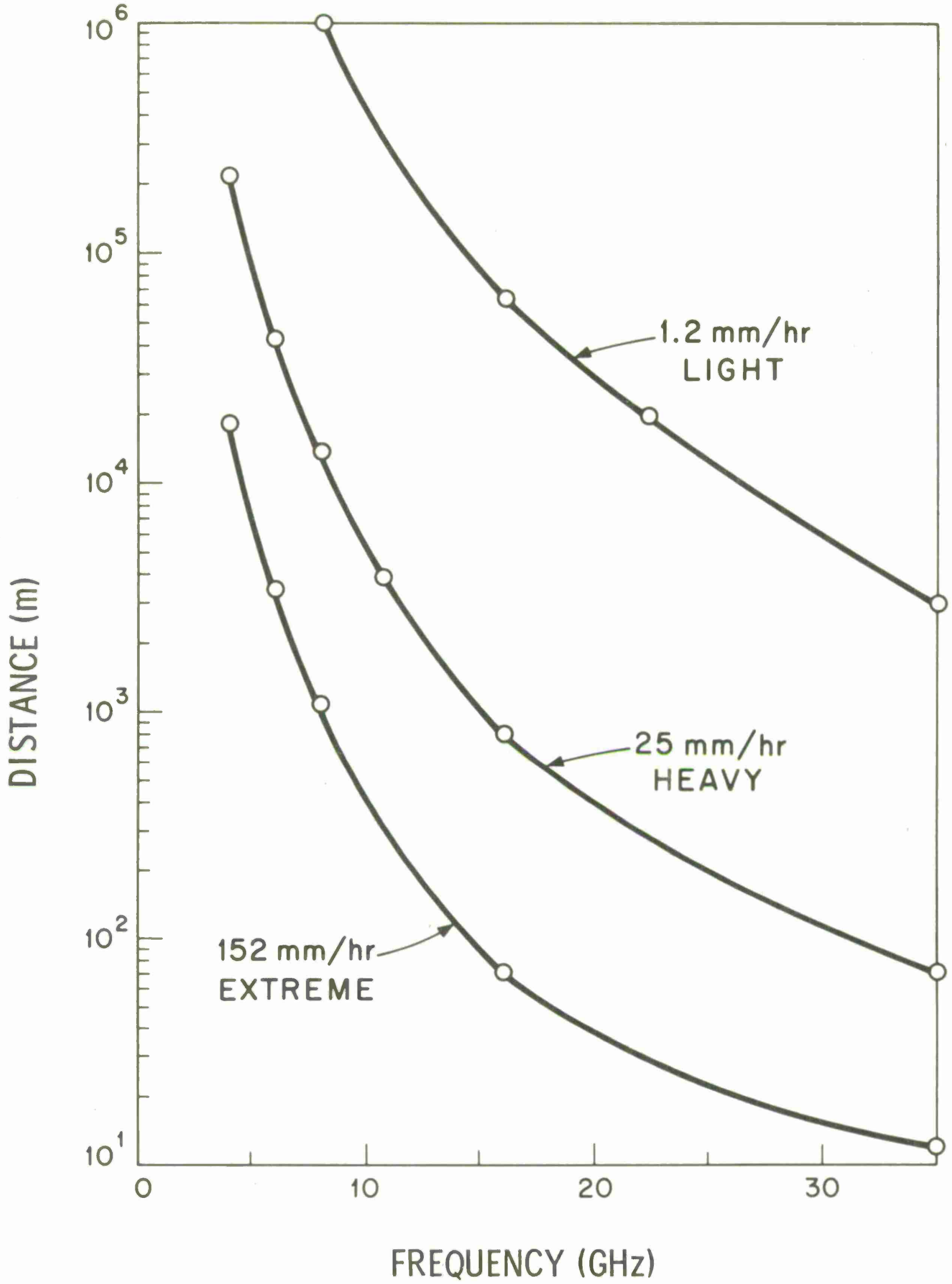


Figure 59. Distance a signal must travel for attenuation due to scattering to become comparable to that due to absorption. (After Crane)

The general results for attenuation are shown in Figure 60. It can be seen that at wavelengths as great as 10 cm the attenuation is small. In fact, the attenuation at wavelengths greater than 10 cm is small enough to be neglected in most practical applications. For the longer wavelengths and the smaller drop sizes the attenuation is mostly due to absorption as it is in the case of fogs and clouds. Comparison of curve C, Figure 57, with curve B, Figure 60, shows that attenuation for the heaviest fogs and clouds, can exceed that of light and moderate precipitation.

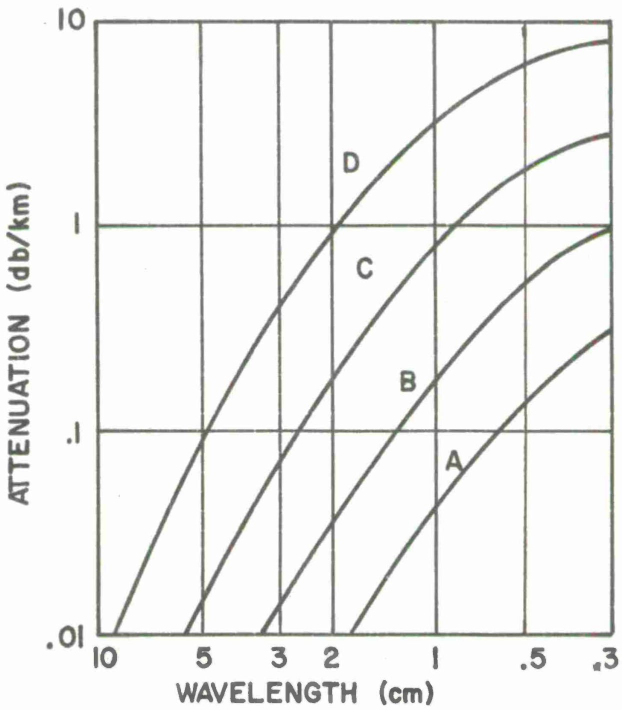


Figure 60. One way attenuation by rain. Curve A represents drizzle (0.25 mm/hr); curve B, light rain (1 mm/hr); curve C, moderate rain (4 mm/hr); and curve D, heavy rain (16 mm/hr). (After Goldstein)

As the wavelength decreases to about 3 cm, attenuation becomes important and in the heaviest showers attenuation can be severe enough to prevent successful operation. At shorter wavelengths, attenuation rapidly becomes so great that at wavelengths of 1 cm or less a heavy rainstorm is very nearly opaque. For comparison, at a wavelength of 1 cm, attenuation is about 3 db/km in heavy rain and 18 db/km for intense rain.

Absorption by rain also varies with temperature, just as with the case of the small particles found in fogs and clouds. Correction values for various temperatures and wavelengths are shown in Table 4.

Table 4. Correction factors for attenuation by precipitation for various temperatures, wavelengths, and precipitation rates. (After Goldstein)

Precipitation rate, mm/hr	λ, cm	0°C	10°C	18°C	30°C	40°C
0.25	0.5	0.85	0.95	1.0	1.02	0.99
	1.25	0.95	1.0	1.0	0.90	0.81
	3.2	1.21	1.10	1.0	0.79	0.55
	10.0	2.01	1.40	1.0	0.70	0.59
2.5	0.5	0.87	0.95	1.0	1.03	1.01
	1.25	0.85	0.99	1.0	0.92	0.80
	3.2	0.82	1.01	1.0	0.82	0.64
	10.0	2.02	1.40	1.0	0.70	0.59
12.5	0.5	0.90	0.96	1.0	1.02	1.00
	1.25	0.83	0.96	1.0	0.93	0.81
	3.2	0.64	0.88	1.0	0.90	0.70
	10.0	2.03	1.40	1.0	0.70	0.59
50	0.5	0.94	0.98	1.0	1.01	1.00
	1.25	0.84	0.95	1.0	0.95	0.83
	3.2	0.62	0.87	1.0	0.99	0.81
	10.0	2.01	1.40	1.0	0.70	0.58
150	0.5	0.96	0.98	1.0	1.01	1.00
	1.25	0.86	0.96	1.0	0.97	0.87
	3.2	0.66	0.88	1.0	1.03	0.89
	10.0	2.00	1.40	1.0	0.70	0.58

c. Ice

The dielectric absorption is considerably less for ice than for water. This may be attributed to the fact that the water molecules are more mobile in the liquid state than in the crystalline state. Thus for the droplet size ranges involved, the absorption and scattering effects are comparable, so the full Mie solution must be used.

Calculations of attenuation for spherical hailstones are possible. Table 5 below gives the attenuation for representative radii of stones and representative wavelengths, assuming the temperature is 0°C and the stones are not wet. The values are db/km per hailstone per m^3. It is quickly obvious that hailstones which reach radii of a few mm will cause significant attenuation at wavelengths near or below the resonance wavelength of 1.35 cm. Longer wavelengths require very large stones to be significantly affected.

Table 5. Attenuation by dry hailstones at $0^{o}C$, db/km per stone per m^3 as a function of radius a, and wavelength λ. (After Goldstein)

a, cm	λ = 0.5 cm	λ = 1.25 cm	λ = 3.2 cm	λ = 10 cm
0.025	$3.75 \cdot 10^{-6}$	$2.81 \cdot 10^{-7}$	$7.43 \cdot 10^{-8}$	$2.29 \cdot 10^{-8}$
0.050	$3.16 \cdot 10^{-4}$	$7.57 \cdot 10^{-6}$	$7.15 \cdot 10^{-7}$	$1.85 \cdot 10^{-7}$
0.10	$1.41 \cdot 10^{-2}$	$4.00 \cdot 10^{-4}$	$1.37 \cdot 10^{-5}$	$1.57 \cdot 10^{-6}$
0.15	8.91	$3.31 \cdot 10^{-3}$	$1.19 \cdot 10^{-4}$	5.99
0.20	$1.79 \cdot 10^{-1}$	$2.53 \cdot 10^{-2}$	6.15	$1.80 \cdot 10^{-5}$
0.25	2.63	9.55	$2.27 \cdot 10^{-3}$	4.64
0.375	4.07	$6.31 \cdot 10^{-1}$	$2.63 \cdot 10^{-2}$	$3.24 \cdot 10^{-4}$
0.50	4.08	$1.59 \cdot 10^{0}$	$1.59 \cdot 10^{-1}$	$1.67 \cdot 10^{-3}$
0.75	. . . . .	2.69	$1.18 \cdot 10^{0}$	$1.70 \cdot 10^{-2}$
1.00	. . . . .	. . . . .	3.98	9.77
1.50	. . . . .	. . . . .	14.5	$1.10 \cdot 10^{0}$
2.00	. . . . .	. . . . .	17.9	5.37

Snow and ice particles are not spherical; so it is very difficult to numerically compute their attenuation. However, some work has been done and it has been found that when the size of the particles is very small compared with the wavelength, the attenuation is proportional to the volume of the particles. At $0^{o}C$ and a wavelength of 1 cm, the attenuation by ice clouds is two magnitudes smaller than the attenuation from a cloud of liquid droplets with the same water content. At $-40^{o}C$ the relative attenuation is still less. Attenuation for various shapes of ice crystals is shown in Table 6 as a function of the water content. Since total water content rarely exceeds 0.5 gm/m^3, attenuation by ice crystals is generally so small that it is negligible compared to other losses. As for snow, electromagnetically it behaves differently from ice, and in general, attenuation for snow will be less even than that for hailstones of similar size and much less that that for rain at equal precipitation rates.

Table 6. Attenuation by ice crystal clouds. M is water content in grams per cubic meter. λ is wavelength in centimeters. (After Goldstein)

Type of Particle	Attenuation (db/km) T = - 40 C	T = 0 C
Spheres	$0.00044M/\lambda$	$0.0035M/\lambda$
Needles	$0.00062M/\lambda$	$0.0050M/\lambda$
Disks	$0.00087M/\lambda$	$0.0070M/\lambda$

4. Scattering Effects in the Troposphere

a. General Discussion

Section 3 was devoted to the effects of hydrometeors on propagating electromagnetic signals. Included were absorption and scattering effects. Scattering also occurs from the interaction of individual gas molecules with the ambient signal, but such scattering does not become important until visible wavelengths are reached. Molecular scattering is an example of Limiting Case 1 discussed at the beginning of the previous section. That is, the ratio of particle size to wavelength, a/λ, is quite small and therefore the effectiveness of scattering depends upon wavelength inversely to the fourth power $(1/\lambda^4)$. The blueness of the sky is related to the fact that shorter wavelengths of the solar incoming radiation are scattered more effectively than the longer. The redness of the sun at sunset is a result of the longer atmospheric path the solar light must pass through as it descends, causing appreciably greater loss of its bluer components by molecular scattering.

b. Tropospheric Scatter Communications

More important and even more difficult to analyze are the scattering effects of the inhomogeneities in the atmosphere caused by turbulence, stable layers, and the density changes between adjacent air masses. These inhomogeneities are found on all size scales and throughout the entire atmosphere. They have a mean activity over which are impressed large local variations in density and therefore electrical properties.

The reason for the importance of the atmospheric inhomogeneities is their ability to scatter enough energy so that narrow beams of radiation directed toward the horizon can be detected hundreds of kilometers below the horizon. This is called troposcatter communication and has become an important means of long range communication in the last ten years. We will investigate the properties of this form of communication and its dependence upon the state of the troposphere.

When radio radiation is beamed just above the horizon as illustrated in Figure 61, one would expect to measure negligible amounts of energy outside the beam if the earth were not present at all. With the earth, one would expect that diffraction would carry some of the energy into the geometrical shadow zone below the horizon. And if the atmosphere were included, the normal refraction would cause a slightly greater amount of energy to propagate beyond the line of sight horizon. However, when actual measurements are performed, with an experimental arrangement, such as shown in Figure 61, the results are more closely represented by Figure 62. The decrease in the signal strength when the separation between receiver and transmitter is increased is less, to be sure, than what would be detected if the receiver and transmitter were in sight of each other (free space curve), but very much smaller than that predicted from the evaluation of diffraction and refraction alone. The difference is due to a complex scattering process of the radiation off of inhomogeneities in the atmosphere in the region of the common volume intercepted by transmitter and receiver (see Figure 61).

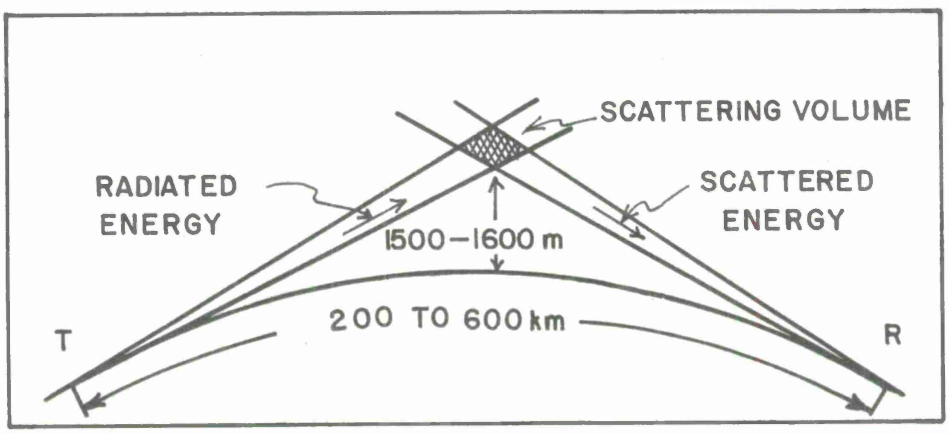


Figure 61. The scattering volume viewed in common by transmitter T and receiver R in a troposcatter communication system.

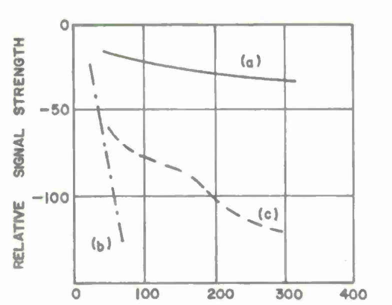


Figure 62. Relative signal strengths. Energies received are depicted for (a) free space propagation, (b) earth diffraction, and (c) actual measurements. (After duCastel)

The actual mechanism through which the troposcatter link operates is very complex. But at least two important processes are at work: true scattering of radiation from the turbulently produced inhomogeneities in the atmosphere; and partial reflection at the discontinuities between stable layers and more turbulent regions. These two processes are schematically illustrated in Figure 63. Most likely, the two processes are operative simultaneously and to talk of them separately is an abstraction. Each is enhanced or diminished by different meteorological conditions but neither are likely to be completely negligible.

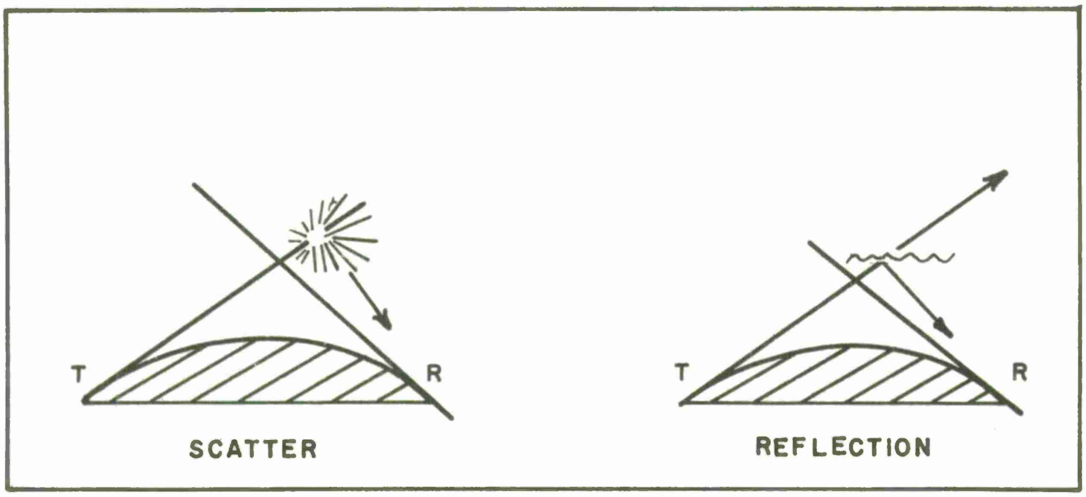


Figure 63. Schematic showing scatter and reflection mechanisms. (After duCastel)

The frequencies which are effectively used for troposcatter communication lie between about 50 MHz and 10 GHz. Lower frequencies begin to interact with the ionosphere and there is present at these frequencies a high level of natural noise. Above 10 GHz, the gases in the atmosphere begin to appreciably attenuate the radiation.

When troposcatter communication was first proposed, it was believed that the bandwidth of the signal would be quite narrow, perhaps of the order of tens of thousands of cycles per second because of the multipath interference which would necessarily occur. That is, parts of the beam would be scattered or reflected from different levels or regions in the atmosphere and would consequently have different path lengths from transmitter to receiver. They would be out of phase and therefore, would interfere. Actual experimental results, however, indicate that bandwidths of several MHz are certainly possible.

With narrower antenna beams and several receiving antennas, even greater bandwidths are possible. This is an important consideration because the amount of information that a system can convey is proportional to bandwidth.

The distances which can be conveniently spanned using troposcatter techniques depend, of course, on the power transmitted, but also upon the structure of the atmosphere, the topography between receiver and transmitter, the beam width, the antenna size, and the frequency being used. Upper limits of perhaps 1200 to 1400 km are generally reported, but most practical communication links are much shorter, mostly in the range from 200 to 800 km.

The terrain between the transmitter and receiver is an important factor in the attenuation experienced by a signal. If the terrain is flat out to the distant horizon, elevating the antenna produces little effect on the signal. However, if the horizon is close by because of hills or raised ground, the antenna's height above the ground will most likely produce large variation in the received signal strength, indicating the importance of the topography.

Because the scattering of energy for troposcatter communication links takes place in the lower atmosphere, the conditions in this region between transmitter and receiver play an important role in overall quality of the received signal.

A diurnal effect is evident, especially in the summer. The strength of the received signal varies from a minimum near midday to a maximum during the night. The pattern is less regular and less pronounced in the winter time.

The general climatic conditions are also important to troposcatter links. Attenuation over a given distance appears to be greatest for the drier regions, i.e., deserts and arctic regions, while moist tropical regions show the least attenuation.

There also exists a strong relationship between the synoptic weather conditions which prevail between terminals and the strength of the received signal. Anticyclones, producing stable, stratified conditions, are associated with stronger signals than the turbulent, less stratified conditions which prevail during the passage of cyclonic disturbances.

Several attempts have been made to establish correlations between atmospheric parameters and the attenuation which will occur over a troposcatter link. So far, the two most promising parameters which appear to be linked to attenuation and which can be measured fairly conveniently are the gradient of the refractive index from the surface up to the level where the horizons of the two antenna beams meet, and the stability of the atmosphere against vertical motions (see Section 2, page 20, for the meaning of atmospheric stability). However, the correlations are imperfect and work is necessary on the exact propagation mechanism to better choose correlation parameters.

IV. THE IONOSPHERE

In the early history of wireless telegraphy, perhaps the most fundamental discovery was the fact that there is a layer, high in the atmosphere, whose conductivity exceeds that of sea water. It was this layer, acting as a boundary from which electromagnetic waves were reflected, which permitted the waves to travel far beyond the horizon. The layer was presumed to be made up of ionized atmospheric gases, and it was therefore called the ionosphere. The importance of the ionosphere has not lessened since the pioneering efforts of Marconi first showed its practical effects. Even with the greatly expanded usable spectrum for electromagnetic propagation and the proliferation of communication techniques, the ionosphere remains of central importance to long range telecommunications. To better understand why this is so it will be helpful first to review the cause, structure, and temporal and geographic characteristics of the ionosphere.

A. DESCRIPTION OF THE IONOSPHERE

The ultimate cause of the abundance of free electrons and positive ions in the ionospheric region is high energy radiation from the sun bombarding the top of the atmosphere. X-rays and ultraviolet photons collide with the rarefied gas particles above about 60 kilometers separating them into heavy, positive ions, and light, free electrons. If there were only one constituent of the atmosphere at a constant temperature throughout, in equilibrium under the force of gravity, being bombarded by only one frequency of ionizing radiation, a simple picture of the ionospheric layer formation could be presented. Under these conditions, the layer formed is called a Chapman Layer, after the scientist who first described it.

1. A Simple Chapman Layer

For a qualitative view of the formation of such a region, suppose that a constant number of photons enter the atmosphere vertically during a given time interval. The uppermost regions of the atmosphere are so thin that the photons do not encounter many particles to ionize. But as they penetrate lower down, into more dense regions, the photons strike more and more particles and the ion concentration increases. However, each photon which ionizes a molecule or atom is lost for further ionization. Eventually, the increase in ionization with decreasing height is halted because of the depletion of the photon supply; thus a very rapid decline in the ion concentration with further decrease in height occurs, giving a fairly flat bottom to the layer profile A graphical representation of the electron or ion distribution as a function of height is given by Figure 64.

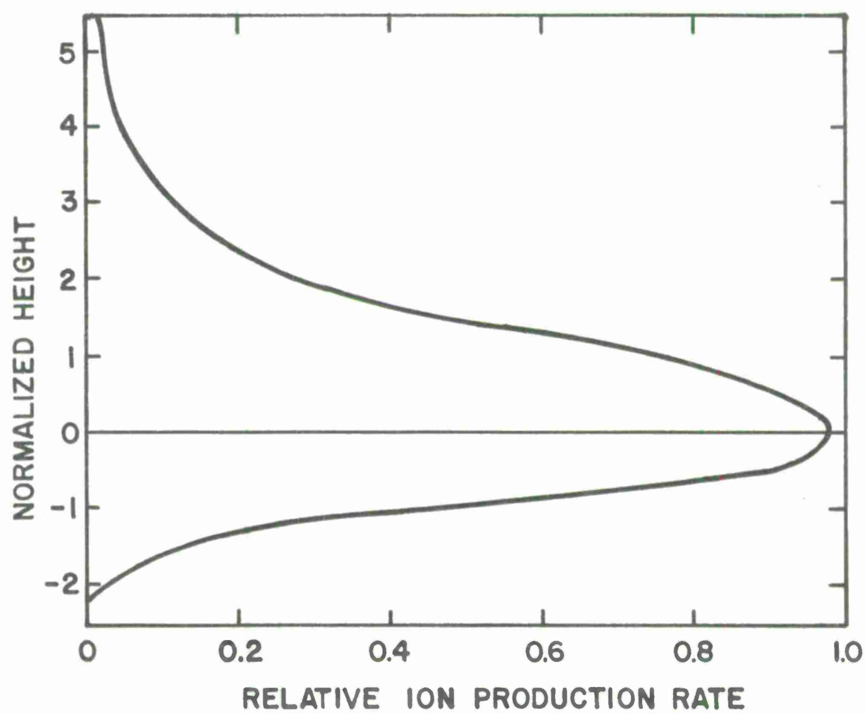


Figure 64. Relative ion production rate versus normalized height in a Chapman Layer. (After Ratcliffe)

The vertical coordinate is a normalized neight, $(Z - Z_m)/H$ centered on the height of the ion production maximum, where Z is the real height, Z_m is the height of the maximum electron concentration or density, and H is the scale height; its value is approximately 6 km at the height of 100 km, and increase to 60 km at the height of 300 km The figure readily shows the sharp maximum and layered appearance of the distribution.

2. Observations of the Vertical Electron Distribution

The atmosphere does not consist of a single constituent, nor does the sun emit only one frequency of ionizing radiation. Consequently, the ionosphere has a much more complicated structure than depicted in Figure 64. In reality, it has several electron concentration maxima which come and go, move vertically, and vary in strength depending upon the elevation of the sun, the season, and the sun's surface activity. A realistic vertical profile of the electron concentration might appear more like that depicted in Figure 65.

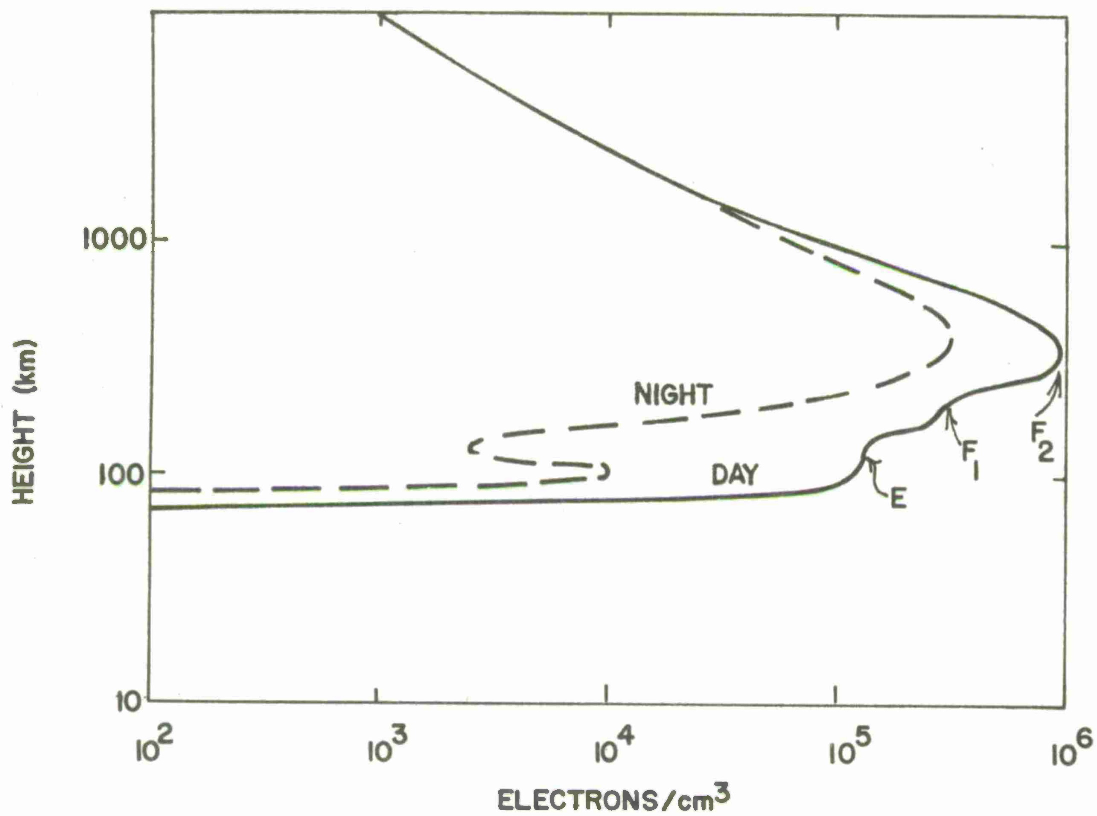


Figure 65. Electron density as a function of height in the earth's ionosphere. (Nawrocki and Papa)

Until the advent of rocket technology profiles such as that presented in Figure 65 could only be inferred. One of the most common ways to do this was to direct a beam vertically upward and vary the frequency. Time of receipt of the reflected signals was traced on a chart and called an <u>ionogram</u>. The equipment making the ionogram was called an <u>ionosonde</u>. Characteristically these diagrams showed two or more sharp inflection points or discontinuities, each of which was interpreted as representing the height of a peak or maximum in the electron density profile. The frequency at which the peak occurred was called the critical frequency for that height or layer. The physical concept of critical frequency will be discussed in more detail in a later section.

The two major layers were, very early in ionospheric research, designated as the E and F layers. Under closer surveillance, other more indistinct and variable layers were discovered and designated the D and Sporadic E layers; and, as more data were collected, the F region was divided into the F_1 and F_2 regions. All of these layers were thought to be separate peaks of maxima of electron concentration or density, but direct measurements have shown them usually to be only changes in the vertical gradient of the electron density, except for the F_2 layer which really is a maximum of electron density.

The heights of the various layers, as has been mentioned, vary considerably in time, but representative values for the various regions would be as follows: D, 75 to 85 km; E, 95 to 110 km; Sporadic E, 100 to 105 km; F_1, about 200 km; and the F_2, about 300 km.

The immediate causes of each ionospheric layer are very complex, and there are widely differing opinions as to the reason for their existence. For completeness in this description of the ionosphere, some of the better established theories are discussed below.

a. The D Region

The lowest layer or region of the ionosphere is particularly ill defined. One reason for this is that conventional ionosondes cannot detect the relatively low electron densities of this region. Below 90 km various effects on radio waves and rocket probes have, nevertheless, led investigators to postulate the D Region. It is believed to be caused primarily by Lyman alpha (1215.7 Å) radiation which originates in the sun and ionizes nitric oxide (NO) in the earth's atmosphere in the region from about 60 to 90 km. X-rays also have sufficient energy to ionize nitric oxide as well as other gases at these altitudes but very few of these very high energy radiations penetrate to the altitudes of the D layer under normal conditions. The D layer exists, essentially, only during the hours when the ultraviolet light enters the atmosphere, and the layer varies in electron density with the intensity of such radiation. Since the sun is the primary source of Lyman alpha radiation, the D layer electron density will change from an undetectable level at night, to a normal maximum at local noon. D layer electron density can also be extraordinarily high during solar flares. Satellite observations have shown that during solar flares there is little change in the Lyman alpha flux, however substantial increases in the X-ray flux, have been detected. It is therefore believed that under solar flare conditions X-rays are responsible for the increased D region ionization.

The closer to the local vertical the incoming radiation is, the more intense the ionizing radiation will be at the D layer height. Seasonal fluctuations are therefore apparent. Solar flare activity waxes and wanes over an 11 year cycle which consequently also causes the long time average of the D region electron density to show similar systematic fluctuations. Occasionally, solar flares are followed some time later by an influx of high energy particles into the atmosphere above both poles. This causes intense ionization in the D region above these areas for periods up to several days.

b. The E Region

The electron density of the E region in the ionosphere also shows systematic variations with time of day, season, phase of the solar sunspot cycle, and position on the earth in ways similar to those of the D region. There also are random variations in time and space superimposed on these regular variations. The E region ionization is found to fluctuate along with fluctuations in the number and sizes of events occurring on the sun. Good correlations also occur between solar radio emission at 10 cm wavelengths and day-to-day variations in the E region. It is quite apparent from the mass of data collected that an intimate relationship exists between the physical processes on the sun and the existence and strength of the E region ionization.

c. Sporadic E

Complicating the description of the E layer is the intermittent and random appearance of a considerable enhancement of the electron density at E layer altitudes. The phenomenon is distinct enough to be given the name Sporadic E (E_s). It shows some relationship to several physical phenomena other than the sun, i.e., lunar time, meteor showers, and geomagnetic activity. E_s ionization extends in patches over several hundred kilometers and lasts generally 1 to 2 hours. So far, no entirely satisfactory theory accounts for the appearance of Sporadic E.

d. The F region

As one ascends from lower to higher altitudes in the ionosphere, the electron density increases from the D region through the E region and reaches its maximum near 300 km in what is called the F region. Then more or less smoothly, the electron density decreases as one ascends further. The region of the maximum is sometimes broken into two regions or layers. The lower layer, or F_1 region, on a plot of electron density as a function of height, appears as an inflection of almost insignificant proportions; however, because of the clearness which this ledge produces in ionosonde records its behavior is fairly well documented.

The F_1 ledge appears to be strictly a daytime phenomenon. It is most pronounced in summer and at the minimum of the solar cycle. During winter and at the height of the solar activity cycle it is never observed. To a rough first approximation, the F_1 ledge varies over the earth as though it were a simple Chapman layer. However, many departures from the expected Chapman layer behavior occur.

The second, or upper layer, includes the maximum of the vertical electron density profile and is called the F_2 layer or peak. The temporal and spatial variations of the so-called F_2 peak, are complex, vary with observation point on the earth, and do not lend themselves to easy summarization. But to gain a general feeling for the layer and its fluctuations, some of its more well known features must be described. Perhaps the best approach to this description is to point out the differences which the F_2 layer shows from what can be expected from the theory of a simple Chapman layer.

3. Deviations of the F Region Maximum from a Simple Chapman Layer

One of the most obvious features of the Chapman layer is the direct dependence of the maximum electron density on the strength of the ionizing radiation. This would mean that the largest concentration of electrons would be expected at the equator when the sun is directly overhead at local noontime i.e., midday at the equinox. However, the F_2 layer shows a distinct minimum over the geomagnetic equator at the equinox at noontime. This is called the "geographic anomaly", and is depicted in Figure 66.

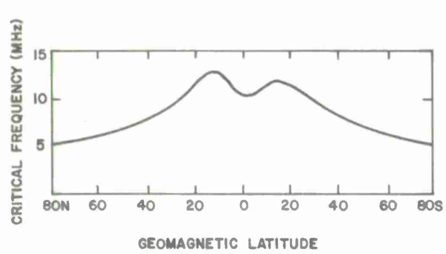


Figure 66. The F_2 geographic anomaly. The critical frequency is proportional to the maximum electron density. (After Ratcliffe)

Further, it would be expected that the peak F_2 electron concentration would occur everywhere at midday, again, because the intensity of radiation would be a maximum during this time.

The F_2 peak, however, is often far removed from noontime; and, at times, even shows a minor minimum near midday. Figure 67 shows an example of such behavior.

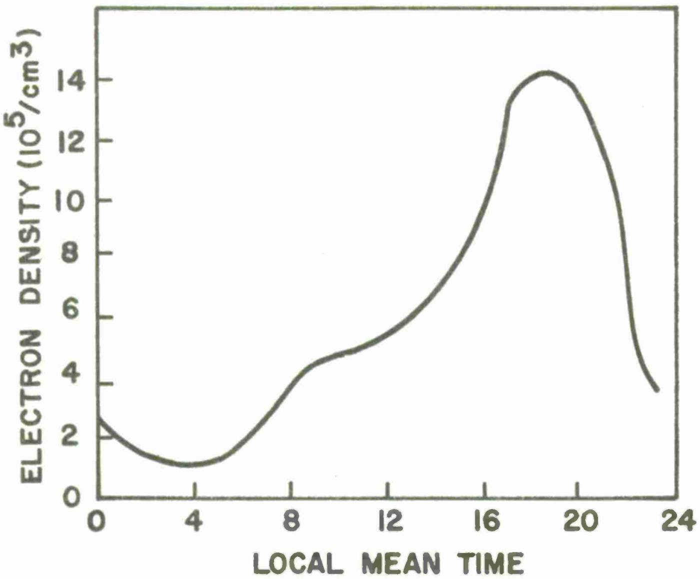


Figure 67. The maximum electron content at the F_2 peak over Maui, Hawaii, as a Function of Local Mean Time Averaged over Five Quiet Sun Days in June 1954. (After Ratcliffe)

The Chapman theory would predict a seasonal variation of the F_2 peak. When the sun is high in the summer sky, the maximum photoionization should occur; in winter, when the sun is closer to the horizon at local noon, there should be a minimum of ion production. This simple view is confounded by actual measurements. Two anomalies occur: the "December" anomaly; and the so-called "winter" anomaly.

The December anomaly manifests itself as an unexpectedly high midday maximum of electrons during the months of November, December, and January, over the latitude belt of 50°N to 35°S (in comparison to a theoretical Chapman layer).

The winter anomaly occurs all over the earth but more markedly around latitudes 50° north and south. This anomaly is characterized by values of electron density in the F_2 region larger than expected in local winter.

The two effects add, in the northern hemisphere, to produce diurnal extremes of electron density. (The anomalous effects are not observed at midnight.) Moreover, they are most pronounced during high sunspot years. Figure 68 shows what has occurred over Slough, England, latitude 54°N, during a high and low sunspot year.

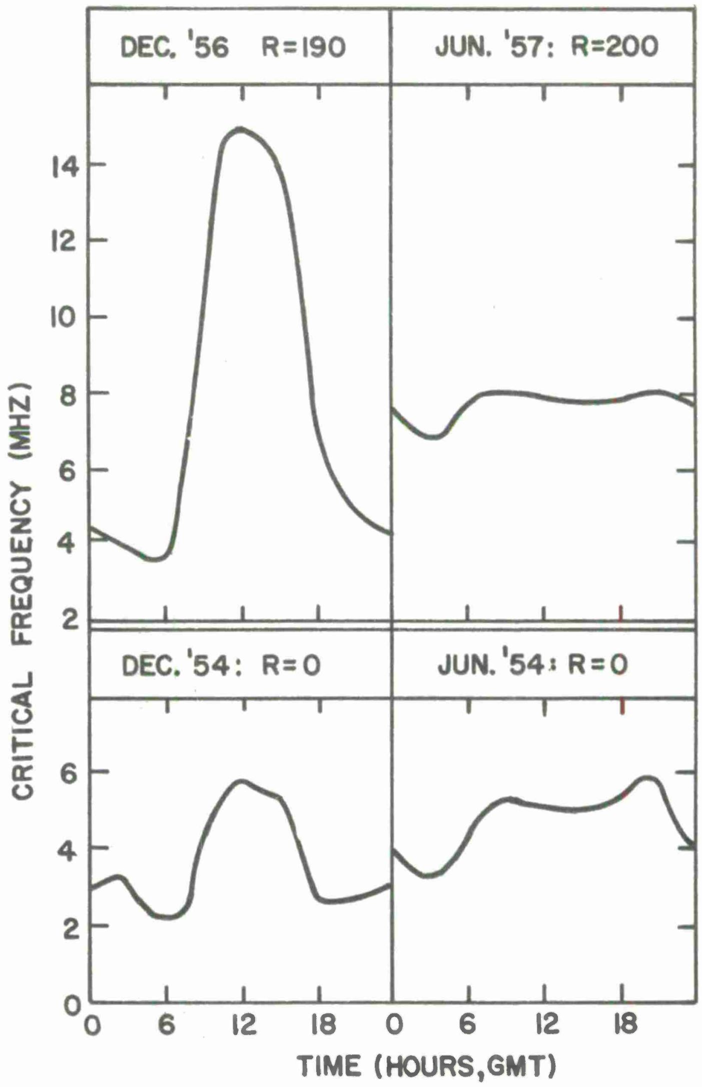


Figure 68. The combined effects of the F_2 winter and December anomalies at Slough, England, during a high sunspot year (upper portion) and during a low sunspot year (lower portion). (After Ratcliffe)

In the above figure, $\overline{R}$ is a measure of sunspot activity (i.e., higher $\overline{R}$ values are associated with higher sunspot activity) and the critical frequencies are proportional to maximum electron density.

Results such as portrayed in Figure 68 hint strongly at an intimate relationship between sunspots, solar activity, and the ionosphere. Indeed, the mean solar sunspot number, $\overline{R}$, is strongly correlated with the F_2 electron density such that the maximum density occurs at maximum solar activity (see Figure 69). More detailed correlations have been carried out which show daily variations associated with changes in solar flocculi. Also, variations show a 27 day recurrence tendency, the period of one solar revolution.

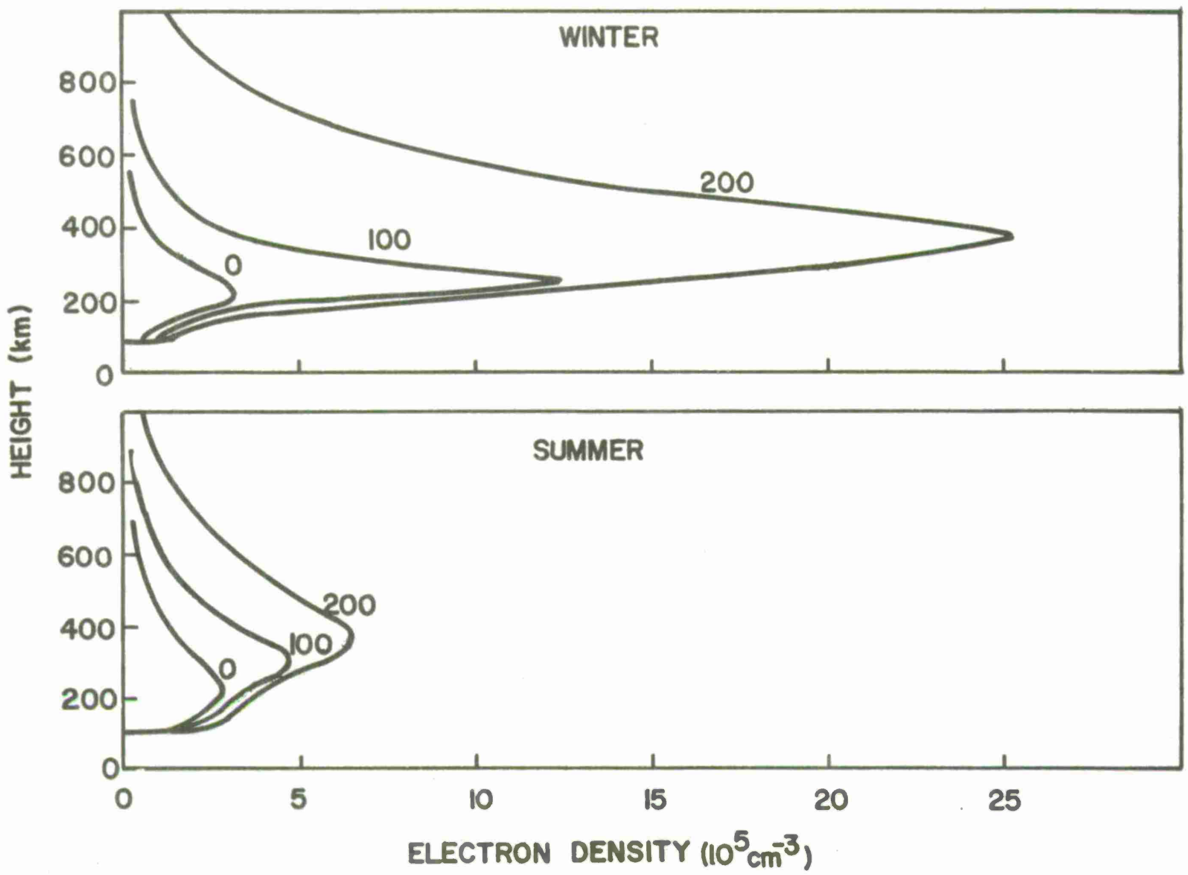


Figure 69. Seasonal and solar cycle variation of F region electron density in mid-latitudes. Numbers on curves represent sunspot number. (After Bean and Dutton)

However, it must not be supposed that the sun is the only influencing factor on the F_2 region., Other semi-regular anomalies occur quite distinct from any known solar disturbance. These are mostly unexplained. Many suggestions have been made to explain certain groups of observations, but no generally acceptable theory has emerged. The influencing factors of most importance are lunar and solar tides, the effect of the geomagnetic field, and the electrostatic forces originating in the atmospheric dynamo in the lower ionosphere.

4. Ionospheric Disturbances

Any perturbation of the normal incoming flux of energetic particles or radiation will disturb the balance of the various factors which work together to create the ionosphere. A disturbance of the structure of the ionosphere results. Since the intensity, duration, location, etc. of external perturbations can vary, the appearance of the resultant disturbance can take on many forms.

Emanations from the sun are the main source of the ionizing radiations. Electromagnetic energy arrives with the same speed as the visible light and radio waves that are often associated with solar disturbances such as flares. The more energetic radiation of the X-ray and ultraviolet part of the spectrum cause sudden, immediate increased ionization in the D, E, and F regions of the sunlit portions of the ionosphere. These manifestations are called sudden ionospheric disturbances (SIDs). The effects on electromagnetic communications will be discussed later.

The energetic particles generated in the same solar disturbance do not travel with the speed of light. The most energetic particles may be first detected at the earth from perhaps 10 or 20 minutes to several hours after the flare has been observed. Energetic particles may continue to arrive for as long as several days after a flare has subsided. About the same time as the arrival of the first energetic solar particle at the earth, it is observed that particles are pouring into the polar auroral regions causing intense ionization and resulting in large increases of electron density at altitudes as low as 40 km. It is not certain that all these auroral particles are the particles which had recently arrived from the sun.

Sometimes the arrival of the solar particles sets off what is called a geomagnetic storm, observed as a disturbance in the geomagnetic field. Typically, the disturbance begins simultaneously all over the earth. The reason for this, it is believed, is that the flux of particles from the sun can compress the geomagnetic field lines toward the earth. This is observed as an increase of the horizontal component of the magnetic field, and occurs simultaneously all over the earth. Normally, the increase lasts for a few hours after which it rapidly decreases to well below normal values. This decrease is attributed to the establishment of electric rings currents in the ionosphere which oppose the increased magnetic field. The recovery to normal conditions from the below normal conditions occurs over a period of several days.

Since the geomagnetic field strongly affects the motion of charged particles in the ionosphere, it follows that when disturbances occur in the geomagnetic field, the ionosphere will also show a disturbed character. Indeed, ionospheric storms are often observed to occur simultaneously with geomagnetic storms. And further, during the geomagnetic storm recovery period, ionospheric storms will be seen to reoccur.

B. THE EFFECTS OF THE IONOSPHERE ON ELECTROMAGNETIC RADIATION

As we now know, there are a great many major and minor structural details to the ionosphere. Also, temporal fluctuations occur over time periods ranging from seconds to solar cycles. Each of these details and fluctuations will have its individual effect upon any penetrating electromagnetic beam. However, at least to begin with, it is most instructive to construct a simplified model of the ionosphere and ask what effects it might have on the propagation of radio signals.

1. The Effects of a Simplified Ionosphere

For our study of the interactions between a simplified ionosphere and electromagnetic radiation, we shall choose the following model. The earth will have above it an ionosphere which will consist of a monotonically increasing electron density up to some maximum, above which it will monotonically decrease. There will be neither horizontal variations in the density nor temporal fluctuations. We will, however, include the geomagnetic field. Even with this patently unrealistic model, we can illustrate several of the most important ionospheric influences upon a penetrating wave. Perhaps the easiest way to study these influences is through examination of the ionospheric index of refraction.

a. The Ionospheric Index of Refraction

If one considers a plane electromagnetic wave, five parameters are needed for its complete specification: its frequency, its amplitude, its direction of propagation, its phase, and its polarization. The ionosphere will affect all of these parameters. This may be seen by studying a somewhat simplified equation for the ionospheric index of refraction, $n = (c/V)$. The equation below ignores the collision of particles which occur in the plasma, but does include the very important effect of the geomagnetic field.

$$n = \left(1 - \frac{X}{1-Y\pm Z}\right)^{\frac{1}{2}} \qquad (14)$$

X here represents the square of the ratio between the plasma frequency ω_η and the penetrating radio wave frequency ω, i.e., $X = (\omega_\eta/\omega)^2$. The plasma frequency is directly proportional to the square root of the electron density η. Y and Z are variables which depend, among other things, upon the geomagnetic field. If there were no magnetic field, Y and Z would be zero and the expression for the index of refraction would be simply

(15) $$n = (1 - X)^{\frac{1}{2}} = \left[1 - \left(\frac{\omega_n}{\omega}\right)^2\right]^{1/2}$$

It is clear from equation (15) that if the electron density did not vary, the higher the frequency of the signal the smaller effect the ionosphere would have on the wave. In other words, as ω gets larger, $(\omega_n/\omega)^2$ gets smaller and n (= c/V) approaches 1, i.e., V approaches the speed of light in empty space.

It can be further seen, through equation (14), that the characteristics of the index of refraction can be summarized as follows: (i) it is ~~directly proportional to the square root of~~ inversely related to the electron density; (ii) it approaches unity as the radio frequency increases ($X \rightarrow 0$); and (iii) it has a different value for each polarization of the electromagnetic wave. (The polarizations of interest are along and transverse to the geomagnetic field.) (see change 1)

Characteristics (i) and (iii) lead to additional effects. If one recalls Snell's Law of Refraction, it will be remembered that the bending of the ray resulted from a change in the index of refraction from one medium to another. If we look at any two infinitesimally thin layers of our model ionosphere, the electron density will differ and consequently so will the index of refraction. Thus, any ray propagating through the medium at any angle except normal (perpendicular) incidence will be bent away from the normal up to the level of maximum electron density, then back towards the normal as it recedes from the level of maximum density. The geometry of this refraction is shown for both a plane and a curved earth in Figure 70.

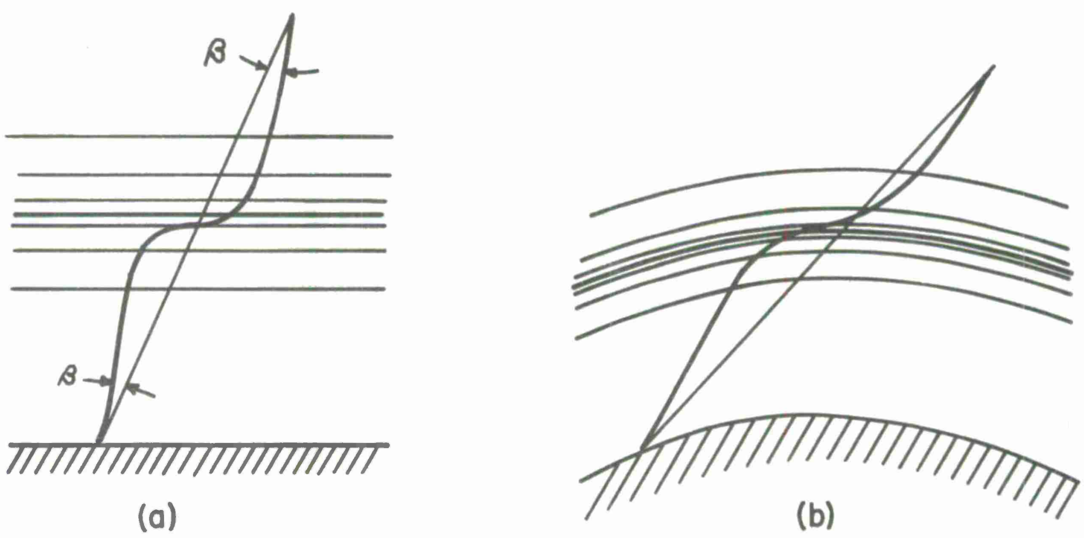


Figure 70. (a) The bending of a radio ray as it passes through an idealized plane atmosphere. (b) The bending of a radio ray as it passes through an idealized spherical shell model of the ionosphere. (After Lawrence)

One peculiarity of Figure 70 must be discussed. Since the index of refraction is n = (c/V), Figure 70, and equation (15) clearly show that n will be less than 1, implying that V is greater than c, a condition some will regard as ruled out by the Special Theory of Relativity. However, it must be remembered that this restriction, i.e., that the maximum velocity attainable in the universe is that of light in a vacuum, applies only to an energy carrying signal. In a dispersive medium like that of the ionosphere, all energy is carried not at the phase velocity but at the group velocity, which is, indeed, less than c.

If the ionospheric distribution of electrons were symmetrical about the level of maximum density, then the ray in the planar case would emerge traveling at the same zenith angle as it entered. However, for the spherical model, the curvature affects the trajectory so that the deviation of the true direction from that detected by the radio receiver (or satellite, for that matter) is somewhat greater in the real case than the planar one.

b. The Ordinary and Extraordinary Rays

In equation (14), it can be seen that two values for the index of refraction are contained in this one equation. They are determined by whether the plus or minus sign between the two terms in the denominator is chosen.. The two values are applied to what are called the <u>ordinary</u> and the <u>extraordinary</u> rays respectively. They are of opposite polarization and arise because of the influence of the geomagnetic field.

Because the ordinary and extraordinary ray have different indices of refraction, they travel essentially independent paths, each being refracted according to its own index. Thus a radio wave with components along and transverse to the magnetic field will be broken into two separate rays which will be refracted differently, travel at different speeds through the ionized medium, and thus emerge with a different phase relationship than when entering the medium. This results in a rotation of the plane of polarization of any linear polarized signal. The apparent zenith angle of the source will, also as a result of the two indices, be slightly different for each wave.

c. Absorption

Besides the splitting of the beam, and the effects of the changing refractive index on each of the resulting rays, our simplified model of the ionosphere also absorbs energy from waves passing through it. The physical mechanism which accounts for this decrease in amplitude of the electromagnetic wave is quite simple.

When an electromagnetic wave enters a region which contains free electrons, these electrons are set into vibration at the frequency of the penetrating wave. If they did not collide with the other constituents of the atmosphere, most particularly the heavy positive ions and neutral particles, the energy imparted to the electrons would be passed back to the propagating wave so that no apparent attenuation of the wave would occur. However, when collisions occur, some of the momentum of the electron is transferred to the particle collided with. Thus, the kinetic energy imparted to the other particles by the electrons is lost to the wave and is seen as increased random thermal energy in the gas. Attenuation of the extraordinary and ordinary rays will be different. This is illustrated in an extreme case by the observations recorded in Figure 71.

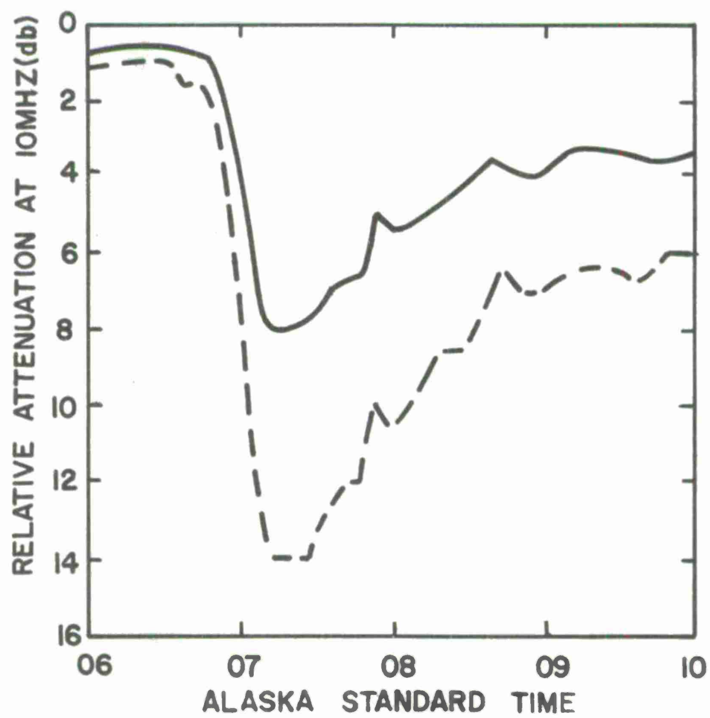


Figure 71. A record of ionospheric absorption of the extraordinary (dashed curve) and ordinary rays (solid curve) during a time period when they showed large differences. The frequency was 10 MHz. (After Lawrence)

2. Added Effects of the Observed Ionosphere on Electromagnetic Radiation

It is possible now to abandon our simplified ionospheric model and look at some new phenomena associated with the observed ionosphere, which affect electromagnetic radiation. Certainly, all those effects which we have already discussed will be present, but what else must now be considered? The most straightforward approach is to consider one at a time the major features of the ionosphere which our model disregarded and ask what effect each of these has upon electromagnetic wave propagation.

Probably the three most important shortcomings of our simple model were the following: its smooth vertical structure, its lack of large scale temporal fluctuations, and its obvious disregard for small horizontal variations and their fluctuations in time.

a. The Unsymmetrical Vertical Electron Density Distribution

The first problem is that of the unsymmetrical vertical electron density distribution. From the simplified theory of the production of the ionosphere through radiation, it is quite obvious that the shape of the ionospheric electron density distribution cannot be symmetrical and observation shows it does not monotonically increase up to the level of maximum electron density. Thus trajectories through the ionosphere such as encountered in radio astronomy or satellite communications will be more complex than those illustrated in Figure 70. Figure 72 shows a more realistic trajectory and Figure 73 shows angular errors at the ground from the true elevation angle of a source at various altitudes.

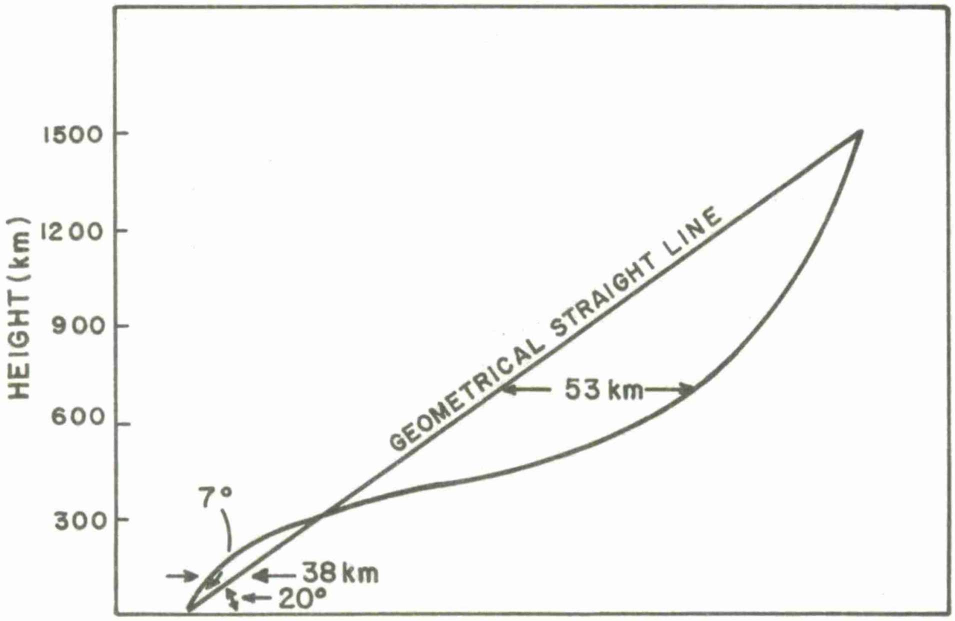


Figure 72. The deviation from a straight line path of a 20 MHz signal as it passes through the ionosphere. (After Lawrence)

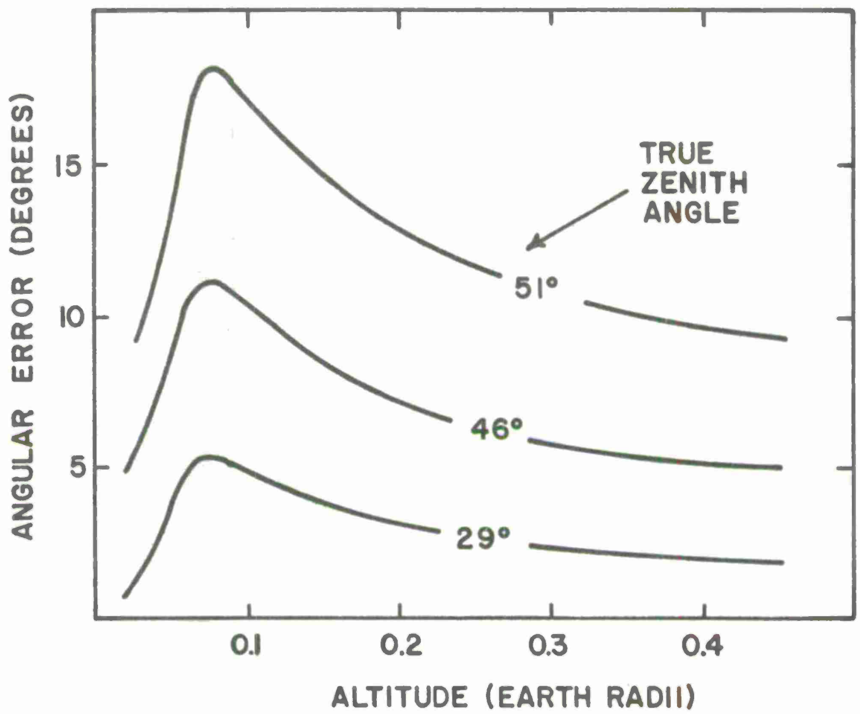


Figure 73. The angular error of a source from its true zenith angle as a function of its distance from the earth. The frequency is 20 MHz. (After Lawrence)

b. Large Scale Temporal Fluctuations

Another shortcoming of our simple model was the disregard for the hourly, daily, and seasonal fluctuations of the electron density distribution. These are the changes which occur over widespread areas on a more or less predictable basis. Their chief effects are to make the true path of the rays more difficult to predict and to alter the absorption of penetrating waves. They will also cause the critical frequency of penetration to vary as the smallest index of refraction (highest density of free electrons) varies. The details of the normal variations have been treated in the section describing the ionosphere and will not be elaborated again.

Closely allied with the large scale fluctuations in the ionospheric structure is the variation in the absorption of radio waves. The effect of absorption is to reduce the amplitude of any signal. Absorption in the ionosphere increases for the following reasons: (i) the electron density increases; (ii) the frequency of the penetrating radio wave decreases; (iii) the collision frequency increases; and (iv) the index of refraction decreases. Also, as indicated above, the absorption is different for the extraordinary and ordinary rays.

During daylight, the D layer of the ionosphere accounts for much of the absorption of VHF (30 to 300 MHz) radio waves. Above 100 MHz, even with oblique incidence, this absorption is normally less than 1 decibel for a path through the ionosphere. However, because absorption in this frequency region depends upon a $(1/\omega^2)$ factor, normal attenuation will be several decibels at 30 MHz, and higher at lower frequencies. At night the D layer almost disappears, and the absorption is much less. These characteristics may be drastically altered by events such as solar flares and geomagnetic storms, as will be discussed later.

So far, only effects on the D region have been considered. The reasons for this are at least two. One is that changes in the D region are most drastic and produce, consequently, the largest anomalous propagation effects. However, some studies have shown that during years of high sunspot activity, absorption in the F region can be comparable to the absorption in the undisturbed daytime D region. Above 100 MHz, the absorption is generally negligible for earth-space communications, being less than 0.3 db at vertical incidence. The correction factor for large zenith angles never exceeds two because of the height of the F region.

c. Small Scale Random Fluctuations

(1) Scintillation

Smaller scale fluctuations in the electron density lead to a phenomenon called scintillation. Evidence of scintillation is gained by monitoring steady signals from outside the ionosphere such as radio stars or satellite transmissions. The amplitude of the signal is seen to vary randomly as the scintillation sources (electron anomalies) pass through the view of the receiving antenna. Physically, the variation in electron density in the viewing path causes variations in the index of refraction which cause the signal to be focused randomly.

Important differences exist between the radio star electromagnetic spectrum and that of satellites. Radio stars are essentially broad band sources of finite size which produce incoherent plane wave fronts at the outer boundary of the ionosphere. Satellites, on the other hand, are narrow band, coherent, point sources of radio waves and have spherical wave fronts at the outer boundary. Radio stars offer a very slowly changing position in space, while satellites move rapidly across the sky. Time periods for scintillation of the radio star source varies from one second to one minute. The scintillations are primarily due to the drift of the electron clouds across the antenna path. A low level satellite moves so swiftly, however, that the scintillations are due to the satellite antenna path sweeping across many different electron density regions. The rate of amplitude fluctuation for this latter case can amount to 10 cycles per second.

(2) Scattering

Electromagnetic radiation also is scattered by ionospheric inhomogeneities. Exactly what the physical mechanism is which causes the scatter is not completely understood. But it is possible to briefly present the most likely physical causes.

The most widely accepted class of theories assumes that an irregular distribution and movement of the electrons and ions exists. Within each small homogeneous blob of ionized gas, the electron density and thus dielectric constant is approximately uniform. These properties differ slightly from blob to blob. The penetrating beam sees, then, a turbulent mass in which the index of refraction changes from eddy to eddy. Any initially plane wave front is refracted and reflected so that energy is distributed into all directions.

The scale of the turbulent elements can be estimated and the resultant values for the mean size of the blobs is of the order of 100 - 200 m. The variation in the electron density as a percentage of the total density can also be estimated. Its order of magnitude appears to be about 1%.

The reason for the existence of the blobs is proposed to be found in the meteorological dynamics of this part of the atmosphere. Theoretically, it can be shown that the most important considerations in the formation and dissolution of the blobs is the random turbulent velocity of the gas combined with the vertical temperature gradient, and the vertical gradient of the electron density. The fluctuations of density alone, however, due to the turbulence, do not appear to be of sufficient magnitude to explain the total electromagnetic variations from blob to blob. The full explanation of the scattering process remains to be found.

The strength of signals sent via the ionospheric scatter method is observed to have slow systematic fluctuations imposed upon the rapid random fluctuations which always characterize the signal. These systematic variations are determined by systematic changes in the ionospheric structure. In mid-latitudes the strongest signals are observed in the summer and the weakest at the equinoxes. The winter midday signals are about as strong as the summer midday signals but during other times of the day comparative signals are much weaker. There appears to be a close association between signal strength and solar zenith angle.

Over paths in the auroral zone, markedly different systematic variations occur. In general, the signals are stronger than over mid-latitude paths and they show considerably different dependence on time for the daily maximum and minimum signal strength.

(3) Fading

If a radio signal which has propagated through the atmosphere is monitored for a length of time, it is invariably true that its strength fluctuates. The periods of these fluctuations can be from fractions of a second to solar cycles. The phenomenon is called fading.

Fading results from several causes. If the absorption changes over a communications path, then the signal strength will fluctuate as a consequence. Changes in the lower ionosphere are especially important to the absorption of signals below 30 MHz. Fading can occur when this part of the ionosphere is disturbed.

Because a received radio signal is made up of several parts of the original wave front by the time it has passed through the turbulence of the atmosphere, fading can occur because of phase difference between these several separate parts. In Section 2 page 11 the effect of adding coherent signals with phase differences was illustrated and discussed. This can be the cause of the rapid fluctuations in a received signal as the multiple paths from transmitter to receiver of several parts of the signal fluctuate.

Fading can be not only due to small differences in the lengths of similar paths, but also may be due to interference between the multi-hop wave and a single-hop signal or other multi-hop signals. (The concept of single or multiple ionosphere reflections will be explained on pages 131 and 132 below.)

Finally, another fading phenomena, selective fading, occurs where slightly different frequencies are attenuated by differing amounts. Severe distortion of wideband signals can result from this type of fading.

Fading depends intimately upon the conditions of the ionosphere and the troposphere. It will change with time, sometimes rapidly if turbulent conditions exist in the region of the atmosphere causing fading, or sometimes very slowly if the governing meteorological conditions are essentially stable and slowly changing.

d. Meteor Trails

Besides the general topic of ionospheric scattering of electromagnetic radiation, one should consider the reflection and scattering from ionized meteor trails. Cosmic particles entering the atmosphere are normally vaporized in the region from 80 to 120 km above the earth's surface. They follow essentially a straight path, decelerating a negligible amount before they disappear. The ionized vapors left behind are in the form of a column which, at first, is very small, but rapidly diffuses into the surrounding environment. Very few of the ambient air molecules are ionized. As a first approximation, the maximum electron density resulting, per unit length of trail, depends only upon the mass of the particle and the cosine of the angle which its trajectory makes with the local vertical.

The ionized columns can be broken up into two classes: underdense and overdense, determined by whether or not the line density in the column is below or above 10^{14} electrons/meter. For columns below this line density, the electrons can be considered as independent scatterers. Columns with this line density or above act much like a metallic reflector.

C. USES OF THE IONOSPHERE FOR COMMUNICATIONS

The preceding brief description of the ionosphere and its general effects on electromagnetic waves gives enough background to proceed with a discussion of some of the more direct consequences of the ionosphere on communications.

The discussion which follows assumes that both transmitter and receiver are on or near the surface of the earth and that the signal of interest is that part of the transmitted signal which is reflected, refracted, or scattered by the ionosphere.

1. The Lowest Frequencies

It is appropriate to begin the discussion of the use of the ionosphere for communications with the lowest frequency signals. For, after years of neglect, a great deal of interest has recently been generated again in very low frequency propagation (1-30 KHz) through the atmosphere. The reasons for this renaissance of interest are bound up with four facts: (i) there is an increasing need for long range navigation; (ii) very low frequencies propagate for great distances without significant attenuation; (iii) signals via these frequencies are not severely degraded when the environment is disturbed; and (iv) new interest is being aroused for the study of atmospheric disturbances generating low frequency radio energy.

The study of these very low frequencies can be done from at least two standpoints. Both approaches adhere to the view that the earth and lower edge of the ionosphere are sharp reflecting surfaces. The wave is then confined between these two surfaces. One viewpoint traces hypothetical electromagnetic rays much like we trace light rays. From the interference of several such rays which have been reflected from the ionosphere and earth one can determine the strength of the electric field out to about 1500 km from the transmitter.

The other approach views the earth-ionosphere system as a large natural waveguide (see Section 2, pg 33). Different modes are excited within this cavity and travel long distances with only slight attenuation. This approach gives excellent results beyond 1000 km and less than 8000 km.

Essentially, attenuation occurs from the spreading of the wave and from absorption by the earth and ionosphere. A graph of the average attenuation of the power of a very low frequency signal as a function of frequency is given in Figure 74.

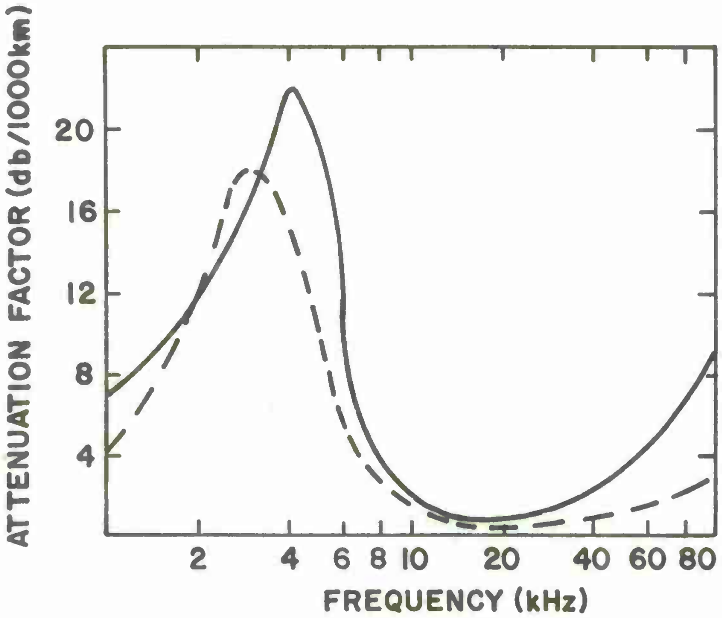


Figure 74. Exponential attenuation factor for the atmosphere as a function of frequency. The solid curve represents a daytime sea water path and the dashed curve represents a nighttime sea water path. (After Watt and Maxwell)

It is quite obvious from the figure that certain frequencies should be preferred over others for long distance transmission. This has been found experimentally to be the case. Twenty KHz radio signals dominate their lower frequency companions more and more as one recedes from the source of the signal.

2. The Ionospheric Wave

It will be recalled from Figure 25 that several possible propagation paths exist for an isotropic signal of medium wavelengths. As described above, the bulk of communications in the mid-frequency range are carried via the ionospheric or sky wave, except for the first few miles away from a transmitter. In the ionosphere, the density of free electrons increases to a maximum at some level, then diminishes above, it; so the index of refraction, which decreases with increasing electron concentration, decreases up to the maximum ion density and increases above.

The effect is greater for lower frequencies. Under the proper conditions, then, a radio wave directed at the ionosphere will return to the earth some distance away. The reason for the bending of the ray was described in an earlier section on refraction.

The ionospheric wave is bent gradually, but for many purposes it is convenient to view the return as a simple reflection and assign a virtual height based on the total transit time. This is illustrated in Figure 75.

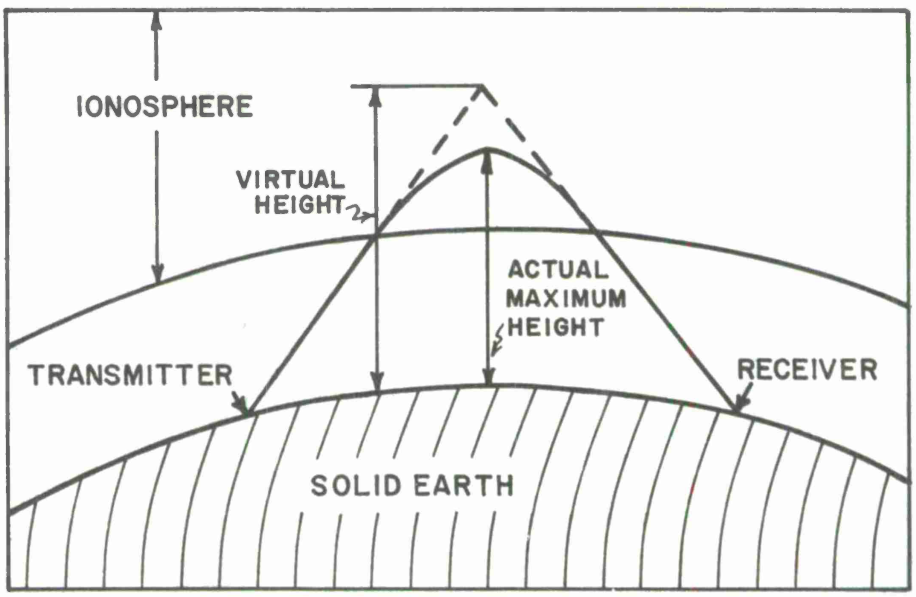


Figure 75. The relationship between virtual height and actual maximum height. The virtual height is determined from the total travel time of the wave from transmitter to receiver.

It is apparent that if the sky wave is directed toward the ionosphere at a smaller angle with respect to a tangent plane touching the earth at the the transmitter site, it will need less bending to return to the earth. Also, more distance will be covered by such a trajectory before it reaches the earth's surface again. This angle is called the angle of radiation or wave angle. It leads us naturally into a discussion of the concepts of skip zone, skip distance, and critical and maximum usable frequencies.

a. Skip Distance and Skip Zone

Consider Figure 76. A wave launched at some small wave angle may be returned at D_2. As the wave angle is increased, the distance D_2 will diminish. If the frequency is high enough for the existing ionospheric conditions, eventually a critical wave angle will be reached which directs the ray at the ionosphere at such an angle it cannot be bent back towards the earth. There will then be a zone of silence or skip zone between the point where the last ray is returned to the earth and the point beyond which the ground wave is no longer effectively detectable. This distance is called the skip distance.

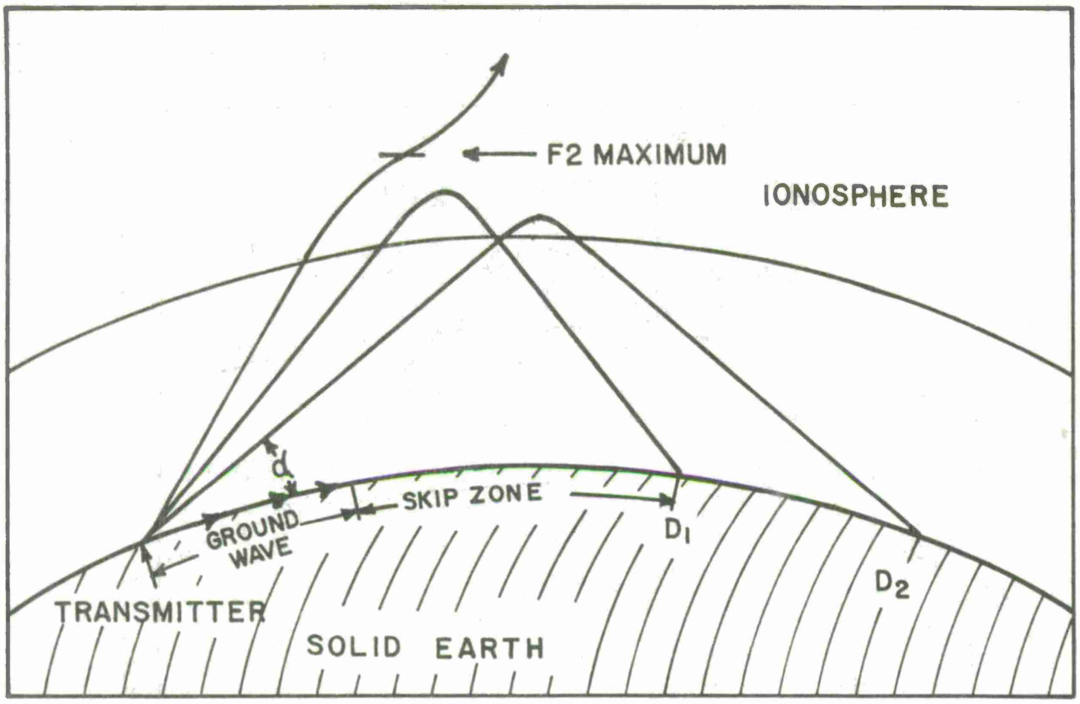


Figure 76. A schematic diagram illustrating the wave angle α, skip zone, and penetration of the ionosphere at a frequency above the critical frequency.

b. Critical Frequency

If one decreases the frequency, the effective index of refraction ($n = c/V$) will increase, and the critical wave angle will be larger. Eventually a limiting frequency will be found at which a reflection will occur for a vertical wave angle ($\alpha = 90^{o}$), and any higher frequency will penetrate the ionosphere. All lower frequencies will be reflected back to earth at all wave angles.

This frequency is called the critical frequency and depends upon the maximum electron density in the ionosphere (normally this is at the peak in the F_2 layer). A great deal of the earliest and most effective study of ionospheric layers was done by following the variations of its critical frequency, which varies systematically with the time of the day and the season, and also randomly. The systematic observation of the critical frequency as a function of time and place, coupled with the theoretical reasoning for its existence, had led to a large body of knowledge upon which most speculations concerning the structure of the real ionosphere were based until the introduction of direct sampling techniques. Rocket and satellite sampling of the ionosphere has, in the large, substantiated these earlier deductions. They have, as may have been expected, also added many refinements.

c. Maximum Usable Frequency

Closely related to the critical frequency is the maximum usable frequency. This will be higher than the critical frequency because it measures the maximum frequency which can be used to transmit over a specified distance. (α will therefore be less than 90 degrees.)

Every sky wave is partially absorbed by its passage through the ionosphere, the degree of absorption decreasing with increasing frequency. The maximum usable frequency is the frequency least absorbed. Below this frequency, absorption increases rapidly. Natural environmental noise also increases rapidly; so the signal to noise ratio decreases until communications cannot be maintained. The lowest frequency that can be used is called the lowest usable frequency. For the most satisfactory results frequencies are chosen as near to the maximum usable frequency as possible.

Typical distances for waves launched tangentially to the earth and the relationship between the maximum usable frequencies for these distances to the critical frequency have been determined from many years of data. For average ionospheric conditions, one reflection from the F_2 layer will return to earth about 4000 kilometers from the source transmitter. If the E layer is the reflecting region, 2000 km is more typical. The MUF (maximum usable frequency) for the 4000 km hop F_2 reflection is about three times the critical frequency; for the 2000 km hop E layer reflection, the MUF is about five times the critical frequency.

For the propagation of signals beyond those distances available from one reflection in the ionosphere, there is the possibility of multiple reflections. A wave, returned from the ionosphere after an initial reflection, can be reflected from the earth, return to the ionosphere, and be reflected a second time. It will encounter the earth's surface again roughly twice the distance of the first return. This can be repeated several times. However, each time the wave is reflected from the ionosphere or the earth, absorption removes some of the energy from the wave. The surface absorption largely depends upon the surface encountered, sea water being the best terrestrial reflecting surface.

For a graphic illustration of the effects described above, consider Figure 77. In it, two important layers are used to show the geometrical relationship between single hop distances and layer heights as well as the concept of multiple hops.

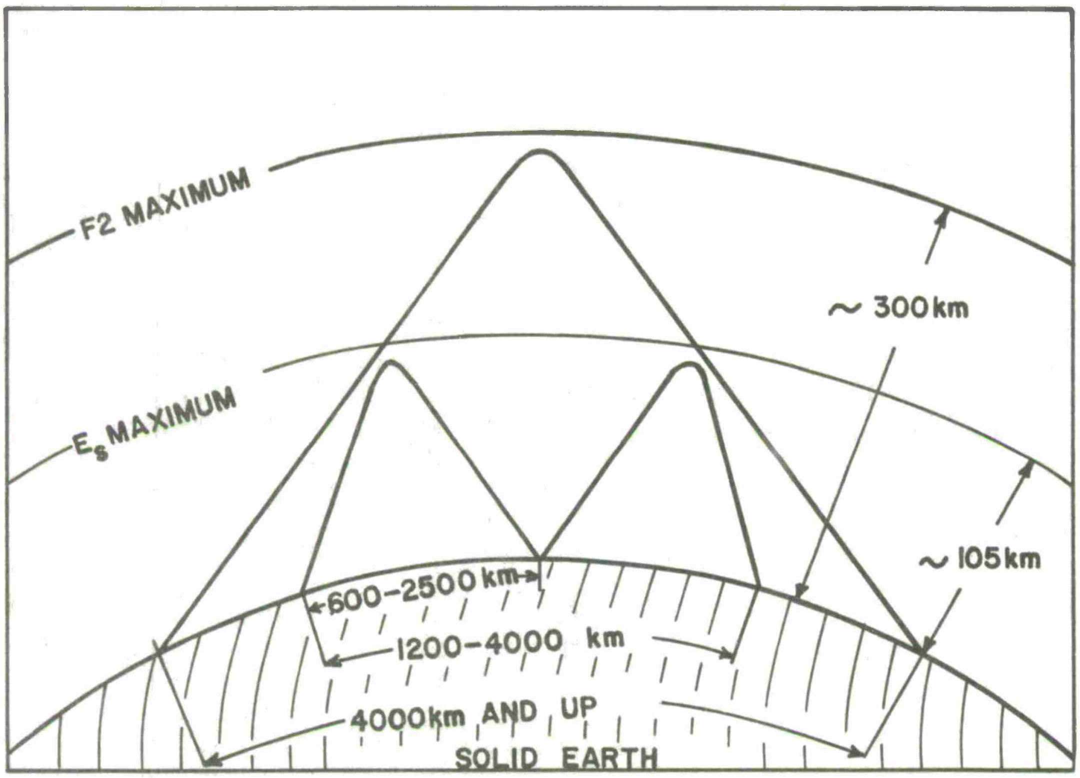


Figure 77. Relationship between F_2 reflection, Sporadic E (E_s) reflection, and single hop and double hop propagation.

3. The Scattered Wave

Below the frequencies where so-called microwaves begin (somewhat arbitrarily at 3 GHz or 10 cm in wavelength), and above those frequencies which are reflected from the ionosphere, (roughly 30 MHz depending upon ionospheric conditions), lies a transition zone of frequencies. At the lower end they will often be usable for long distance communication via the ionospheric wave and at the upper end will be propagated with most of the characteristics of microwaves. The most important geophysical effects on these frequencies gradually change from those characteristic of the mid-frequency region to those most important to microwaves. However, in certain respects, the interaction of this band of frequencies with the environment exhibits unique properties which have proven to be of value for an entirely new technique of communication -- scatter communication.

At least two distinct scattering mechanisms exist to propagate energy beyond the normal radio horizon: one depends upon the existence of inhomogeneities in the ionosphere; the other upon inhomogeneities in the troposphere (see page 104).

In the ionospheric case, scatter communication is usable over distances from about 1000 to 2000 km. Frequencies available are in the region from 25 to 60 MHz but bandwidths seldom exceed 10 KHz. Teletype channels are primarily transmitted by this means. The physical mechanism which supports this mode was discussed in the section describing scattering in the ionosphere (see page 125).

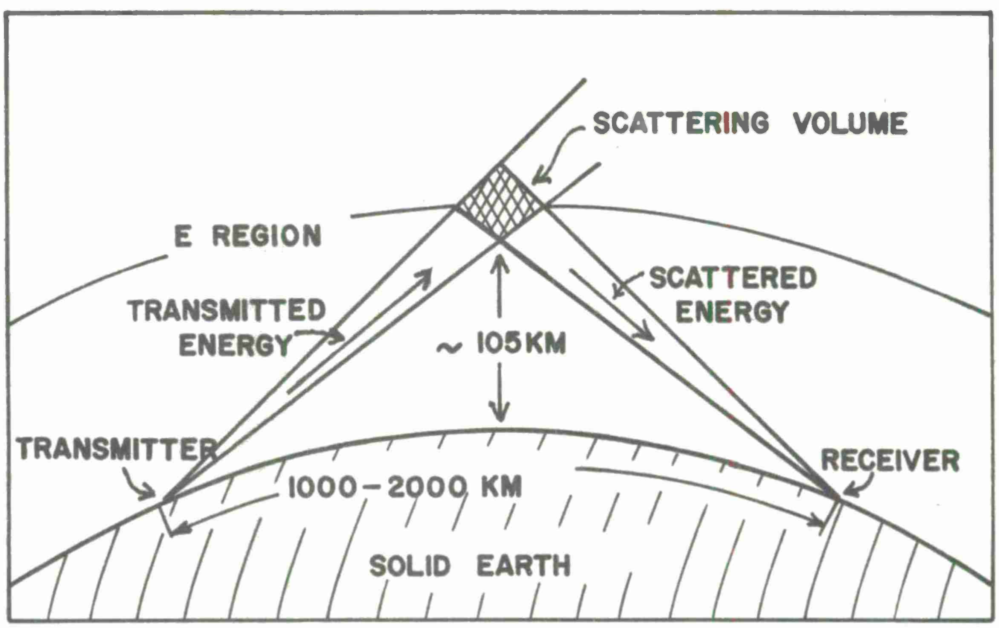


Figure 78. The scattering volume viewed in common by transmitter and receiver in an ionoscatter communication mode.

As described before, ionized meteor trails can reflect and scatter electromagnetic energy. And it has been possible to use this phenomenon for communications. Distances up to approximately 2200 km can be reliably linked by this method. The effective frequency range of this propagation mode is from about 6 to 75 MHz. The physical processes were considered in the section describing meteor trails (see page 126). Briefly the electron density in the trail determines the reflection characteristics. Overdense trails act much like metallic reflectors, but they are so rare it is necessary to rely on specular reflections from underdense columns for a continuous signal. In this case it is necessary that the scattering column be perpendicular to the line of sight. Even though perhaps only five percent of the underdense meteor trails satisfy this requirement, the overall meteor frequency is high enough to make a continuous signal possible.

4. The Satellite Mode

With the advent of Sputnik I, the subsequent increased emphasis on space research, and advances in the communications state of the art, it has become feasible to use satellites as relay stations. In principle a ground station transmits a narrow beam signal to the satellite. The satellite receives the signal, and either amplifies it and retransmits it back to earth or merely reflects some of the energy towards the receiving station. By using a satellite in stationary orbit, some 22,000 miles above the earth, it is possible to achieve reliable one hop communications between stations up to approximately 10,000 miles apart.

It is clear that any frequency considered for satellite relay must be above the maximum usable frequency of penetration in order to be able to continue into space. This will be somewhere in the vicinity of 5 - 20 MHz for the real ionosphere at vertical incidence. Further, the general statement can be made that the shorter the wavelength one considers (higher frequency), the smaller will be the effect of the ionosphere upon its passage through it. Practical usable frequencies extend from about 100 MHz to several GHz. At lower frequencies ionospheric effects are too strong, while at higher frequencies signals may be strongly absorbed in the troposphere by oxygen, water vapor, and hydrometeors.

The order of magnitude and the variation with frequency of various effects of the ionosphere on this mode of propagation are listed in Table 7. The listed values can be extrapolated downward to about 30 MHz. At lower frequencies the variation with frequency cannot be approximated by such simple relationships. At frequencies above about 100 MHz the values diminish rapidly and generally become unimportant.

Table 7. Frequency dependence and order of magnitude of various ionospheric propagation effects. (After Lawrence)

PROPAGATION EFFECT	FREQUENCY DEPENDENCE	ORDER OF MAGNITUDE
Refraction	$1/\omega^2$	$4x10^{-4}$ radians
Absorption	$1/\omega^2$	
night		0.005 db
day		0.05 db
daytime polar cap event		1.0 db
Differential Absorption	$1/\omega^3$	
night		$5x10^{-5}$db
day		$5x10^{-4}$db
daytime polar cap event		$5x10^{-2}$db
Phase path length change	$1/\omega^2$	-400 meters
Differential phase path length change	$1/\omega^3$	-6.4 meters
Polarization rotation	$1/\omega^2$	6.6 radians

In addition to these effects there is interference caused by ionospheric and tropospheric scattering of energy into the receiving antenna from sources radiating at the same, or nearly the same frequency, as the satellite. This mechanism had not been important until recently because relatively few sources had existed, but with the proliferation of radiating sources this scattering must be considered in the design of future systems.

5. Effects of Ionospheric Disturbances on Electromagnetic Wave Propagation

In our description of the ionosphere we described two general classes of disturbances of the ionosphere. Both were related to events on the sun. The first class results from the enhancement of the X-ray and ultraviolet portions of the solar electromagnetic spectrum which produces the almost immediate disturbances generally classed as sudden ionospheric disturbances (SIDs). The second class is produced by solar particles expelled from the sun at the same time as the enhanced electromagnetic radiation is produced. These travel at speeds less than the speed of light and arrive later than the radiation. They cause geomagnetic storms and other associated effects.

As was pointed out before, the disturbances have been identified or named in accordance with the appearance of the disturbance on the output of the particular sensor being used.

The presence of a SID may be deduced from measurements at the earth's surface in several ways: (i) Because of enhancement of the ionization in the E or F regions, a shift of the frequency occurs in stable signals passing through them. This shift may be detected and is called a sudden frequency deviation (SFD). (ii) Enhancement of ionization also occurs in the D region of the ionosphere. This may be detected by the increased absorption of cosmic noise and regular surface to surface radio transmissions. A sudden cosmic noise absorption event (SCNA) or in severe enhancement conditions short wave fade-outs (SWFs) are the result of the increased ionization in the D region. (iii) While absorption is observed at higher frequencies, increased electron density in the D region increases its reflectivity for the lower frequencies and can therefore change the virtual reflecting height of the signals. The result is better propagation of low and very low frequencies. Since the intensity of atmospherics is a maximum in the LF and VLF bands, their signal strength increases under these conditions. This is called a sudden enhancement of atmospherics (SEA). The change in the D region reflecting height also causes changes in the phase of LF and VLF signals. These are called sudden phase anomalies (SPA). (iv) VHF band signals are scattered in the ionosphere. During a SID the scattering is enhanced, but the associated increased absorption can more than compensate for it. The relative strength of these two effects is very frequency dependent, and therefore reception of scattered energy might be enhanced at one frequency but fade-out at another frequency. (v) Because the atmosphere is moving rapidly at ionospheric altitudes, the increase in the number of moving electrons induces an increase in the geomagnetic field. This is called a magnetic crochet. This effect may be detected by instruments at the surface of the earth.

As mentioned above, the second class of disturbance is manifested in two somewhat different ways. First, the flux of particles into polar regions causes large increases in electron density at low altitudes. This results in strong absorption of high frequency (HF) band communications. These events are called polar cap absorption (PCA) event, or simply, polar cap events (PCEs). HF band communication typically is not possible during these events because of the high degree of absorption. This complication is called an HF radio blackout. Secondly, there are the effects induced by geomagnetic storms. For example, generally increased auroral activity with frequent intense aurora, frequent appearance of Sporadic E layers, decreased F_2 layer critical frequencies, and occasional ionospheric storms with SEAs, SPAs, etc..

6. Summary

From the information above, it should be clear that the propagation of radio waves in the frequency region below about 30 MHz and above about 30 KHz is quite dependent upon the state of the ionosphere. The absorption, reflection, and refraction of waves act so differently for the various frequency bands that certain of them are known to be usable only at night, while others are best propagated in the daytime. This knowledge plus information concerning the seasonal solar epochal effects are important factors to consider when choosing a frequency for long distance communication. It may be said however, that in general the lower the frequency of a band, excepting the very lowest, the more susceptible it is to outage during ionospheric disturbances, auroral displays, and magnetic anomalies. Conversely, some of the bands around and above 30 MHz are only available for distant communication during times when the electron concentrations in the ionosphere are unusually high; this would mean during the presence of the phenomena mentioned above. Further, lower frequencies are propagated best, on the average, during winter and nighttime. Very low frequencies, in the region below 30 KHz, begin to show very greatly enhanced propagation characteristics due to the waveguide effect that the earth-ionosphere presents. Essentially, the wavelengths become so long as frequency decreases that only a few or perhaps one mode of propagation can exist in this natural waveguide. In this latter case, very long distance signals without multi-path interference or significant attenuation are possible.

The dependence of long distance communication on the ionosphere has led the U. S. National Bureau of Standards to develop forecasting techniques to identify in advance the periodic and well known variations in the ionosphere. This information is published, including predictions of maximum usable frequencies and other related data. Supplementary real time forecasts tailored to the needs of various Department of Defense Agencies has recently been implemented by the Air Weather Service.

V. NATURAL GEOPHYSICAL RADIO NOISE

One of the chief concerns in this paper has been to provide some feeling for how the natural environment modifies and limits radio propagation. To be faithful to this purpose some mention of the natural geophysical radio noise bombarding an exposed receiving antenna must be made. Although much has been written on the various parts of the natural radio noise spectrum, in this brief resume we will only outline the major features of the spectrum and identify their sources as best is known.

If one were to go out into the open air, away from man made noise sources such as power lines and ignition systems, and measure the electrical field intensity as a function of frequency, he might obtain a diagram similar to that shown in Figure 79. Further investigation would reveal that the noise at the lower frequencies (below 10 MHz) is predominantly generated by lightning discharges, that the noise at the higher frequencies (above 100 MHz) is predominantly generated by atmospheric thermal emissions and by the equipment itself, and that in the mid-range of frequencies (10 MHz - 100 MHz) the predominant source is outside the earth, i.e. the galaxy.

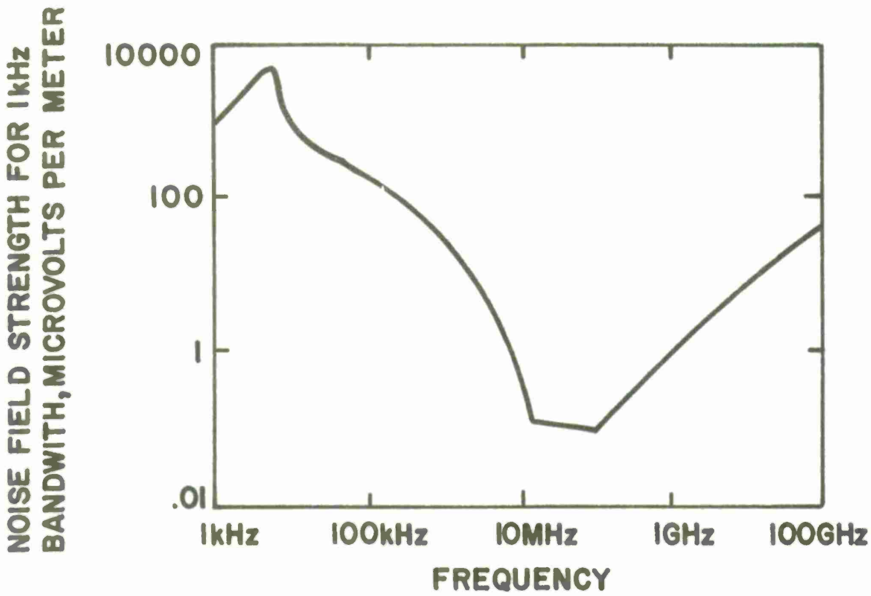


Figure 79. General spectrum of natural geophysical radio noise. (After Cottony and Johler)

A. NOISE CAUSED BY LIGHTNING DISCHARGES

At the lower end of the frequency spectrum the noise is mostly caused by lightning discharges. A typical spectrum showing the variation of noise intensity versus frequency is shown in Figure 80. Note that the maximum intensity is near 4 KHz, and that the intensity decreases with increasing frequency.

When lightning is examined closely it is found that there is not just one stroke associated with a bolt of lightning. Instead, one observes first numerous leaders or precursors going from the cloud toward the ground. They develop an ionized channel which is then utilized as the path for the main stroke, a tremendous surge of current from the ground back to the cloud.

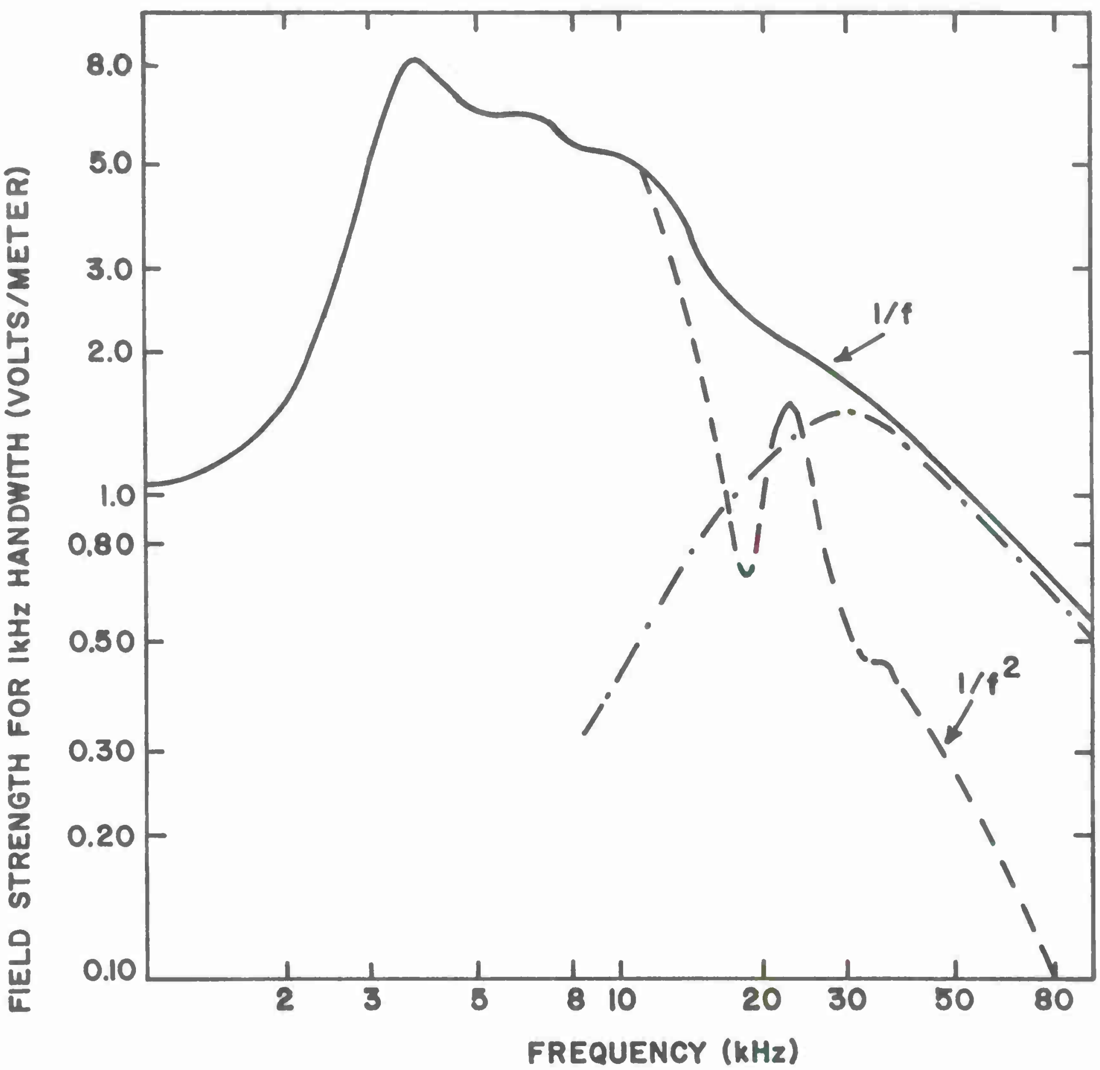


Figure 80. Frequency spectrum of lightning discharges at a distance of one mile. The dot-dashed curve represents the spectrum due to precursors; the dashed curve, the spectrum due to the main stroke; and the solid curve the sum. (After Watt and Maxwell)

Most of the radiated energy is produced by the massive return stroke peaking around a frequency near 4 KHz. The precursors account for most of the higher frequency energy, above about 20 KHz. Figure 80 provides some idea of the energy found at various frequencies measured in 1 kHz intervals from both the main stroke and the combined precursors. For frequencies above the maximum for the main stroke energy curve, it can be noted that the energy production decreases approximately as $1/f^2$ where f is the frequency; the energy associated with the precursors decreases less rapidly, approximately as $1/f$.

Over a 24 hour period, and over the entire earth, on the average somewhere near 50,000 thunderstorms occur. Each of these has many lightning bolts associated with it. Since this noise is concentrated at low frequencies, it can propagate great distances. As a result, at any one place or time, noise which originates from lightning can be quite intense. The exact intensity observed, however, is highly variable. Systematic variations are observed for geographical location, time of the day, weather conditions, and season. These variations are closely associated with the world-wide distribution of thunderstorms.

Study of the noise has revealed some of the systematic variations. Graphs exist for calculating the mean noise level by season and time of day over the entire earth. These calculations or predictions, however, must be used with caution. The data sources, upon which the graphs are based, were sparse. While the predictions are fairly good for sites near those of the data base, the results for more remote sites can be as much as plus 40 to minus 25 db from that predicted. Also, noise from nearby storms usually exceeds the values estimated.

At any given time, the noise generally originates from localized geographical regions which do not change significantly over time intervals up to three hours. Therefore, by using a narrow beam and if it can be directed away from source regions, noise intensity can be greatly reduced.

Lightning generated noise can also be propagated via the so called whistler mode. This propagation mode is a feature of frequencies below 30 KHz. When energy in this band is produced in a lightning flash, some of the energy leaks through the ionosphere and is guided by the earth's magnetic field. If the electron density in the ionosphere is high enough, the wave will be returned to the earth at a point in the opposite hemisphere of the earth. The speed of propagation of the wave is dependent upon the frequency, being faster for the higher frequency. Therefore, when a receiver is connected to an audio output a descending tone is heard lasting 1-2 seconds; hence the name whistler.

Because the ionosphere affects propagation of low frequency signals, it strongly influences the spectral characteristics of the noise at long distances from the sources of lightning generated energy. Absorption over the frequency interval from 1 to 100 KHz was shown in Figure 74 and it may be noted that sharp maximum of absorption exists at 4 kHz, very close to the maximum of the spectral energy curve for lightning. The resulting spectrum for low frequency noise is shown in Figure 81, and it shows that a relative maximum is induced between 10 and 20 kHz.

It should be noted that the above discussion applies to the sky wave portion of the energy produced by lightning. It is possible that the ground wave, also excited by a lightning discharge, may at times, be as strong or stronger than the sky wave.

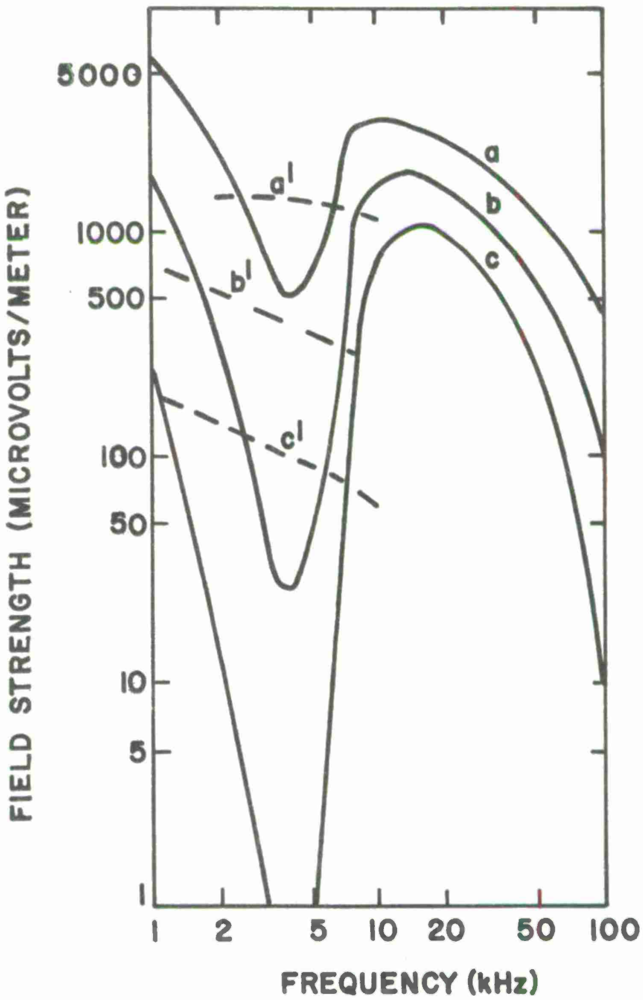


Figure 81. Variation of field strength with frequency at three distances from an idealized source. Curve a shows the variation at a distance of 1000 km; curve b, 2000 km; and curve c, 4000 km. The primed curves show the field strength due to the ground wave at the same distances as curves a, b, and c. (After Watt and Maxwell)

B. NOISE OF COSMIC ORIGIN

In the previous section we discussed the fact that the noise which prevails at the lower frequencies was due to lightning discharges and that the intensity decreased with increasing frequency. At about 20 MHz noise from outside the earth (cosmic noise) generally becomes more important than noise produced by lightning. As the frequency increases further, cosmic noise is the most important natural environmental noise source up to frequencies of about 200 MHz.

Cosmic noise may be broken down into several sources; noise coming from the center of our own galaxy; noise received from the sun; and noise from high intensity discrete sources like radio stars. Generally, these sources are highly directional, and as the earth turns some diurnal variation is evident. The bounds of this variation, that is the average daily upper and lower limits are shown in Figure 82 below. Cosmic noise sources generally subtend small solid angles so that a narrow antenna beam can usually be helpful in minimizing this source of noise.

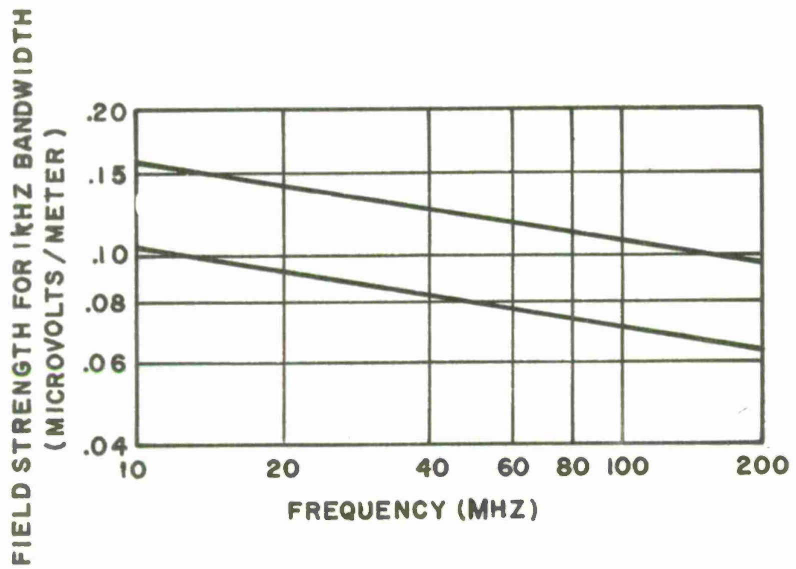


Figure 82. Average daily upper and lower limits of the normal cosmic radio noise field intensities. (After Cottony and Johler)

C. NOISE CAUSED BY ATMOSPHERIC THERMAL EMISSIONS

It is well known that the emissivity of a black body at a given frequency and in thermal equilibrium is equal to its absorptivity at the same frequency. Absorption by atmospheric constituents was discussed in the sections found on pages 85 through 107 where it was shown that atmospheric absorption becomes increasingly important as frequency increases beyond 100 MHz. Therefore, because atmospheric gases absorb, they also emit radiation. This radiation is called atmospheric thermal noise. It varies with temperature and frequency. For a standard atmospheric temperature distribution the variation with frequency as detected by directional antenna is shown in Figure 83.

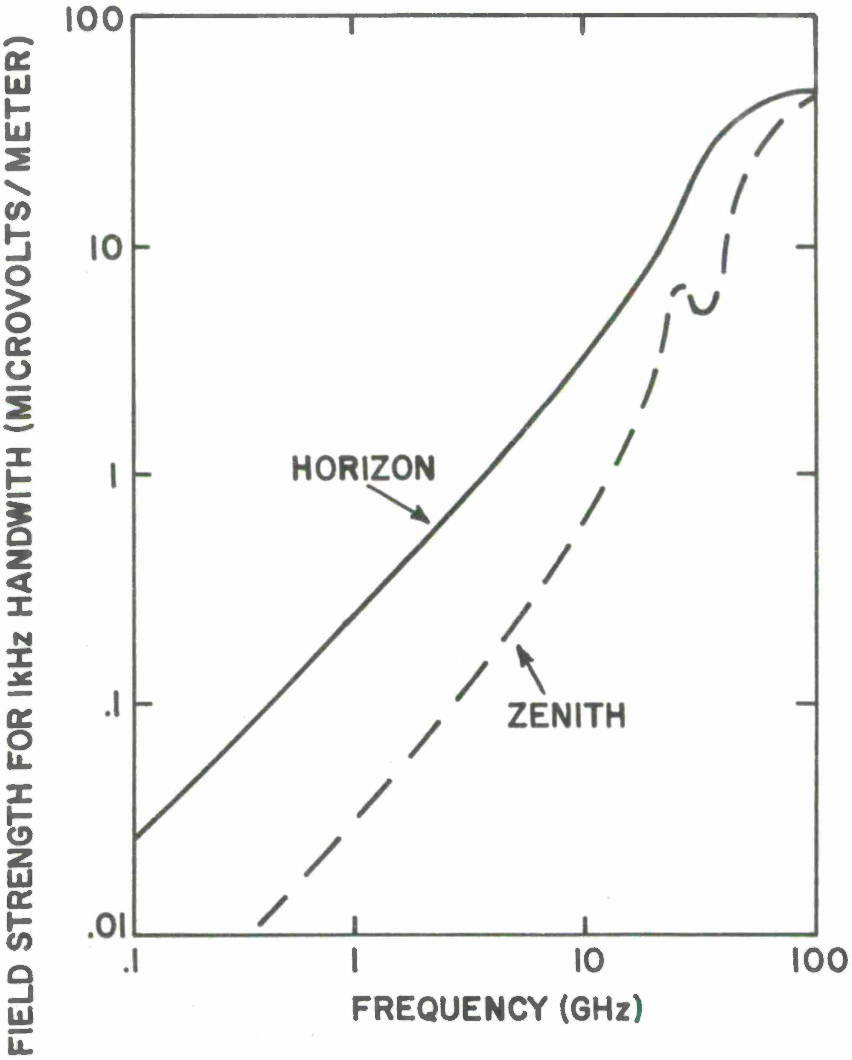


Figure 83. The variation of noise with frequency due to atmospheric thermal emissions. (After Wachowski)

D. NATURAL RADIO NOISE IN THE SEA

Figure 84 is a plot of the electromagnetic noise spectrum in the sea. Since attenuation at frequencies above 100 Hz is much too large to allow significant propagation, the spectrum shows only frequencies below this rate of oscillation.

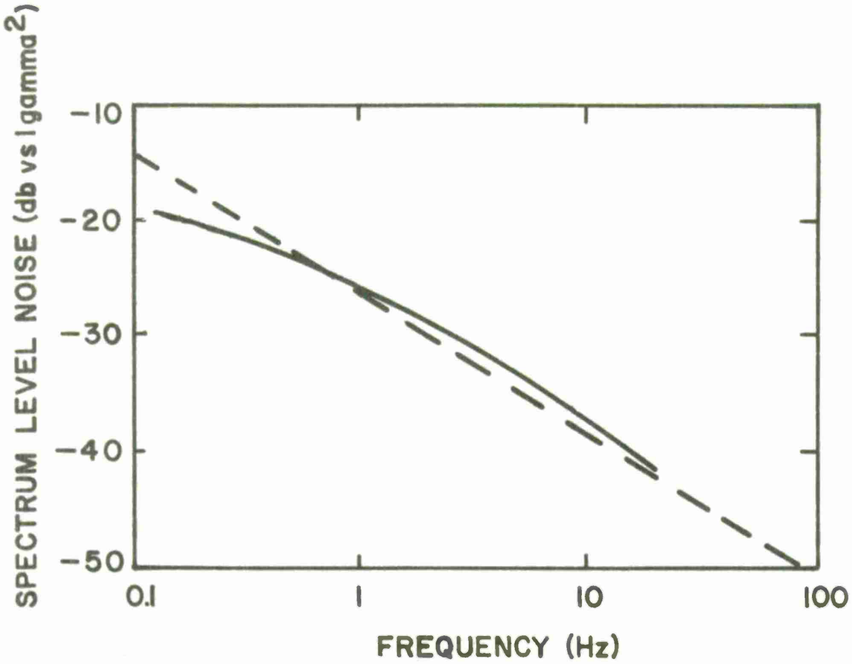


Figure 84. The electromagnetic noise spectrum in the sea. The solid line shows measured values, and the dashed line shows an approximation assuming noise varies as $f^{1.2}$. (After Hill)

Between 5 and 100 Hz, the source of the noise is predominantly lightning. As explained above this source exists somewhere on earth practically all the time, and the noise can travel great distances with little attenuation. Consequently this part of the noise spectrum is relatively constant.

Below 5 Hz lightning no longer contributes much energy to the oceanic noise spectrum. The origin of the noise between 5 and 0.001 Hz remains unknown, but some evidence exists linking it to sunspot activity. Besides the steady component, there are intermittent sinusoidal fluctuations of noise and also short bursts of noise.

The noise in the sea is latitude dependent, increasing toward the auroral zones. Further north, because of the nearly vertical geomagnetic field, induced electrical currents dampen the noise considerably.

BIBLIOGRAPHY

Ames, L A., J. W. Frazier, and A. S. Orange, Geological and geophysical considerations in radio propagation through the earth's crust, IEEE Trans. Ant. Prop., AP-11, (3), 1963.

Bailey, D. K., and R. Bateman, Radio transmission at VHF by scattering and other processes in the lower ionosphere, Proc. IRE, 43, (10), 1955.

Bean, B. R., B. A. Cahoon, C. A. Samson, and G. D. Thayer, A World Atlas of Atmospheric Radio Refractivity, ESSA Monograph 1, U. S. Gov. Print. Office, 1966.

Bean. B. R., and E. J. Dutton, Radio Meteorology, NBS Monograph 92, U. S. Gov. Print. Office, 1966.

Blackband, W. T., Propagation of Radio Waves at Frequencies Below 300 kc/s, MacMillan Company, New York, 1964.

Budden, K. G., The Waveguide Mode Theory of Wave Propagation, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1961.

Burrows, C. R., Radio communication within the earth's crust, IEEE Trans. Ant. Prop., AP-11, (3), 1963.

Canby, E. T., A History of Electricity, Hawthorne Books, New York, 1963.

C.C.I.R. Report 322, World Distribution and Characteristics of Atmospheric Radio Noise, International Telecommunication Union, Geneva, 1963.

Cottony, H. V., and J. R. Johler, Cosmic radio noise intensities in the VHF band, Proc., IRE, 40, (9), 1952.

Crane, R., Coherent pulse transmission through rain, IEEE Trans. Ant. Prop., AP-15, (2), 1967.

Davies, K., Ionospheric Radio Propagation, NBS Monograph 80, U. S. Gov. Print. Office, 1965.

Debye, P., Polar Molecules, Dover, New York, 1929.

Derr, V. E., Propagation of Millimeter and Submillimeter Waves, NASA CR 863, Nat. Aero. Sp. Adm., Washington, D. C., 1967.

duCastel, F., Tropospheric Radiowave Propagation Beyond the Horizon, Pergamon, New York, 1966.

Fischer, W. H., The radio noise spectrum from ELF to EHF, J. Atm. Terr. Phys., 27, (4) 1965.

Fischer W. H., (Correction to above reference.), J. Atm. Terr. Phys., 28, (4), 1966.

Freehafer, J. E., in Propagation of Short Radio Waves, (ed. Kerr), Radiation Laboratory Series, M.I.T., 13, McGraw-Hill, New York, 1951.

Gassman, G. J. (ed.), The Effect of Disturbances of Solar Origin on Communications, MacMillan Company, New York, 1963.

Goldstein, H., in Propagation of Short Radio Waves, (ed. Kerr), Radiation Laboratory Series, M.I.T., 13, McGraw-Hill, New York, 1951.

Goody, R. M., Atmospheric Radiation, I. Theoretical Basis, Oxford, University Press, London, 1964.

Hansen, R. E., Radiation and reception with buried and submerged antennas, IEEE Trans. Ant. Prop., AP-11, (3), 1963.

Herman, J. R., Reliability of atmospheric radio noise predictions, J. Res. NBS - D. Radio Propagation, 65D, (6), November - December 1961.

Herzberg, G., Molecular Spectra and Molecular Structure: I. Spectra of Diatomic Molecules, Van Nostrand, Princeton, New Jersey, 1950.

Hill, M. N., The Sea, 1, Interscience, New York, 1963.

Lawrence, R. S., C. G. Little, and H. J. A. Chivers, A survey of ionospheric effects upon earth-space radio propagation, Proc. IEEE, 52, (1), 1964.

Malone, T. F. (ed.), Compendium of Meteorology, American Meteorology Society, Boston, Mass., 1951.

Marion, J. B., Classical Electromagnetic Radiation, Academic Press, New York, 1965.

Meeks, M. L., Atmospheric emission and opacity at millimeter wavelengths due to oxygen, J. Geo. Res., 66, (11), 1961.

Mellen, G. L., et al, UHF long range communication system, Proc. IRE, 43, (10), 1955.

Menzel, D. H., Elementary Manual of Radio Propagation, Prentice-Hall Inc., New York, 1948.

Meyer, H. K., in Meteorological and Astronomical Influences on Radio Wave Propagation, NATO Conf. Series, (3), Pergamon Press, New York, 1963.

Mitra, S. K., The Upper Atmosphere, the Asiatic Society, Calcutta, 1952.

Moreland, W. B., Estimating Meteorological Effects on Radio Propagation, Technical Report 183, 1, AWS, USAF, Scott AFB, Illinois, 1965.

Mott, H., VLF propagation below the sea, IEEE Trans. Ant. Prop., AP-11, (3), 1963.

Nawrocki, P. J., and R. Papa, Atmospheric Processes, Prentice-Hall, Inc., New Jersey, 1963.

Ornstein, E., Attenuation of Electromagnetic Radiation by Sea Water, NRL Report 5280, U. S. Naval Res. Lab., Washington, D. C., 1959.

Petterssen, S., Weather Analysis and Forecasting, McGraw-Hill, New York, 1940.

Petterssen, S., Weather Analysis and Forecasting, II, McGraw-Hill, New York, 1956.

Radio Amateurs Handbook, American Rad. Relay League, Hartford, Connecticut, 1965.

Ratcliffe, J. A., Physics of the Upper Atmosphere, Academic Press, New York, 1960.

Riehl, H., Tropical Meteorology, McGraw-Hill, New York, 1954.

Saxton, J. A., (ed.), Meteorological Factors in Radio Wave Propagation. Report on a conference held by Phys. Soc. of London and Roy. Met. Soc. in April 1946, Phys. Soc., 1947.

Saxton, J. A., Monograph of Radio Wave Propagation in the Troposphere, Elsevier, 1962.

Saxton, J. A., (ed.), Advances in Radio Research, 1, Academic Press, London, 1964.

Scott, W. I., The Physics of Electricity and Magnetism, John Wiley & Sons, New York, 1959.

Slater, J. C., and N. H. Frank, Electromagnetism, McGraw-Hill, New York, 1947.

Stratton, J. A., Electromagnetic Theory, McGraw-Hill Book Company, 1941.

Terman, Frederick E., Electronic and Radio Engineering, McGraw-Hill, New York, 1955.

Townes, C. H., and A. L. Schawlow, Microwave Spectroscopy, McGraw-Hill, New York, 1955.

Valley, Shea L., Handbook of Geophysics and Space Environments, Air Force Cambridge Research Laboratory, Bedford, Mass., 1965.

VanDeHulst, H. C., Light Scattering by Small Particles, John Wiley & Sons, New York, 1957.

VanVleck, J. H., Propagation of Short Radio Waves,(ed. Kerr), 13, Rad. Lab. Series, M.I.T., McGraw-Hill, New York, 1951.

Vaughn, W. W., Distribution of Hydrometeors with Altitude for Missile Design and Performance Studies, Report DA-TM-138-59, Dept. of the Army, 1959.

Villard, O. G., et al, The role of meteors in extended range VHF propagation, Proc. IRE, 43, (10), 1955.

vonHippel, A. R., Dielectrics and Waves, John Wiley & Sons, Inc., New York, 1954.

Wachowski, H. M., Atmospheric Propagation Effects, Internal Report 1953.4-60 of the Aerospace Corporation, El Segundo, Calif., 1963.

Wait, J. R., The mode theory of VLF ionospheric propagation for finite ground conductivity, Proc. IRE, 45, (6), 1957.

Watt, A. D., and E. L. Maxwell, Characteristics of atmospheric noise from 1 to 100 kc/s, Proc. IRE, 45, (6), 1957.

Westman, H. D., (ed.), Reference Data for Radio Engineers, (fourth edition), International Tel. and Tel., American Book Stratford Press, Inc., New York, 1956.

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Technical Requirements and Standards Office Electronic Systems Division L G Hanscom Field, Bedford, Mass.	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP N/A

3. REPORT TITLE

A GENERAL ANALYSIS OF NATURAL ENVIRONMENTAL EFFECTS ON ELECTROMAGNETIC RADIATION UTILIZED FOR COMMUNICATIONS

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

None

5. AUTHOR(S) *(First name, middle initial, last name)*

Dr. Norman E. Gaut
Major Kenneth E. German

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
January 1968	148	60

8a. CONTRACT OR GRANT NO. IN-HOUSE b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) ESD-TR-67-631 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Electronic Systems Division Air Force Systems Command L. G. Hanscom Field, Bedford, Mass.

13. ABSTRACT

The earth and its atmosphere provide a highly complex environment for the propagation of electromagnetic waves. A convenient method to study the interplay between this environment and propagating electromagnetic signals is to isolate various quasi-homogeneous regions of the earth-atmosphere system, and study the effects they have on different frequency signals. The troposphere, the ionosphere, the oceans and the solid crust are such quasi-homogeneous subdivisions. This report presents a short review of the physical make-up of each of the subdivisions important to radio propagation, together with the most important effects each has upon the propagation of electromagnetic signals.

DD FORM 1 NOV 65 1473

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT