

THEORY OF IONOSPHERIC WAVES

Volume 17

K. C. Yeh &

C. H. Liu

Theory of Ionospheric Waves

This is Volume 17 in INTERNATIONAL GEOPHYSICS SERIES A series of monographs Edited by J. VAN MIEGHEM, Royal Belgian Meteorological Institute

A complete list of the books in this series appears at the end of this volume.

Theory of Ionospheric Waves

K. C. Yeh and C. H. Liu

DEPARTMENT OF ELECTRICAL ENGINEERING UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN URBANA, ILLINOIS



ACADEMIC PRESS New York and London

COPYRIGHT © 1972, BY ACADEMIC PRESS, INC. ALL RIGHTS RESERVED. NO PART OF THIS PUBLICATION MAY BE REPRODUCED OR TRANSMITTED IN ANY FORM OR BY ANY MEANS, ELECTRONIC OR MECHANICAL, INCLUDING PHOTOCOPY, RECORDING, OR ANY INFORMATION STORAGE AND RETRIEVAL SYSTEM, WITHOUT PERMISSION IN WRITING FROM THE PUBLISHER.

ACADEMIC PRESS, INC. 111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by ACADEMIC PRESS, INC. (LONDON) LTD. 24/28 Oval Road, London NW1

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 76-182661

PRINTED IN THE UNITED STATES OF AMERICA

To the memory of our Fathers

This page is intentionally left blank

1.	Introducti	on
	1.1	Nature of the Ionosphere
	1.2	Progress in the Study of Ionospheric Waves
	1.3	Scope of the Book
	1.4	Notations
		References
2.	Review o	f Electromagnetic Theory
	2.1	Maxwell's Equations
	2.2	Vector and Scalar Potentials
	2.3	Electric and Magnetic Polarizations
	2.4	Slow and Fast Processes
	2.5	Kramers-Kronig Relations
	2.6	Onsager Relation
	2.7	Plane Waves
	2.8	Refractive Indices
	2.9	Characteristic Polarizations
	2.10	Energy and Power
	2.11	Group and Energy Velocities
	2.12	Geometric Interpretation of Group Velocity 59
	2.13	Excitation of Fields
	2.14	Dyadic Green's Functions
		Problems
		References 83

3.	Waves in	Fluid Plasma	
	3.1	Introduction.	85
	3.2	Charge Neutrality	86
	3.3	Oscillation	89
	3.4	Screening	91
	3.5	Electron and Ion Plasma Waves	94
	3.6	Plasma Density Fluctuations	99
	3.7	Two-Stream Instability	103
	3.8	Interaction of Charged Particles with Longitudinal	
		Waves	109
	3.9	Excitation of Fields by a Test Particle	113
		Problems	123
		References	129
4.	Waves in	Fluid Plasma with a Steady Magnetic Field	
	4.1	Transverse Dielectric Constant and Index of Refraction	130
	4.2	Reflection of a Plane Transient Wave from the Plasma	
		Half-Space	135
	4.3	Signal Propagation in Lossless, Isotropic Plasma	137
	4.4	Gyrofrequency in the Ionosphere	148
	4.5	Dielectric Tensor of a Cold Magnetoplasma	149
	4.6	Effect of Collisional Loss and DC Conductivity	153
	4.7	Longitudinal Oscillations	155
	4.8	Refractive Indices and Polarizations	159
	4.9	Propagation Parallel to Steady Magnetic Field	165
	4.10	Faraday Effect	170
	4.11	Electron and Ion Whistlers	173
	4.12	Propagation Perpendicular to Steady Magnetic Field	176
	4.13	Hydromagnetic Waves-Low Frequency Approxima-	
		tion	179
	4.14	Appleton-Hartree Formula-High Frequency Approx-	
		imation	183
	4.15	Some Properties of the Appleton-Hartree Formula	188
	4.16	Cutoffs and Resonances in Parameter Space	193
	4.17	Index Circle and Index Surface	197
	4.18	Dielectric Tensor of a Warm Magnetoplasma	204
	4.19	Warm Plasma Correction to the High Frequency Waves	207

4.20	Plasma	Wa	ves	а	ind	1 '	Ти	/0-	St	rea	am	I	ns	tal	bil	iti	es			212
	Problem	ns.															•			215
	Referen	ces																		221

5. Wave Propagation in Inhomogeneous Media

5.1	Introduction.	223
5.2	Foundations of Geometrical Optics-Isotropic Media	224
5.3	Amplitude Variation along the Ray	228
5.4	Fermat's Principle	231
5.5	Ray Equations in Anisotropic Media	234
5.6	Effect of Boundary on the Ray and Generalized Snell's	
	Law	240
5.7	Reflection and Transmission of Waves at Sharp Bound-	
	aries	246
5.8	Wave Propagation in Stratified MediaIsotropic Case	250
5.9	The WKB Solution	252
5.10	The Matrix Method	255
5.11	The Stokes Phenomenon	260
5.12	An Example	267
5.13	Reflection Coefficients for Stratified Media-High	
	Frequency Approximation	270
5.14	Reflection Coefficients for Stratified Media-Very Low	
	Frequency Approximation	276
5.15	Signal Propagation and Reflection in Stratified Media	279
5.16	The True Height Problem—Ionosonde	281
5.17	Wave Propagation in Stratified Magnetoplasma-För-	
_	sterling's Coupled Equations	284
5.18	An Application of Försterling's Coupled Equations	287
5.19	Wave Propagation in Stratified Anisotropic Media-	••••
5 9 9	General Coupled Equations.	292
5.20	Application of the Coupled Equations Method to Wave	200
	Propagation in a Stratified Magnetoplasma	298
		305
		500

6. Wave Propagation in Random Media

6.1	Mathematical Background	308
6.2	Wave Propagation in Random Media	316

6.3	Scattering of Electromagnetic Wayes by Irregularities	317
6.4	Fluctuation of Electromagnetic Waves in Random	517
	Media-Geometrical Optics	329
6.5	Fluctuation of Electromagnetic Wayes in Random	
	Media—Wave Theory	333
6.6	Correlations of Fluctuations and Application to the	
	Ionosphere	343
6.7	Higher Order Approximations-Perturbation Tech-	
	niques	349
6.8	Effective Dielectric Tensor for Coherent Waves	358
	Problems	364
	References	365
Nonlinear	Wave Propagation	
7.1	Introduction.	367
7.2	Breaking of Waves	368
7.3	Nonlinear Effects in a Plasma in an Electromagnetic	
	Field	370
7.4	Self-Interaction of Waves	374
7.5	Cross-Modulation Phenomenon	379
7.6	Wave-Wave Interaction	382
7.7	An Averaged Variational Principle	392
	Problems	399
	References	400
Interactio	n of Atmospheric Waves with the Ionosphere	
8.1	Structure of the Atmosphere	403
8.2	Buoyancy Oscillations	407
8.3	Acoustic Gravity Waves in an Isothermal Atmosphere	409
8.4	Properties of Internal Waves	413
8.5	Propagation in a Wind-Stratified Isothermal Atmo-	
	sphere	419
8.6	Effect of Ion Drag	423
8.7	Attenuation due to Thermal Conduction and Viscosity	429
8.8	Effect of Internal Waves in the Ionospheric F Region	430
8.9	Impulse Response of an Isothermal Atmosphere	432
8.10	Effect of a Wind Shear in the Ionospheric E Region	436
	Problems	439
	References	441

7.

8.

Appendix A.	The Method of Steepest Descents	
A.1	The Method of Steepest Descents	443
A.2	Modified Method of Saddle Point	449
	References	450
Appendix B.	The Distant Radiation Field from a Localized Source	451
	References	458
SUBJECT INDE	x	459

This page is intentionally left blank

Preface

This book is an outgrowth of material covered in a graduate course on ionospheric radio propagation offered since 1959 by the Department of Electrical Engineering, University of Illinois at Urbana-Champaign. Some material was also used in a special problems course on propagation in inhomogeneous media. The book is intended primarily as a graduate text for students in radio science, plasma physics, wave propagation, ionospheric physics, and atmospheric science. As such, theories and ideas are stressed more than engineering practice. Problems have been included to illustrate and amplify these theories and ideas. Some of these problems have been taken from the published literature and are by no means trivial; some others are fairly simple. We anticipate that the book may be of interest as well to researchers in the field. For their benefit, a list of references are attached at the end of each chapter; this list is far from being complete. We wish to apologize to those authors upon whose work we may have drawn but neglected to acknowledge because of oversight.

The reader is assumed to have background in elementary electromagnetic theory and complex analysis. Some understanding of matrix theory is helpful. An understanding of time series is desirable for Chapter 6, but not necessary.

In the process of writing this book we have received assistance from many people. All the graduate students who took our course over the years have been a constant source of inspiration to us. It is fair to say that we have learned as much from them as they have from us. We would also like to thank all of the researchers who have contributed to the field. At times, keeping abreast of developments has been frustrating; there are always unexpected turns before we are able to see the light at the end of the dark tunnel. Many colleagues have provided us with encouragement. Special mention should be made of Professors G. W. Swenson, Jr., and G. A.

Preface

Deschamps of the University of Illinois at Urbana-Champaign, and Professor O. G. Villard, Jr., of Stanford University. We would like to take this opportunity to thank our wives who have provided us with their loving support. Portions of the manuscript were typed by Mrs. Marilena Stone, Audrey Owens, and Diana Poole.

xiv

1. Introduction

A wave can be broadly defined as a phenomenon whereby the spatial distribution of energy travels from one point of the space to another point of the space. As such, the wave can be thought of as a means by which energy and information can be transmitted from a source to an observer and user. This definition of wave is more restrictive than the usual concept of wave since it excludes a class of waves known as the standing waves. The restriction in definition does not necessarily indicate a handicap however since, for example, standing waves can be decomposed into traveling waves in a linear medium. Figuratively speaking, waves can be likened to "fingers" since both can "stretch out" and "feel" the environment. In this sense a wave can be considered as a device which is capable of remote sensing. Indeed, man's knowledge about the ionosphere was derived almost exclusively up to the 1950's from probing with radio waves on the ground. These wave-probing techniques are still being used not only in connection with ionospheric studies but also in the laboratory plasma diagnosis.

According to the IEEE Standard (1969) the ionosphere is defined as "that part of a planetary atmosphere where ions and electrons are present in quantities sufficient to affect the propagation of radio waves." Therefore, the ionosphere is defined from the viewpoint of its effect on radio waves. This definition is quite adequate for this book and shall be adopted. However, the same standard defines the ionospheric wave as the sky wave. We prefer to adopt a broader definition and define ionospheric waves as those wave motions, natural or driven, that can be sustained in the ionosphere. This broader definition would include waves that are reflected, refracted,

1. Introduction

scattered, and guided as well as those forced into oscillation by a source such as excitation of traveling ionospheric disturbances by the propagating internal acoustic-gravity waves.

1.1 Nature of the Ionosphere

The ionosphere is a part of the upper atmosphere. To discuss it, we must start with a brief discussion of the properties of the upper atmosphere. Workers in the atmospheric studies like to use words such as zones, belts, spheres, and regions to classify the physical space of interest. This is necessary because of the large variation of parameters involved. In terms of thermal structure the atmosphere is classified into, in the order of increasing altitude, troposphere, tropopause, stratosphere, stratopause, mesosphere, mesopause, and thermosphere. In terms of composition it is classified into homosphere (up to 85 km height) in which the mean molecular mass is constant due to mixing and heterosphere (above 85 km) in which the mean molecular mass varies due to diffusive separation. In terms of the escape properties of neutral particles, the atmosphere is called the exosphere. A brief discussion on the structure of the atmosphere can be found in Section 8.1. More detailed treatment can be found in Ratcliffe (1960) and Hines *et al.* (1965).

The atmospheric mass density varies over large orders of magnitude. For example, the mass density is 1.2 kg/m³ at the ground level and this value is reduced to 4.1×10^{-1} kg/m³ at a height of 10 km, to 5.0×10^{-7} kg/m³ at 100 km height, to 3.6×10^{-7} kg/m³ at 300 km height, and to 1.5×10^{-13} kg/m³ at 700 km height. Therefore, in a height range of 700 km the density is reduced by thirteen orders of magnitude. This reduction is caused mainly by gravity. Its static distribution can be discussed by solving the hydrostatic equation if the temperature profile is known.

When the solar radiation falls on the atmosphere, there may be creation of electrons and ions through the photoionization process. The rate of production of electron-ion pair per unit volume is given by

$$Q = \sigma n S \tag{1.1.1}$$

where σ is the effective ionization cross section of a given constituent gas, *n* its number density, and S the local intensity of the ionizing solar radiation. The solar radiation has the highest intensity when it is outside the atmosphere. When the radiation penetrates into the atmosphere, the rate of production increases due to increasing n. When the number density is increased further the absorption of radiation by the overlaying atmosphere may become appreciable and S starts to decrease. The rate of production, being proportional to the product of n and S, must therefore exhibit a peak. This problem was first worked out in a classical paper by Chapman (1931) in an isothermal atmosphere. If the atmosphere is a nonisothermal mixture, and the radiation spectrum is not monochromatic, the computation of Qcan be very tedious and is usually carried out on an electronic computer. Once created, the ionization is subjected to control by many physical and chemical processes. The ionization may be transported due to collision with other moving particles or due to the presence of electric field. Ionization of a particular kind may appear or disappear through chemical processes. In general, for each ionization species there is an equation of continuity of the form

$$\partial N/\partial t + \nabla \cdot (N\mathbf{v}) = Q - L$$
 (1.1.2)

where N is the ionization density, \mathbf{v} its velocity, Q the rate of production per unit volume, and L the rate of loss per unit volume through chemical processes. Rishbeth and Garriott (1969) have discussed in detail the solution of such an equation.

The terrestrial ionosphere is usually classified into three regions. The lowest region is called the D region which extends in height from about 40 km to 90 km. This region is responsible for absorption of radio waves. Its electron density is about $2.5 \times 10^9/m^3$ by day and diminishes to negligible value at night. The middle region is called the *E* region. It is the region of the ionosphere between about 90 km and 160 km altitude. The electron density in this region behaves regularly so far as its dependence on the solar zenith angle and the solar activity are concerned. The density can have a value 2×10^{11} /m³ in the daytime and this value is high enough to reflect radio waves with a frequency of several megahertz. At night the E region electron density is more than one order of magnitude lower. Above the E region is the F region. The F region behavior is fairly irregular and is usually classified into a number of anomalies such as equatorial anomaly and seasonal anomaly. The electron density at the peak has an average value 2×10^{12} /m³ by day and 2×10^{11} /m³ by night. This region is responsible for reflection of radio waves. In some literature the F region stops at about 800 km above which is called either a protonosphere if classified according to its composition or a magnetosphere if classified according to its dynamic property. But according to the IEEE definition, the F region extends all the way to the magnetospheric shock boundary at several earth radii away. Two idealized ionization profiles, one for daytime and one for nighttime, are shown in Fig. 1.1-1.

The ionosphere is an extremely interesting medium for purposes of studying wave propagation. As shown in Fig. 1.1-1, the electron density varies over four orders of magnitude. In going from the *D* region to the upper *F* region, the medium changes from a collision-dominated behavior (electron collision frequency 8×10^7 /sec) to a collisionless behavior so that the frozen-in magnetic field concept is applicable. Additionally, the medium is inhomogeneous and anisotropic. Its nonlinear properties can also be revealed easily; the outstanding example is the well-known cross-modulation or Luxembourg effect.

Historically, Heaviside and Kennelly postulated the existence of a "conducting" layer in 1902 to explain the reflection of radio waves. Such a layer was called, at that time, the Kennelly-Heaviside layer. The first demonstration that proved the existence of the ionosphere was carried out by Appleton



Fig. 1.1-1. Idealized ionization profiles of the terrestrial ionosphere at temperate latitudes near sunspot maximum.

1.2 Progress in the Study of Ionospheric Waves

and Barnett in 1925. They used the method of interference between the ground wave and the sky wave. Their conclusion was verified the next year by Breit and Tuve by measuring the time of flight of a radio pulse reflected vertically from the ionosphere. This technique is in principle retained even at present and becomes one of the most powerful tools in modern ionospheric investigation.

1.2 Progress in the Study of Ionospheric Waves

Like many fields, the progress in the study of ionospheric waves has been uneven. As is often the case, a massive effort in one field speeds up the progress in many related fields. Ionospheric research is no exception.

The chief interest in ionospheric research in the 1930's was motivated by the capability of the ionosphere to reflect radio waves. The existence of the ionosphere extended the frequency to the short wave spectrum and changed the mode of propagation from ground waves to sky waves. It was soon realized that the earth's magnetic field played an important role in influencing wave propagation and the magnetoionic theory was developed. It is interesting to note that Appleton developed the theory by using Lorentz's results (1915) applicable to light propagation in solids, and generalized it to arbitrary propagation angle with respect to the magnetic field. Several other fields also helped. The work in electric discharges showed how the transport coefficients can be computed (Loeb, 1961). The work in astrophysics showed how the guiding center approach can be used to describe the dynamic properties of an ionized gas (Alfvén and Fälthammer, 1963).

In addition to electric discharges and astrophysics, the recent massive efforts in space research and thermonuclear containment problems have helped ionospheric research directly or indirectly. Through these efforts, plasma physics has become a fullblown discipline. A chronology of plasma physics has been prepared by Allis (1962) and he went back as early as 1733. As if to complete the circle started by Lorentz, these processes have assisted and have now extended back to the solid-state plasma. A non-mathematical review on this subject can be found in Chynoweth and Buchsbaum (1965).

Steady progress has been made in the study of ionospheric waves since the development of the magnetoionic theory. In addition to the obvious interest in connection with the long range communication, research in ionospheric waves has long been proven to be very fruitful in gaining information about the upper atmosphere. More sophisticated mathematical

1. Introduction

techniques in wave theory and better understanding of the properties of waves in plasma from plasma research have helped in this area. In the last decade or so, more and more evidence pointed to the fact that the dynamics of the neutral atmosphere is coupled closely to that of the ionosphere. Ionospheric waves play an important role in this relationship both as a participant in the coupling process and as a probing tool in the observation. This adds a new aspect to the study of ionospheric waves.

1.3 Scope of the Book

The book starts in Chapter 2 with a review on electromagnetics, especially those aspects essential in a nonisotropic medium. The discussion is centered on the dielectric constant, although parallel treatment on the permeability can be done similarly. These discussions are approached from a fairly general point of view and do not have any specific medium in mind. Several concepts in temporal and spatial dispersion are brought out, depending on the time scale and spatial scale of the electromagnetic process. Restrictions on the form of dielectric tensor due to symmetry properties, causality, passitivity are elucidated next. It is then seen that in a uniform anisotropic medium, there exist characteristic waves which can propagate without change in their states of wave polarization. Energy considerations show that wave energy propagates with the group velocity in a transparant medium. This velocity can be interpreted geometrically by relating it to the refractive index surface. Such an interpretation is useful in evaluating asymptotic Green's functions which have applications on radiation and scattering problems. Even though the discussion in this chapter is centered on the electromagnetic waves, the mathematical techniques and the interpretations can be equally applied to other waves. As a matter of fact, some of the points were first developed by fluid dynamists in studying magnetohydrodynamic waves, ocean waves, and atmospheric waves.

In exploring the ionospheric waves in subsequent chapters, we can find that there are essentially five areas: waves in plasmas, inhomogeneous media, random media, nonlinear media, and interaction with atmospheric waves. Because of the possible existence of many kinds of particles, a plasma may have a large degree of freedom. Consequently, as a medium, the plasma is extremely rich in sustaining wave motions. These are discussed in Chapters 3 and 4. When the medium becomes inhomogeneous, as in the ionosphere, the discussion on wave propagation depends on our ability to solve the partial differential equations with variable coefficients. In the high

1.4 Notations

frequency limit the ray concept can be used and the medium may be said to be locally homogeneous. When this is not the case—the phenomena of reflection—coupling can be treated most easily by choosing a horizontally stratified model.

When the medium has many irregularities, the structure becomes extremely complicated. In this case it is convenient to introduce a statistical approach and treat the medium as a random function of position. Phenomena discussed include scattering, fluctuations of amplitude and phase, and propagation of coherent waves. As is well known, the nonlinear wave is extremely complicated to treat. Several problems discussed in Chapter 7 are wave breaking, cross modulation, and wave-wave interaction. The last chapter starts with the propagation of acoustic-gravity waves. The interaction of these waves with the ionosphere is treated next.

Two appendixes are included at the end of the book. These appendixes deal with techniques of asymptotic expansion. In Appendix A, the method of steepest descents is discussed briefly. Appendix B is devoted to the topic of asymptotic evaluation of radiation field due to localized source. The results are used at various places throughout the book.

It should be remarked that throughout the book the boundary effect on wave propagation is not emphasized. This is so because we wish to confine our attention mainly to ionospheric applications. We also do not bring in the propagation of discontinuity and shock since it seems to be more properly a province of fluid dynamics.

1.4 Notations

In concluding this chapter we wish to say a few words about notations.

We have found it convenient to use dyadic notations for several reasons. In the more elementary texts on electromagnetic theory vector notation is almost universally adopted. It seems to us that dyad is a natural extension of a vector. Fortunately, we usually stop at dyadics, since triadics and tetradics are seldom needed. A dyad is nothing but a Cartesian tensor of rank two. Let \hat{x}_i , i = 1, 2, 3, be three unit vectors, mutually orthogonal so that $\hat{x}_1 \times \hat{x}_2 = \hat{x}_3$. For any two vectors $\mathbf{A} = A_i \hat{x}_i$, $\mathbf{B} = B_i \hat{x}_i$ where repeated indices are summed from 1 to 3, the scalar and vector products are defined,

$$\mathbf{A} \cdot \mathbf{B} = A_i B_i$$
$$(\mathbf{A} \times \mathbf{B})_i = \varepsilon_{ijk} A_j A_k$$

1. Introduction

where ε_{ijk} is 1 if *i*, *j*, *k* are in cyclic order 1, 2, 3; -1 if *i*, *j*, *k* are in anticyclic order 1, 2, 3; and 0 if *i*, *j*, *k* are not all different. Let **P** be a dyad; then

$$\mathbf{P} = P_{ij} \hat{x}_i \hat{x}_j$$

We note that a unit dyad is

$$\mathbf{I} = \delta_{ij} \hat{x}_i \hat{x}_j$$

The dot product between two dyads or between a dyad and a vector can be defined by natural extension of vector dot product. For example,

$$\mathbf{P} \cdot \mathbf{A} = P_{ij}A_j\hat{x}_i$$
$$\mathbf{A} \cdot \mathbf{P} = A_iP_{ij}\hat{x}_j$$
$$\mathbf{P} \cdot \mathbf{R} = P_{ij}R_{jk}\hat{x}_i\hat{x}_k$$
$$\mathbf{I} \cdot \mathbf{P} = \mathbf{P} \cdot \mathbf{I} = \mathbf{P}$$

The cross product can be defined similarly. Note the parallel of dot products with matrix multiplication. In matrix notation if \mathbf{P} is a 3 \times 3 matrix and \mathbf{A} a three-dimensional column vector, the post multiplication of \mathbf{P} by \mathbf{A} is \mathbf{PA} , but in order to make matrix multiplication conformable the premultiplication of \mathbf{P} by \mathbf{A} requires \mathbf{A} to be transposed, i.e., $\mathbf{A}^T \mathbf{P}$. In dyadic notation it is not necessary to take transposition. Operations involving dyad can be defined by writing

$$\mathbf{P} = \mathbf{P}_i \hat{x}_i$$

where $\mathbf{P}_i = P_{ij}\hat{x}_j$ and is a vector. Then

$$\nabla \cdot \mathbf{P} = (\nabla \cdot \mathbf{P}_i)\hat{x}_i$$
$$\nabla \times \mathbf{P} = (\nabla \times \mathbf{P}_i)\hat{x}_i$$

A full discussion of dyadic analysis can be found in Morse and Feshback (1953).

In order to avoid flooding the text with an unnecessary number of notations, we find it convenient to use the argument to distinguish the transform pair. A Fourier transform pair is denoted by the same symbol but distinguished by their arguments. For example, $F(\mathbf{r}, \omega)$ is the Fourier transform in time of $F(\mathbf{r}, t)$ and $F(\mathbf{k}, \omega)$ is the Fourier transform in space of $F(\mathbf{r}, \omega)$. For an

References

electromagnetic process starting at t = 0, we have

$$\mathbf{F}(\mathbf{k},\omega) = \int_{-\infty}^{\infty} \int_{0}^{\infty} \mathbf{F}(\mathbf{r},t) e^{-j(\omega t - \mathbf{k} \cdot \mathbf{r})} dt d\mathbf{r}$$

and its inversion

$$\mathbf{F}(\mathbf{r},t) = \frac{1}{(2\pi)^4} \int_{-\infty}^{\infty} \int_{c} \mathbf{F}(\mathbf{k},\omega) e^{j(\omega t - \mathbf{k} \cdot \mathbf{r})} \, d\omega \, d\mathbf{k}$$

The contour c is sometimes referred to as the Laplace contour and is parallel to the real ω -axis and below all singularities of $F(\mathbf{k}, \omega)$ in ω -space.

References

- ALFVÉN, H., and FÄLTHAMMER, C. G. (1963). "Cosmical Electrodynamics," 2nd. ed. Oxford Univ. Press, London and New York.
- ALLIS, W. P. (1962). A chronology of plasma physics, Phys. Today 15, No. 12, 12.
- ALLIS, W. P., BUCHSBAUM, S. J., and BERS A. (1963). "Waves in Anisotropic Plasmas." MIT Press, Cambridge, Massachusetts.
- APPLETON, E. V., and BARNETT, M. A. F. (1925). On some direct evidence for downward atmospheric reflection of electric rays, *Proc. Roy. Soc.* 109A, 621-641.
- BREIT, G. and TUVE, M. A. (1926). A test of the existence of the conducting layer, *Phys. Rev.* 28, 554.
- BUDDEN, K. G. (1961). "Radio Waves in the Ionosphere." Cambridge Univ. Press, London and New York.
- CHAPMAN, S. (1931). The absorption and dissociative of ionizing effect of monochromatic radiation in an atmosphere on a rotating earth, Part I; Part II; Grazing incidence, *Proc. Phys. Soc. London* 43, 26, 484.
- CHYNOWETH, A. G., and BUCHSBAUM, S. J. (1965). Solid-state plasma, *Phys. Today* 18, No. 11, 26-37.
- GINZBURG, V. L. (1964). "Propagation of Electromagnetic Waves in Plasmas," Pergamon, Oxford.
- HEAVISIDE, O. (1902) Telegraphy, I. Theory. *Encyclopedia Brittannica*, 9th ed., Vol. 33, p. 215.
- HINES, C. O., PAGHIS, I., HARTZ, T. R., and FEJER, J. A. ed. (1965). "Physics of the Earth's Upper Atmosphere," Prentice-Hall, Englewood Cliffs, New Jersey.
- IEEE Stand., Definitions of Terms for Radio Wave Propagation No. 211, January 1969.
- KELSO, J. M. (1964). "Radio Ray Propagation in the Ionosphere," McGraw-Hill, New York.
- KENNELLY, A. E. (1902). On the Elevation of the Electrically Conducting Strata of the Earth's Atmosphere, *Elect. World and Eng.* **39**, 473.
- LEONTOVICH, M. A. (1967). "Reviews of Plasma Physics," Vols. 1–4, Consultants Bureau, New York.
- LOEB, L. B. (1961). "The Kinetic Theory of Gases," 3rd ed. Dover, New York.
- LORENTZ, L. A. (1915). "The Theory of Electrons," 2nd ed. Dover, New York.

- MITRA, S. K. (1952). "The Upper Atmosphere," 2nd ed. Asiatic Soc., Calcutta.
- MORSE, P. M., and FESHBACK, H. (1953) "Methods of Theoretical Physics," McGraw-Hill, New York.
- RATCLIFFE, J. A. (1959). "The Magneto-Ionic Theory and Its Application to the Ionosphere," Cambridge Univ. Press, London and New York.
- RATCLIFFE, J. A., ed. (1960). "Physics of the Upper Atmosphere," Academic Press, New York.
- RAWER, K., and SUCHY, K. (1967). Radio-observations of the ionosphere, *In* "Handbuch der Physik" (S. Flügge, ed.), Vol. 69/2, "Geophysics," 3/2, pp. 1-546. Springer-Verlag, Berlin and New York.
- RISHBETH, H., and GARRIOTT, O. K., (1969). "Introduction to Ionospheric Physics," Academic Press, New York.

STIX, T. H. (1962). "The Theory of Plasma Waves," McGraw-Hill, New York.

2. Review of Electromagnetic Theory

In this chapter we review electromagnetic theory, especially those aspects that deal with propagation in an anisotropic medium. The presence of the electromagnetic field in the medium is to induce polarization charges and magnetization currents. If the electromagnetic process is slow in time, the medium can respond instantaneously. However, if the process is fast, the response is no longer instantaneous and in such a case, the medium is said to possess temporal (or frequency) dispersion. Similarly the absence or presence of spatial dispersion depends on whether the thermal effects are of importance. These different ways of classifying the medium are discussed in fairly general terms in this chapter. We also note that the medium may possess physical properties such as causality and passiveness. The presence of these physical properties means that the Hermitian and anti-Hermitian part of the dielectric tensor are not arbitrary, they must be related through the Kramers-Kronig relations. Further, irreversible thermodynamic considerations show that the dielectric tensor must possess certain symmetry properties known as the Onsager relation which implies that only six of the nine elements in the dielectric tensor are independent.

The propagation of plane waves in a general anisotropic medium is discussed next. It is found that only those waves called characteristic waves can propagate without change of wave polarization. Energy considerations show that in a lossless medium the group velocity can be interpreted as the velocity of energy flow. The group velocity is also seen geometrically related to the refractive index surface. The geometric relation is further amplified when excitation of radiation field is considered.

No specific medium is discussed in this chapter. The emphasis is on the general techniques. The basic idea in the approach adopted in this book is that the medium is described macroscopically by the dielectric tensor. The excitation and propagation of waves in this medium are then discussed accordingly.

2.1 Maxwell's Equations

We shall accept Maxwell's equations as a set of physical laws which govern the behavior of the macroscopic electromagnetic fields. Historically these laws were established from a consistent induction of a vast wealth of experiments with notable contribution from Coulomb, Ampere, Faraday, and many others. The reader is supposed to be familiar with such an approach from a more elementary course.

The physical quantities dealt with in this macroscopic theory are smoothed over a volume in which the density function is meaningful. This is necessary because of the discrete nature of the charged particles.

Maxwell's equations deal with the interdependence of four field vectors: **E** (electric field vector), **H** (magnetic field vector), **D** (electric flux density), and **B** (magnetic flux density), and their dependence on the sources. In the limit of static case the electric fields and the magnetic fields become decoupled, the electric charge is the source only of the electric field, and the current is the source only of the magnetic field. This is no longer the case when the sources are time dependent. The time-dependent electric field and magnetic field become coupled and it is more conventional to speak of electromagnetic fields.

In the international system of units Maxwell's equations are written as

$$\nabla \mathbf{X} \mathbf{E} = -\mathbf{\dot{B}} \tag{2.1.1a}$$

$$\nabla \mathbf{X} \mathbf{H} = \mathbf{J} + \dot{\mathbf{D}}$$
(2.1.1b)

$$\nabla \cdot \mathbf{D} = \varrho \tag{2.1.1c}$$

 $\nabla \cdot \mathbf{B} = 0 \tag{2.1.1d}$

where an overdot is used to denote the partial differentiation in time, $\partial/\partial t$. The first equation is the Faraday's induction law which states that electric field may be induced by the time-varying magnetic field. The second

equation was originally due to Ampere but generalized to the time-varying case by Maxwell by adding the displacement current, $\dot{\mathbf{D}}$; it expresses the fact that the source of the magnetic field is current and that the magnetic field is coupled back to the electric field. The third equation is known as Gauss's theorem and states that electric flux starts and ends on charges. The fourth equation shows that magnetic flux lines are closed.

Since the current is manifested by the motion of charged particles it may be expected that J and ϱ must be related in some way. In fact that this is so may be seen by taking the divergence of Eq. (2.1.1b) and making the substitution of Eq. (2.1.1c). The resulting equation is

$$\partial \varrho / \partial t + \nabla \cdot \mathbf{J} = 0 \tag{2.1.2}$$

which is known as the equation of continuity. The physical meaning of Eq. (2.1.2) is quite clear. It says that the time rate of increase of charge density at any point must be equal to the inward flow of current density to the same point, which is consistent with the law of conservation of charge.

Actually the current and charge appearing in Eqs. (2.1.1) deserve some comment. A more thorough discussion on them can be found in a later section. For the moment it is sufficient to say that the current in Eq. (2.1.1b) includes the externally applied current, the conduction current (if the medium is a conductor), and the convection current. The charge in Eq. (2.1.1c) includes the externally applied charge and the free charge. In a material medium there may be other currents and charges.

Now, Maxwell's equations (2.1.1a) through (2.1.1d) are equivalent (at most) to eight equations for twelve scalar unknowns even if the current and charge are assumed given quantities. Other relations between the fields are needed if the system of equations is to be determinate. These subsidiary relations are called material equations and they are supposed to describe completely the behavior of the matter in the electromagnetic field. In free space the material equations take the simplest forms

$$\mathbf{D} = \varepsilon_0 \mathbf{E}, \qquad \mathbf{B} = \mu_0 \mathbf{H} \tag{2.1.3}$$

where the dielectric constant and the permeability of free space are given, respectively, by

$$\varepsilon_0 = (\mu_0 c^2)^{-1} \quad \text{F/m}$$

 $\mu_0 = 4\pi \times 10^{-7} \quad \text{H/m}$
(2.1.4)

The velocity of light in free space appearing in the first equation of (2.1.4) has been measured to a high degree of accuracy. The most recent value

adopted by URSI is

$$c = 299,792.5 \pm 0.4$$
 km/sec
 $\simeq 3 \times 10^8$ m/sec (2.1.5)

Substituting the approximate value of c into the first equation of (2.1.4) gives

$$\varepsilon_0 \simeq (1/36\pi) \times 10^{-9} \,\mathrm{F/m}$$
 (2.1.6)

to a high degree of approximation.

2.2 Vector and Scalar Potentials

As mentioned in the last section Maxwell's equations alone are not enough to solve for all the unknown quantities, and there is need of subsidiary relations. In free space these relations are given by (2.1.3). The two vector equations in (2.1.3) are equivalent to six scalar equations. When added to the eight equations due to Maxwell we have a total of fourteen equations for twelve scalar unknowns. Therefore, it seems that we may be overspecifying our system. In order to resolve this question let us suppose that the fields are excited by the given charge and current sources in free space. For this case Maxwell's equations take the form

$$\nabla \mathbf{X} \mathbf{E} + \dot{\mathbf{B}} = 0 \tag{2.2.1a}$$

$$(\nabla/\mu_0) \times \mathbf{B} - \varepsilon_0 \dot{\mathbf{E}} = \mathbf{J} \tag{2.2.1b}$$

 $\nabla \cdot \mathbf{E} = \varrho/\varepsilon_0 \tag{2.2.1c}$

$$\nabla \cdot \mathbf{B} = 0 \tag{2.2.1d}$$

By vector identity, the divergence of curl of a vector is identically zero. If we take divergence of the entire equation (2.2.1a) and apply this vector identity, we obtain

$$(\partial/\partial t)\nabla \cdot \mathbf{B} = 0 \tag{2.2.2}$$

Integration of (2.2.2) with respect to time tells us that the divergence of **B** is a temporal constant which, according to (2.2.1d), must be zero. Therefore, Maxwell's equation (2.2.1d) can be viewed as the initial condition on div **B**, i.e., if the magnetic fields is initially solenoidal, it must be so for all times as indicated by (2.2.2). Similarly, by taking the divergence of (2.2.1b)

14

and applying the equation of continuity (2.1.2), we obtain

$$(\partial/\partial t)(\nabla \cdot \mathbf{E} - \varrho/\varepsilon_0) = 0 \tag{2.2.3}$$

In the same vein, Gauss's theorem (2.2.1c) can also be viewed as the initial condition of (2.2.3).

Now we wish to relate the field directly to sources. The simplest approach makes use of vector and scalar potentials. We proceed in the following.

The magnetic field is seen to be solenoidal by (2.2.1d). There then must exist a vector potential **A** whose curl is **B**. That is

$$\mathbf{B} = \mathbf{\nabla} \times \mathbf{A} \tag{2.2.4}$$

Substitution of (2.2.4) in (2.2.1a) tells us that $\mathbf{E} + \dot{\mathbf{A}}$ is irrotational so that we can define a potential function whose negative gradient is equal to this vector, i.e.,

$$\mathbf{E} = -\nabla V - \dot{\mathbf{A}} \tag{2.2.5}$$

The fields when defined by potentials through (2.2.4) and (2.2.5) satisfy the homogeneous equations (2.2.1a) and (2.2.1d) automatically. However, we note that **B** and **E** are unchanged by the transformations

where ψ is an arbitrary function of coordinates and time. The transformation (2.2.6) is known as the gauge transformation and the invariance of fields under such a transformation is called gauge invariance. Evidently **B** and **E** defined by (2.2.4) and (2.2.5) are gauge invariant.

So far we have made use only of the two homogeneous Maxwell's equations (2.2.1a) and (2.2.1d). Substitution of (2.2.4) and (2.2.5) in the inhomogeneous Maxwell's equations (2.2.1b) and (2.2.1c) gives

$$\nabla^{2}\mathbf{A} - \ddot{\mathbf{A}}/c^{2} - \nabla(\nabla \cdot \mathbf{A} + \dot{V}/c^{2}) = -\mu_{0}\mathbf{J} \qquad (2.2.7)$$

$$\nabla^2 V - \ddot{V}/c^2 + (\partial/\partial t)(\nabla \cdot \mathbf{A} + \dot{V}/c^2) = -\varrho/\varepsilon_0 \qquad (2.2.8)$$

where the vector identity $\nabla \times \nabla \times \mathbf{A} = \nabla (\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$ has been used. The velocity of light in free space is related to the dielectric constant and the permeability through (2.1.4). According to Helmholtz's theorem of vector analysis any vector field which is finite, uniform, and continuous and which vanishes at infinity, is uniquely specified only when both its curl and its divergence are known. The curl of **A** is defined by (2.2.4). But the divergence of A is yet undefined. Inspection of (2.2.7) and (2.2.8) indicates that it is convenient to define it by a relation known as the Lorentz condition

$$\nabla \cdot \mathbf{A} + \dot{V}/c^2 = 0 \tag{2.2.9}$$

Introduction of the Lorentz condition in (2.2.7) and (2.2.8) reduces the equations for vector potential and scalar potential to a symmetric form

$$\nabla^2 \mathbf{A} - \ddot{\mathbf{A}}/c^2 = \mu_0 \mathbf{J} \tag{2.2.10}$$

$$\nabla^2 V - \ddot{V}/c^2 = -\varrho/\varepsilon_0 \tag{2.2.11}$$

This condition also requires that the gauge in (2.2.6) satisfy the homogeneous wave equation

$$\nabla^2 \psi - \ddot{\psi} / c^2 = 0 \tag{2.2.12}$$

The inhomogeneous wave equations (2.2.10) and (2.2.11) can be solved by several different methods, among them the Fourier transform technique. This is done in books on electromagnetic theory and the reader should consult these books for detailed step-by-step derivation. The particular solutions are expressed in terms of integrals over sources. They are

$$\mathbf{A}(\mathbf{r},t) = \frac{\mu_0}{4\pi} \int_{V'} \frac{\mathbf{J}(\mathbf{r},t \pm |\mathbf{r} - \mathbf{r}'|/c)}{|\mathbf{r} - \mathbf{r}'|} d\nu' \qquad (2.2.13a)$$

$$V(\mathbf{r}, t) = \frac{1}{4\pi\varepsilon_0} \int_{V'} \frac{\varrho(\mathbf{r}', t \pm |\mathbf{r} - \mathbf{r}'|/c)}{|\mathbf{r} - \mathbf{r}'|} d\nu' \qquad (2.2.13b)$$

where $\mathbf{r} = (x, y, z)$ are the coordinates of the observational point and $\mathbf{r}' = (x', y', z')$ the coordinates of the source point. The integration is carried out over the volume of all sources. Note that the potential at time t depends on the behavior of the source at times $t \pm t_0(\mathbf{r}, \mathbf{r}')$ where t_0 is the time required for the wave to travel from the source point to the observation point. Mathematically, both signs are valid. Physically, because of principle of causality, we expect the cause to precede the effect. The study of waves in free space has shown that perturbations in fields travel with a velocity c. It is, therefore, reasonable to expect that the field measured at present is caused by the behavior of the source at a time $t_0(\mathbf{r}, \mathbf{r}') = |\mathbf{r} - \mathbf{r}'|/c$ earlier. This means that we take the lower sign in (2.2.13a) and (2.2.13b). Solutions with the minus (lower) sign are known as the retarded solutions and those with the plus sign the advanced solutions. The advanced solution appears to have no physical significance, except occasionally it is used as a mathematical aid.

If the source functions J and ϱ are known, the integrations (2.2.13a) and (2.2.13b) give us the vector and scalar potentials. The electromagnetic fields can be computed when these potentials are substituted into (2.2.4) and (2.2.5).

2.3 Electric and Magnetic Polarizations

In free space the material equations take particularly simple forms given by (2.1.3) and Maxwell's equations simplify to (2.2.1). In a material medium such simple relations are generally not valid. Under the influence of electromagnetic fields, microscopic charge distributions associated with molecules and atoms are distorted, producing electric and magnetic dipole moments. To describe the effects on the medium, we introduce the electric and magnetic polarization vectors as quantities of departure from those given by (2.1.3) i.e.,

$$\mathbf{P} = \mathbf{D} - \varepsilon_0 \mathbf{E}, \qquad \mathbf{M} = \mathbf{B}/\mu_0 - \mathbf{H}$$
(2.3.1)

In free space both P and M vanish. The presence of P and M must, therefore, indicate the effect of medium.

Let us substitute (2.3.1) into Maxwell's equations (2.1.1). By a slight rearrangement, the equations can be put in a form

$$\nabla \times \mathbf{E} + \dot{\mathbf{B}} = \mathbf{0}$$

$$(\nabla/\mu_0) \times \mathbf{B} - \varepsilon_0 \dot{\mathbf{E}} = \mathbf{J}_T$$

$$\nabla \cdot \mathbf{E} = \varrho_T/\varepsilon_0$$

$$\nabla \cdot \mathbf{B} = \mathbf{0}$$
(2.3.2)

The form (2.3.2) for material media is identical to form (2.2.1) for free space, except that total current and total charge have been used here instead of free charge and free current in (2.2.1). The total current and total charge are defined, respectively, by

$$\mathbf{J}_T = \mathbf{J} + \mathbf{\nabla} \times \mathbf{M} + \dot{\mathbf{P}}$$
(2.3.3)

$$\varrho_T = \varrho - \nabla \cdot \mathbf{P} \tag{2.3.4}$$

Maxwell's equations written in the form (2.3.2) is not useful in obtaining solutions since the total current and charge given by (2.3.3) and (2.3.4) involve the polarization and magnetization which depend on the unknown

fields **E** and **B**. The form is useful to physically interpret the origin and meaning of terms appearing in (2.3.3) and (2.3.4) in a material medium. So far we only know that $\nabla \times \mathbf{M}$ and $\dot{\mathbf{P}}$ have dimensions of a current density and $-\nabla \cdot \mathbf{P}$ that of a charge density. We can also show by substitution that when defined this way the equation of continuity

$$\partial \varrho_T / \partial t + \nabla \cdot \mathbf{J}_T = 0 \tag{2.3.5}$$

is satisfied. We must still show the physical significance of such a definition. This is done in the following.

We note the similarity in form of Maxwell's equations (2.3.2) in material media and those equations (2.2.1) in free space. As before define a vector and a scalar potential by

$$\mathbf{B} = \nabla \times \mathbf{A} \tag{2.3.6}$$

$$\mathbf{E} = -\nabla V - \dot{\mathbf{A}} \tag{2.3.7}$$

If again the Lorentz condition is prescribed, i.e.,

$$\nabla \cdot \mathbf{A} + \dot{V}/c^2 = 0 \tag{2.3.8}$$

a set of inhomogeneous wave equations can be obtained in an identical manner. They are

$$\nabla^2 \mathbf{A} - \ddot{\mathbf{A}}/c^2 = -\mu_0 \mathbf{J}_T \tag{2.3.9}$$

$$\nabla^2 V - \ddot{V}/c^2 = -\varrho_T/\varepsilon_0 \tag{2.3.10}$$

The retarded solutions to these equations are, respectively,

$$\mathbf{A}(\mathbf{r},t) = \frac{\mu_0}{4\pi} \int_{V'} \frac{\mathbf{J}_T(\mathbf{r}',t-|\mathbf{r}-\mathbf{r}'|/c)}{|\mathbf{r}-\mathbf{r}'|} d\nu' \qquad (2.3.11)$$

$$V(\mathbf{r},t) = \frac{1}{4\pi\varepsilon_0} \int_{V'} \frac{\varrho_T(\mathbf{r}',t-|\mathbf{r}-\mathbf{r}'|/c)}{|\mathbf{r}-\mathbf{r}'|} d\nu' \qquad (2.3.12)$$

The expressions (2.3.11) and (2.3.12) show very clearly that equivalent sources of the field in a material medium consist of not only free charges and free currents but also other quantities arising from polarization of the medium. Our purpose is to identify the physical meaning of these other quantities. To do this we introduce the square brackets to indicate a quantity at the source coordinates and evaluated at a retarded time. The potentials given by (2.3.11) and (2.3.12) in this new notation can be written as

$$\mathbf{A}(\mathbf{r},t) = \frac{\mu_0}{4\pi} \int_{V'} \left(\frac{[\mathbf{J}]}{R} + \frac{[\mathbf{\vec{V}'} \times \mathbf{M}]}{R} + \frac{[\mathbf{\vec{P}}]}{R} \right) dv' \quad (2.3.13)$$

$$V(\mathbf{r}, t) = \frac{1}{4\pi\varepsilon_0} \int_{V'} \left(\frac{[\varrho]}{R} - \frac{[\overline{V}' \cdot \mathbf{P}]}{R}\right) d\nu' \qquad (2.3.14)$$

where

$$\mathbf{R} = \mathbf{r} - \mathbf{r}' \tag{2.3.15}$$

and ∇' operates on the source coordinates r'.

By using the chain rule in differential calculus we obtain

$$abla' \times [\mathbf{M}] = [
abla' \times \mathbf{M}] + \mathbf{R} \times [\dot{\mathbf{M}}]/cR$$
 $abla' \cdot [\mathbf{P}] = [
abla' \cdot \mathbf{P}] + \mathbf{R} \cdot [\dot{\mathbf{P}}]/cR$

Substituting these identities into expressions (2.3.13) and (2.3.14) results in

$$\mathbf{A}(\mathbf{r},t) = \frac{\mu_0}{4\pi} \int_{V'} \left\{ \frac{[\mathbf{J}]}{R} + \frac{1}{R} \nabla' \times [\mathbf{M}] - \frac{1}{cR^2} \mathbf{R} \times [\dot{\mathbf{M}}] + \frac{1}{R} [\dot{\mathbf{P}}] \right\} d\nu'$$
(2.3.16)

$$V(\mathbf{r}, t) = \frac{1}{4\pi\varepsilon_0} \int_{V'} \left\{ \frac{[\varrho]}{R} - \frac{1}{R} \nabla' \cdot [\mathbf{P}] + \frac{1}{cR^2} \mathbf{R} \cdot [\dot{\mathbf{P}}] \right\} d\nu' \qquad (2.3.17)$$

The second term in the integrands of (2.3.16) and (2.3.17) can be reexpressed through the use of vector identities

$$\overline{\nu}' \times \left(\frac{\mathbf{M}}{R}\right) = \frac{1}{R} \overline{\nu}' \times \mathbf{M} - \mathbf{M} \times \overline{\nu}' \left(\frac{1}{R}\right)$$
$$\overline{\nu}' \cdot \left(\frac{\mathbf{P}}{R}\right) = \frac{1}{R} \overline{\nu}' \cdot \mathbf{P} + \mathbf{P} \cdot \overline{\nu}' \left(\frac{1}{R}\right)$$

Integrate these equations and make use of Gauss's theorems to change volume integrals to surface integrals

$$\int_{V'} \nabla' \times \left(\frac{\mathbf{M}}{R}\right) d\nu' = -\int_{S} \frac{\mathbf{M}}{R} \times d\mathbf{S}$$
 (2.3.18)

$$\int_{V'} \nabla' \cdot \left(\frac{\mathbf{P}}{R}\right) dv' = \int_{S} \frac{\mathbf{P}}{R} \cdot d\mathbf{S}$$
 (2.3.19)

We obtain, after rearrangement of terms, the following relations

$$\int_{V'} \frac{\overline{V'} \times \mathbf{M}}{R} d\nu' = \int_{V'} \mathbf{M} \times \overline{V'} \left(\frac{1}{R}\right) d\nu' - \int_{S} \frac{\mathbf{M}}{R} \times d\mathbf{S}$$
$$\int_{V'} \frac{\overline{V'} \cdot \mathbf{P}}{R} d\nu' = -\int_{V'} \mathbf{P} \cdot \overline{V'} \left(\frac{1}{R}\right) d\nu' + \int_{S} \frac{\mathbf{P}}{R} \cdot d\mathbf{S}$$

The surface integrals in these relations vanish if we choose the surface of the integration so that the material substance lies inside of the region. Replace the second terms of (2.3.16) and (2.3.17) by using these relations. The result is

$$\mathbf{A}(\mathbf{r},t) = \frac{\mu_0}{4\pi} \int_{V'} \left\{ \frac{[\mathbf{J}]}{R} + [\mathbf{M}] \times \overline{V'} \left(\frac{1}{R} \right) - \frac{1}{cR^2} \mathbf{R} \times [\dot{\mathbf{M}}] + \frac{[\dot{\mathbf{P}}]}{R} \right\} dv'$$
(2.3.20)

$$V(\mathbf{r},t) = \frac{1}{4\pi\varepsilon_0} \int_{V'} \left\{ \frac{[\varrho]}{R} + [\mathbf{P}] \cdot \nabla' \left(\frac{1}{R} \right) + \frac{1}{cR^2} \mathbf{R} \cdot [\dot{\mathbf{P}}] \right\} d\nu' \qquad (2.3.21)$$

The expressions (2.3.20) and (2.3.21) are in convenient forms for physical identification. We shall do this first for the scalar potential.

The first term in the integrand of (2.3.21) is $[\varrho]/R$. It represents the usual Coloumb potential of free charges in vacuum given by (2.2.13b). The next two terms show the effect of the material medium to the potential. Their physical significance can be investigated by considering a dipole arrangement of charges.

Let two point charges of equal magnitude but of opposite sign be placed a small distance $\boldsymbol{\xi}$ apart as shown in Figure 2.3-1. The potential due to this dipole arrangement can be written simply as



Since ξ is small, we may expand the second term in Taylor series and obtain

$$\frac{q(t-|\mathbf{R}+\boldsymbol{\xi}|/c)}{4\pi\varepsilon_{0}|\mathbf{R}+\boldsymbol{\xi}|} = \frac{q(t-R/c)}{4\pi\varepsilon_{0}R} - [q]\boldsymbol{\xi}\cdot\boldsymbol{\nabla}'(1/R) - ([\dot{q}]/cR^{2})\mathbf{R}\cdot\boldsymbol{\xi} + \text{higher order terms}$$
(2.3.23)

In the limit of vanishing ξ , we can define

$$\lim_{\xi \to 0} [q] \boldsymbol{\xi} = [\mathbf{p}] \tag{2.3.24}$$

where \mathbf{p} is the dipole moment. The potential arising from such a dipole arrangement is then obtained from substituting (2.3.23) back to (2.3.22), giving

$$V = \frac{1}{4\pi\varepsilon_0} \left\{ [\mathbf{p}] \cdot \nabla' \left(\frac{1}{R} \right) + \frac{1}{cR^2} \mathbf{R} \cdot [\dot{\mathbf{p}}] \right\}$$
(2.3.25)

This is the potential due to a single dipole. Suppose such dipoles are distributed throughout the volume with a density **P**. We must then find the resultant contribution by integrating over the volume, i.e.,

$$V_D(\mathbf{r}, t) = \frac{1}{4\pi\varepsilon_0} \int \left\{ [\mathbf{P}] \cdot \nabla' \left(\frac{1}{R} \right) + \frac{1}{cR^2} \mathbf{R} \cdot [\dot{\mathbf{P}}] \right\} dv' \quad (2.3.26)$$

Comparing (2.3.26) with (2.3.21), we see that the last two terms of (2.3.21) are identical to (2.3.26). Therefore, the scalar potential in material substances given by (2.3.21) is arising from two sources of contributions: the free charge ρ as the case in free space and a distribution of dipole moments of volume density **P**. Due to presence of the electric field, microscopic charge arrangements are perturbed. These perturbations set up equivalent dipole moments as shown in Figure 2.3-1. When the electric field is weak, the dipole moment **P** is usually linearly dependent on **E**. When the field is strong the relation may become nonlinear. Our purpose is to derive such relations for plasmas under different conditions.

The terms appearing in (2.3.20) for the vector potential can be physically identified in the same manner. The first term [J]/R is the usual contribution from the free current and is the only term in free space. The remaining three terms indicate contributions associated with the material medium. Microscopically a molecule or an atom has electrons orbiting about a nucleus. Effective current may result on a macroscopic scale if there exists imperfect orbit cancellation on an atomic scale. Such a current is called the magne-
tization current and is the origin of the second and third terms in the integrand of (2.3.20). In a plasma the magnetization effect is negligible and we shall not choose to discuss this in detail here. The interested reader should consult books on electromagnetic theory for a more complete discussion. The last term in (2.3.20) is related to the time rate of change of the electric polarization. Microscopically, it comes from the motion of dipole moments. We proceed to prove it in the following.

The microscopic dipole moment is given by

$$\mathbf{p} = \varrho_p \mathbf{\xi} \tag{2.3.27}$$

where ϱ_p is the microscopic polarization charge and ξ the separation between the negative charge and the positive charge. Since (2.3.27) is a microscopic expression it is highly irregular on an atomic scale. The time rate of change of the microscopic dipole moment may be caused by the motion of these dipoles or by changes in the separation distance. Hence,

$$\partial \mathbf{p} / \partial t = \mathbf{\xi} \partial \varrho_p / \partial t + \varrho_p \dot{\mathbf{\xi}}$$

= $-\mathbf{\xi} \nabla \cdot (\varrho_p \mathbf{u}) + \varrho_p \dot{\mathbf{\xi}}$ (2.3.28)

where the equation of continuity has been applied to obtain the last expression in (2.3.28). The quantity **u** is the velocity of the polarization charge. The vector identity

can be used to recast (2.3.28) into the following form:

$$\partial \mathbf{p}/\partial t = -\nabla \cdot (\varrho_p \mathbf{u} \mathbf{\xi}) + \varrho_p \mathbf{u} + \varrho_p \mathbf{\xi}$$
 (2.3.29)

The expression (2.3.29) is still for microscopic quantities. The macroscopic expression can be obtained by integrating (2.3.29) over a volume ΔV which is large compared with the atomic scale but small compared with the macroscopic scale. The integration of the divergence term vanishes since there is no surface charge density in the medium. The time rate of change of the macroscopic polarization density is then given by this integral divided by the volume ΔV , i.e.,

$$\frac{\partial \mathbf{P}}{\partial t} = \frac{1}{\Delta V} \int_{\Delta V} \frac{\partial \mathbf{p}}{\partial t} \, d\nu = \langle \varrho_p \mathbf{u} + \varrho_p \dot{\mathbf{\xi}} \rangle \qquad (2.3.30)$$

where the angular brackets are used to denote the averaging process. The expression (2.3.30) states that the time rate of change of polarization density arises from transport of polarization charges and change in the separation of polarization charges.

In summary we see that Maxwell's equations in a material medium can be written in a form identical to that in free space, provided we take into account all sources. These sources include true current, magnetization current, and polarization current as current sources, and true charge and polarization charge as charge sources. Comparing (2.1.1c) with the third equation of (2.3.2) we see that the source for the electric flux density is free charge alone while that for the electric intensity is the total charge. For this reason **D** is referred to as the partial field sometimes. However, it should be mentioned that the Maxwell's equations in the form (2.3.2) are useful only to provide physical insights and not convenient for problem solving purposes. As a matter of fact (2.3.13) and (2.3.14) are not true solutions since **P** and **M** are related to the unknown electromagnetic fields. What we need are material equations. A general discussion on these relations is carried out in the next three sections.

2.4 Slow and Fast Processes

The material equations in free space are particularly simple. They are given by

$$\mathbf{D}(\mathbf{r}, t) = \varepsilon_0 \mathbf{E}(\mathbf{r}, t), \qquad \mathbf{B}(\mathbf{r}, t) = \mu_0 \mathbf{H}(\mathbf{r}, t)$$
(2.4.1)

Such relations are linear, nondispersive, isotropic, and local. These adjectives are used very often in describing material equations. We shall explain their meaning more carefully in the following.

Consider a certain medium in which there exists a relation between \mathbf{D}_1 and \mathbf{E}_1 . The relation may be that given by (2.4.1) or one that is more complicated. Similarly for the same medium we find that there exists a relation between \mathbf{D}_2 and \mathbf{E}_2 . The relation is said to be linear if $\mathbf{D}_1 + \mathbf{D}_2$ and $\mathbf{E}_1 + \mathbf{E}_2$ satisfy the relation. The medium is defined as a linear medium if linear relations exist between \mathbf{D} , \mathbf{B} and \mathbf{E} , \mathbf{H} . Obviously material equations (2.4.1) are linear. The relations (2.4.1) are the simplest linear relations since they are algebraic. A more complicated linear relation would be one in which ε_0 in (2.4.1) is replaced, for example, by a linear operator. We shall have occasion to discuss some of these cases later on in this section. In a material medium the relations (2.4.1) are in general no longer valid. The difference is caused by the electric polarization and magnetization of the medium. The presence of the electromagnetic fields distort the charge distribution on a molecular scale. This distortion is the origin of the electric polarization charges, magnetization currents, etc. If the fields vary in time sufficiently slowly the establishment of the polarization effect is nearly instantaneous. This means that the electric polarization at time t depends only on the electromagnetic field at the same instant t. In this case the medium is said to be nondispersive.

The material relations in a nondispersive medium (as well as linear, isotropic, and local) are expected to be similar to (2.4.1). Since the response of the medium is instantaneous, we have

$$\mathbf{P}(\mathbf{r}, t) = \varepsilon_0 \chi(\mathbf{r}) \mathbf{E}(\mathbf{r}, t), \qquad \mathbf{M}(\mathbf{r}, t) = \chi_m(\mathbf{r}) \mathbf{H}(\mathbf{r}, t) \qquad (2.4.2)$$

where χ and χ_m are, respectively, electric and magnetic susceptibility. They depend on the thermodynamic state of the medium. Substitution of (2.4.2) into (2.3.1) gives

$$\mathbf{D} = \varepsilon_0 (1 + \chi) \mathbf{E} = \varepsilon_0 K \mathbf{E}, \qquad \mathbf{B} = \mu_0 (1 + \chi_m) \mathbf{H} = \mu_0 K_m \mathbf{H} \quad (2.4.3)$$

where K and K_m are, respectively, the relative dielectric constant and relative permeability.

However, as the time variation of the electromagnetic fields increases this instantaneous dependence is no longer expected. Due to finite mass of these charges, there always exists at least one frequency, ω_m , which characterizes the speed of establishing the polarization state. As the frequency of the fields approaches or exceeds ω_m , the response of the medium cannot keep up with the change in the fields. The instantaneous relations such as those given by (2.4.1) are no longer valid. The medium is then said to be time dispersive.

The material equations (2.4.1) are isotropic because they are invariant under rotational transformations about any axis. This is no longer true, for instance, in a magnetoplasma in which ε_0 must be replaced by a tensor and the medium is said to be anisotropic.

In most cases, material equations are derived by averaging over a volume whose linear dimension is large when compared with the characteristic length λ_c of the medium and small when compared with the wavelength λ of the electromagnetic field in the medium. Suitable choices of characteristic length are the atomic dimensions, lattice constant, and Debye length. The wavelength in the medium is related to wavelength in free space, λ_0 , by the relation $\lambda = \lambda_0/n = c/nf$, where *n* is the refractive index, *c* the velocity of light in free space, and *f* the frequency. Therefore, our requirement $\lambda_c/\lambda \ll 1$ becomes

$$\lambda_c nf/c \ll 1 \tag{2.4.4}$$

When the inequality (2.4.4) is valid, the polarization of the medium is expected to depend on the electromagnetic fields at the same location. The resulting material equations must be local since **D** at **r** is related to **E** at the same r and similarly between B and H. Inspection of (2.4.4) shows that there are two cases in which (2.4.2) may be violated: (i) the frequency f is sufficiently large or (ii) the refractive index becomes very large. The violation of (2.4.2) means that the material equation cannot be a local one. The vector **D** at **r** depends not only on **E** at **r** but also on **E** at points in the neighborhood of r. If this is the case, the medium is said to be spatially dispersive. The local and nonlocal relations can also be looked at from a different point of view. When the medium is cold, the particles are essentially stationary. The effect imparted by the fields is not likely to be carried to a neighboring point. Therefore, the material equation is expected to be a local relation. However, as the temperature of the medium increases thermal agitation may start to carry the effect felt at a point to a neighboring point. This is then the beginning of spatial dispersion.

The consideration of more and more general material media will be carried out in steps. In plasma media the permeability is nearly always that of the free space while the dielectric constant may take different forms. In order to simplify our discussion in the following we shall consider the dielectric constant only, although similar arguments and reasoning can be applied to permeability. We proceed to consider the time dispersion first.

For sufficiently weak fields we may assume a linear relation between **P** and **E**. As the time variation increases the instantaneous relation between **P** and **E** is no longer expected. The polarization of the medium will depend on the value of the electric field at present as well as at previous times. The principle of causality is a very basic physical principle and it states that the effect must always come after the cause. This is essentially the basis for adopting the retarded potentials in (2.2.13a) and (2.2.13b) as the solutions of the inhomogeneous wave equations. The application of this principle in the present case indicates that **P** at *t* must depend on **E** at all times previous to *t*. Let the field be applied at t = 0; the most general such relation we can write is

$$\mathbf{P}(\mathbf{r}, t) = \varepsilon_0 \int_0^t \chi(\mathbf{r}, t, \tau) \mathbf{E}(\mathbf{r}, \tau) d\tau + (\text{contribution related to initial conditions}) \quad (2.4.5)$$

The polarization density depends on the value of the electric field from the initial time t = 0 to the present t. The contribution is integrated over this time interval with a weighting function $\chi(\mathbf{r}, t, \tau)$. In general there is also a contribution related to the initial state of polarization. For t large the initial conditions may have decayed sufficiently with time in a lossy medium. We shall assume that this is the case and ignore the contribution due to initial conditions.

If the property of the medium does not change with time the weighting function $\chi(\mathbf{r}, t, \tau)$ should be invariant with respect to the time translation, i.e.,

$$\chi(\mathbf{r}, t, \tau) = \chi(\mathbf{r}, t + t_1, \tau + t_1)$$
(2.4.6)

In particular if $t_1 = -\tau$, the third argument reduces to 0 which can be suppressed and we have

$$\chi(\mathbf{r}, t, \tau) = \chi(\mathbf{r}, t - \tau) \qquad (2.4.7)$$

Substitute (2.4.7) into (2.4.5), giving

$$\mathbf{P}(\mathbf{r}, t) = \varepsilon_0 \int_0^t \chi(\mathbf{r}, t - \tau) \mathbf{E}(\mathbf{r}, \tau) d\tau \qquad (2.4.8)$$

Here we have assumed that the electric field has been applied for a time of sufficient length so that initial conditions do not contribute. When (2.4.8) is substituted into the Maxwell's equations, we have a set of integrodifferential equations. To solve them, we note that (2.4.8) is in the form of a convolution integral and we shall therefore apply the convolution theorem in the theory of Fourier transform. In order to avoid flooding the text with an unnecessary number of notations we shall use the argument to distinguish the transform pair. For example,

$$\mathbf{E}(\mathbf{r},\omega) = \int_{0}^{\infty} \mathbf{E}(\mathbf{r},t) e^{-j\omega t} dt$$

$$\mathbf{E}(\mathbf{r},t) = (1/2\pi) \int_{\sigma} \mathbf{E}(\mathbf{r},\omega) e^{j\omega t} d\omega$$
(2.4.9)

Therefore, $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{E}(\mathbf{r}, \omega)$ form a Fourier transform pair. Similar notations are used for other quantities. The contour *c* in the second equation of (2.4.10) is sometimes referred to as the Laplace contour and is parallel to the real ω -axis and below all singularities. In the transformed domain, (2.4.8) can be written as

$$\mathbf{P}(\mathbf{r},\omega) \coloneqq \varepsilon_0 \chi(\mathbf{r},\omega) \mathbf{E}(\mathbf{r},\omega)$$
(2.4.10)

This in turn gives, through (2.3.1),

$$\mathbf{D}(\mathbf{r},\omega) = \varepsilon(\mathbf{r},\omega)\mathbf{E}(\mathbf{r},\omega)$$
(2.4.11)

where the dielectric constant is given by

$$\varepsilon(\mathbf{r}, \omega) = \varepsilon_0 [1 + \chi(\mathbf{r}, \omega)]$$
$$= \varepsilon_0 K(\mathbf{r}, \omega) \qquad (2.4.12)$$

where $K(\mathbf{r}, \omega)$ is the relative dielectric constant and is given by

$$K(\mathbf{r},\omega) = 1 + \int_0^\infty \chi(\mathbf{r},t) e^{-j\omega t} dt \qquad (2.4.13)$$

The integration is from zero to $+\infty$ because of the causality principle. With the application of Fourier transform, the original set of Maxwell's equations becomes a set of partial differential equations which will be studied in detail. But before we go into the discussion of the solution of the set of equations, we shall first consider the properties of more general materials. As mentioned above, the property of the medium may be anisotropic such that the polarization is not in the same direction as the electric field, as in most crystals as well as magnetoplasmas. Furthermore, if the condition (2.4.4) is not satisfied, a nonlocal relation may exist between **D** and **E**. Under these conditions, the most general linear relation between the components of the polarization and electric field can be written as

$$P_i(\mathbf{r}, t) = \varepsilon_0 \int_0^t d\tau \int \chi_{ij}(\mathbf{r}, \mathbf{r}', t, \tau) E_j(\mathbf{r}', \tau) d\mathbf{r}' \qquad (2.4.14)$$

where χ_{ij} is the *ij*th component of a tensor and summation must be carried out over all subscripts that occur twice. Contributions from initial conditions are neglected as before. If the medium is temporally stationary and spatially homogeneous so that its property is invariant under translation in time and space, then

$$\chi_{ij}(\mathbf{r}, \mathbf{r}', t, \tau) = \chi_{ij}(\mathbf{r} - \mathbf{r}', t - \tau) \qquad (2.4.15)$$

and (2.4.14) becomes

$$P_i(\mathbf{r}, t) = \varepsilon_0 \int_0^t d\tau \int \chi_{ij}(\mathbf{r} - \mathbf{r}', t - \tau) E_j(\mathbf{r}', \tau) d\mathbf{r}' \qquad (2.4.16)$$

which is again in the form of a convolution integral. When time and space

Fourier transform is applied to (2.4.16), we obtain

$$P_{i}(\mathbf{k},\omega) = \varepsilon_{0}\chi_{ij}(\mathbf{k},\omega)E_{j}(\mathbf{k},\omega) \qquad (2.4.17)$$

The Fourier transform pair is defined as

$$E_{i}(\mathbf{k},\omega) = \int_{0}^{\infty} dt \int_{-\infty}^{+\infty} E_{i}(\mathbf{r},t) e^{-j[\omega t - \mathbf{k} \cdot \mathbf{r}]} d\mathbf{r}$$

$$E_{i}(\mathbf{r},t) = \frac{1}{(2\pi)^{4}} \int_{e} d\omega \int_{-\infty}^{+\infty} E_{i}(\mathbf{k},\omega) e^{j[\omega t - \mathbf{k} \cdot \mathbf{r}]} d\mathbf{k}$$
(2.4.18)

where the contour c is the same as that in (2.4.9). Similar expressions exist for other quantities. From (2.4.17) and (2.3.1), we have

$$D_{i}(\mathbf{k},\omega) = \varepsilon_{ij}(\mathbf{k},\omega)E_{j}(\mathbf{k},\omega)$$
$$= \varepsilon_{0}K_{ij}(\mathbf{k},\omega)E_{j}(\mathbf{k},\omega) \qquad (2.4.19)$$

where

$$\varepsilon_{ij}(\mathbf{k},\omega) = \varepsilon_0[\delta_{ij} + \chi_{ij}(\mathbf{k},\omega)]$$

= $\varepsilon_0\left[\delta_{ij} + \int_0^\infty dt \int_{-\infty}^{+\infty} \chi_{ij}(\mathbf{r},t) e^{-j[\omega t - \mathbf{k} \cdot \mathbf{r}]} d\mathbf{r}\right]$ (2.4.20)

Here δ_{ij} is the Kronecker delta. Equations (2.4.19) and (2.4.20) may be thought of as the definition of the dielectric tensor $\varepsilon_{ij}(\mathbf{k}, \omega)$ and relative dielectric tensor $K_{ij}(\mathbf{k}, \omega)$. This is the general material equation relating the two field quantities **D** and **E** in a linear, stationary, and homogeneous medium. Similar consideration will lead to the corresponding relation between **B** and **H**. We shall not discuss it here.

We note that in the limit $k \to 0$, $K_{ij}(\omega, \mathbf{k} \to 0)$ becomes the relative dielectric tensor for a medium in the absence of spatial dispersion. If furthermore, the medium is isotropic, then $K_{ij}(\omega) = \delta_{ij}K(\omega)$. In general, for an isotropic medium with spatial dispersion, the relative dielectric tensor should be symmetric and invariant under rotational transformation about the vector **k**. Under these conditions, it can be shown that (see problem at the end of this chapter)

$$K_{ij}(\mathbf{k},\omega) = K_{\perp}(k^{2},\omega)(\delta_{ij} - k_{i}k_{j}/k^{2}) + K_{\parallel}(k^{2},\omega)k_{i}k_{j}/k^{2} \quad (2.4.21)$$

where K_{\perp} and K_{\parallel} are functions of ω and k^2 only. They are referred to as the transverse and longitudinal relative dielectric constants in an isotropic medium. We will come back to this in a later section.

2.5 Kramers-Kronig Relations

In this section we shall discuss some important general properties of the relative dielectric tensor $K_{ij}(\mathbf{k}, \omega)$ based on its definition and certain very general physical laws.

 $K_{ii}(\mathbf{k}, \omega)$ by its definitions is, in general, a complex function of the two independent variables k and ω . The dependence on ω indicates the time dispersion of the medium while the \mathbf{k} dependence indicates the spatial dispersion. From the principal of causality we see that $\chi_{ii}(\mathbf{r}, t) = 0$ for t < 0. Also $\chi_{ii}(\mathbf{r}, t)$ is finite and approaches zero when $t \to \infty$. This simply means that the establishment of the polarization at present cannot be appreciably affected by the electric field at the remote past. Further, for an electric field given by $E_j(\mathbf{r}, t) = E_{j0} \,\delta(\mathbf{r} - \mathbf{r}_1) \,\delta(t - t_1)$, E_{j0} a constant, the polarization given by (2.4.16) is $\varepsilon_0 \chi_{ij}(\mathbf{r} - \mathbf{r}_1, t - t_1) E_{i0}$. Therefore, $\chi_{ii}(\mathbf{r}-\mathbf{r}',t-\tau)$ can be interpreted as the response of the medium to an impulse applied at time $t = \tau$ and position $\mathbf{r} = \mathbf{r}'$. In a passive medium on physical ground, we expect this "impulse" response to decay with increasing time for $t > \tau$ and with increasing distance $|\mathbf{r} - \mathbf{r}'|$. Alternatively, we expect $\chi_{ij}(\mathbf{r}, t)$ to be a monotonically decreasing function of t and $|\mathbf{r}|$. From the definition (2.4.20) it follows that $\chi_{ii}(\mathbf{k}, \omega)$ and hence also $K_{ii}(\mathbf{k}, \omega)$ must be regular in the lower half-plane of the complex $\omega = \omega' + j\omega''$ plane. This can be seen easily by letting $\omega'' < 0$ in the integrand in (2.4.20). In the upper half-plane of ω the definition of $K_{ij}(\mathbf{k}, \omega)$ has to be extended by analytical continuation and in general it does have singularities. As $|\omega| \rightarrow \infty$ in any manner in the lower half-plane, K_{ij} tends to the unity tensor δ_{ii} as can be seen easily from (2.4.20). For $\omega \to 0$, K_{ii} is finite for dielectric material but has a simple pole for a conducting medium.

 $K_{ij}(\mathbf{k}, \omega)$ is a regular function of **k** in the whole region of the complex variable **k** since in general $\chi_{ij}(\mathbf{r} - \mathbf{r}', t - t')$ approaches zero as $\mathbf{r} - \mathbf{r}'$ increases to infinite.

In Maxwell's equations (2.1.1), the fields $\mathbf{E}(\mathbf{r}, t)$, etc., are all physical quantities and hence are all real. It follows from (2.3.1), and (2.4.14) and (2.4.20) that $\varepsilon_{ij}(\mathbf{r}, t)$ must also be real. However, its transform $\varepsilon_{ij}(\mathbf{k}, \omega)$ is in general not real. The properties of $\varepsilon_{ij}(\mathbf{k}, \omega)$ can be found by noting (2.4.20) which gives

$$K_{ij}(\mathbf{k}^*, \omega^*) = K_{ij}^*(-\mathbf{k}, -\omega)$$
(2.5.1)

where * is used to denote complex conjugate. We denote the relative dielec-

tric tensor by its Hermitian and anti-Hermitian parts as

$$K_{ij}(\mathbf{k},\omega) = K'_{ij}(\mathbf{k},\omega) - jK''_{ij}(\mathbf{k},\omega)$$
(2.5.2)

where

$$\begin{aligned} K_{ij}'(\mathbf{k},\omega) &= \frac{1}{2}(K_{ij} + K_{ji}^*) \\ K_{ij}''(\mathbf{k},\omega) &= \frac{1}{2}(K_{ij} - K_{ji}^*) \end{aligned}$$
(2.5.3)

are both Hermitian tensors (jK_{ij} is anti-Hermitian). From (2.5.2), we have

$$\operatorname{Re} K_{ij} = \operatorname{Re} K'_{ij} + \operatorname{Im} K''_{ij}$$

$$\operatorname{Im} K_{ij} = \operatorname{Im} K'_{ij} - \operatorname{Re} K''_{ij}$$
(2.5.4)

where Re denotes real part and Im denotes imaginary part.

If the medium is not spatially dispersive, i.e., K_{ij} is independent of k, it is easy to see from (2.5.1) that

$$\operatorname{Re} K_{ij}(\omega') = \operatorname{Re} K_{ij}(-\omega'), \qquad \operatorname{Im} K_{ij}(\omega') = -\operatorname{Im} K_{ij}(-\omega') \quad (2.5.5)$$

and

$$\operatorname{Im} K_{ij}(j\omega'') = -\operatorname{Im} K_{ij}(j\omega'') = 0 \qquad (2.5.6)$$

which means that the real part of K_{ij} is an even function while the imaginary part of K_{ij} is an odd function on the real axis of the complex ω -plane and K_{ij} is real on the imaginary axis of the ω -plane.

From the analytic properties of K_{ij} discussed above, it is possible to derive some general relations between the Hermitian and anti-Hermitian parts of the tensor. To see this, let us consider the following integration in the complex ω -plane.

$$I = \int_{c} \frac{K_{ij}(\mathbf{k},\omega) - \delta_{ij}}{\omega - \omega_{0}} \, d\omega \tag{2.5.7}$$

where the contour c is shown in Fig. 2.5-1. Here we shall assume that K_{ij} may have a simple pole at the origin but otherwise it is analytic on the real axis and the lower half-plane. The simple pole at the origin is to take care of the case where the medium may possess a dc conductivity. The integral is zero since the integrand is regular inside the contour. The contour can be broken into several parts: the principal part integrating from $-\infty$ to $+\infty$ on real axis, the small indentation about the pole at ω_0 , the small indentation about the pole at the origin, and the big semicircle. The contribution from the big semicircle vanishes at the limit. The remaining contributions



Fig. 2.5-1. The contour c for the integral in (2.5.7).

can be evaluated by using the residue theorem in complex variables, giving

$$P\int_{-\infty}^{+\infty}\frac{K_{ij}(\mathbf{k},x)-\delta_{ij}}{x-\omega'}\,dx+\pi j[K_{ij}(\mathbf{k},\omega')-\delta_{ij}]-\frac{\pi\sigma_{ij}(\mathbf{k},0)}{\varepsilon_0\cdot\omega'}=0$$
(2.5.8)

where we have changed ω_0 to ω' . The symbol *P* in front of the integral sign denotes the Cauchy principal value which means in our case

$$P\int_{-\infty}^{+\infty} \frac{K_{ij}(\mathbf{k}, x) - \delta_{ij}}{x - \omega'} dx$$

=
$$\lim_{\epsilon \to 0, \eta \to 0} \int_{-\infty}^{-\epsilon} + \int_{+\epsilon}^{\omega' - \eta} + \int_{\omega' + \eta}^{+\infty} \frac{[K_{ij}(\mathbf{k}, x) - \delta_{ij}]}{x - \omega'} dx \quad (2.5.9)$$

 $\sigma_{ij}(\mathbf{k}, 0)$ is the dc conductivity of the medium defined as

$$\sigma_{ij}(\mathbf{k},0) = j\varepsilon_0 \lim_{\omega' \to 0} \omega' K_{ij}(\mathbf{k},\omega')$$
(2.5.10)

Equating real and imaginary parts of (2.5.8), we obtain

$$\operatorname{Re} K_{ij}(\mathbf{k},\omega') - \delta_{ij} = \frac{-1}{\pi} P \int_{-\infty}^{+\infty} \frac{\operatorname{Im} K_{ij}(\mathbf{k},x)}{x-\omega'} dx \qquad (2.5.11)$$

$$\operatorname{Im} K_{ij}(\mathbf{k},\omega') = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \left[\frac{\operatorname{Re} K_{ij}(\mathbf{k},x) - \delta_{ij}}{x - \omega'} - \frac{\sigma_{ij}(\mathbf{k},0)}{\varepsilon_0 \omega'} \right] dx \quad (2.5.12)$$

These relations are known as Kramers-Kronig relations first derived by them in 1927 for an isotropic medium without spatial dispersion so that $K_{ij}(\omega, \mathbf{k}) = \delta_{ij}K(\omega)$. These relations follow directly from the analytic property of K_{ij} that it is regular in the lower half-plane of ω . Therefore we can conclude that the Kramers-Kronig relations are the consequence of the principle of causality. Since K'_{ij} and K''_{ij} given by (2.5.3) are both Hermitian, we have

$$K'_{ij} = \operatorname{Re} K'_{ij} + j \operatorname{Im} K'_{ij} = (K'_{ji})^* = \operatorname{Re} K'_{ji} - j \operatorname{Im} K'_{ji}$$

$$K''_{ij} = \operatorname{Re} K''_{ij} + j \operatorname{Im} K''_{ij} = (K''_{ji})^* = \operatorname{Re} K''_{ji} - j \operatorname{Im} K''_{ji}$$
(2.5.13)

Therefore Re K'_{ij} and Re K''_{ij} are symmetric tensors while Im K'_{ij} and Im K''_{ij} are antisymmetric. Substituting (2.5.4) into (2.5.11) and making use of the symmetric and antisymmetric properties of the tensors, the Kramers-Kronig relations can be put into the form

$$\operatorname{Re} K_{ij}'(\mathbf{k}, \omega') - \delta_{ij} = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{\operatorname{Re} K_{ij}'(\mathbf{k}, x)}{x - \omega'} dx$$

$$\operatorname{Re} K_{ij}''(\mathbf{k}, \omega') - \frac{\sigma_{ij}(\mathbf{k}, 0)}{\varepsilon_0 \omega'} = \frac{-1}{\pi} P \int_{-\infty}^{+\infty} \frac{\operatorname{Re} K_{ij}'(\mathbf{k}, x) - \delta_{ij}}{x - \omega'} dx$$

$$\operatorname{Im} K_{ij}''(\mathbf{k}, \omega') = \frac{-1}{\pi} P \int_{-\infty}^{+\infty} \frac{\operatorname{Im} K_{ij}'(\mathbf{k}, x)}{x - \omega'} dx$$

$$\operatorname{Im} K_{ij}'(\mathbf{k}, \omega') = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{\operatorname{Im} K_{ij}''(\mathbf{k}, x)}{x - \omega'} dx$$

where $\sigma_{ij}(\mathbf{k}, 0)$ is taken as a real symmetric tensor. Equation (2.5.14) can also be written as

$$K_{ij}'(\mathbf{k},\omega') - \delta_{ij} = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{K_{ij}''(\mathbf{k},x)}{x-\omega'} dx$$

$$K_{ij}''(\mathbf{k},\omega') - \frac{\sigma_{ij}(\mathbf{k},0)}{\varepsilon_0\omega'} = -\frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{K_{ij}'(\mathbf{k},x) - \delta_{ij}}{x-\omega'} dx$$
(2.5.15)

which relates the Hermitian and anti-Hermitian parts of the relative dielectric tensor. For the special case of an isotropic medium without spatial dispersion so that $K_{ij} = \delta_{ij} K(\omega)$, we have

$$K_{ij}^{\prime\prime}(\omega^{\prime}) = \frac{1}{2} \delta_{ij} [K(\omega^{\prime}) + K^{*}(\omega^{\prime})] = \delta_{ij} \operatorname{Re} K(\omega^{\prime})$$

$$K_{ij}^{\prime\prime}(\omega^{\prime}) = \frac{1}{2} j \delta_{ij} [K(\omega^{\prime}) - K^{*}(\omega^{\prime})] = -\delta_{ij} \operatorname{Im} K(\omega^{\prime})$$
(2.5.16)

Equation (2.5.15) can then be written for $K(\omega')$ as

$$\operatorname{Re} K(\omega') - 1 = \frac{-2}{\pi} P \int_{0}^{\infty} \frac{x \operatorname{Im} K(x)}{x^{2} - \omega'^{2}} dx$$

$$\operatorname{Im} K(\omega') + \frac{\sigma(0)}{\varepsilon_{0}\omega'} = \frac{2\omega'}{\pi} P \int_{0}^{\infty} \frac{\operatorname{Re} K(x) - 1}{x^{2} - \omega'^{2}} dx$$
(2.5.17)

where $\sigma(0)$ is the corresponding dc conductivity and (2.5.5) has been used. In an isotropic medium, the loss of electromagnetic energy is characterized by the anti-Hermitian part of the dielectric tensor (this will be discussed in detail in Section 10). Therefore, from absorption experiments, it is often possible to obtain approximate values of $K_{ij}^{\prime\prime}$ for all frequencies. Then from (2.8.15) K_{ij}^{\prime} can be computed. On the other hand, if K_{ij}^{\prime} is known $K_{ij}^{\prime\prime}$ may also be computed, but here care must be taken as it is possible that if K_{ij}^{\prime} is known only approximately, we may obtain a physically unstable condition in the computation of $K_{ij}^{\prime\prime}$. We will come back to this point in a later section.

If in addition to the principle of causality, we also take into account the fact that all interactions are propagated at finite velocities, then for the spatial dispersive medium, additional relations between the Hermitian and anti-Hermitian parts of the dielectric tensor can be derived.

2.6 Onsager Relation

If an external magnetic field is applied to the medium, then the polarization and, hence, the dielectric property of the medium will depend on the external field, as in the case of a plasma in a static magnetic field. From the very general irreversible thermodynamic principle, the Onsager reciprocal relations, we can show that for any nonactive medium in an external static magnetic field \mathbf{B}_0 ,

$$\varepsilon_{ij}(\mathbf{k},\omega,\mathbf{B}_0) = \varepsilon_{ji}(-\mathbf{k},\omega,-\mathbf{B}_0) \tag{2.6.1}$$

If in addition the medium also satisfies the relation (corresponding to the so called nongyrotropic medium),

$$\varepsilon_{ij}(\mathbf{k},\omega,\mathbf{B}_0) = \varepsilon_{ij}(-\mathbf{k},\omega,\mathbf{B}_0)$$
(2.6.2)

then

$$\varepsilon_{ij}(\mathbf{k},\omega,\mathbf{B}_0) = \varepsilon_{ji}(\mathbf{k},\omega,-\mathbf{B}_0)$$
(2.6.3)

Suppose we first choose a Cartesian coordinate such that \mathbf{B}_0 is in the zdirection and k in the xz-plane (Fig. 2.6-1a). According to the Onsager relation (2.6.1) the dielectric tensor for the coordinates shown in Fig. 2.6-1a is the transpose of the dielectric tensor for the coordinates shown in



Fig. 2.6-1. Coordinate system showing the relative directions of k and B_0 .

Fig. 2.6-1b in which both directions of \mathbf{B}_0 and k have been reversed. Reversing directions of \mathbf{B}_0 and k is equivalent to reversing the x and z coordinates in the original coordinates shown in Fig. 2.6-1a. The new coordinates are shown in Figure 2.6-2 and the dielectric tensor expressed in them must be the transposed dielectric tensor expressed in the coordinates shown in Fig. 2.6-1a. Note that the new coordinate system Fig. 2.6-2 is still right-handed. In each coordinate system (Figs. 2.6-1a and 2.6-2) we have

$$\mathbf{D} = \mathbf{\epsilon} \cdot \mathbf{E}$$
$$\mathbf{D}' = \mathbf{\epsilon}' \cdot \mathbf{E}'$$
(2.6.4)

where the primed system is related to the original system through the transform \mathbf{A} .

$$\mathbf{D}' = \mathbf{A} \cdot \mathbf{D}, \qquad \mathbf{E}' = \mathbf{A} \cdot \mathbf{E} \tag{2.6.5}$$



Fig. 2.6-2. Transformed coordinate system.

where

$$\mathbf{A} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$
(2.6.6a)

and with an inverse

$$\mathbf{A}^{-1} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$
(2.6.6b)

Therefore,

$$\boldsymbol{\varepsilon}' = \mathbf{A} \cdot \boldsymbol{\varepsilon} \cdot \mathbf{A}^{-1} = \begin{bmatrix} \varepsilon_{xx} & -\varepsilon_{xy} & \varepsilon_{xz} \\ -\varepsilon_{yx} & \varepsilon_{yy} & -\varepsilon_{yz} \\ \varepsilon_{zy} & -\varepsilon_{zy} & \varepsilon_{zz} \end{bmatrix}$$
(2.6.7)

But $\mathbf{\epsilon}'$ is obtained by reversing **k** and **B**₀; therefore, from the Onsager relation, we have

$$\boldsymbol{\varepsilon}'(\mathbf{k},\,\omega,\,\mathbf{B}_0) = \boldsymbol{\varepsilon}^{\mathrm{T}}(\mathbf{k},\,\omega,\,\mathbf{B}_0) \tag{2.6.8}$$

where $\boldsymbol{\varepsilon}^{T}$ denotes the transpose of the tensor $\boldsymbol{\varepsilon}$. From (2.6.7) and (2.6.3), it follows

$$\varepsilon_{xy} = -\varepsilon_{yx}$$

 $\varepsilon_{yz} = -\varepsilon_{zy}$

 $\varepsilon_{zz} = \varepsilon_{xz}$
(2.6.9)

Hence, in general, the Onsager relation reduces ε to six independent elements. For the original Cartesian coordinates (Fig. 2.6-1a), the dielectric tensor must have the form

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ -\varepsilon_{xy} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{xz} & -\varepsilon_{yz} & \varepsilon_{zz} \end{bmatrix}$$
(2.6.10)

If, in addition, the medium is lossless so that $\boldsymbol{\varepsilon}$ is Hermitian, then ε_{xx} , ε_{yy} , ε_{zz} , and ε_{xz} are purely real, ε_{xy} and ε_{yz} are purely imaginary. Also, since $\varepsilon_{xy}(\mathbf{B}_0) = -\varepsilon_{yx}(\mathbf{B}_0)$, $\varepsilon_{yz}(\mathbf{B}_0) = -\varepsilon_{zy}(\mathbf{B}_0)$ and $\varepsilon_{zz}(\mathbf{B}_0) = \varepsilon_{zx}(\mathbf{B}_0)$ from (2.6.9), and according to the Onsager relation $\varepsilon_{ij}(\mathbf{B}_0) = \varepsilon_{ji}(-\mathbf{B}_0)$, we may conclude that ε_{xx} , ε_{yy} , ε_{zz} and ε_{xz} are even in \mathbf{B}_0 while ε_{xy} and ε_{yz} are odd in \mathbf{B}_0 .

If the medium is not spatially dispersive the only axis of symmetry is the z-axis along which is the external magnetic field \mathbf{B}_0 (see Fig. 2.6-1a). For such coordinates $\boldsymbol{\varepsilon}$ must be rotationally symmetric about z-axis. This requires that $\varepsilon_{xz} = \varepsilon_{yz} = 0$ and $\varepsilon_{xx} = \varepsilon_{yy}$ (see problem at the end of this chapter).

2.7 Plane Waves

In the above, we have discussed the material equations for a homogeneous, stationary medium. With the help of these relations, we are ready to proceed to solving the set of Maxwell's equations for the case of unbounded region. As mentioned earlier, because of the spatial homogeneity and temporal stationarity in the medium, it makes it possible to apply the Fourier transform to the problem. This actually is equivalent to the method of plane wave solution. The Fourier components of the field quantities such as $E(\mathbf{k}, \omega)$, $D(\mathbf{k}, \omega)$, etc., are the corresponding amplitudes of the different plane waves, $E(\mathbf{k}, \omega)e^{j(\omega t-\mathbf{k}\cdot\mathbf{r})}$, $D(\mathbf{k}, \omega)e^{j(\omega t-\mathbf{k}\cdot\mathbf{r})}$.

Let us apply the Fourier transform of the type in (2.4.18) to (2.1.1a) and (2.1.1b). We have

$$-j\mathbf{k} \times \mathbf{E}(\mathbf{k},\omega) = -j\omega\mathbf{B}(\mathbf{k},\omega) \qquad (2.7.1a)$$

$$-j\mathbf{k} \times \mathbf{H}(\mathbf{k},\omega) = \mathbf{J}(\mathbf{k},\omega) + j\omega\mathbf{D}(\mathbf{k},\omega) \qquad (2.7.1b)$$

In this section, we shall study the system in the absence of external current so that J = 0 in (2.7.1b). We observe first the following relations between the Fourier components of the field quantities. From (2.7.1a), (2.7.1b), we have

$$\mathbf{H}(\mathbf{k},\omega) = -\frac{1}{\omega\mu_0} \mathbf{k} \times \mathbf{E}(\mathbf{k},\omega)$$
(2.7.2a)

$$\mathbf{D}(\mathbf{k},\omega) = -\frac{1}{\omega} \mathbf{k} \times \mathbf{H}(\mathbf{k},\omega) \qquad (2.7.2b)$$

In general all vectors appearing in (2.7.2a) and (2.7.2b) are complex even for a real angular frequency ω . If the medium is unbounded and lossless, the vector **k** is real in the propagating region (see proof in Section 10). For this case these equations tell us that Re H \perp **k**, Re D \perp Re H, Re D \perp **k**; and **k**, Re D and Re H form a right-handed rectangular coordinate system. Similar relations also exist among **k**, Im D, and Im H. Also, since Re H \perp **k**, Re H \perp Re D, Re H \perp Re E, the three vectors **k**, Re D and Re E must be in the same plane, the plane perpendicular to Re H. In Fig. 2.7-1, we demonstrate these relations graphically. Similar diagrams can be drawn for **k**, Im E, etc. For a medium there may exist relations between Re E and Im E. If such a relation exists, it describes completely the state of polarization of the wave. Such a polarization is called the characteristic polarization and is discussed in Section 9. Waves with the characteristic polarization are called normal modes or characteristic waves.



Fig. 2.7-1. Vector relation of plane waves in the nonisotropic medium.

In characteristic waves the energy flow is given by the Poynting vector

$$\mathbf{S} = \operatorname{Re} \frac{1}{2} \mathbf{E} \times \mathbf{H}^* = [\operatorname{Re} \mathbf{E} \times \operatorname{Re} \mathbf{H} + \operatorname{Im} \mathbf{E} \times \operatorname{Im} \mathbf{H}]/2 \quad (2.7.3)$$

We see that in general S is not in the same direction as k. In an isotropic medium without spatial dispersion, D and E are in phase and in the same direction and hence k and S must be parallel.

With the help of the material equations, we can eliminate H from (2.7.1a) and (2.7.1b) and obtain

$$\mathbf{k} \times (\mathbf{k} \times \mathbf{E}) + (\omega^2/c^2)\mathbf{K} \cdot \mathbf{E} = 0$$
 (2.7.4)

Or, in component form

$$[k^2 \,\delta_{ij} - (\omega^2/c^2)K_{ij} - k_i k_j]E_j = 0 \tag{2.7.5}$$

This is a set of three homogeneous algebraic equations. We shall study them in detail in the following. Define the index of refraction

$$\mathbf{n} = \mathbf{k}/k_0 = \mathbf{k}c/\omega = n\mathbf{\hat{s}} \tag{2.7.6}$$

where $k_0 = \omega/c$ is the free-space wave number and \mathfrak{s} is a unit vector in the direction of **k**. Since it is for an infinite medium, we shall consider only homogeneous plane waves such that \mathfrak{s} is a real vector. Consider the equation

$$L_{ij}a_j = [n^2(\delta_{ij} - s_i s_j) - K_{ij}]a_j = 0$$
(2.7.7)

Equation (2.7.7) has a nontrivial solution if and only if the determinant $|L_{ij}|$ vanishes. This condition

$$|L_{ij}| = |n^2(\delta_{ij} - s_i s_j) - K_{ij}| = 0$$
(2.7.8)

is the dispersion relation which gives the relations between ω and **n** (or ω

and **k**) for which (2.7.7) has a nontrivial solution. More than one relation between ω and **n** may exist which correspond to more than one solutions of (2.7.7). Each of these solutions is defined as one normal mode or characteristic mode. We have, for the solution of (2.7.8)

$$n_{\alpha}^{2} = n_{\alpha}^{2}(\omega), \qquad \alpha = 1, 2, \dots$$
 (2.7.9)

where the subscript α denotes the different modes. The corresponding solution of (2.7.7) is then of the form $a_{j\alpha}e^{j[\omega t-k_{\alpha}\cdot r]}$ Physically, this corresponds to one particular mode of plane waves that can propagate in the medium. A general wave in the medium can then be expanded in terms of these normal plane waves. In the following we shall discuss some general properties of these modes for arbitrary K_{ij} . Let us consider (2.7.7) again for the α th mode

$$[n_{\alpha}^{2}(\delta_{ij}-s_{i}s_{j})-K_{ij}(\mathbf{n}_{\alpha},\omega)]a_{j\alpha}=0 \qquad (2.7.10)$$

Since the determinant of L^T is the same as that of L itself, we can introduce the following system of equations for the conjugate vectors b^*

$$[n_{\beta}^{2}(\delta_{ij} - s_{i}s_{j}) - K_{ji}]b_{j\beta}^{*} = 0$$
(2.7.11)

where the set of values *n* is obtained from equation $|\mathbf{L}^T| = 0$ and is the same as shown in (2.7.9). We note that if **K** is Hermitian, then **a** and **b** are the same. Multiplying (2.7.10) by $b_{i\beta}^*$ and (2.7.11) by $a_{i\alpha}$, we have

$$n_{\alpha}^{2}(\delta_{ij} - s_{i}s_{j})b_{i\beta}^{*}a_{j\alpha} - K_{ij}b_{i\beta}^{*}a_{j\alpha} = 0 \qquad (2.7.12)$$

$$n_{\beta}^{2}(\delta_{ij} - s_{i}s_{j})b_{j\beta}^{*}a_{i\alpha} - K_{ji}b_{j\beta}^{*}a_{i\alpha} = 0 \qquad (2.7.13)$$

Interchanging the indices i and j in (2.7.13) and subtracting it from (2.7.12), we obtain

$$(n_{\alpha}^{2} - n_{\beta}^{2})(\delta_{ij} - s_{i}s_{j})b_{i\beta}^{*}a_{j\alpha} = 0 \qquad (2.7.14)$$

If (2.7.8) has distinct roots, then for $\alpha \neq \beta$, $n_{\alpha}^2 \neq n_{\beta}^2$, the factor multiplying $(n_{\alpha}^2 - n_{\beta}^2)$ in (2.7.14) must be zero. For $\alpha = \beta$, this factor can be any arbitrary number since **a** and **b** are only determined within some constant factor. We shall choose the constant factor by the following orthonormality relation

$$(\delta_{ij} - s_i s_j) b^*_{i\beta} a_{j\alpha} = \delta_{\alpha\beta}$$
 (2.7.15)

and from (2.7.12) it follows that

$$K_{ij}a_{j\alpha}b_{i\beta}^* = n_{\alpha}^2 \,\delta_{\alpha\beta} \tag{2.7.16}$$

Equations (2.7.15) and (2.7.16) are important relations between the different characteristic modes of (2.7.7). They will be used later on in the computation of fields generated by external sources in an anisotropic medium. These normalization relations are obtained on the assumption that the dispersion equation has distinct roots. In an isotropic medium, multiple roots do exist. Then, instead of the normalization procedure described above, some other means must be used.

2.8 Refractive Indices

In order to study the characteristic modes of (2.7.7), we must first obtain the relations n_{α} of (2.7.8). When (2.7.8) is expanded, the following algebraic equation results

$$K_{ij}s_is_jn^4 - [(K_{ij}s_is_j)K_{kk} - K_{ik}K_{kj}s_is_j]n^2 + |K_{ij}| = 0 \qquad (2.8.1)$$

where $|K_{ij}|$ denotes the determinate for the tensor K_{ij} and once again the reader is reminded that repeated subscripts represent summation. From (2.8.1), we note that if there is no spatial dispersion so that K_{ij} does not depend on *n*, then (2.8.1) is quadratic in n^2 and will yield two independent solutions for n^2 in general which correspond to two characteristic modes in the medium. If spatial dispersion is taken into account, however, (2.8.1) will have more than two roots for n^2 .

Before we go into the solutions of (2.8.1) for some specific examples, we want to point out one important general property of n^2 . If the tensor K is Hermitian which corresponds to a lossless medium (see Section 10) so that $K_{ij} = K_{ji}^*$, then from (2.7.7) we have

$$a_i^*[n^2(\delta_{ij}-s_is_j)-K_{ij}]a_j=0$$

or

$$n^{2}(a^{2} - s_{i}s_{j}a_{i}^{*}a_{j}) = K_{ij}a_{i}^{*}a_{j}$$
(2.8.2)

Also, taking the complex conjugate of (2.7.7), we have

$$(n^2)^*(\delta_{ij} - s_i s_j) a_i^* = K_{ji}^* a_i^* = K_{ij} a_i^*$$
(2.8.3)

where the Hermitian property of **K** has been used. Multiplying (2.8.3) by a_j , we obtain

$$(n^2)^*(a^2 - s_i s_j a_i^* a_j) = K_{ij} a_i^* a_j \tag{2.8.4}$$

Comparing (2.8.2) with (2.8.4), we conclude that

$$n^2 = (n^2)^* \tag{2.8.5}$$

so that for a lossless medium, n^2 must be real. Consequently, *n* must be purely real (propagating mode) or purely imaginary (evanescent mode). The wave number being given by $k = k_0 n$ must be similarly purely real (propagating mode) or purely imaginary (evanescent mode).

Now, let us consider some special cases for (2.8.1) which will be of interest to us later on. First, for an isotropic medium without spatial dispersion so that $K_{ij} = \delta_{ij} K(\omega)$, Eq. (2.8.2) becomes

$$n^4 - 2K(\omega)n^2 + K^2 = 0 \tag{2.8.6}$$

The equation has double roots:

$$n^2 = K(\omega) \tag{2.8.7}$$

Second, for an isotropic medium with spatial dispersion, the relative dielectric tensor can be written in the form given by (2.4.20a):

$$K_{ij}(k,\omega) = K_{\perp}(n^2,\omega)(\delta_{ij} - s_i s_j) + K_{\parallel}(n^2,\omega)s_i s_j \qquad (2.4.20a)$$

For this case, (2.8.1) reduces to

$$K_{\parallel}(n^4 - 2K_{\perp}n^2 + K_{\perp}^2) = 0 \qquad (2.8.8)$$

If K_{\perp} does not depend on n^2 , (2.8.8) again yields two double roots:

$$n^2 = K_\perp(\omega) \tag{2.8.9}$$

We will see in Section 9 that both (2.8.8) and (2.8.9) correspond to propagation of waves with electric field perpendicular to the direction of $k(E \perp \hat{s})$, called the transverse waves, in an isotropic medium. The other solution of (2.8.8) is

$$K_{\parallel}(n^2,\omega) = 0 \tag{2.8.10}$$

which corresponds to longitudinal waves (E $\parallel \hat{s}$) in the medium. We shall discuss this point again in Section 9.

Finally, let us study the anisotropic case. We shall orient the coordinates such that the vector **k** is in the *xz*-plane and the external magnetic field is in the *z*-direction. The angle between **k** and **B**₀ is θ (Fig. 2.6-1a). The dielectric tensor is given by (2.6.10). Substituting the components of (2.6.10) into

(2.8.1), the following equation is obtained:

$$a_4n^4 + a_2n^2 + a_0 = 0 \tag{2.8.11}$$

where

$$a_{4} = K_{xx} \sin^{2} \theta + 2K_{xz} \cos \theta \sin \theta + K_{zz} \cos^{2} \theta$$

$$a_{2} = -K_{xx}K_{zz} + K_{xz}^{2} - (K_{xx}K_{yy} + K_{xy}^{2}) \sin^{2} \theta$$

$$+ 2(K_{xy}K_{yz} - K_{xz}K_{yy}) \cos \theta \sin \theta - (K_{yz}^{2} + K_{yy}K_{zz}) \cos^{2} \theta$$

$$a_{0} = |\mathbf{K}|$$
(2.8.12)

and the relations $s_1 = \sin \theta$, $s_2 = 0$, $s_3 = \cos \theta$ have been used.

For a medium without spatial dispersion, the components K_{ij} 's are all independent of *n*. The solutions to (2.8.11) can be written as

$$n^{2} = \frac{-a_{2} \pm (a_{2}^{2} - 4a_{4}a_{0})^{1/2}}{2a_{4}}$$
(2.8.13)

or

$$n^{2} = 1 - \frac{2(a_{4} + a_{2} + a_{0})}{2a_{4} + a_{2} \pm (a_{2}^{2} - 4a_{4}a_{0})^{1/2}}$$
(2.8.14)

As discussed at the end of Section 6, for the spatially nondispersive case, $K_{xz} = K_{yz} = 0$ and $K_{xx} = K_{yy}$. Inspection of the coefficients given by (2.8.12) shows that the refractive index has the same value at $180^{\circ} - \theta$ as at θ . Equations (2.8.13) and (2.8.14) represent two independent solutions of the dispersion equation and correspond to two characteristic waves that can propagate in the medium. For spatial dispersive media, as mentioned earlier, there may exist more than two solutions for (2.8.11).

The condition for which n = 0 is called cutoff because it divides the propagating region from the attenuating region in the lossless case. The phase velocity v_p of the plane wave for this case becomes infinite. From (2.8.11) we see that cutoff occurs when $a_0 = |\mathbf{K}| = 0$. At cutoff, $\mathbf{H} = \mathbf{D} = 0$. Since the determinant $|\mathbf{K}|$ is invariant under rotation of transformation of the coordinates, the cutoff condition does not depend on the direction of propagation.

The condition for which $n^2 \to \infty$ is called resonance. It occurs when $a_4 = 0$ in (2.8.13). At resonance, $v_p = 0$. Since a_4 in (2.8.12) depends on θ , the resonance condition depends on the direction of propagation. From (2.8.12), we obtain, for resonance

$$K_{xx}\tan^2\theta_r + 2K_{xz}\tan\theta_r + K_{zz} = 0 \qquad (2.8.15)$$

which represents in general two resonance cones. For angles near θ_r , the phase velocity is much below the velocity of light in free space and the Čerenkov radiation is possible.

2.9 Characteristic Polarizations

We have seen that for plane waves, Maxwell's equations can be written in the form

$$\mathbf{D}(\mathbf{k},\omega) = -(n/c)\mathbf{\hat{s}} \times \mathbf{H}(\mathbf{k},\omega)$$
(2.9.1a)

$$\mathbf{H}(\mathbf{k},\omega) = (\varepsilon_0/\mu_0)^{1/2} n \hat{s} \times \mathbf{E}(\mathbf{k},\omega)$$
(2.9.1b)

$$\hat{\mathbf{s}} \cdot \mathbf{D}(\mathbf{k}, \omega) = 0 \tag{2.9.1c}$$

$$\hat{s} \cdot \mathbf{B}(\mathbf{k}, \omega) = 0 \tag{2.9.1d}$$

$$\mathbf{D}(\mathbf{k},\omega) = n^2 \varepsilon_0 [\mathbf{E} - \hat{\mathbf{s}}(\hat{\mathbf{s}} \cdot \mathbf{E})]$$
(2.9.1e)

Sometimes for specific media, for certain directions, it is possible to have independent transverse modes or longitudinal modes propagating in the media. For transverse modes, $\mathbf{E} \perp \hat{s}$, so that $\hat{s} \cdot \mathbf{E} = 0$. From (2.9.1e), we have

$$\mathbf{D} = n^2 \varepsilon_0 \mathbf{E} \tag{2.9.2}$$

Hence, **D** and **E** are in the same direction. For these modes to exist, **D** and **E** must simultaneously satisfy the condition $\hat{s} \cdot \mathbf{E} = 0$ and Eq. (2.9.2) as well as the material equation $D_i = \varepsilon_{ij}E_j$ which amounts to four equations $(\hat{s} \cdot \mathbf{E} = 0, n^2\varepsilon_0E_i = \varepsilon_{ij}E_j)$ for the three components E_x , E_y , and E_z . Therefore, in general there is no solution. Only for special cases, purely transverse modes can occur.

For longitudinal modes, $\mathbf{E} \parallel \hat{s}$, so that $\mathbf{E} = (\mathbf{E} \cdot \hat{s})\hat{s}$. From (2.9.1e) and (2.9.1b), we have

$$\mathbf{D} = \mathbf{0}, \qquad \mathbf{H} = \mathbf{0} \tag{2.9.3}$$

The condition $\mathbf{D} = 0$ can be written as

$$\mathbf{D} = \varepsilon_0 \mathbf{K} \cdot \mathbf{E} = 0 \tag{2.9.4}$$

Therefore in order to have nontrivial longitudinal modes, it is necessary to require

$$|\mathbf{K}| = 0 \tag{2.9.5}$$

Note that (2.9.5) is also the condition for cutoff. It should be emphasized that the condition (2.9.5) is only a necessary condition but not a sufficient condition. When (2.9.5) is satisfied, it guarantees that **D** vanishes, but it does not necessarily follow that **E** is longitudinal. In a nonisotropic, spatially dispersive medium, strict longitudinal waves may not exist. In this case, we can expect the waves to be, at best, approximately longitudinal. This is discussed at the end of this section.

Now, let us study the characteristic modes for an isotropic medium. In this case it is possible to decompose the field into transverse and longitudinal components

$$\mathbf{a} = \mathbf{a}_{\perp} + \mathbf{a}_{\parallel} \tag{2.9.6}$$

Then for K_{ij} given by (2.4.20a), (2.7.7) can be reduced to

$$(n^2 - K_{\perp})\mathbf{a}_{\perp} = 0$$

$$K_{\parallel}\mathbf{a}_{\parallel} = 0$$
(2.9.7)

We see that in an isotropic medium, it is possible to have independent transverse and longitudinal modes. The dispersion relation for the transverse modes is

$$n^2 - K_{\perp} = 0 \tag{2.9.8}$$

and the dispersion relation for the longitudinal modes is

$$K_{\parallel}(\mathbf{k},\omega) = 0 \tag{2.9.9}$$

We note that (2.9.9) also satisfies the necessary condition (2.9.5). We now turn to the case of anisotropic media.

In an anisotropic medium the normal modes are neither transverse nor longitudinal in general. This can be seen by substituting (2.9.6) into (2.7.7) for general K_{ij} . It is no longer possible to separate \mathbf{a}_{\perp} and \mathbf{a}_{\parallel} in this case.

Let us now find the characteristic modes for general anisotropic media. A characteristic or normal mode is defined as a wave whose polarization remains the same as propagating in the homogeneous medium. To derive an expression for the wave polarization, it is more convenient to change to a coordinate system in which k coincides with the z-axis. This is done by rotating the coordinate system in Fig. 2.9-1a about the y-axis by an angle θ in the clockwise sense. The transformation is represented by the matrix

$$\mathbf{T} = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix}$$
(2.9.10)



Fig. 2.9-1. Coordinate systems.

In the primed system the relative dielectric tensor is obtained by applying the transformation to \mathbf{K} in (2.6.10)

$$\mathbf{K}' = \mathbf{T} \cdot \mathbf{K} \cdot \mathbf{T}^{-1} \tag{2.9.11}$$

Its components are given by

$$\begin{aligned} K'_{xx} &= K_{xx}\cos^2\theta + K_{zz}\sin^2\theta - 2K_{xz}\sin\theta\cos\theta\\ K'_{xy} &= -K'_{yx} = K_{xy}\cos\theta + K_{yz}\sin\theta\\ K'_{zz} &= K'_{zx} = (K_{xx} - K_{zz})\sin\theta\cos\theta + K_{xz}(\cos^2\theta - \sin^2\theta) \quad (2.9.12)\\ K'_{yy} &= K_{yy}, \qquad K'_{yz} = -K'_{zy} = K_{yz}\cos\theta - K_{xy}\sin\theta\\ K'_{zz} &= K_{xx}\sin^2\theta + K_{zz}\cos^2\theta + 2K_{xz}\sin\theta\cos\theta \end{aligned}$$

Equations (2.8.11), (2.8.12), and the solutions for the refractive indices are unchanged in the primed system. The wave equation (2.7.7) becomes

$$\begin{bmatrix} K'_{xx} - n^2 & K'_{xy} & K'_{xz} \\ -K'_{xy} & K'_{yy} - n^2 & K'_{yz} \\ K'_{xz} & -K'_{yz} & K'_{zz} \end{bmatrix} \begin{bmatrix} a'_{x'} \\ a'_{y'} \\ a'_{z'} \end{bmatrix} = 0$$
(2.9.13)

In the following we shall study the case without spatial dispersion. Let us write the characteristic mode in the form

$$\mathbf{a}' = a_{y}'(R_{x}'\hat{x}' + \hat{y}' + R_{z}'\hat{z}') \qquad (2.9.14)$$

where

$$R_x' = a_x'/a_y', \qquad R_z' = a_z'/a_y'$$
 (2.9.15)

Substituting (2.9.14) into (2.9.13), the third equation gives us

$$R_{z}' = (-R_{x}'K_{xz}' + K_{yz}')/K_{zz}'$$
(2.9.16)

 R_z' can be eliminated from the first two equations of (2.9.13) by using (2.9.16). They give, respectively,

$$[n^{2} - (K'_{xx} - K'^{2}_{xz}/K'_{zz})]R'_{x} - K'_{xy} - K'_{xz}K'_{yz}/K'_{zz} = 0$$

$$n^{2} + (K'_{xy} + K'_{yz}K'_{xz}/K'_{zz})R'_{x} - K'_{yy} - K'^{2}_{yz}/K'_{zz} = 0$$
(2.9.17)

Each of these two equations gives us a relation connecting the refractive index to the transverse polarization of the wave. They are

$$n^{2} = \frac{K'_{xy} + K'_{xz}K'_{yz}/K'_{zz}}{R'_{x'}} + K'_{xx} - K'^{2}_{xz}/K'_{zz}$$

$$n^{2} = K'_{yy} + K'^{2}_{yz}/K'_{zz} - (K'_{xy} + K'_{yz}K'_{xz}/K'_{zz})R'_{x'}$$
(2.9.18)

Setting the two expressions in (2.9.18) equal to each other, we obtain an equation for R_x'

$$R_{x}^{\prime 2} + \frac{K_{xx}^{\prime} - K_{yy}^{\prime} - (K_{xz}^{\prime 2} + K_{yz}^{\prime 2})/K_{zz}^{\prime}}{K_{xy}^{\prime} + K_{yz}^{\prime}K_{xz}^{\prime}/K_{zz}^{\prime}} R_{x}^{\prime} + 1 = 0 \qquad (2.9.19)$$

The two roots of (2.9.19) for R_x' are referred to as characteristic polarizations for the two normal modes. They also correspond to the two values of *n*. From (2.9.19), it follows that

$$R_{x1}'R_{x2}' = 1 \tag{2.9.20}$$

The transverse part of the characteristic modes is

$$\mathbf{a}_{\perp} = a_{y}'(R_{x}'\hat{x}' + \hat{y}') \tag{2.9.21}$$

Since for lossless medium, K'_{xx} , K'_{yy} , K'_{zz} and K'_{xz} are real and K'_{xy} and K'_{yz} are imaginary, R_x' is purely imaginary, the vector \mathbf{a}_{\perp}' is elliptically polarized in the x'y'-plane. The sense of rotation and the ratio between the major axis to minor axis of the ellipse depend on the value of R_x' . Since $R'_{x1}R'_{x2} = 1$, the two ellipses for the two characteristic modes are perpendicular to each other, one with major axis aligned with the x'-axis, one with major axis aligned with the y'-axis, and they have opposite senses of rotation. In fact, the two polarization ellipses are mirror images about a line making 45° with respect to the x'-axis.

Similar arguments show that R_z' must be purely imaginary for both characteristic modes.

The two solutions of (2.9.19) can be expressed in terms of the components of K_{ij} . Since the medium is assumed to be without spatial dispersion, the

tensor K_{ij} in the unprimed system must be invariant under rotation about the external magnetic field \mathbf{B}_0 , which is the z-axis. Therefore, $K_{xx} = K_{yy}$, $K_{xz} = K_{yz} = 0$. Under these conditions

$$R_{x}' = \frac{K_{xy}K_{zz}\cos\theta}{n^{2}(K_{xx}\sin^{2}\theta + K_{zz}\cos^{2}\theta) - K_{xx}K_{zz}}$$
(2.9.22)

$$R_{z}' = \frac{K_{xy} \sin \theta (K_{zz} - n^{2})}{n^{2} (K_{xx} \sin^{2} \theta + K_{zz} \cos^{2} \theta) - K_{xx} K_{zz}}$$
(2.9.23)

It is interesting to note that in an isotropic medium, $K_{xx} = K_{yy} = K_{zz}$ and all the off-diagonal components vanish; also $n^2 = K_{zz}$. Then (2.9.22) and (2.9.23) become indeterminant. Physically, this implies that any polarization is a characteristic polarization in an isotropic medium which certainly is what one would expect.

Next, we want to find the amplitude of the characteristic modes that satisfy the orthonormal condition (2.7.15). To do this, we make use of the orthonormal relations derived in Section 7. We first must find the conjugate vectors b^* in (2.7.11). This can be done in a manner similar to the case for finding the vectors **a**. In the primed system, solutions of (2.7.11) can be written in the form

$$\mathbf{b}^{\prime *} = a_{y}^{\prime}(-R_{x}^{\prime}\hat{x}^{\prime} + \hat{y}^{\prime} - R_{z}^{\prime}\hat{z}^{\prime}) \qquad (2.9.24)$$

where R_x' and R_z' are given in (2.9.22) and (2.9.23). Substituting (2.9.14) and (2.9.24) into (2.7.15), and remembering that **k** is parallel to z'-axis the following is obtained:

$$a'_{y\alpha}a'_{y\beta}(1-R'_{x\alpha}R'_{x\beta})=\delta_{\alpha\beta} \qquad (2.9.25)$$

This orthonormal relation is automatically satisfied for $\alpha \neq \beta$ as it should be, since $R'_{x\alpha}R'_{x\beta} = 1$. For $\alpha = \beta$, we have

$$a_{y}' = 1/(1 - R_{x}'^{2})^{1/2}$$
 (2.9.26)

Therefore, the normalized characteristic modes are given by

$$\mathbf{a}_{\alpha}' = (R'_{x\alpha} \hat{x}' + \hat{y}' + R'_{z\alpha} \hat{z}') / (1 - R'^{2}_{x\alpha})^{1/2}, \quad \alpha = 1, 2 \quad (2.9.27)$$

This formula is valid for a lossy anisotropic medium in the absence of spatial dispersion. In the original unprimed system, the characteristic modes can be obtained from (2.9.27) by coordinate transformation.

$$\mathbf{a}_{\alpha} = T^{-1} \cdot \mathbf{a}_{\alpha}'$$

= $[(R'_{x\alpha}\cos\theta + R'_{z\alpha}\sin\theta)\hat{x} + \hat{y} + (R'_{z\alpha}\cos\theta - R'_{x\alpha}\sin\theta)\hat{z}]/(1 - R'^{2}_{x\alpha})^{1/2}$
(2.9.28)

In the following we demonstrate the utility of the expressions just derived, by considering two special cases for a lossless medium.

(i) Parallel Case ($\mathbf{k} \parallel \mathbf{B}_0$). In this case $\theta = 0$. The coefficients in the dispersion relation are given by (2.8.12) and they simplify to

$$a_4 = K_{zz}, \quad a_2 = -2K_{xx}K_{zz}, \quad a_0 = K_{zz}(K_{xx}^2 + K_{xy}^2)$$
 (2.9.29)

The solutions of the refractive index equation (2.8.11) become

$$n_{1,2}^2 = K_{xx} \pm jK_{xy} \tag{2.9.30}$$

where $K_{zz} \neq 0$ is assumed. If $K_{zz} = 0$, the electric field is polarized in the z-direction corresponding to the longitudinal mode.

Equations (2.9.22) and (2.9.23) yield the polarizations for the characteristic modes in the primed system.

$$R'_{x1} = -j, \qquad R'_{x2} = j, \qquad R'_{z1} = R'_{z2} = 0$$
 (2.9.31)

In the unprimed system, we have

$$R_{x1} = -j, \qquad R_{z1} = 0$$

$$R_{x2} = j, \qquad R_{z2} = 0$$
(2.9.32)

Therefore the two characteristic modes are

$$\mathbf{a}_{1} = (-j\hat{x} + \hat{y})/\sqrt{2}$$

$$\mathbf{a}_{1} = (j\hat{x} + \hat{y})/\sqrt{2}$$

(2.9.33)

Both are circularly polarized. The first one is in the left-handed sense while the second one is right-handed. The two modes are both pure transverse waves.

(ii) Perpendicular Case $(\mathbf{k} \perp \mathbf{B}_0)$. The angle between the external magnetic field and the wave vector is 90°. The refractive indices are now

$$n_1^2 = K_{zz}$$

$$n_2^2 = K_{xx} + K_{xy}^2 / K_{xx}$$
(2.9.34)

For $n_1^2 = K_{zz}$, (2.9.22) shows that R_x' becomes indeterminant. Actually, for this case the formulas derived are no longer valid. Since (2.9.14) assumed that the y'-component of the wave does not vanish, while in this case, it

does. To find the correct characteristic mode corresponding to $n_1^2 = K_{zz}$, we have to go back to (2.9.13). Under the present conditions, (2.9.13) becomes

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & K_{xx} - K_{zz} & -K_{xy} \\ 0 & K_{xy} & K_{xx} \end{bmatrix} \begin{bmatrix} a'_x \\ a'_y \\ a'_z \end{bmatrix} = 0$$
(2.9.35)

which implies $a_{y'}$ and $a_{z'}$ are zero and $a_{x'}$ is arbitrary. The wave is linearly polarized in the \hat{x}' -direction (the direction of the external field \mathbf{B}_0). This is sometimes referred to as the ordinary mode. The fact that the wave is polarized along the magnetic field indicates that it is not affected by the magnetic field so that the refractive index is just K_{zz} .

For the second mode given by (2.9.34) we have,

$$R'_{x2} = R_{z2} = 0$$

$$R'_{z2} = R_{x2} = \frac{K_{xy}(K_{zz} - K_{xx} - K^2_{xy}/K_{xx})}{K^2_{xx} + K^2_{xy} - K_{xx}K_{zz}} = -\frac{K_{xy}}{K_{xx}}$$
(2.9.36)

Therefore, the normalized expression is given by

$$\mathbf{a}_2 = -(K_{xy}/K_{xx})\hat{x} + \hat{y} \tag{2.9.37}$$

Since this is a lossless medium, K_{xy} is purely imaginary, K_{xx} is purely real. The second wave given by (2.9.37) is therefore elliptically polarized in the xy-plane, rotating in a left-handed sense.

We summarize the discussion in this section as follows. In an isotropic medium with spatial dispersion, there exist independent transverse and longitudinal characteristic modes. While in an anisotropic medium, this is not true in general. For anisotropic media in the absence of spatial dispersion there are two independent characteristic modes. The transverse components of these characteristic waves are in general elliptically polarized. These two characteristic polarization ellipses form mirror images about a line making a 45° angle with the magnetic meridian plane. Even though the exact longitudinal wave may not be obtainable in a nonisotropic medium, it is possible that the component of the electric field transverse to **k** is so small when compared with the longitudinal component, the wave is then approximately longitudinal. Let

$$\mathbf{E} = \mathbf{E}_{\parallel} + \mathbf{E}_{\perp} \tag{2.9.38}$$

where \mathbf{E}_{\parallel} is parallel to **k** and \mathbf{E}_{\perp} is perpendicular to **k**. Substitute (2.9.38) in (2.7.4) to obtain

$$k^{2}\mathbf{E}_{\perp} = (\omega^{2}/c^{2})\mathbf{K} \cdot (\mathbf{E}_{\parallel} + \mathbf{E}_{\perp})$$
(2.9.39)

The wave is approximately longitudinal if

$$|\mathbf{E}_{\parallel}| \gg |\mathbf{E}_{\perp}| \tag{2.9.40}$$

It follows from (2.9.39) that (2.9.40) can be so only if

$$k^2 \gg (\omega^2/c^2) \parallel \mathbf{K} \parallel \tag{2.9.41}$$

Equation (2.9.41) implies that approximate longitudinal waves are associated with the large values of the refractive index for which the propagation velocity is slow. When referenced with (2.9.39), the condition (2.9.40) also implies that

$$\mathbf{k} \cdot \mathbf{K} \cdot \mathbf{k} = 0 \quad \text{or} \quad a_4 = 0 \tag{2.9.42}$$

where a_4 is given by (2.8.12). This condition is used in Section 4.20 where longitudinal waves in a warm magnetoplasma are discussed.

2.10 Energy and Power

One of the most important topics in electrodynamics is the relation for energy conservation. In order to discuss it, we shall first derive the fundamental Poynting's theorem.

Let us take the scalar products of **H** with (2.1.1a) and **E** with (2.1.1b) and subtract one from the other. Applying the vector identity $\nabla \cdot (\mathbf{E} \times \mathbf{H})$ = $\mathbf{H} \cdot \nabla \times \mathbf{E} - \mathbf{E} \cdot \nabla \times \mathbf{H}$, we obtain

$$\nabla \cdot (\mathbf{E} \times \mathbf{H}) = -(\mathbf{H} \cdot \dot{\mathbf{B}} + \mathbf{E} \cdot \dot{\mathbf{D}}) - \mathbf{E} \cdot \mathbf{J}$$
(2.10.1)

Integrating over the volume V and applying the divergence theorem, we have

$$\int_{V} (\mathbf{H} \cdot \dot{\mathbf{B}} + \mathbf{E} \cdot \dot{\mathbf{D}}) \, dv = -\int_{S} (\mathbf{E} \times \mathbf{H}) \cdot d\mathbf{S} - \int_{V} \mathbf{E} \cdot \mathbf{J} \, dv \quad (2.10.2)$$

We note that $\mathbf{E} \times \mathbf{H} = \mathbf{S}$ is the instantaneous Poynting vector which defines the flux of electromagnetic energy. Equation (2.10.2) is referred to as the Poynting theorem which essentially is the statement of the principle of conservation of energy. In the absence of loss, the two terms in the lefthand side represent the instantaneous rates of change of magnetic and electric energies, respectively. And the two terms on the right-hand side represent the power carried away by the wave and the power supplied by the source, respectively. But in a time-varying field in the presence of absorption, the meaning of each term in (2.10.2) is in general not clear. Let us now first discuss (2.10.2) for monochromatic plane waves of the form

$$\mathbf{E}(\mathbf{r}, t) = \frac{1}{2} [\mathbf{E}_0 e^{j(\omega t - \mathbf{k} \cdot \mathbf{r})} + \mathbf{E}_0^* e^{-j(\omega t - \mathbf{k} \cdot \mathbf{r})}]$$

$$\mathbf{H}(\mathbf{r}, t) = \frac{1}{2} [\mathbf{H}_0 e^{j(\omega t - \mathbf{k} \cdot \mathbf{r})} + \mathbf{H}^* e^{-j(\omega t - \mathbf{k} \cdot \mathbf{r})}]$$
(2.10.3)

etc., where ω and **k** are both assumed to be real. The plane waves are assumed to be sustained by the external source. Let the volume in (2.10.2) increase to infinity, then the surface integral in (2.10.2) can be neglected with respect to the volume integrals. Averaging (2.10.2) over a time interval large compared with the period of the wave $(2\pi/\omega)$, the left-hand side of (2.10.2) can be identified as the average heat dissipation of the fields in the medium per unit time which is supplied by the average power $-\int \mathbf{J} \cdot \mathbf{E} \, dv$ of the external source. If we define the average heat dissipation per unit time per unit volume as Q, then

$$Q = \left\langle \frac{1}{V} \int_{V} \left(\mathbf{H} \cdot \dot{\mathbf{B}} + \mathbf{E} \cdot \dot{\mathbf{D}} \right) dv \right\rangle$$

= $(-j\omega/4) [\varepsilon_{ij}^{*} E_{0j}^{*} E_{0i} - \varepsilon_{ij} E_{0i}^{*} E_{0j}]$
= $(-j\omega/4) [\varepsilon_{ij}^{*} - \varepsilon_{ji}] E_{0j}^{*} E_{0i}$
= $(-j\omega/4) [\varepsilon_{ij}^{'*} + j\varepsilon_{ij}^{''*} - \varepsilon_{ji}^{'} + j\varepsilon_{ji}^{''}] E_{0j}^{*} E_{0i}$
= $(\omega/2) \varepsilon_{ij}^{''} E_{0j}^{*} E_{0j}$ (2.10.4)

where the material relations and Hermitian properties of ε'_{ij} and ε''_{ij} have been used. On time averaging, terms involving $e^{\pm j2\omega t}$ all vanish.

From (2.10.4), we see that the heat dissipation or energy loss for plane waves in the medium is given by the anti-Hermitian part of the dielectric tensor. For a stable medium from second law of thermodynamics, Q must be greater then zero. Therefore,

$$\varepsilon_{ij}^{\prime\prime} E_{0i}^* E_{0j} > 0$$
 (2.10.4a)

or $\varepsilon_{ij}^{\prime\prime}$ must be positive definite. If the medium is isotropic (i.e., $\varepsilon_{ij}^{\prime\prime} = \varepsilon^{\prime\prime} \delta_{ij}$) and $\varepsilon = \varepsilon^{\prime} - j\varepsilon^{\prime\prime}$, the condition (2.10.4a) reduces to $\varepsilon^{\prime\prime} > 0$. For a lossless medium, $\varepsilon_{ij}'' = 0$. In an unstable medium, Q may be less than zero and the converse of (2.10.4a) is true. If the unstable medium is also isotropic, then $\varepsilon'' < 0$.

The discussion above for plane waves is in general not realistic in most practical situations. Plane waves are infinite in time duration and spatial extent while all observed waves are turned on for a finite time and exist in finite spatial domain. Besides, the average energy in a plane wave is independent of spatial coordinates and hence there is no way to follow its motion and measure, for example, its velocity. It is, therefore, necessary to consider the propagation of a wave packet. Assuming the fields are of the type

$$\mathbf{E}(\mathbf{r}, t) = \frac{1}{2} [\mathbf{E}_{00}(\mathbf{r}, t) e^{j(\omega_0 t - \mathbf{k}_0 \cdot \mathbf{r})} + \text{c.c.}]$$

= $\frac{1}{2} [\mathbf{E}_0(\mathbf{r}, t) e^{j(\omega_0 t - \mathbf{k}_0' \cdot \mathbf{r})} + \text{c.c.}]$
= $[1/(2\pi)^4] \int [\mathbf{E}(\mathbf{k}, \omega) e^{j(\omega t - \mathbf{k} \cdot \mathbf{r})} + \text{c.c.}] d\omega d\mathbf{k}$ (2.10.5)

where

$$\mathbf{E}_{0}(\mathbf{r}, t) = \mathbf{E}_{00}(\mathbf{r}, t)e^{-\mathbf{k}_{0}^{\prime\prime}\cdot\mathbf{r}}$$

$$\mathbf{k}_{0} = \mathbf{k}_{0}^{\prime} - j\mathbf{k}_{0}^{\prime\prime} \qquad (2.10.6)$$

and c.c. denotes complex conjugate of the preceding term. The Fourier transform is used in writing the third line. If we assume that $\mathbf{E}_0(\mathbf{r}, t)$ is a slowly varying function both in time and space with respect to $(2\pi/\omega_0)$ and (1/k), (2.10.5) represents a real wave packet with the carrier frequency ω_0 . $\mathbf{E}(\mathbf{k}, \omega)$ therefore has a sharp peak at $\omega = \omega_0$ and $\mathbf{k} = \mathbf{k}_0$. Similar expressions can be written for other quantities. From (2.10.5) we have

$$\mathbf{E}_{0}(\mathbf{r},t) = \left[\frac{2}{(2\pi)^{4}}\right] \int \mathbf{E}(\mathbf{k},\omega) e^{j\left[\left(\omega-\omega_{0}\right)t-\left(\mathbf{k}-\mathbf{k}_{0}'\right)\cdot\mathbf{r}\right]} d\omega d\mathbf{k} \qquad (2.10.7)$$

Applying the material relation, we have

$$\mathbf{D}(\mathbf{r},t) = \frac{1}{(2\pi)^4} \int \left[\mathbf{\epsilon}(\mathbf{k},\omega) \cdot \mathbf{E}(\mathbf{k},\omega) e^{j(\omega t - \mathbf{k} \cdot \mathbf{r})} + \text{c.c.} \right] d\omega \, d\mathbf{k} \quad (2.10.8)$$

We note that because of the fact that the signal is no longer monochromatic we have to take into account the dispersive effects of the medium in (2.10.8). For the wave packet we are now considering, since $\mathbf{E}(\mathbf{k}, \omega)$ is highly peaked at $\omega = \omega_0$ and $\mathbf{k} = \mathbf{k}_0$, it is possible to expand any function of ω and \mathbf{k} under the integral sign (integrals over ω and \mathbf{k}) about the values ω_0 and \mathbf{k}_0 . For instance

$$\omega \varepsilon_{ij}(\mathbf{k}, \omega) \cong (\omega \varepsilon_{ij})_0 + (\partial \omega \varepsilon_{ij}/\partial \omega)_0 \Omega + (\omega \nabla_k \varepsilon_{ij})_0 \cdot \mathbf{q} \qquad (2.10.9)$$

where $\Omega = \omega - \omega_0$, $\mathbf{q} = \mathbf{k} - \mathbf{k}_0$, the subscript 0 indicates that values at $\omega = \omega_0$, $\mathbf{k} = \mathbf{k}_0$ are taken. We note that only first-order terms are retained.

Let us now consider (2.10.2) for this quasi-monochromatic wave packet. We first examine the term

$$E_{i}(\mathbf{r}, t)(\partial D_{i}(\mathbf{r}, t)/\partial t) = [1/(2\pi)^{8}] \iint [E_{i}(\mathbf{k}_{1}, \omega_{1})e^{j(\omega_{1}t-\mathbf{k}_{1}\cdot\mathbf{r})} + \text{c.c.}]$$

$$\times [j\omega_{2}\varepsilon_{ij}(\mathbf{k}_{2}, \omega_{2})E_{j}(\mathbf{k}_{2}, \omega_{2})e^{j(\omega_{2}t-\mathbf{k}_{2}\cdot\mathbf{r})} + \text{c.c.}]$$

$$\times d\omega_{1} d\omega_{2} d\mathbf{k}_{1} d\mathbf{k}_{2} \qquad (2.10.10)$$

Taking the average in time, the interval being large compared to $(2\pi/\omega_0)$ but small compared to the characteristic time for $\mathbf{E}_0(\mathbf{r}, t)$, (2.10.10) can be written approximately in the form

$$\langle E_i \,\partial D_i / \partial t \rangle = \frac{1}{(2\pi)^8} \int e^{-2\mathbf{k}_0^{\prime\prime} \cdot \mathbf{r}} \left\{ j \omega \varepsilon_{ij}(\mathbf{q}_2, \Omega_2) E_i^*(\mathbf{q}_1, \Omega_1) E_j(\Omega_2, \mathbf{q}_2) \right. \\ \left. \times e^{j[(\Omega_2 - \Omega_1)t - (\mathbf{q}_2 - \mathbf{q}_1^{\bullet}) \cdot \mathbf{r}]} + \text{c.c.} \right\} d\mathbf{q}_1 \, d\mathbf{q}_2 \, d\Omega_1 \, d\Omega_2$$
(2.10.11)

where the transformations $\Omega_1 = \omega_1 - \omega_0$, $\Omega_2 = \omega_2 - \omega_0$, $\mathbf{q}_1 = \mathbf{k}_1 - \mathbf{k}_0$, $\mathbf{q}_2 = \mathbf{k}_2 - \mathbf{k}_0$ have been used. When (2.10.9) is substituted into (2.10.11), we obtain

$$\left\langle E_{i} \frac{\partial D_{i}}{\partial t} \right\rangle$$

$$\approx \frac{1}{(2\pi)^{8}} \int e^{-2\mathbf{k}_{0}^{\prime\prime} \cdot \mathbf{r}} \left\{ j(\omega\varepsilon_{ij})_{0} E_{i}^{*}(\mathbf{q}_{1}, \Omega_{1}) E_{j}(\mathbf{q}_{2}, \Omega_{2}) e^{j[(\Omega_{2}-\Omega_{1})t-(\mathbf{q}_{2}-\mathbf{q}_{1}^{*}) \cdot \mathbf{r}]} \right.$$

$$+ \mathrm{c.c.} \right\} d\mathbf{q}_{1} d\mathbf{q}_{2} d\Omega_{1} d\Omega_{2}$$

$$+ \frac{1}{(2\pi)^{8}} \int e^{-2\mathbf{k}_{0}^{\prime\prime} \cdot \mathbf{r}} \left\{ j\left(\frac{\partial \omega\varepsilon_{ij}}{\partial \omega}\right)_{0} \Omega_{2} E_{i}^{*}(\mathbf{q}_{1}, \Omega_{1}) E_{j}(\mathbf{q}_{2}, \Omega_{2}) \right.$$

$$\times e^{j[(\Omega_{2}-\Omega_{1})t-(\mathbf{q}_{2}-\mathbf{q}_{1}^{*}) \cdot \mathbf{r}]} + \mathrm{c.c.} \right\} d\mathbf{q}_{1} d\mathbf{q}_{2} d\Omega_{1} d\Omega_{2}$$

$$+ \frac{1}{(2\pi)^{8}} \int e^{-2\mathbf{k}_{0}^{\prime\prime} \cdot \mathbf{r}} \left\{ j(\omega\nabla_{k}\varepsilon_{ij})_{0n}q_{2n}E_{i}^{*}(\mathbf{q}_{1}, \Omega_{1}) E_{j}(\mathbf{q}_{2}, \Omega_{2}) \right.$$

$$\times e^{j[(\Omega_{2}-\Omega_{1})t-(\mathbf{q}_{2}-\mathbf{q}_{1}^{*}) \cdot \mathbf{r}]} + \mathrm{c.c.} \right\} d\mathbf{q}_{1} d\mathbf{q}_{2} d\Omega_{1} d\Omega_{2}$$

$$\times e^{j[(\Omega_{2}-\Omega_{1})t-(\mathbf{q}_{2}-\mathbf{q}_{1}^{*}) \cdot \mathbf{r}]} + \mathrm{c.c.} \right\} d\mathbf{q}_{1} d\mathbf{q}_{2} d\Omega_{1} d\Omega_{2}$$

$$(2.10.12)$$

Using (2.10.7), it is easy to show the following relations:

$$E_{0i}(\mathbf{r}, t)E_{0j}^{*}(\mathbf{r}, t)$$

$$= \frac{4}{(2\pi)^{8}} \int e^{-2\mathbf{k}_{0}^{\prime\prime}\cdot\mathbf{r}} E_{i}(\mathbf{q}_{1}, \Omega_{1})E_{j}^{*}(\mathbf{q}_{2}, \Omega_{2})e^{j[(\Omega_{1}-\Omega_{2})t-(\mathbf{q}_{1}-\mathbf{q}_{2}^{*})\cdot\mathbf{r}]}$$

$$\times d\Omega_{1} d\Omega_{2} d\mathbf{q}_{1} d\mathbf{q}_{2}$$

$$E_{0i}(\mathbf{r}, t) \frac{\partial E_{0j}^{*}(\mathbf{r}, t)}{\partial t}$$

$$= \frac{-4j}{(2\pi)^{8}} \int e^{-2\mathbf{k}_{0}^{\prime\prime}\cdot\mathbf{r}} \Omega_{2}E_{i}(\mathbf{q}_{1}, \Omega_{1})E_{j}^{*}(\mathbf{q}_{2}, \Omega_{2})e^{j[(\Omega_{1}-\Omega_{2})t-(\mathbf{q}_{1}-\mathbf{q}_{2}^{*})\cdot\mathbf{r}]}$$

$$\times d\Omega_{1} d\Omega_{2} d\mathbf{q}_{1} d\mathbf{q}_{2}$$

$$\frac{\partial E^{*}(\mathbf{r}, t)}{\partial t}$$

$$E_{0i}(\mathbf{r}, t) \frac{\partial E_{0j}(\mathbf{r}, t)}{\partial x_n}$$

$$= \frac{4j}{(2\pi)^8} \int e^{-2\mathbf{k}_0^{\prime\prime} \cdot \mathbf{r}} (q_{2n} + jk_{0n}) E_i(\mathbf{q}_1, \Omega_1) E_j^{\ast}(\mathbf{q}_2, \mathbf{U}_2) e^{j[(\Omega_1 - \Omega_2)t - (\mathbf{q}_1 - \mathbf{q}_2^{\ast}) \cdot \mathbf{r}]}$$

$$\times d\Omega_1 d\Omega_2 d\mathbf{q}_1 d\mathbf{q}_2 \qquad (2.10.13)$$

Using the first of (2.10.13), we see that the first term in (2.10.12) can be written as

$$\frac{1}{4} \left[-j(\omega \varepsilon_{ij}^{*})_{0} E_{0i}(\mathbf{r}, t) E_{0j}^{*}(\mathbf{r}, t) + j(\omega \varepsilon_{ij})_{0} E_{0i}^{*}(\mathbf{r}, t) E_{0j}(\mathbf{r}, t) \right] \\
= (-j\omega_{0}/4) \left[\varepsilon_{ij}^{\prime *}(\mathbf{k}_{0}, \omega_{0}) + j\varepsilon_{ij}^{\prime \prime *}(\mathbf{k}_{0}, \omega_{0}) \right] E_{0i}(\mathbf{r}, t) E_{0j}^{*}(\mathbf{r}, t) \\
+ (j\omega_{0}/4) \left[\varepsilon_{ji}^{\prime}(\mathbf{k}_{0}, \omega_{0}) - j\varepsilon_{ji}^{\prime \prime}(\mathbf{k}_{0}, \omega_{0}) \right] E_{0j}^{*}(\mathbf{r}, t) E_{0i}(\mathbf{r}, t) \\
= \frac{1}{2} (\omega \varepsilon_{ij}^{\prime \prime})_{0} E_{0i}^{*} E_{0j} \qquad (2.10.14)$$

The second term in (2.10.12) can be written as

$$\frac{1}{4} \left(\frac{\partial \omega \varepsilon_{ij}}{\partial \omega} \right)_{0} E_{0i}^{*} \frac{\partial E_{0j}}{\partial t} + \frac{1}{4} \left(\frac{\partial \omega \varepsilon_{ij}^{*}}{\partial \omega} \right)_{0} E_{0i} \frac{\partial E_{0j}^{*}}{\partial t}$$

$$= \frac{1}{4} \left(\frac{\partial \omega \varepsilon_{ij}^{\prime}}{\partial \omega} \right)_{0} \frac{\partial}{\partial t} \left(E_{0i}^{*} E_{0j} \right)$$

$$- \frac{j}{4} \left(\frac{\partial \omega \varepsilon_{ij}^{\prime \prime}}{\partial \omega} \right)_{0} \left(E_{0i}^{*} \frac{\partial E_{0j}}{\partial t} - E_{0j} \frac{\partial E_{0i}^{*}}{\partial t} \right) \qquad (2.10.15)$$

and the third term becomes

$$-\frac{1}{4} \left(\omega \frac{\partial \varepsilon_{ij}}{\partial k_n} \right)_0 \left[E_{0i}^* \frac{\partial E_{0j}}{\partial x_n} + k_{0n}^{\prime\prime} E_{0i}^* E_{0i} \right] \\ -\frac{1}{4} \left(\omega \frac{\partial \varepsilon_{ij}^*}{\partial k_n} \right)_0 \left[E_{0i} \frac{\partial E_{0j}^*}{\partial x_n} + k_{0n}^{\prime\prime} E_{0i} E_{0j}^* \right] \\ = -\frac{1}{4} \left(\omega \frac{\partial \varepsilon_{ij}^\prime}{\partial k_n} \right)_0 \left[\frac{\partial}{\partial x_n} \left(E_{0i}^* E_{0j} \right) + 2k_{0n}^{\prime\prime} E_{0i}^* E_{0j} \right] \\ + \frac{j}{4} \left(\omega \frac{\partial \varepsilon_{ij}^\prime}{\partial k_n} \right)_0 \left[E_{0i}^* \frac{\partial E_{0j}}{\partial x_n} - E_{0j} \frac{\partial E_{0i}^*}{\partial x_n} \right]$$
(2.10.16)

Combining (2.10.12)–(2.10.16), we have the time averaged $\langle E_i \partial D_i / \partial t \rangle$. Similar computation can be applied to the other terms in (2.10.2). The energy relation, after time averaging for the wave packet of the form in (2.10.5), can then be written as

$$\frac{1}{4} \frac{\partial}{\partial t} \left[\left(\frac{\partial \omega \varepsilon_{ij}'}{\partial \omega} \right)_{0} (E_{0}^{*} E_{0j}) + \mu_{0} H_{0i}^{*} H_{0i} \right] + \frac{1}{2} (\omega \varepsilon_{ij}'')_{0} E_{0i}^{*} E_{0j} \\
- \frac{j}{4} \left(\frac{\partial \omega \varepsilon_{ij}'}{\partial \omega} \right)_{0} \left(E_{0i}^{*} \frac{\partial E_{0j}}{\partial t} - E_{0j} \frac{\partial E_{0i}}{\partial t} \right) \\
+ \frac{j}{4} \left(\omega \frac{\partial \varepsilon_{ij}'}{\partial k_{n}} \right)_{0} \left(E_{0i}^{*} \frac{\partial E_{0j}}{\partial x_{n}} - E_{0j} \frac{\partial E_{0i}}{\partial x_{n}} \right) \\
= - \langle \mathbf{E} \cdot \mathbf{J} \rangle - \frac{1}{4} \nabla \cdot (\mathbf{E}_{0}^{*} \times \mathbf{H}_{0} + \mathbf{E}_{0} \times \mathbf{H}_{0}^{*}) \\
+ \frac{1}{4} \left(\omega \frac{\partial \varepsilon_{ij}'}{\partial k_{n}} \right)_{0} \left[\frac{\partial}{\partial x_{n}} (E_{0i}^{*} E_{0j}) + 2k_{0n}'' E_{0i}^{*} E_{0j} \right] \quad (2.10.17)$$

Recall that $E_{0i}(\mathbf{r}, t) = E_{00i}(\mathbf{r}, t)e^{-\mathbf{k}_{0}^{\prime\prime}\cdot\mathbf{r}}$, we can reduce the last term on the right-hand side of (2.10.17) to the form

$$-e^{-2\mathbf{k}_0^{\prime\prime}\cdot\mathbf{r}}(\partial/\partial x_n)S_n^{(1)}$$

where

$$S_n^{(1)} = -\frac{1}{4} (\omega \,\partial \varepsilon_{ij}^{\prime} / \partial k_n)_0 (E_{00i}^* E_{00j}) \tag{2.10.18}$$

We note that for a monochromatic plane wave where $\mathbf{E}_{00} = \text{constant}$, integration of (2.10.17) throughout the volume gives us the same result as (2.10.4).

Equation (2.10.17) is the general equation for the conservation of energy for electromagnetic waves in an anisotropic, dispersive medium. To see the physical significance of this equation, we first consider several special cases. For a lossless medium ($\varepsilon_{ij}'' = 0$) in the absence of external sources, (2.10.17) may be put in the form

$$\partial \langle W \rangle / \partial t = -\nabla \cdot [\langle \mathbf{S}^{(0)} \rangle + \langle \mathbf{S}^{(1)} \rangle]$$
 (2.10.19)

where we have defined

$$\langle W \rangle = \frac{1}{4} [\mathbf{E}_0^* \cdot (\partial \omega \mathbf{\epsilon}' / \partial \omega)_0 \cdot \mathbf{E}_0 + \mu_0 \mathbf{H}_0^* \cdot \mathbf{H}_0] \qquad (2.10.20)$$

$$\langle \mathbf{S}^{(0)} \rangle = \frac{1}{4} [\mathbf{E}_0^* \times \mathbf{H}_0 + \mathbf{E}_0 \times \mathbf{H}_0^*]$$
 (2.10.21)

$$\langle \mathbf{S}^{(1)} \rangle = -\frac{1}{4} \mathbf{E}_0^* \cdot (\omega \nabla_k \boldsymbol{\varepsilon}') \cdot \mathbf{E}_0 \qquad (2.10.22)$$

W is the time average (with respect to $2\pi/\omega_0$) of the energy density in the medium. The first term in $\langle W \rangle$ contains the usual electric energy density and the portion of the kinetic energy of the particles in the medium which is associated with the coherent wave motion while the second term is the magnetic energy density. $\langle S^{(0)} \rangle$ is the usual Poynting vector which represents the average energy flux of the electromagnetic field and $\langle S^{(1)} \rangle$ is the average energy flux connected to the spatial dispersion of the medium. Physically $\langle S^{(1)} \rangle$ represents the energy flux transferred by the motion of the particles in the medium. In a cold medium where there is no spatial dispersion, $S^{(1)}$ vanishes. For the case where the medium has no spatial nor temporal dispersion, (2.10.19) reduces to the familiar form of Poynting's theorem.

For a general medium with loss and dispersion, (2.10.17) is rewritten as

$$\frac{\partial \langle W \rangle}{\partial t} + 2(\omega \varepsilon_{ij}^{"})_{0} E_{0i}^{*} E_{0j} - j \left(\frac{\partial \omega \varepsilon_{ij}^{"}}{\partial \omega}\right)_{0} \left(E_{0i}^{*} \frac{\partial E_{0j}}{\partial t} - E_{0j} \frac{\partial E_{0j}^{*}}{\partial t}\right) + j \left(\omega \frac{\partial \varepsilon_{ij}^{"}}{\partial k_{n}}\right)_{0} \left(E_{0i}^{*} \frac{\partial E_{0j}}{\partial x_{n}} - E_{0j} \frac{\partial E_{0i}^{*}}{\partial x_{n}}\right) = -\nabla \cdot \left[\langle \mathbf{S}^{(0)} \rangle + \langle \mathbf{S}^{(1)} \rangle\right] - \langle \mathbf{E} \cdot \mathbf{J} \rangle$$
(2.10.23)

where $\langle W \rangle$, $\langle S^{(0)} \rangle$ and $\langle S^{(1)} \rangle$ are the same as defined in (2.10.20), (2.10.21) and (2.10.22), respectively. However, it is no longer possible now to identify $\langle W \rangle$ as the energy density, nor is it possible to define $\frac{1}{2}\omega \varepsilon_{ij}'' E_{0i}^* E_{0j}$ as the heat dissipation in the medium. To see this we first show that $\langle W \rangle$ may be negative in the case of lossy medium. Let us consider the simple case of an isotropic plasma with collisional loss. The relative permittivity is given by (see Chapter 4)

$$\varepsilon(\omega) = 1 - \frac{\omega_p^2}{\omega(\omega - j\nu)}$$
(2.10.24)

where ω_p is the plasma frequency and ν is the collision frequency. Simple calculation shows that for this medium

$$\langle W \rangle = \frac{1}{4} \left\{ \left[1 + \frac{\omega_p^2 (\omega_0^2 - \nu^2)}{(\omega_0^2 + \nu^2)^2} \right] E_0^2 + \mu_0 H_0^2 \right\}$$
(2.10.25)

which can become negative for large collision frequencies.

Similar computation shows that $\frac{1}{2}\omega\varepsilon_{ij}E_{0i}^*E_{0j}$ is not equal to the heat dissipation caused by the collisions.

Thus, we see that in general we do not have distinct physical interpretation attached to each of the terms appearing in (2.10.23), even though as a whole, (2.10.23) is a general statement of conservation of energy.

2.11 Group and Energy Velocities

In Section 6 we have shown that for characteristic waves, the following equation is valid

$$\mathbf{D} \cdot \mathbf{E} = (\mathbf{k}\mathbf{k} - k^2 I + k_0^2 \mathbf{K}) \cdot \mathbf{E} = 0 \qquad (2.11.1)$$

For lossless medium and real values of ω and k (transparent region), the tensor **D** is Hermitian. Now consider the case in which small perturbations take place in k (due to attenuation), in ω (due to damping or because of our consideration of a wave packet), and in K (due to loss or change of parameters in the medium), then the tensor **D** becomes **D**₁ and the corresponding electric field becomes **E**₁, satisfying

$$\mathbf{D}_1 \cdot \mathbf{E}_1 = 0 \tag{2.11.2}$$

where D_1 can be written approximately as

$$\mathbf{D}_{1} = \mathbf{D} + \delta \omega \, \frac{\partial \mathbf{D}}{\partial \omega} + \, \delta \mathbf{k} \cdot \boldsymbol{\nabla}_{k} \mathbf{D} + k_{0}^{2} \, \delta \mathbf{K} \qquad (2.11.3)$$

From (2.11.1), since **D** is Hermitian, we have

$$\mathbf{E}^* \cdot \mathbf{D} = 0 \tag{2.11.4}$$

Multiply (2.11.2) on the left by E^* and (2.11.4) on the right by E_1 and subtract, we obtain

$$\mathbf{E}^{*} \cdot \left(\delta \omega \, \frac{\partial \mathbf{D}}{\partial \omega} + \, \delta \mathbf{k} \cdot \boldsymbol{\nabla}_{k} \mathbf{D} + k_{0}^{2} \, \delta \mathbf{K} \right) \cdot \mathbf{E}_{1} = 0 \qquad (2.11.5)$$

Since only first-order terms are considered, we can approximate E_1 in (2.11.5) by **E**. We shall discuss the individual terms of (2.11.5) in the following.

The first term

$$\mathbf{E}^{*} \cdot \left(\delta\omega \frac{\partial \mathbf{D}}{\partial\omega}\right) \cdot \mathbf{E} = \frac{\delta\omega}{c^{2}} \mathbf{E}^{*} \cdot \frac{\partial\omega^{2}\mathbf{K}}{\partial\omega} \cdot \mathbf{E}$$
$$= \mu_{0}\delta\omega \Big[\mathbf{E}^{*} \cdot \omega \mathbf{\epsilon} \cdot \mathbf{E} + \omega \mathbf{E}^{*} \cdot \frac{\partial\omega \mathbf{\epsilon}}{\partial\omega} \cdot \mathbf{E} \Big]$$
$$= \mu_{0}\omega \,\delta\omega \Big[\mathbf{E}^{*} \cdot \mathbf{D} + \mathbf{E}^{*} \cdot \frac{\partial\omega \mathbf{\epsilon}}{\partial\omega} \cdot \mathbf{E} \Big] \qquad (2.11.6)$$

But from Maxwell's equations

$$\mathbf{H} = \frac{1}{\omega \mu_0} \mathbf{k} \times \mathbf{E}, \qquad \mathbf{D} = -\frac{1}{\omega} \mathbf{k} \times \mathbf{H}$$

Therefore

$$\mathbf{E}^* \cdot \mathbf{D} = -\frac{1}{\omega} \mathbf{E}^* \cdot (\mathbf{k} \times \mathbf{H}) = -\frac{1}{\omega} \mathbf{H} \cdot (\mathbf{E} \times \mathbf{k}) = \mu_0 \mathbf{H} \cdot \mathbf{H}^*$$

Substituting the last equation into (2.11.6), we obtain

$$\mathbf{E}^{*} \cdot \left(\delta\omega \frac{\partial \mathbf{D}}{\partial\omega}\right) \cdot \mathbf{E}$$
$$= \delta\omega\mu_{0}\omega \left[\mathbf{E}^{*} \cdot \frac{\partial\omega\mathbf{\epsilon}}{\partial\omega} \cdot \mathbf{E} + \mu_{0}\mathbf{H}^{*} \cdot \mathbf{H}\right] = 4 \ \delta\omega\mu_{0}\omega \langle W \rangle \quad (2.11.7)$$

where W is defined in (2.10.20). The second term in (2.11.5) is, in component form,

$$E_{i}^{*} \delta k_{n} \frac{\partial D_{ij}}{\partial k_{n}} E_{j} = (\delta_{in}k_{j} + \delta_{jn}k_{i} - 2k_{n} \delta_{ij} + k_{0}^{2}(\partial K_{ij}/\partial k_{n}))E_{i}^{*}E_{j} \delta k_{n}$$

$$= E_{n}^{*}(\delta k_{n})k_{j}E_{j} + E_{n}(\delta k_{n})k_{i}E_{i}^{*} - 2k_{n} \delta k_{n} E_{i}^{*}E_{i}$$

$$+ k_{0}^{2}E_{i}^{*}(\delta k_{n})(\partial K_{ij}/\partial k_{n})E_{j}$$

$$= \delta k_{n}[(\mathbf{E} \cdot \mathbf{k})E_{n}^{*} + (\mathbf{E}^{*} \cdot \mathbf{k})E_{n} - 2k_{n}(\mathbf{E} \cdot \mathbf{E}^{*})]$$

$$+ \omega^{2}\mu_{0} \delta k_{n} E_{i}^{*} \frac{\partial \varepsilon_{ij}}{\partial k_{n}} E_{j}$$

Put into vector form, the second term becomes

$$-4\omega\mu_0 \,\delta\mathbf{k} \cdot \left[\langle \mathbf{S}^{(0)} \rangle + \langle \mathbf{S}^{(1)} \rangle\right] \tag{2.11.8}$$
where $\langle S^{(0)} \rangle$ and $\langle S^{(1)} \rangle$ are defined in (2.10.21) and (2.10.22), respectively. The third term in (2.11.5) can be written as

$$k_0^2 \mathbf{E}^* \cdot \delta \mathbf{K} \cdot \mathbf{E} = \omega^2 \mu_0 \mathbf{E}^* \cdot \delta \boldsymbol{\epsilon} \cdot \mathbf{E}$$

= $\omega^2 \mu_0 \mathbf{E}^* \cdot \delta \boldsymbol{\epsilon}' \cdot \mathbf{E} - j \omega^2 \mu_0 \mathbf{E}^* \cdot \delta \boldsymbol{\epsilon}'' \cdot \mathbf{E}$ (2.11.9)

where we have assumed that

$$\delta \boldsymbol{\varepsilon} = \delta \boldsymbol{\varepsilon}' - j \, \delta \boldsymbol{\varepsilon}'' \tag{2.11.10}$$

The first term on the right-hand side of (2.11.9) is related to the change of the average energy stored in the electric field due to change in ε' while the second term is related to the change of heat dissipation in the medium, δQ .

Combining (2.11.5), (2.11.7), (2.11.8) and (2.11.9), we have

$$\delta\omega\langle W\rangle - \delta\mathbf{k} \cdot [\langle \mathbf{S}^{(0)} \rangle + \langle \mathbf{S}^{(1)} \rangle] - \frac{1}{2}j\,\delta Q + \omega\,\delta\langle\mathscr{E}_e\rangle = 0 \quad (2.11.11)$$

where

$$\delta Q = \frac{1}{2}\omega \mathbf{E}^* \cdot \delta \mathbf{\epsilon}'' \cdot \mathbf{E}, \qquad \delta \langle \mathscr{C}_e \rangle = \frac{1}{4}\mathbf{E}^* \cdot \delta \mathbf{\epsilon}' \cdot \mathbf{E}$$

We shall interpret the terms in (2.11.11) for different cases.

(i) Lossless Medium without Change of Parameters in the Medium, i. e., $\delta Q = \delta \langle \mathscr{E}_e \rangle = 0$. From (2.11.11)

$$\mathbf{v}_{W} = \frac{\langle \mathbf{S}^{(0)} \rangle + \langle \mathbf{S}^{(1)} \rangle}{\langle W \rangle} = \frac{\delta \omega}{\delta \mathbf{k}} = \nabla_{k} \omega(\mathbf{k}) \qquad (2.11.12)$$

 \mathbf{v}_W is the energy velocity, the velocity that the energy propagates in the medium. We now define a group velocity

$$\mathbf{v}_g = \nabla_k \omega(\mathbf{k}) \tag{2.11.13}$$

which is the velocity a wave packet propagates without distortion in the medium. This can be seen by considering a characteristic wave packet with "carrier" frequency ω_0 and wave vector \mathbf{k}_0 . The field can be expressed as

$$\mathbf{E}(\mathbf{r}, t) = \frac{1}{(2\pi)^3} \int \mathbf{E}(\mathbf{k}) e^{j[\omega(\mathbf{k})t - \mathbf{k} \cdot \mathbf{r}]} d\mathbf{k}$$
(2.11.14)

where the dispersion relation for the characteristic waves $\omega = \omega(\mathbf{k})$ has been used. For the wave packet, $\mathbf{E}(\mathbf{k})$ has a sharp peak at k_0 ; therefore we approximate (2.11.4) by expanding the exponential about ω_0 and \mathbf{k}_0 and take the first-order terms

$$\mathbf{E}(\mathbf{r}, t) \simeq e^{j(\omega_0 t - \mathbf{k}_0 \cdot \mathbf{r})} \frac{1}{(2\pi)^3} \int \mathbf{E}(\mathbf{q}) e^{j[t(V_k \omega)_0 - \mathbf{r}] \cdot \mathbf{q}} d\mathbf{q}$$
$$= \mathbf{A}(\mathbf{r} - (\nabla_k \omega)_0 t) e^{j(\omega_0 t - \mathbf{k}_0 \cdot \mathbf{r})}$$
(2.11.15)

where A is the amplitude of the wave packet; in (2.11.15) we see that this packet propagates with velocity $\mathbf{v}_g = \nabla_k \omega$ without changing its shape. The same result can be obtained by applying the stationary phase method to (2.11.14). More about the geometrical interpretation of the group velocity will be discussed in a later section.

From (2.11.12), we see that in a lossless medium for real values of ω and k (transparent medium), the energy velocity and group velocity coincide.

In a lossy or lossless but nontransparent medium, \mathbf{k} is complex or purely imaginary; the definition of group velocity looses its physical meaning. In many practical cases, however, if the loss is small the definition may still be used.

(ii) Lossy Medium with $\delta \mathbf{\epsilon}' = 0$, Forced Monochromatic Oscillation $\delta \omega = 0$. In this case (2.11.11) becomes

$$\delta \mathbf{k} \cdot \left[\langle \mathbf{S}^{(0)} \rangle + \langle \mathbf{S}^{(1)} \rangle \right] = -\frac{1}{2} j \, \delta Q \qquad (2.11.16)$$

Equation (2.11.16) gives the spatial rate of decay of the signal for positive δQ .

(iii) Lossy Medium with $\delta \mathbf{\epsilon}' = 0$, Initial Value Problem for Wave Propagation, $\delta \mathbf{k} = 0$. Equation (2.11.11) becomes

$$\delta\omega\langle W\rangle = \frac{1}{2}\,\delta Q \tag{2.11.17}$$

which gives the time rate of decay of oscillation for positive δQ .

2.12 Geometric Interpretation of Group Velocity

In a lossless, transparent medium, we have seen that the energy propagates along the direction of the group velocity \mathbf{v}_g . This direction is called the ray direction. In general in an anisotropic medium the ray direction is different from that of the wave vector **k**. To see this, let us recall the definition of the group velocity

$$\mathbf{v}_g = \nabla_k \omega(\mathbf{k}) \tag{2.12.1}$$

From this equation, it is obvious that \mathbf{v}_g is normal to the surface $\omega(\mathbf{k}) = \text{constant}$ which is the solution of the dispersion relation for a fixed frequency. This surface can be plotted in k-space by solving for $k_z = k_z(k_x, k_y, \omega)$. The surface then contains the endpoints of the wave vector $k = |\mathbf{k}|$ and is called the wave vector surface, or the dispersion surface. Sometimes it is more convenient to introduce the refractive index surface through the definition

$$n = kc/\omega$$

which contains the same information as does the wave vector surface at a fixed frequency.

Let us write the dispersion relation in the form

$$f(\mathbf{k},\omega) = k - n\omega/c = 0 \qquad (2.12.2)$$

Applying the formula for differentiation of implicit functions we obtain

$$\mathbf{v}_{g} = \nabla_{k} \omega(\mathbf{k}) = -\nabla_{k} f / (\partial f / \partial \omega) \qquad (2.12.3)$$

The expression for the refractive index *n* appearing in (2.12.2) can be related to the dielectric tensor as discussed in Section 8. In the absence of spatial dispersion the refractive index satisfies a certain biquadratic algebraic equation (2.8.11) where the coefficients are expressed in terms of angles of the wave vector with the coordinate axes and the angular frequency. Let θ and ϕ be the polar and azimuthal angles, respectively; then $n = n(\theta, \phi, \omega)$. Since $\theta = \arccos k_z/(k_x^2 + k_y^2 + k_z^2)^{1/2}$ and $\phi = \arctan k_y/k_x$, we can show that

$$\frac{\partial n}{\partial k_x} = \frac{\partial n}{\partial \theta} \frac{\partial \theta}{\partial k_x} + \frac{\partial n}{\partial \phi} \frac{\partial \phi}{\partial k_x}$$
$$= \left(\frac{\partial n}{k \partial \theta}\right) \cos \theta \cos \phi - \left(\frac{\partial n}{k \partial \phi}\right) \left(\frac{\sin \phi}{\sin \theta}\right)$$

The x-component of the group velocity is obtained, by using the above formula (2.12.2) and (2.12.3),

$$\nu_{gx} = \frac{\partial \omega}{\partial k_x} = -\frac{\partial k/\partial k_x - (\omega/c)(\partial n/\partial k_x)}{[-\partial(n\omega)/\partial\omega]/c}$$
$$= \frac{c}{\partial(n\omega)/\partial\omega} \left[\sin\theta\cos\phi - \frac{1}{n} \left(\frac{\partial n}{\partial\theta}\cos\theta\cos\phi - \frac{\partial n}{\partial\phi}\frac{\sin\phi}{\sin\theta}\right)\right]$$
(2.12.4a)

Similarly, we can derive expressions for the y- and z-components of the group velocity,

$$\nu_{\sigma\nu} = \frac{\partial\omega}{\partial k_{\nu}} = \frac{c}{\partial(n\omega)/\partial\omega} \left[\sin\theta \sin\phi - \frac{1}{n} \left(\frac{\partial n}{\partial\theta} \cos\theta \sin\phi + \frac{\partial n}{\partial\phi} \frac{\cos\phi}{\sin\theta} \right) \right]$$
(2.12.4b)

and

$$v_{gz} = \frac{\partial \omega}{\partial k_z} = \frac{c}{\partial (n\omega)/\partial \omega} \left[\cos \theta + \frac{1}{n} \frac{\partial n}{\partial \theta} \sin \theta \right]$$
(2.12.4c)

The magnitude of the group velocity is just the square root of the sum of the squares of (2.12.4a-c), giving

$$\nu_g = \frac{c}{\partial (n\omega)/\partial \omega} \left[1 + \frac{1}{n^2} \left(\frac{\partial n}{\partial \theta} \right)^2 + \frac{1}{n^2 \sin^2 \theta} \left(\frac{\partial n}{\partial \phi} \right)^2 \right]^{1/2} \quad (2.12.4d)$$

Sometimes it is desirable to express the group velocity components in spherical coordinates. We list the resulting expression in the following for later reference.

$$v_{gk} = \frac{c}{\partial(n\omega)/\partial\omega}$$
(2.12.5a)

$$v_{g\theta} = -\frac{c}{\partial (n\omega)/\partial \omega} \frac{1}{n} \frac{\partial n}{\partial \theta}$$
 (2.12.5b)

$$v_{g\phi} = -\frac{c}{\partial (n\omega)/\partial \omega} \frac{1}{n\sin\theta} \frac{\partial n}{\partial \phi}$$
 (2.12.5c)



We see from (2.12.5) that the group velocity is parallel to **k** if the medium is isotropic. Further, the expression for group velocity along **k** is unaffected by the fact the medium may be anisotropic. The anisotropy comes in through the appearance of the θ - and ϕ -components of the group velocity. Suppose that there is an axial symmetry in the medium and we orient the z-axis to coincide with this axis. The refractive index for such a case is no longer a function of ϕ . Let α be the angle between **k** and **v**_g as shown, then

$$v_{gk} = v_g \cos \alpha, \qquad v_{g\theta} = v_g \sin \alpha$$

The angle α can be related to the refractive index by taking the ratio of the above two expressions and making use of (2.12.5a) and (2.12.5b)

$$\tan \alpha = \frac{v_{g\theta}}{v_{gk}} = -\frac{1}{n} \frac{\partial n}{\partial \theta}$$
(2.12.6)

The magnitude of the group velocity as given by (2.12.4d) in this case reduces to

$$\nu_g = \frac{c}{\cos \alpha \left(\partial n \omega / \partial \omega \right)} \tag{2.12.7}$$

The foregoing discussion shows that the group velocity and the wave vector in general do not lie in the same direction unless both $\partial n/\partial \theta$ and $\partial n/\partial \phi$ vanish. When this happens, the refractive index does not have angular dependence and the refractive index surface becomes the surface of a sphere. This corresponds to the isotropic case.

For a transparent medium without spatial dispersion, we have seen that

$$\langle \mathbf{S}^{(0)} \rangle = W \mathbf{v}_{q} \tag{2.11.12}$$

But by definition [Eq. (2.10.21)]

$$\langle \mathbf{S}^{(0)} \rangle = \frac{1}{4} [\mathbf{E}_0^* \times \mathbf{H}_0 + \mathbf{E}_0 \times \mathbf{H}_0^*]$$

$$= \frac{1}{4\omega\mu_0} [\mathbf{E}_0^* \times (\mathbf{k} \times \mathbf{E}_0) + \mathbf{E}_0 \times (\mathbf{k} \times \mathbf{E}_0^*)]$$

Therefore

$$\mathbf{k} \cdot \langle \mathbf{S}^{(0)} \rangle = \frac{1}{4\omega\mu_0} \left[2k^2 (\mathbf{E}_0 \cdot \mathbf{E}_0^*) - 2(\mathbf{E}_0^* \cdot \mathbf{k}) (\mathbf{E}_0 \cdot \mathbf{k}) \right] \ge 0 \quad (2.12.8)$$

From (2.12.8) we see that the angle between **k** and $\langle S^{(0)} \rangle$ is acute. But in (2.11.12), we have the relation that the group velocity in a transparent medium without spatial dispersion is in the same direction as the average Poynting vector $\langle S^0 \rangle$. Therefore, we can conclude that the angle α between \mathbf{v}_g and **k** is acute for this medium, $|\alpha| < \pi/2$. Since \mathbf{v}_g in this medium actually coincides with the energy velocity, it follows from the theory of relativity that $|\mathbf{v}_g| \leq c$. From (2.12.7), we have

$$\partial(\omega n)/\partial \omega \ge 1/\cos \alpha > 1$$
 (2.12.9)

If there is spatial dispersion, however, the above conclusions are no longer valid. Since for this case,

$$\mathbf{v}_{g} = (1/W)[\langle \mathbf{S}^{(0)} \rangle + \langle \mathbf{S}^{(1)} \rangle]$$
(2.11.12)

 \mathbf{v}_q is no longer in the direction of $\langle \mathbf{S}^0 \rangle$.

The geometrical relations between \mathbf{v}_{q} , k and the refractive index surface for a transparent medium without spatial dispersion are shown in Fig. 2.12-2.



Fig. 2.12-2. The refractive index surface and its relation to the group velocity.

We see that depending on the shape of the refractive index surface, there may be focusing or defocusing effect on the rays. Also, in some cases, the normals to the surface at more than one point may lie in the same direction. In this case, the observer in the given direction will find waves with different wave vector directions in a given mode.

The concept of group velocity can also be elucidated with a kinematic approach. This approach is very general and can be used to study all types of waves, including electromagnetic waves. As it turns out, insights gained in such an approach are useful in understanding the asymptotic behavior of waves, especially in radiation problems such as the evaluation of the asymptotic dyadic Greens' function to be discussed in Section 14 of this chapter. The approach starts with the assumption that the wave function can be written in the form

$$A \exp j\psi \qquad (2.12.10)$$

where A is the amplitude assumed slowly varying, and ψ the phase. In a uniform medium, the uniform plane wave solution is required to have a

constant amplitude and a phase given by

$$\boldsymbol{\psi} = \boldsymbol{\psi}_0 + \boldsymbol{\omega} \boldsymbol{t} - \mathbf{k} \cdot \mathbf{r} \tag{2.12.11}$$

where ψ_0 is a constant reference phase. Such a wave extends to all space and exists for all time and hence is a highly idealized situation. An arbitrary wave, of course, does not possess these properties. In many cases these arbitrary waves may be described as locally plane, i.e., the plane wave properties can be applied only locally. This is generally the case for any arbitrary wave as $t \to \infty$ or the case of radiation field from a localized source. When this is the case, we assume that the wave function still has the form (2.12.10) except now both A and ψ are functions of the coordinates and time with A varying very slowly as compared with ψ . The surface

$$\psi(\mathbf{r}, t) = \text{constant}$$
 (2.12.12)

defines the surface of constant phase or phase front. Define the local (angular) frequency ω and wave number vector **k** by

$$\omega = \partial \psi / \partial t, \quad \mathbf{k} = -\nabla \psi \qquad (2.12.13)$$

The fact that \mathbf{k} is expressible as a gradient implies that \mathbf{k} is irrotational. Consequently, by making use of Stokes's theorem, we arrive at

$$\oint \mathbf{k} \cdot d\mathbf{l} = 0 \tag{2.12.14}$$

where the integration is along an arbitrary closed path. Since k gives the number of waves per unit length, the above relation is a statement of the conservation of number of waves. It implies that if there exists a phase function ψ to describe the wave, the total number of waves along any closed stationary curve must be zero.

The wave front surface (2.12.12) is generally not stationary. In order for an observer $\mathbf{r}(t)$ to stay on the same phase front, he must move with a velocity $d\mathbf{r}/dt$ to satisfy the relation

$$d\psi/dt = \partial\psi/\partial t + \dot{\mathbf{r}} \cdot \nabla \psi = 0 \qquad (2.12.15)$$

The phase velocity is defined as the velocity of travel of the phase front and its direction is normal to the phase front. Hence (2.12.15) immediately produces the well-known result for the velocity.

$$\mathbf{v}_p = \hat{s}\omega/k \tag{2.12.16}$$

where $\hat{s} = \mathbf{k}/k$. A phase ray is a streamline obtained by integrating \mathbf{v}_p in **r**-space. As seen in (2.12.16), the phase rays are straight lines only in a homogeneous medium but these rays are not necessarily parallel. When the medium is inhomogeneous, phase rays are generally curved.

The definition (2.12.13) implies

$$\partial \mathbf{k} / \partial t + \nabla \omega = 0 \tag{2.12.17}$$

We shall discuss the implication of this relation first in a homogeneous medium and later in an inhomogeneous medium.

In a homogeneous medium, the dispersion relation

$$\omega = \omega(\mathbf{k}) \tag{2.12.18}$$

is assumed to have been obtained for a certain wave in a given mode. Substituting (2.12.18) in (2.12.17), we obtain

$$\partial \mathbf{k} / \partial t + (\nabla_k \omega) \cdot (\nabla \mathbf{k})^{\mathrm{T}} = 0 \qquad (2.12.19)$$

where the superscript T denotes a transposition. In our case, because of (2.12.13) $\nabla \mathbf{k}$ is symmetric (i.e., $(\nabla \mathbf{k})^{T} = \nabla \mathbf{k}$) and (2.12.19) reduces to

$$\partial \mathbf{k} / \partial t + \mathbf{v}_{\mathbf{a}} \cdot \nabla \mathbf{k} = 0 \tag{2.12.20}$$

The group velocity $\mathbf{v}_q = \nabla_k \omega$ in the homogeneous medium is uniform. The group rays in r-space obtained by integrating $\dot{\mathbf{r}} = \mathbf{v}_{q}$ are therefore straight lines, but they are not necessarily parallel. According to (2.12.20) an observer moving along the group ray with a speed v_q sees waves with the same wave number and consequently, through the dispersion relation (2.12.8), the same frequency. Hence k and ω propagate with \mathbf{v}_a , while the phase front propagates with v_p . As we have shown earlier in this section, the two velocities do not in general have equal magnitude nor equal direction. These kinematic properties of group velocity are very important and their usefulness can be demonstrated by looking at the following two examples. In the first case we assume the initial wave perturbation or equivalently, the initial distribution in k, is known. Then the future distribution in k can be obtained by letting each value of k be displaced by a vector \mathbf{v}_{at} . In the second case let us consider a radiation problem. In this problem waves of various k are continuously created by the source and they propagate out with a velocity v_q . These ideas are further amplified in Section 14.

If the medium is inhomogeneous, we shall assume that a local dispersion relation for each mode

$$\omega = \omega(\mathbf{k}, \mathbf{r}) \tag{2.12.21}$$

still exists. In this case (2.12.17) reduces to

$$d\mathbf{k}/dt = \partial \mathbf{k}/\partial t + \mathbf{v}_{g} \cdot \nabla \mathbf{k} = -\partial \omega/\partial \mathbf{r} \qquad (2.12.22)$$

The group velocity in this case varies with **r** and the group rays are generally curved. Along these group rays, values of **k** are not conserved as in the homogeneous medium; its change is prescribed by (2.12.22). But it is interesting to note that ω is still constant along the group ray. This can be proved by premultiplying (2.12.17) by \mathbf{v}_g and by noting $\mathbf{v}_g \cdot \partial \mathbf{k}/\partial t = \partial \omega/\partial t$ to produce

$$\partial \omega / \partial t + \mathbf{v}_{g} \cdot \nabla \omega = 0 \tag{2.12.23}$$

Hence ω is convected with the group velocity.

In concluding this section we wish to draw an analogy between the above treatment of wave propagation and the study of mechanics. For this purpose we note that (2.12.22) is

$$d\mathbf{k}/dt = -\partial\omega/\partial\mathbf{r}, \quad d\mathbf{r}/dt = \partial\omega/\partial\mathbf{k}$$
 (2.12.24)

which are just Hamilton's equations with **k** analogous to momentum and ω to Hamilton's function.

The kinematic approach to wave propagation problems will be discussed in more detail in Chapter 5 when we introduce the method of geometric optics.

2.13 Excitation of Fields

Up to this point, we have confined our discussions to the propagation of characteristic plane waves in the absence of external sources. In this section we shall take up the subject of excitation of the electromagnetic fields by external sources. From Maxwell's equations for a homogeneous medium, after the application of the Fourier transform, we obtain the wave equation

$$(k^{2}\mathbf{I} - \mathbf{k}\mathbf{k} - k_{0}^{2}\mathbf{K}) \cdot \mathbf{E}(\mathbf{k}, \omega) = -j\mu_{0}\omega\mathbf{J}(\mathbf{k}, \omega) \qquad (2.13.1)$$

where $J(\mathbf{k}, \omega)$ is the Fourier component of the external current density. Note that J must satisfy the charge continuity equation. Let us first consider the case for an isotropic medium in which

$$\mathbf{K} = K_{\perp}(\mathbf{I} - \mathbf{k}\mathbf{k}/k^2) + K_{\parallel}\mathbf{k}\mathbf{k}/k^2$$

Substituting into (2.13.1), we obtain

$$[(k^{2} - k_{0}^{2}K_{\perp})\mathbf{I} + (K_{\perp}k_{0}^{2} - k_{0}^{2}K_{\parallel} - k^{2})\mathbf{k}\mathbf{k}/k^{2}] \cdot \mathbf{E} = j\mu_{0}\omega\mathbf{J} \quad (2.13.2)$$

We decompose E into

$$\mathbf{E} = \mathbf{E}_{\perp} + \mathbf{E}_{\parallel} \tag{2.13.3}$$

where $\mathbf{E}_{\perp} \cdot \mathbf{k} = 0$, $\mathbf{E}_{\parallel} \cdot \mathbf{k} = kE_{\parallel}$ are the transverse component and longitudinal component, respectively. Substituting (2.13.3) into (2.13.2), we have

$$(k^{2} - k_{0}^{2}K_{\perp})\mathbf{E}_{\perp} - k_{0}^{2}K_{\parallel}\mathbf{E}_{\parallel} = -j\mu_{0}\omega\mathbf{J}$$
(2.13.4)

Decomposing (2.12.4) to the two components, the electric field now can be solved in terms of the current. We obtain

$$\mathbf{E}_{\perp} = -j\mu_0 \omega \mathbf{J}_{\perp}(\mathbf{k}, \omega) / (k^2 - k_0^2 K_{\perp})$$
(2.13.5a)

$$\mathbf{E}_{\parallel} = j \mathbf{J}_{\parallel}(\mathbf{k}, \omega) / \omega \varepsilon_0 K_{\parallel} \tag{2.13.5b}$$

where $\mathbf{J} = \mathbf{J}_{\perp} + \mathbf{J}_{\parallel}$ has been used.

As we have mentioned earlier, in an isotropic medium, there exist independent longitudinal and transverse modes. The electric field $E(\mathbf{r}, \mathbf{t})$ can then be expressed as the inverse transform of (2.13.5).

$$\mathbf{E}(\mathbf{r},t) = \frac{1}{(2\pi)^4} \int e^{j(\omega t - \mathbf{k} \cdot \mathbf{r})} \left\{ \frac{j \mathbf{J}_{\parallel}(\mathbf{k},\omega)}{\varepsilon_0 \omega K_{\parallel}(\mathbf{k},\omega)} - \frac{j \mu_0 \omega \mathbf{J}_{\perp}(\mathbf{k},\omega)}{k^2 - k_0^2 K_{\perp}(\mathbf{k},\omega)} \right\} d\mathbf{k} \, d\omega$$
(2.13.6)

The explicit form of the electric field depends on our ability to evaluate the fourfold integral. We discuss certain asymptotic techniques in the next section. Other field quantities such as H, D, B can be computed from (2.13.6) through Maxwell's equations.

Next, we consider a general anisotropic medium. In Section 6, we have seen that in such a medium, there exist in general several characteristic modes, \mathbf{a}_{α} , and we have derived an orthonormality relation among the different modes in a coordinate system with **k** in the direction of the z-axis, the primed system; i.e.,

$$(\delta_{ij} - s_i' s_j') b_{i\alpha}'^* a_{j\beta} = \delta_{\alpha\beta}$$
(2.7.15)

or

$$K'_{ij}a'_{j\beta}b'^*_{i\alpha} = n_{\alpha}^2 \,\delta_{\alpha\beta} \tag{2.7.16}$$

Since in later calculation of the inverse transform, we will integrate over **k** space, we must express the characteristic modes in a system with \mathbf{B}_0 in the fixed z-axis and the wave vector in an arbitrary direction $(\sin \theta \times \cos \phi, \sin \theta \sin \phi, \cos \theta)$.



Fig. 2.13-1. Three coordinate systems.

As shown in Fig. 2.13-1, the new coordinate system is obtained from the primed system by two rotational transformations. First, for a rotation about the y'-axis in a left-handed sense by θ , the transform is governed by

$$\mathbf{T}_{1} = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}$$
(2.13.7)

This transformation is followed by a rotation about the z_0 -axis in a righthanded sense by an angle ϕ , for which the transform is

$$\mathbf{T}_{2} = \begin{bmatrix} \cos \phi & -\sin \phi & 0\\ \sin \phi & \cos \phi & 0\\ 0 & 0 & 1 \end{bmatrix}$$
(2.13.8)

The characteristic mode in the new coordinates (shown as III) is given by

$$a_{x} = \cos\phi\cos\theta \ a_{x}' - \sin\phi \ a_{y}' + \cos\phi\sin\theta \ a_{z}'$$

$$a_{y} = \sin\phi\cos\theta \ a_{x}' + \cos\phi \ a_{y}' + \sin\phi\sin\theta \ a_{z}' \qquad (2.13.9)$$

$$a_{z} = -\sin\theta \ a_{x}' + \cos\theta \ a_{z}'$$

68

where the components a_x' , a_y' , and a_z' are given in Section 6. The relative dielectric tensor in the new system is given by

$$\mathbf{K} = \mathbf{T}_2 \cdot \mathbf{T}_1 \cdot \mathbf{K}' \cdot \mathbf{T}_1^{-1} \cdot \mathbf{T}_2^{-1}$$
(2.13.10)

which will have the same form as \mathbf{K} in a II-coordinate system if the medium does not have spatial dispersion.

Since all the transformation is orthogonal, the characteristic values n_{α}^{2} in the new system will not be changed. The orthonormal relations are also valid for the new system, i.e.,

$$(\delta_{ij} - s_i s_j) b_{i\alpha}^* a_{j\beta} = \delta_{\alpha\beta}$$
(2.13.11)

or

$$K_{ij}a_{j\beta}b_{i\alpha}^* = n_{\alpha}^2 \,\delta_{\alpha\beta} \tag{2.13.12}$$

Now, in the new coordinates, we expand the solution of (2.13.1) in terms of the normal modes

$$\mathbf{E}(\mathbf{k},\omega) = \sum_{\alpha} E_{\alpha} \mathbf{a}_{\alpha} \qquad (2.13.13)$$

Substituting (2.13.13) into (2.13.1), the equation becomes in component form

$$(k^{2} \delta_{ij} - k_{i}k_{j} - k_{0}^{2}K_{ij}) \sum_{\alpha} E_{\alpha}a_{j\alpha} = -j\mu_{0}\omega J_{i} \qquad (2.13.14)$$

Multiplying (2.13.14) by $b_{i\beta}^*$ and applying the orthonormality conditions (2.13.11) and (2.13.12), we have

$$(k^2/k_0^2 - n_\alpha^2)E_\alpha = (-j/\omega\varepsilon_0)(\mathbf{J}\cdot\mathbf{b}_\alpha^*) \qquad (2.13.15)$$

The total electric field in the transform domain is the sum of all the characteristic modes as given by (2.13.13). Each of these characteristic modes can be solved in (2.13.15). Our result is

$$\mathbf{E}(\mathbf{k},\omega) = \sum_{\alpha} \frac{-j(\mathbf{J} \cdot \mathbf{b}_{\alpha}^{*})\mathbf{a}_{\alpha}}{\omega\varepsilon_{0}(k^{2}/k_{0}^{2} - n_{\alpha}^{2})}$$
(2.13.16)

The total electric field in the space-time domain is just the inverse transform of (2.13.16) given by

$$\mathbf{E}(\mathbf{r},t) = \frac{1}{(2\pi)^4} \sum_{\alpha} \int \frac{-j}{\omega\varepsilon_0} \cdot \frac{[\mathbf{J}(\mathbf{k},\omega) \cdot \mathbf{b}_{\alpha}^*] \mathbf{a}_{\alpha}}{k^2/k_0^2 - n_{\alpha}^2(\mathbf{k},\omega)} e^{j(\omega t - \mathbf{k} \cdot \mathbf{r})} d\mathbf{k} d\omega \quad (2.13.17)$$

where, for any spatially nondispersive medium, $n_{\alpha}(\mathbf{k}, \omega)$ is a function of ω and the direction of \mathbf{k} but not $|\mathbf{k}|$. For the case of forced monochromatic sinusoidal oscillations, the current density is given by

$$\mathbf{J}(\mathbf{r},t) = \mathbf{J}(\mathbf{r})e^{j\omega_0 t}$$
(2.13.18)

with transformed current

$$\mathbf{J}(\mathbf{k},\omega) = 2\pi\delta(\omega - \omega_0)\mathbf{J}(\mathbf{k})$$
(2.13.19)

where

$$\mathbf{J}(\mathbf{k}) = \int e^{j\mathbf{k}\cdot\mathbf{r}} \,\mathbf{J}(\mathbf{r}) \,d\mathbf{r} \qquad (2.13.20)$$

Substituting (2.13.20) into (2.13.6) results in

$$\mathbf{E}(\mathbf{r},t) = \frac{j}{(2\pi)^3} \int e^{j(\omega_0 t - \mathbf{k} \cdot \mathbf{r})} \left\{ \frac{\mathbf{J}_{\parallel}(\mathbf{k})}{\varepsilon_0 \omega K_{\parallel}(\mathbf{k},\omega_0)} - \frac{\mu_0 \omega_0 \mathbf{J}_{\perp}(\mathbf{k})}{k^2 - k_0^2 K_{\perp}(\mathbf{k},\omega_0)} \right\} d\mathbf{k}$$
(2.13.21)

for isotropic media. Similarly for the anisotropic case with force oscillations we obtain from (2.13.17)

$$\mathbf{E}(\mathbf{r},t) = \frac{-j}{(2\pi)^3} \sum_{\alpha} \int \frac{1}{\omega_0 \varepsilon_0} \cdot \frac{[\mathbf{J}(\mathbf{k}) \cdot \mathbf{b}_{\alpha}^*] \mathbf{a}_{\alpha}}{k^2 / k_0^2(\omega_0) - n_{\alpha}^2(\mathbf{k},\omega_0)} e^{j(\omega_0 t - \mathbf{k} \cdot \mathbf{r})} d\mathbf{k} (2.13.22)$$

We want to point out here that the method we used in deriving (2.13.21) and (2.13.22) is equivalent to the Green's function method. It can be shown, for example, that the dyadic Green's function $\Gamma(\mathbf{r}, \mathbf{r}')$ satisfying the equation

$$\nabla \times \nabla \times \Gamma - k_0^2 \mathbf{K} \cdot \mathbf{\Gamma} = \mathbf{I} \delta(\mathbf{r} - \mathbf{r}')$$
(2.13.23)

is given by

$$\mathbf{\Gamma}(\mathbf{r},\mathbf{r}') = \frac{1}{(2\pi)^3} \sum_{\alpha} \int \frac{\mathbf{a}_{\alpha} \mathbf{b}_{\alpha}^* e^{-j\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}}{k^2 - k_0^2 n_{\alpha}^2(\mathbf{k},\omega_0)} d\mathbf{k} \qquad (2.13.24)$$

In (2.13.23) K is the relative dielectric tensor operator in the sense discussed in Section 4. Therefore the solution for the original wave equation after the Fourier transform in time

$$\nabla \times \nabla \times \mathbf{E} - k_0^2 \mathbf{K} \cdot \mathbf{E} = -j\mu_0 \omega_0 \mathbf{J}(\mathbf{r})$$
(2.13.25)

can be written as

$$\mathbf{E}(\mathbf{r},\omega_{0}) = \int -j\mu_{0}\omega_{0}\mathbf{\Gamma}(\mathbf{r},\mathbf{r}')\cdot\mathbf{J}(\mathbf{r}')\,d\mathbf{r}'$$

$$= \frac{-j}{(2\pi)^{3}}\sum_{\alpha}\int \frac{1}{\omega_{0}\varepsilon_{0}} \frac{\mathbf{a}_{\alpha}\mathbf{b}_{\alpha}^{*}\cdot\mathbf{J}(\mathbf{r}')}{k^{2}/k_{0}^{2}-n_{\alpha}^{2}(\mathbf{k},\omega_{0})}e^{-j\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}\,d\mathbf{r}'\,d\mathbf{k}$$

$$= \frac{-j}{(2\pi)^{3}}\sum_{\alpha}\int \frac{1}{\omega_{0}\varepsilon_{0}} \frac{\mathbf{a}_{\alpha}[\mathbf{b}_{\alpha}^{*}\cdot\mathbf{J}(\mathbf{k})]}{k^{2}/k_{0}^{2}-n_{\alpha}^{2}(\mathbf{k},\omega_{0})}e^{-j\mathbf{k}\cdot\mathbf{r}}\,d\mathbf{k} \qquad (2.13.26)$$

which combining with the time variation $e^{j\omega_0 t}$ is identical with (2.13.22).

The analytic computation of (2.13.6) and (2.13.22) in general is very complicated if not impossible. Approximate methods developed for some special cases are usually employed. One of the most important approximation procedures is the asymptotic evaluation of the fields at large distances, or the far field computation. A general asymptotic technique is given in Appendix B. In the next section we shall apply this technique to (2.13.22).

If the time variation is not monochromatic, then we have to go back to (2.13.6) and (2.13.17) to include the inverse Fourier transform in time as well as in space (see Appendix B).

2.14 Dyadic Green's Functions

As mentioned in the last section, the normal mode expansion method of computing the excitation of fields due to external sources is equivalent to the technique of the Green's functions. In many radiation and scattering problems it is useful to have explicit expressions of the Green's function in an infinite region. In the following, we shall discuss the general procedure of evaluating these integrals asymptotically.

In general, the dyadic Green's function for the wave equation in a medium without spatial dispersion satisfies the equation (time variation $e^{j\omega t}$)

$$\nabla \times \nabla \times \Gamma(\mathbf{r}, \mathbf{r}') - k_0^2 \mathbf{K} \cdot \mathbf{\Gamma}(\mathbf{r}, \mathbf{r}') = + \mathbf{I} \,\,\delta(\mathbf{r} - \mathbf{r}') \qquad (2.14.1)$$

where $\Gamma(\mathbf{r}, \mathbf{r}')$ is the dyadic Green's function and for homogeneous medium is a function of $\mathbf{R} = \mathbf{r} - \mathbf{r}'$.

Let us first consider the case of the isotropic medium so that $\mathbf{K} = \mathbf{I}K(\omega)$. Equation (2.14.1) then becomes

$$\{\nabla \nabla - [\nabla^2 + k_0^2 K(\omega)] \mathbf{I}\} \cdot \mathbf{\Gamma} = \mathbf{I} \,\,\delta(\mathbf{r} - \mathbf{r}') \tag{2.14.2}$$

Equation (2.14.2) can be written in the form

$$[\nabla^2 + k_0^2 K(\omega)]\mathbf{\Gamma} = -\mathbf{I} \ \delta(\mathbf{r} - \mathbf{r}') + \nabla \nabla \cdot \mathbf{\Gamma} \qquad (2.14.3)$$

But from (2.14.1) for the present case, we have

$$\nabla \cdot \mathbf{\Gamma} = -(1/k_0^2 K) \nabla \,\delta(\mathbf{r} - \mathbf{r}') \tag{2.14.4}$$

Equation (2.14.3) then becomes

$$(\nabla^2 + k_0^2 K)\mathbf{\Gamma} = -(\mathbf{I} + (1/k_0^2 K)\nabla\nabla) \,\delta(\mathbf{r} - \mathbf{r}') \qquad (2.14.5)$$

Now define a function $G(\mathbf{r}, \mathbf{r}')$ such that

$$(\nabla^2 + k_0^2 K)G(\mathbf{r}, \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}')$$
 (2.14.6)

Operate on both sides of (2.14.6) by $(\mathbf{I} + (1/k_0^2 K) \nabla \nabla)$; since the operators commute with each other, we have

$$(\nabla^2 + k_0^2 K) \cdot (\mathbf{I} + (1/k_0^2 K) \nabla \nabla) G(\mathbf{r}, \mathbf{r}') = -(\mathbf{I} + (1/k_0^2 K) \nabla \nabla) \delta(\mathbf{r} - \mathbf{r}')$$
(2.14.7)

Comparing with (2.14.5), we obtain

$$\mathbf{\Gamma}(\mathbf{r},\mathbf{r}') = [\mathbf{I} + (\nabla \nabla / k_0^2 K)] G(\mathbf{r},\mathbf{r}') \qquad (2.14.8)$$

Therefore the procedure of obtaining the dyadic Green's function in an isotropic medium is to solve (2.14.6) for $G(\mathbf{r}, \mathbf{r}')$ first, then substitute it into (2.14.8). Here we note that generalized functions have been used in our derivation since $G(\mathbf{r}, \mathbf{r}')$ is singular at $\mathbf{r} = \mathbf{r}'$.

Equation (2.14.6) can be solved by many different methods. The method of Fourier transform has proved to be a powerful method to deal with equations more complex than (2.14.6). We shall use this method to demonstrate its technique. The transformed equation of (2.14.6) is

$$(-k^2 + k_0^2 K)G(\mathbf{k}) = -1$$
 (2.14.9)

The Green's function $G(\mathbf{r}, \mathbf{r}')$ is obtained by taking the inverse transform,

$$G(\mathbf{r},\mathbf{r}') = \frac{1}{(2\pi)^3} \int \frac{e^{-j\mathbf{k}\cdot\mathbf{R}}}{k^2 - k_0^2 K} d\mathbf{k}$$
(2.14.10)

The solution of (2.14.6) is defined uniquely only if we apply the radiation condition which requires that for a localized source, the excited waves must

be outgoing away from the source. If the medium is lossy, then radiation condition requires that the wave decay away from the source region. In evaluating (2.14.10) this condition must be used.

In the following we shall discuss two different approaches in this evaluation. First, the integral will be computed in a spherical coordinate system. Let the coordinates be so directed that the polar axis is in the direction of **R**. Equation (2.14.10) becomes

$$G(\mathbf{r}, \mathbf{r}') = \frac{1}{(2\pi)^3} \int_0^\infty k^2 \, dk \int_0^\pi \sin \theta \, d\theta \int_0^{2\pi} \frac{e^{-jkR \cos \theta}}{k^2 - k_0^2 K} \, d\phi$$

= $\frac{1}{2\pi^2 R} \int_0^\infty \frac{\sin kR}{k^2 - k_0^2 K} \, k \, dk$
= $\frac{1}{4\pi^2 R} \int_{-\infty}^{+\infty} \frac{p \sin p}{p^2 - \sigma^2} \, dp$ (2.14.11)

where the third line is obtained by substituting p = kR and $\sigma^2 = k_0^2 KR^2$. The *p*-integration is carried out by contour integral method. The integral can be written in the form

$$\frac{1}{2j} \int_{-\infty}^{+\infty} \frac{pe^{jp}}{p^2 - \sigma^2} dp - \frac{1}{2j} \int_{-\infty}^{+\infty} \frac{pe^{-jp}}{p^2 - \sigma^2} dp \qquad (2.14.12)$$

In the integrand, there are two poles, $\sigma = \pm k_0 R \sqrt{K} = \pm \sigma_1$. For a lossy medium K has a negative imaginary part. Choosing the branch such that \sqrt{K} also has a negative imaginary part, the positions of the two poles



of the integrand in (2.14.12) are shown in Fig. 2.14-1. They approach the real axis as shown by arrows for the lossless case. The first integral can now be integrated by completing the contour as shown in Fig. 2.14-2.

$$\int_{-\infty}^{+\infty} \frac{p e^{jp}}{p^2 - \sigma^2} dp = \oint () dp - \oint () dp$$



Fig. 2.14-2. Contour for the first integral (2.14.12).

The second term on the right-hand side of the above equation vanishes as the radius of the contour approaches infinity. The contribution from the first term of (2.14.12) comes from the pole at $-\sigma_1$.

It gives us

$$\int_{-\infty}^{+\infty} \left[p e^{jp} / (p^2 - \sigma^2) \right] dp = (2\pi j/2) e^{-j\sigma_1} = \pi j e^{-jk_0 \sqrt{KR}} \quad (2.14.13)$$

which is in a form of outgoing waves.

We note that either by including some loss in the medium as in Fig. 2.14-1 or by choosing the contour as in Fig. 2.14-2 for lossless case, the radiation condition is satisfied. For the second term in (2.14.12), the contour is shown in Fig. 2.14-3, and we obtain

$$\int_{-\infty}^{+\infty} \left[p e^{-jp} / (p^2 - \sigma^2) \right] dp = -(2\pi j/2) e^{-j\sigma_1} = -\pi j e^{-jk_0 \sqrt{KR}} \quad (2.14.14)$$

Fig. 2.14-3. Contour for the second integral of (2.14.12).



Substituting (2.14.12), (2.14.13), and (2.14.14), into (2.14.11), we obtain the well-known result,

$$G(\mathbf{r}, \mathbf{r}') = -\frac{1}{4\pi R} e^{-jk_0\sqrt{K}R}$$
(2.14.15)

Next, we apply the technique discussed in Appendix B to obtain the asymptotic form of the Green's function in (2.14.10). The idea of the

asymptotic technique is based on the fact that the major contribution to the integral (2.14.10) at a large distance comes from those groups of plane waves with group velocity in the direction of **R** (see discussion in Section 2.12). The asymptotic form of (2.14.10) is written as (see B.1.17)

$$G(\mathbf{r},\mathbf{r}') \sim \frac{1}{2\pi} \sum_{i} \frac{C^{(i)} \exp(-j\mathbf{k}^{(i)} \cdot \mathbf{R})}{|K^{(i)}|^{1/2} \cdot |\nabla_{k} D|_{\mathbf{k}^{(i)}}R}$$
(2.14.16)

where

$$D = k^2 - k_0^2 K \tag{2.14.17}$$

and D = 0 corresponds to the wave vector surface (or dispersion surface). The summation is over those $\mathbf{k}^{(i)}$ on the wave vector surface such that the group velocities there are along the direction of **R**. $K^{(i)}$ is the Gaussian curvature of the wave vector surface at $\mathbf{k}^{(i)}$ (not to be confused with the relative permittivity K). For the isotropic medium, a case of concern at the moment, the wave vector surface is a sphere defined by $k = k_0 \sqrt{K}$. Therefore the Gaussian curvature is just $1/k^2$, and

$$\nabla_k D = 2\mathbf{k} \tag{2.14.18}$$

which is also the direction of the group velocity for this case (see Section 12). This means that for a given **R**, only one group of waves contributes to the far field, the group that centered around **k** where **k** || **R**. Hence $\mathbf{k}^{(i)} \cdot \mathbf{R} = k_0 \sqrt{KR}$ in (2.14.16). The constant $C^{(i)}$ takes the value +1 since the spherical surface is convex to the direction $\nabla_k D$, or **k** (see Appendix B).

Substituting all these into (2.14.16), we obtain immediately

$$G(\mathbf{R}) \sim \frac{1}{4\pi R} \exp(-jk_0\sqrt{KR})$$
 (2.14.19)

which happens to agree with the exact expression (2.14.15). Thus we see that for this particular case the asymptotic evaluation yields an exact solution. This, of course, is not to be expected in general.

We can now substitute (2.14.15) or (2.14.19) into (2.14.7) to obtain the dyadic Green's function $\Gamma(\mathbf{r}, \mathbf{r}')$. But we note that $G(\mathbf{r}, \mathbf{r}')$ is singular at the source point $|\mathbf{R}| = 0$; any differentiation makes the resultant dyadic Green's function even more singular at the source point. Therefore if $\Gamma(\mathbf{r}, \mathbf{r}')$ is used to compute the field inside the source region, we have to be careful about the operations. For this purpose let us define the principal

value of a volume integral as

$$\lim_{\epsilon \to 0} \int_{|\mathbf{R}| \ge \epsilon} \mathbf{J}(\mathbf{r}') \cdot \mathbf{\Gamma}(\mathbf{r}, \mathbf{r}') \, d\mathbf{r}' \qquad (2.14.20)$$

where Γ is singular at $|\mathbf{R}| = 0$. It can be proved that when defined this way, the principal value is uniquely determined. It is now convenient to separate the dyadic Green's function into two parts: a regular part and a singular part in the sense of generalized functions. Let us consider the *ikth* component of the dyadic Green's function Γ_{ik} . In the sense of generalized functions the inner product is given by

$$(\Gamma_{ik}, \phi) = \int \Gamma_{ik}(\mathbf{R})\phi(\mathbf{R}) d\mathbf{R}$$
$$= \lim_{\epsilon \to 0} \left\{ \int_{|\mathbf{R}| > \epsilon} \Gamma_{ik}(\mathbf{R})\phi(\mathbf{R}) d\mathbf{R} + \int_{|\mathbf{R}| < \epsilon} \Gamma_{ik}(\mathbf{R})\phi(\mathbf{R}) d\mathbf{R} \right\} \quad (2.14.21)$$

where ϕ is in the space of the so-called test functions. The integral is separated into two parts. The first part involves integration outside of an infinitesimal sphere centered at $|\mathbf{R}| = 0$. The singularity of the Green's dyadic is not in the region of integration and the result is just the principal value of the integral. The second part of the integral involves integration of the singularity. We proceed to evaluate it by first defining a Fourier transform:

$$\Gamma_{ik}(\mathbf{R}) = \frac{1}{(2\pi)^3} \int \Gamma_{ik}(\mathbf{k}) e^{-j\mathbf{k}\cdot\mathbf{R}} d\mathbf{k} \qquad (2.14.22)$$

Substituting (2.14.22) into the second term of (2.14.21), we have

$$\lim_{\epsilon \to 0} \int_{|\mathbf{R}| < \epsilon} \frac{1}{(2\pi)^3} \int \Gamma_{ik}(\mathbf{k}) e^{-j\mathbf{k} \cdot \mathbf{R}} \phi(\mathbf{R}) d\mathbf{k} d\mathbf{R}$$

$$= \lim_{\epsilon \to 0} \frac{1}{(2\pi)^3} \int \Gamma_{ik}(\mathbf{k}) d\mathbf{k} \int_{|\mathbf{R}| < \epsilon} e^{-j\mathbf{k} \cdot \mathbf{R}} \phi(\mathbf{R}) d\mathbf{R}$$

$$= \lim_{\epsilon \to 0} \phi(0) \frac{1}{2\pi^2} \int \Gamma_{ik}(\mathbf{k}) \frac{d\mathbf{k}}{k} \int_0^{\epsilon} R \sin kR dR$$

$$= \frac{1}{2\pi^2} \lim_{\epsilon \to 0} \phi(0) \int \frac{\sin k\epsilon - k\epsilon \cos k\epsilon}{k^2} \cdot \frac{\Gamma_{ik}(\mathbf{k})}{k} d\mathbf{k}$$

$$= \frac{\phi(0)}{2\pi^2} \lim_{\epsilon \to 0} \epsilon^2 \int \frac{\sin k\epsilon - k\epsilon \cos k\epsilon}{(k\epsilon)^2} \cdot \frac{\Gamma_{ik}(\mathbf{k})}{k} d\mathbf{k} \quad (2.14.23)$$

Therefore, the inner product (2.14.21) becomes

$$(\Gamma_{ik}, \phi) = \lim_{\epsilon \to 0} \int_{|\mathbf{R}| > \epsilon} \Gamma_{ik}(\mathbf{R}) \phi(\mathbf{R}) \, d\mathbf{R} + \lim_{\epsilon \to 0} \frac{\phi(0)}{2\pi^2} \, \epsilon^2 \int \frac{\sin k\epsilon - k\epsilon \cos k\epsilon}{(k\epsilon)^2} \cdot \frac{\Gamma_{ik}(\mathbf{k})}{k} \, d\mathbf{k} \qquad (2.14.24)$$

In the sense of generalized functions, the dyadic Green's function can be expressed as

$$\Gamma_{ik}(\mathbf{R}) = P\Gamma_{ik}(\mathbf{R}) + \frac{\delta(\mathbf{R})}{2\pi^2} \lim_{\epsilon \to 0} \epsilon^2 \int \frac{\sin k\epsilon - k\epsilon \cos k\epsilon}{(k\epsilon)^2} \frac{\Gamma_{ik}(\mathbf{k})}{k} d\mathbf{k}$$
(2.14.25)

where P denotes the principal value. Equation (2.14.25) is a general and useful formula which separates the regular and singular parts of the dyadic Green's function.

We wish now to apply (2.14.25) to the case of isotropic media. From (2.14.7) and (2.14.8), we have in the transformed domain

$$\Gamma_{ij}(\mathbf{k}) = \left(\delta_{ij} - \frac{k_i k_j}{k_0^2 K}\right) \frac{1}{k^2 - k_0^2 K}$$
(2.14.26)

Substituting (2.14.26) into the integral in (2.14.25), it is easy to show that all components except the diagonal ones of the second term in (2.14.25) vanish due to integration over the angle. The diagonal components are

$$-\frac{\delta_{ij}\,\delta(\mathbf{R})}{2\pi^2}\cdot\frac{4\pi}{3k_0^2K}\int_0^\infty\frac{\sin\nu-c\,\cos\nu}{\nu}\,d\nu=\frac{-\delta_{ij}}{3k_0^2K}\,\delta(\mathbf{R})\ (2.14.27)$$

Therefore, the Green's dyadic in an isotropic medium is given by

$$\mathbf{\Gamma}(\mathbf{R}) = P\left(\mathbf{I} + \frac{\nabla \nabla}{k_0^2 K}\right) G(\mathbf{R}) - \frac{\mathbf{I}}{3k_0^2 K} \,\delta(\mathbf{R}) \qquad (2.14.28)$$

When this dyadic Green's function is applied outside the source region, it yields the usual result. When it is applied to the source region, the singularity at the origin is taken care of by the delta function part, and the principal value defined by (2.14.20) must be used to all integrals involving the first term of (2.14.28). Therefore by admitting the generalized functions in our solution, (2.14.28) gives a standard procedure for taking care of the singularity.

We next consider the case for an isotropic medium with spatial dispersion.

For this case, the relative dielectric tensor is given by

$$\mathbf{K} = K_{\!\perp}\!\!\left(\mathbf{I} - rac{\mathbf{kk}}{k^2}
ight) + K_{\scriptscriptstyle\parallel} rac{\mathbf{kk}}{k^2}$$

Equation (2.14.1) may be solved by the Fourier transform technique, and the dyadic Green's function can be put in the form (2.14.22) with

$$\Gamma_{ik}(\mathbf{k}) = \frac{1}{k^2 - k_0^2 K_\perp} \left\{ \delta_{ik} - \left[1 - \frac{k_0^2}{k^2} \left(K_\perp - K_\parallel \right) \right] \frac{k_i k_j}{k_0^2 K_\parallel} \right\}$$
(2.14.29)

Integration of (2.14.22) with (2.14.29) as its integrand depends on the explicit forms of K_{\perp} and K_{\parallel} . Asymptotic technique may be used to compute the far field.

On the other hand, from the discussion in Section 13, we know that for the isotropic medium, it is possible to decompose the field into independent transverse and longitudinal modes. Therefore, we can define independent transverse and longitudinal Green's functions. As a matter of fact, from (2.13.6) and the definition of the Green's function, it is obvious that

$$G_{\parallel}(\mathbf{r},\mathbf{r}') = \frac{1}{(2\pi)^3} \int \frac{\exp(-j\mathbf{k}\cdot\mathbf{r})}{k_0^2 K_{\parallel}(\mathbf{k},\omega)} d\mathbf{k}$$
(2.14.30)

$$G_{\perp}(\mathbf{r},\mathbf{r}') = \frac{1}{(2\pi)^3} \int \frac{\exp(-j\mathbf{k}\cdot\mathbf{r})}{k_0^2 K_{\perp}(\mathbf{k},\omega) - k^2} \, d\mathbf{k}$$
(2.14.31)

where G_{\parallel} and G_{\perp} are the longitudinal and transverse Green's functions, respectively. Again, the evaluation of these two integrals depend on the functions $K_{\parallel}(\mathbf{k}, \omega)$ and $K_{\perp}(\mathbf{k}, \omega)$. Such an evaluation for the warm plasma is carried out in Chapter 3.

Finally we consider the problem of finding the dyadic Green's function for an anisotropic medium with no spatial dispersion. The dyadic Green's function is given by (2.13.24)

$$\mathbf{\Gamma}(\mathbf{r},\mathbf{r}') = \frac{1}{(2\pi)^3} \sum_{\alpha} \int \frac{\mathbf{a}_{\alpha} \mathbf{b}_{\alpha}^* e^{-j\mathbf{k} \cdot (\mathbf{r}-\mathbf{r}')}}{k^2 - k_0^2 n_{\alpha}^2} d\mathbf{k}$$
(2.14.32)

where \mathbf{a}_{α} and \mathbf{b}_{α} are the polarization vectors and n_{α} is the refractive index for the α th characteristic mode. For the lossless medium without spatial dispersion $\mathbf{b}_{\alpha} = \mathbf{a}_{\alpha}$ and n_{α} depends only on the direction but not the magnitude of the wave vector \mathbf{k} . The exact evaluation of (2.14.32) in general is not possible. In the following, the asymptotic technique discussed in Appendix B is applied to find the far field expression for the dyadic Green's function. We have, for (2.14.32)

$$\mathbf{\Gamma}(\mathbf{r},\mathbf{r}') \sim \frac{1}{2\pi} \sum_{\alpha,i} \frac{C^{(i)} \mathbf{a}_{\alpha}^{(i)} \mathbf{b}_{\alpha}^{(i)} \exp[-j\mathbf{k}^{(i)} \cdot \mathbf{R}]}{|K_{\alpha}^{(i)}|^{1/2} |\nabla_{k} D|_{\mathbf{k}_{\alpha}^{(i)}} R} \qquad (2.14.33)$$

where

$$D = k^2 - k_0^2 n_{\alpha}^2 \tag{2.14.34}$$

 $K_{\alpha}^{(i)}$ is the Gaussian curvature of the dispersion surface D = 0 at the point $\mathbf{k} = \mathbf{k}_{\alpha}^{(i)}$. Again, the summation is over those points $k_{\alpha}^{(i)}$ on the dispersion surface such that its corresponding group velocity is in the direction of **R**. Only those groups of waves centered around $\mathbf{k}_{\alpha}^{(i)}$ will contribute to the far field at **R**.

Since the medium is anisotropic, the group velocity at $\mathbf{k}_{\alpha}^{(i)}$ is in general not in the same direction as $\mathbf{k}_{\alpha}^{(i)}$. Let us assume that the angle between them is $\alpha_{\alpha}^{(i)}$ (Fig. 2.14-4).





For the case in which the medium has a symmetry axis along z-direction, the angle α is given by (2.12-6)

$$\tan \alpha = -\frac{1}{n} \frac{\partial n}{\partial \theta}$$
(2.12.6)

for each mode. In general, it may be found once the refractive index is given. Therefore, for a given **R**, $\mathbf{k}_{x}^{(i)}$ may be obtained by solving simultaneously the dispersion relation D = 0 and (2.12.6) or its equivalent.

The exponential in (2.14.33) now may be expressed in terms of $\alpha_{\alpha}^{(i)}$ as

$$\mathbf{k}_{\alpha}^{(i)} \cdot \mathbf{R} = k_{\alpha}^{(i)} R \cos \alpha_{\alpha}^{(i)}$$
(2.14.35)

where $k_{\alpha}^{(i)} = | \mathbf{k}_{\alpha}^{i} |$.

The factor $|\nabla_k D|_{\mathbf{k}_{\alpha}^{(i)}}$ can be evaluated in the following manner. Since n_{α} does not depend on $k = |\mathbf{k}|$, the derivative of D with respect to k is

just 2k. This is the component of $\nabla_k D$ along the direction **k**. From Fig. 2.14-4 it is obvious that

$$|\nabla_k D| = 2k \sec \alpha \qquad (2.14.36)$$

Therefore, when evaluated at $\mathbf{k}_{\alpha}^{(i)}$, we have

$$\left| \nabla_k D \right|_{\mathbf{k}_{\alpha}^{(i)}} = 2k_{\alpha}^{(i)} \sec \alpha_{\alpha}^{(i)} \tag{2.14.37}$$

Substituting (2.14.35) and (2.14.37) into (2.14.35), we obtain

$$\mathbf{\Gamma}(\mathbf{r},\mathbf{r}') \sim \frac{1}{4\pi R} \sum_{\alpha,i} \frac{C_{\alpha}^{(i)} \mathbf{a}_{\alpha}^{(i)} \mathbf{a}_{\alpha}^{(i)}}{k_{\alpha}^{(i)} | K_{\alpha}^{(i)} |^{1/2} \sec \alpha_{\alpha}^{(i)}} \quad (2.14.38)$$

The value of the constant $C_{\alpha}^{(i)}$ is determined in the following way. For $K_{\alpha}^{(i)} > 0$, it takes the values ± 1 where the dispersion surface D = 0 is convex to the direction $\pm \nabla_k D$. For $K^{(i)} < 0$, it takes the values $\pm j$ according as the direction $\nabla_k D$ is parallel or antiparallel to **R**.

Equation (2.14.38) gives the asymptotic expression for the dyadic Green's function for an anisotropic medium with no spatial dispersion. Further reduction of the formula can be made only when the relative dielectric tensor of the medium is given explicitly.

Thus, in this section, we have derived the dyadic Green's functions for three different media: the isotropic with no spatial dispersion (2.14.15); the isotropic with spatial dispersion (2.14.29), (2.14.30), and (2.14.31); and the anisotropic with no spatial dispersion given by (2.14.38). These expressions will be used in later chapters for the discussion of various different topics.

Problems

1. From (2.5.11) and (2.5.12), derive (2.5.15) and (2.5.17).

2. For the dielectric tensor ε in (2.6.10), if it is invariant under rotational transformation about the z-axis, prove that

$$\varepsilon_{xz} = \varepsilon_{yz} = 0$$
 and $\varepsilon_{xx} = \varepsilon_{yy}$.

3. If a tensor K(k) is symmetric and invariant under rotational transformation about the vector k, prove that

Problems

(a) for \mathbf{k} in the direction of the z-axis, the tensor can be written as

$$\mathbf{K}(\mathbf{k}) = \begin{bmatrix} K_{\perp}(k^2) & 0 & 0 \\ 0 & K_{\perp}(k^2) & 0 \\ 0 & 0 & K_{\mathbb{H}}(k^2) \end{bmatrix}$$

(b) for k in any arbitrary direction, the tensor can be put into the form

$$K_{ij}(\mathbf{k}) = K_{\perp}(k^2)(\delta_{ij} - k_i k_j/k^2) + K_{\parallel}(k^2)k_i k_j/k^2$$

where $K_{\parallel}(k^2)$ and $K_{\parallel}(k^2)$ are arbitrary functions of k^2 .

4. Given the relative dielectric constant for a cold, homogeneous, isotropic plasma $K(\omega)$ as

$$K(\omega) = 1 - \omega_p^2 / \omega(\omega - i\nu)$$

where ω_p and ν are constants, verify that Kramers-Kronig relations [Eq. (2.5.17)] are satisfied by this function $K(\omega)$.

5. For a uniaxial medium, the dielectric tensor is given by

$$\boldsymbol{\varepsilon}(\omega) = \begin{bmatrix} \varepsilon_1(\omega) & 0 & 0\\ 0 & \varepsilon_1(\omega) & 0\\ 0 & 0 & \varepsilon_3(\omega) \end{bmatrix}$$

(a) Find the refractive indices for plane wave propagation in this medium.

(b) Find the normal modes and discuss their polarizations.

6. In the coordinate system shown, the dielectric tensor for a magnetoplasma can be written as

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{xx} & \varepsilon_{xy} & 0 \\ -\varepsilon_{xy} & \varepsilon_{xx} & 0 \\ 0 & 0 & \varepsilon_{zz} \end{bmatrix}$$

(a) Find the principle axes for this tensor. (The principle axes are obtained by solving the eigenvalue problem.

$$\mathbf{\epsilon} \cdot \mathbf{A}^{\alpha} = \lambda_{\alpha} \mathbf{A}^{\alpha}, \qquad \alpha = 1, 2, 3$$

where λ^{α} is the eigenvalue for the α th eigen vector \mathbf{A}^{α} . The \mathbf{A}^{α} are the principle axes.)

(b) In the new coordinate system where the principle axes are the new coordinate axes, what is the form of the dielectric tensor? Also find the form of the wave equation [Eq. (2.7.5)] in the new system.

7. If $\mathbf{A} \times \mathbf{K} = \mathbf{K} \times \mathbf{A}$ where **K** is symmetric and **A** is any given vector, in what form must **K** be?

8. In a lossless anisotropic medium the relative dielectric tensor is given by

$$\mathbf{K} = \operatorname{Re} \mathbf{K} + j \operatorname{Im} \mathbf{K}$$

where Re denotes real part and Im denotes imaginary part.

(a) Show that Re K is symmetric and Im K is antisymmetric.

(b) Since the antisymmetric tensor is equivalent to some axial vector, the relation between the electric displacement vector and the electric field can be written in the form of

$$\mathbf{D} = \varepsilon_0 [\operatorname{Re} \mathbf{K} \cdot \mathbf{E} + j(\mathbf{E} \times \mathbf{g}_1)]$$

where \mathbf{g}_1 is a real vector called a gyration vector. A medium in which \mathbf{D} and \mathbf{E} are related this way is sometimes called a gyrotropic medium. How is \mathbf{g}_1 related to Im \mathbf{K} ?

(c) If the inverse of **K** exists and letting $\Gamma = \mathbf{K}^{-1}$. Γ must have the same properties as **K**. By a similar argument we can write

$$\varepsilon_0 \mathbf{E} = \operatorname{Re} \mathbf{\Gamma} \cdot \mathbf{D} + j(\mathbf{D} \times \mathbf{g}_2).$$

Now how is g_2 related to Im Γ ? And to Re K and g_1 ? Also, how is Re Γ related to Re K and g_1 ?

9. In an isotropic dispersive medium, the electric field for the longitudinal mode can be expressed by a scalar potential

$$E_{\parallel}(\mathbf{r},t) = -\nabla V(\mathbf{r},t).$$

In the static limit, i.e., as $\omega \rightarrow 0$, show that

(a) for a point source

$$\varrho(\mathbf{r}) = e \,\,\delta(\mathbf{r} - \mathbf{r}_0)$$

the potential is given by

$$V(\mathbf{r}) = \frac{e}{(2\pi)^3} \int d\mathbf{k} \, \frac{e^{[-j\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}_0)]}}{\varepsilon_0 k^2 K_{\parallel}(\mathbf{k},0)}$$

where $K_{\parallel}(\mathbf{k}, 0)$ is the zero frequency limit of the longitudinal relative dielectric constant.

(b) For $K_{ii}(\mathbf{k}, 0)$ of the form

$$K_{\parallel}(\mathbf{k},0) = 1 + 1/k^2 r_{\rm scr}^2$$

and using the Fourier inverse transform, prove that

$$V(\mathbf{r}) = \frac{e}{4\pi\varepsilon_0 |\mathbf{r} - \mathbf{r}_0|} e^{-|\mathbf{r} - \mathbf{r}_0|/r_{\rm scr}}$$

where $r_{\rm scr}$ is the screening distance.

10. Consider a lossy anisotropic medium. In the frequency region in which the medium is transparent, the time-averaged electric energy stored in a wave packet is

$$\langle U_E \rangle_t = \frac{1}{4} \mathbf{E}^* \cdot (d\omega \mathbf{\epsilon}/d\omega) \cdot \mathbf{E}$$

Show that this energy is equal to or greater than the energy in free space of a wave packet with the identical electric field.

References

AGRANOWICH, V. M., and GINZBURG, V. L. (1966). "Spatial Dispersion in Crystal Optics and the Theory of Excitons," Wiley (Interscience), New York.

- BUNKIN, F. V. (1957). On radiation in anisotropic media, Sov. Phys. JEPT 5, 277-283. DESCHAMPS, G. A., and KESLER, O. B. (1967). Radiation of an antenna in a compressible magnetoplasma, Radio Sci. 2 (N. S.), 757-767.
- LANDAU, L. D., and LIFSHITZ, E. M. (1961). "Electrodynamics of Continuous Media," Addison-Wesley, Reading, Massachusetts.
- PANOFSKY, W. K. H., and PHILLIPS, M. (1955). "Classical Electricity and Magnetism," Addison-Wesley, Reading, Massachusetts.

- RUKHADZE, A. A., and SILIN, V. P. (1961). Electrodynamics of media with spatial dispersion, Sov. Phys. Usp. 4, 459-484,
- STRATTON, J. A. (1941). "Electromagnetic Theory," McGraw-Hill, New York.
- VAN BLADEL, J. (1961). Some remarks on Green's dyadic for infinite space, *IRE Trans.* Antennas Propagat. AP-9, 563-566.
- WHITHAM, G. B. (1961). Group velocity and energy propagation for three-dimensional waves, *Commun. Pure Appl. Math.* 14, 675-691.

3. Waves in Fluid Plasma

3.1 Introduction

We have seen in Chapter 2 that because of principles of causality and Onsager relation, the dielectric tensor must satisfy certain properties. These properties are general and must be satisfied by almost all media. When we come to derive a dielectric constant expression, it is necessary to know the response of a given medium to electromagnetic fields. One of the tasks in this chapter and the next is to obtain the dielectric tensor for various plasmas.

A plasma is a collection of free electrons and various ions. The spatial distribution of electrons and ions may be homogeneous, inhomogeneous, or irregular. A steady magnetic field may exist to influence the motion of charged particles. There may be neutral molecules with which the electrons and ions make collisions. In order for the plasma to exist without a strong external force, it is necessary that the plasma be electrically neutral to a high degree of approximation. When forced into an electrically nonneutral state, the plasma has a tendency to recover the neutral state and consequently may lead into oscillations.

There is a hierarchy of equations that can be used to describe plasma kinetics. The choice of equations depends on the nature of the plasma and the accuracy with which we wish to describe the plasma response. Because of the long-range nature of Coulomb force, the plasma may be described as a collection of noninteracting charged particles except through the self-consistent electric field. Such a plasma may be referred to as being "cold" since the pressure force is ignored. When thermal effects are important, hydrodynamic equations can be used. The inclusion of the pressure force in hydrodynamic equations brings in plasma waves. The existence of plasma waves implies that a "warm" plasma is in constant agitation, giving rise to density as well as velocity fluctuations. The hydrodynamic description of a plasma is adequate if the thermal speed is small in comparison with wave speed. If this is not so, it is then necessary to use the distribution function approach. The exact description of the many-particle system is given by the BBGKY hierarchy or equivalently by the generalized stochastic equations. These equations are extremely difficult to work with and besides they are not closed. Generally, approximations can be used by ordering some parameters. Equations such as the kinetic equation, Boltzmann equation, Fokker-Planck equation, and Vlasov equation can be obtained, depending on approximations.

In this book, we assume that the hydrodynamic plasma model is good enough for our purposes. As a result, phenomena such as wave-particle interaction that lead to Landau damping are excluded. The hydrodynamic model is adequate for wave studies except near the region of refractive index poles. When the refractive index approaches the infinite value, the phase velocity of the wave is reduced to zero and is hence comparable with the thermal velocity. This is the condition for strong wave-particle interaction, and the problem can be dealt with conveniently only by using the distribution functions.

3.2 Charge Neutrality

A plasma is a collection of charged particles. The forces involved are Coulomb forces. We shall demonstrate that if charge neutrality is not fullfilled macroscopically, potential energy due to Coulomb forces can be enormous when compared with the thermal energy of the particle. Unless there is a strong external force to maintain this potential, charged particles will move in such a way as to reduce the potential difference and to restore electrical neutrality. A semiquantitative estimate on this problem can be made by considering a simple one-dimensional problem. Let there be a plasma occupying the space $z \ge 0$. It is initially neutral, and we wish to compute the energy required per particle to deplete all positive ions in a planar region. In the depleted region we have just electrons of density N. The electric potential produced by these electrons satisfies the Poisson's equation

$$d^2 V/dz^2 = Ne/\varepsilon_0 \tag{3.2.1}$$

Integrating (3.2.1) twice and making the simplifying assumptions V = dV/dz= 0 at z = 0, we obtain

$$V = Nez^2/2\varepsilon_0 \tag{3.2.2}$$

The potential given by (3.2.2) increases without limit like z^2 . The energy required to move a positively charged particle through such a potential hill is

$$U = eV = Ne^{2}z^{2}/2\varepsilon_{0} = \frac{1}{2}T(z^{2}/\lambda_{D}^{2})$$
(3.2.3)

This energy can be compared with the thermal energy of each particle $\frac{1}{2}T$ as done in (3.2.3). In order to avoid confusion all temperatures will be expressed in energy units since k is reserved as a symbol for wave number. Equation (3.2.3) also defines the Debye length λ_D given by

$$\lambda_D = (\varepsilon_0 T/Ne^2)^{1/2} \simeq 69.0 (T/N)^{1/2}$$
 m (3.2.4)

Note that this quantity depends only on temperature and density but not on the mass of particles.

The Debye length just defined has the meaning of a distance for which the energy required to deplete the positive charge is equal to the thermal energy. If external perturbations are not violent on the plasma, the energy per particle associated with random motions in the z-direction must still be approximately $\frac{1}{2}T$. This means that the perturbation on the plasma, such as by introducing a boundary, cannot extend much beyond a distance of the order of Debye length. Accordingly, near any boundary in a plasma, we would expect a sheath region of the order of Debye length within which the charge neutrality condition need not be maintained. Beyond the sheath region is the plasma region where charge neutrality is maintained to a good degree.

The tendency to maintain charge neutrality also gives rise to other interesting properties in a plasma. We list them and discuss them very briefly in the following.

(i) Oscillation. Suppose that an external force is applied and it produces charge separation in some region. Then suddenly this external force is removed. Immediately charged particles will be set into motion to restore charge neutrality and convert all its potential energy into kinetic energy. This kinetic energy will continue to drive particles away from charge neutral condition even when restored. As a result oscillation about the charge neutral condition occurs in the absence of any damping mechanism. We will discuss this problem more fully with a concrete example in the next section. (ii) Screening. If we introduce a charged particle in the plasma, the potential due to the charged particle will alter the distribution of plasma particles. The result is that the potential will decay much faster than the Coulomb potential and we say that the particle is screened. This problem is discussed in Section 4.

(iii) Ambipolar Diffusion. For a plasma of equal electron and ion temperatures electrons will diffuse faster than ions due to difference in mass. As electrons diffuse away and leave ions behind, an electric field is set up due to space charge. This electric field is so strong as to maintain quasineutrality condition. As a result the plasma diffuse as a whole with an effective mass equal to the average of electron and ion mass. We shall not discuss this problem further since we are more concerned with waves. Readers should consult some books on plasma physics on this subject.

(iv) Scattering. The screening idea advanced in (ii) also applies to particles in the plasma. Since ions are very massive neutrality is maintained by electrons forming shielding clouds about ions. As ions move so do the shielding clouds. These shielding clouds have a radius of the order of Debye length. If a strong electromagnetic wave is incident on the medium, scattering will take place. The nature of the scattering depends on the wavelength of the wave. If the wavelength is much larger than the Debye length, the scattering is from the shielding cloud and the Doppler broadening



Fig. 3.2-1. Density, temperature, and Debye length of a few sample plasmas.

3.3 Oscillation

corresponds to the ionic thermal speed. If the wavelength is much smaller than the Debye length, the scattering is from individual electrons.

In Fig. 3.2-1 the dependence of Debye length on temperature and number density is shown. Also indicated are regions of typical plasmas.

3.3 Oscillation

We mentioned in Section 2 that the tendency to restore the charge neutrality condition is also the mechanism for producing oscillations. To illustrate this let us consider a slab of plasma. When in equilibrium the plasma is electrically neutral. Suppose we apply an external force which uniformly displaces all electrons by a small distance ξ . The excess charge at two parallel faces sets up an electric field. Since ξ is small, the excess surface charge density is given approximately by $Ne\xi$ where N is the charge density. The uniform electric field so produced is then given by

$$E = Ne\xi/\varepsilon_0 \tag{3.3.1}$$

The situation is illustrated by Fig. 3.3-1. Now let us suppose that this external force is removed. Under the influence of the electric field charged particles must begin to move. Since ions are so massive, we shall ignore entirely their motion. The equation of motion for each electron is given by

Fig. 3.3-1. A slab of plasma. Oscillations are possible if all electrons are displaced by a

small distance relative to ions.

$$m\xi = -eE \tag{3.3.2}$$

Substituting the expression for the electric field given by (3.3.1) into (3.3.2), we obtain

$$\ddot{\xi} = -\omega_p^2 \xi \tag{3.3.3}$$



Equation (3.3.3) is an equation describing a simple harmonic motion. It predicts oscillations at an angular frequency

$$\omega_p = (Ne^2/m\varepsilon_0)^{1/2} \tag{3.3.4}$$

Numerically the plasma frequency squared is proportional to the density through a relation

$$f_p^2 = 80.6N \tag{3.3.5}$$

Plasma frequency is the characteristic frequency of the medium. The inverse of plasma frequency gives the rate at which the electrostatic restoring forces in a plasma tend to eliminate deviations from neutrality. It is through this restoring force that modifies transmission of electromagnetic energy from the free-space condition and gives rise to frequency dispersion phenomenon. Actually the frequency of oscillation depends on the geometric configuration of the plasma as well as on the modes. For example, one of the oscillation modes of a cylindrical plasma column has an angular frequency $\omega_p/\sqrt{2}$ (see problem at the end of the chapter).

We observe that there are several interesting properties in the plasma oscillation just described. The electric field given by (3.3.1) is in the same direction as the displacement of particles and hence the oscillation is longitudinal. For longitudinal oscillations we obtained $\nabla \times \mathbf{E} = 0$ and so it is often referred to as electrostatic oscillations even though the electric field may be time dependent and not strictly static. The fact that the electric field is irrotational also means that there is no associated magnetic field. Consequently the Poynting's vector vanishes for longitudinal waves. In a medium which is not spatially dispersive this also means that the energy velocity is zero. That is, that any perturbation introduced in the region will not propagate away as waves but only oscillate locally. This is the case for the present model since the dispersion relation obtained by taking the Fourier transform of (3.3.3) is

$$\omega^2 = \omega_p^2 \tag{3.3.6}$$

which is not \mathbf{k} dependent. Studies in Chapter 2 showed that the dispersion relation is obtained by setting the dielectric constant to zero. Therefore, the dispersion relation (3.3.6) must be the condition of vanishing dielectric constant for our simple model. Later as we allow for thermal effects the dispersion relation (3.3.6) is modified and becomes \mathbf{k} dependent. The medium then is spatially dispersive. Even though the Poynting's vector is still zero, due to spatial dispersion, there is energy flux associated with the coherent motion of particles and hence perturbations on the plasma can then propagate away as waves.

It is convenient to define v_T by the relation

$$\gamma T = m v_T^2 \tag{3.3.7}$$

Here v_T^2 differs from the usual definition of thermal velocity squared if γ is different from 3. The use of γ in (3.3.7) is purely for later notational convenience. When defined as (3.3.7), we find immediately from (3.2.4) and (3.3.4) the useful relation

$$v_T = \lambda_D \omega_p \sqrt{\gamma} \tag{3.3.8}$$

Equation (3.3.8) states that a particle with a velocity v_T (roughly thermal velocity) travels one Debye length in $1/\omega_p \sqrt{\gamma}$ sec.

3.4 Screening

Consider a plasma with different species of positive and negative ions. The charge on α th species is $Z_{\alpha}e$. We shall use the convention that e is the magnitude of charge on an electron and hence is a positive constant. The sign of charge is carried by Z_{α} so that Z_{α} may take values ± 1 , ± 2 , etc. When the plasma is unperturbed, the particle density of α th species is $N_{\alpha 0}$. The condition for charge neutrality is then given by

$$\sum_{\alpha} N_{\alpha 0} Z_{\alpha} = 0 \tag{3.4.1}$$

Into this plasma we introduce a positive test charge of charge e at the origin. As a result the distribution of charged particles will be changed slightly and a cloud of electrons will form to effectively screen out the Coulomb potential produced by the test charge. The particles are distributed according to the Boltzmann distribution,

$$N_{\alpha} = N_{\alpha 0} e^{-Z_{\alpha} e^{V/T_{\alpha}}}$$
$$= N_{\alpha 0} (1 - Z_{\alpha} e^{V/T_{\alpha}} + \cdots)$$
(3.4.2)

where T_{α} is the temperature, in energy units, of α th kind particles. In (3.4.2) we have also assumed that the exponent is small so that we may expand as shown. We shall ignore higher order terms. This is effectively equivalent to linearization and is certainly convenient mathematically. The physical

3. Waves in Fluid Plasma

interpretation of such an approximation will be taken up later on in this section.

Equation (3.4.2) gives us a relation showing the dependence of density on potential. The dependence of potential on density is through the Poisson's equation

$$\nabla^2 V = (-1/\varepsilon_0) [e\delta(\mathbf{r}) + e \sum_{\alpha} N_{\alpha} Z_{\alpha}]$$
(3.4.3)

Substituting (3.4.2) in (3.4.3) and making use of the neutrality condition (3.4.1), we obtain

$$\nabla^2 V - k_D^2 V = -e\delta(\mathbf{r})/\varepsilon_0 \tag{3.4.4}$$

The Debye wave number appearing in (3.4.4) is defined by

$$k_D = \left[\sum_{\alpha} \frac{Z_{\alpha}^2 e^2 N_{\alpha 0}}{\varepsilon_0 T_{\alpha}}\right]^{1/2}$$
(3.4.5)

which simplifies to the inverse Debye length given by (3.2.4) if all ions are ignored except electrons. Our interest is to solve (3.4.4). We note that (3.4.4) is identical in form to the scalar wave equation satisfied by the Green's function discussed in Section 14 of the last chapter if $-k_D^2$ is replaced by $k_0^2 K$. As before, the transform method can be used to solve it. However, because of spherical symmetry, (3.4.4) can also be solved more simply as it reduces to

$$\frac{1}{r} \frac{d^2 r V}{dr^2} - k_D^2 V = 0$$
(3.4.6)

for $r \neq 0$. Equation (3.4.6) can be integrated easily. The solution that is finite at $r \rightarrow \infty$ is

$$V = (A/r)e^{-k_D r} (3.4.7)$$

The constant A can be determined by substituting (3.4.7) back to the original differential equation and integrating over an infinitesimal sphere centered at the origin. Here V behaves like A/r and the left-hand side of (3.4.4) integrates to

$$\int d\mathbf{x} \, \nabla^2 V = \oint \nabla V \cdot d\mathbf{S} = \oint \nabla (A/r) \cdot d\mathbf{S}$$
$$= -A \int d\Omega = -4\pi A$$

3.4 Screening

where $d\Omega$ is a solid angle element. The integration of the right-hand side of (3.4.4) gives $-e/\varepsilon_0$ which determines the constant A. Finally the potential is found to be

$$V = \frac{e}{4\pi\varepsilon_0} \frac{1}{r} e^{-k_D r}$$
(3.4.8)

We see that for an observer inside the Debye sphere $(r < 1/k_D)$ he sees a potential of a unit positive charge. However, this Coulomb potential of a unit charge is screened for distances larger than the Debye length. This screening comes about from an electron cloud with charge density $-\varepsilon_0 k_D^2 V = -(k_D^2 e/4\pi r)e^{-k_D r}$ as can be seen from (3.4.4).

We would like now to reexamine the linearization procedure in (3.4.2) and interpret the physical meaning attached to such an approximation. In the following we list these physical interpretations which seem to be different on the surface but actually are all equivalent. They all refer to weak interactions.

(i) $eV \ll T$. In order to expand the exponent in (3.4.2), we require that the potential energy be much less than the thermal energy of particles. This means that motions connected with thermal fluctuations must dominate. Such a requirement is certainly resonable, for otherwise electrons would simply fall into ions and we no longer have a plasma left.

(ii) $(\delta N/N_0) \ll 1$. As can be seen from (3.4.2), the fractional change in density $\delta N/N_0$ is just eV/T which is small from (i). Therefore, we require density fluctuations be small when compared with the background value.

(iii) $4\pi\lambda_D^3 N_e \gg 1$. We take $r = \lambda_D$ in (3.4.8) as the value of potential. The requirement (i) now becomes $(e/4\pi\epsilon_0\lambda_D) \ll T$. Making use of the expression (3.2.4) for Debye length, we obtain the desired expression $4\pi\lambda_D^3 N_e \gg 1$. Let us imagine a sphere of radius λ_D as the Debye sphere. Then approximately our requirement is that there be a large number of particles in the Debye sphere. Stated in a different way, the approximation involved requires that the interparticle distance be small as compared with the Debye length.

The potential given by (3.4.8) is singular at r = 0. For r small the potential may be so large that the linearization procedure is no longer valid. Examination of the potential behavior in this small region requires use of nonlinear analysis.
3.5 Electron and Ion Plasma Waves

The use of basic equations to describe the plasma depends on the degree of accuracy and sophistication we wish to have. For example the individual particle model used in Section 3 assumes that the particle-particle interaction takes place only through the macroscopic electric field. Such a model is valid only for extremely tenuous plasma. We wish now to improve our model by adopting a fluid description. For many purposes a fluid model is satisfactory because of its simplicity and ease of getting physical insights. But we should be reminded that the fluid model is not valid for waves with phase velocity near the thermal velocity of particles.

The basic equations in a fluid model can be interpreted as conservation laws. They are particle conservation

$$\partial N_{\alpha}/\partial t + \operatorname{div}(N_{\alpha}\mathbf{v}_{\alpha}) = 0 \tag{3.5.1}$$

and momentum conservation

$$m_{\alpha}N_{\alpha}\partial \mathbf{v}_{\alpha}/\partial t + m_{\alpha}N_{\alpha}\mathbf{v}_{\alpha} \cdot \text{grad } \mathbf{v}_{\alpha} = -\text{grad } p_{\alpha} + Z_{\alpha}eN_{\alpha}\mathbf{E} \qquad (3.5.2)$$

where N_{α} is the number density of α th kind of particles with charge $Z_{\alpha}e$, fluid velocity \mathbf{v}_{α} , and pressure p_{α} . These equations are supplemented by the ideal gas law

$$p_{\alpha} = N_{\alpha}T_{\alpha} \tag{3.5.3}$$

the equation of state

$$D(p_{\alpha}N_{\alpha}^{-\gamma})/Dt = 0 \tag{3.5.4}$$

and the Maxwell equation $\varepsilon_0 \mathbf{\dot{E}} + \mathbf{J}_T = (1/\mu_0) \operatorname{curl} \mathbf{B} = 0$ or

$$\varepsilon_{0}\dot{\mathbf{E}} + \sum_{\alpha} N_{\alpha} Z_{\alpha} e \mathbf{v}_{\alpha} = 0 \qquad (3.5.5)$$

The symbol D/Dt in (3.5.4) stands for convective derivative and is equal to $\partial/\partial t + \mathbf{v}_{\alpha} \cdot \mathbf{g}$ rad. The value of the specific heat ratio γ in (3.5.4) has always been a sore spot in the fluid theory. Physically we would expect the process to be adiabatic if the phase velocity of the wave is much faster than the thermal velocity of the particles and to be isothermal if the opposite is true. This means $\gamma = 3$ for fast one-dimensional plasma waves and $\gamma = 1$ for slow ionic plasma waves. However, when in doubt, the more accurate approach of distribution theory must be used.

As mentioned earlier, for longitudinal waves there is no associated magnetic field and thus one of the Maxwell's equations takes the particularly simple form given by (3.5.5).

Consider a plasma which is spatially uniform, time stationary, and electrically neutral. Each species is allowed to have its own temperature, but each temperature is assumed to be constant throughout the space. This plasma is perturbed so that

$$N_{\alpha} = N_{\alpha}^{(0)} + N_{\alpha}', \qquad p_{\alpha} = p_{\alpha}^{(0)} + p_{\alpha}'$$

$$\mathbf{v} = \mathbf{0} + \mathbf{v}_{\alpha}, \qquad \mathbf{E} = \mathbf{0} + \mathbf{E}$$
(3.5.6)

Our convention is that each of the first terms on the right of (3.5.6) represents the unperturbed quantity and each of the second terms on the right represents the perturbed quantity. According to our assumptions (uniform and stationary when unperturbed) $N_{\alpha}^{(0)}$ and $p_{\alpha}^{(0)}$ are constants. If the perturbations are small, we may then linearize our equations about perturbed quantities. Linearization of the equation of state (3.5.4) gives

$$p_{\alpha}' = \gamma T_{\alpha} N_{\alpha}' \tag{3.5.7}$$

The linearized equations of (3.5.1), (3.5.2), and (3.5.5) are

$$\partial N_{\alpha}'/\partial t + N_{\alpha}^{(0)} \operatorname{div} \mathbf{v}_{\alpha} = 0 \tag{3.5.8}$$

$$m_{\alpha} N_{\alpha}^{(0)} \partial \mathbf{v}_{\alpha} / \partial t = -\gamma T_{\alpha} \operatorname{grad} N_{\alpha}' + Z_{\alpha} e N_{\alpha}^{(0)} \mathbf{E}$$
(3.5.9)

$$\varepsilon_{0}\dot{\mathbf{E}} + \sum_{\alpha} N_{\alpha}^{(0)} Z_{\alpha} e \mathbf{v}_{\alpha} = 0 \qquad (3.5.10)$$

Equations (3.5.8)-(3.5.10) are our basic equations of interest. We shall show in the following that they yield an expression for energy conservation. Solve for **E** in (3.5.9) and dot with (3.5.10). We obtain

$$\varepsilon_{0} \dot{\mathbf{E}} \cdot \mathbf{E} + \sum_{\alpha} \left[N_{\alpha}^{(0)} m_{\alpha} \mathbf{v}_{\alpha} \cdot \dot{\mathbf{v}}_{\alpha} + \gamma T_{\alpha} \mathbf{v}_{\alpha} \cdot \text{grad } N_{\alpha}' \right] = 0 \quad (3.5.11)$$

By vector identity and (3.5.8) the last term of (3.5.11) can be manipulated to give

$$\gamma T_{\alpha} \mathbf{v}_{\alpha} \cdot \operatorname{grad} N_{\alpha}' = \gamma T_{\alpha} [\operatorname{div}(N_{\alpha}' \mathbf{v}_{\alpha}) - N_{\alpha}' \operatorname{div} \mathbf{v}_{\alpha}] = \gamma T_{\alpha} \operatorname{div}(N_{\alpha}' \mathbf{v}_{\alpha}) + (\gamma T_{\alpha} N_{\alpha}' / N_{\alpha}^{(0)}) \partial N_{\alpha}' / \partial t \quad (3.5.12)$$

Substituting (3.5.12) back to (3.5.11), we get

$$\operatorname{div}(\sum_{\alpha} \gamma T_{\alpha} N_{\alpha}' \mathbf{v}_{\alpha}) + \frac{\partial}{\partial t} \left[\sum_{\alpha} \left(\frac{\gamma_{\alpha} T_{\alpha}}{2 N_{\alpha}^{(0)}} N_{\alpha}'^{2} + \frac{1}{2} N_{\alpha}^{(0)} m_{\alpha} v_{\alpha}^{2} \right) + \frac{1}{2} \varepsilon_{0} E^{2} \right] = 0$$
(3.5.13)

Equation (3.5.13) can be interpreted as energy conservation. The divergence term represents the energy flux while the terms in the square bracket represent energy density.

We wish now to derive the longitudinal dielectric constant for our model. Since our equations (3.5.8)-(3.5.10) are linear in perturbed quantities, we may Fourier analyze and work in the transformed domain. This is equivalent to assuming a dependence $e^{j(\omega t-\mathbf{k}\cdot\mathbf{r})}$ if initial conditions can be ignored. Such an assumption reduces (3.5.8) and (3.5.9), respectively, to

$$j\omega N_{\alpha}' - jN_{\alpha}^{(0)}\mathbf{k} \cdot \mathbf{v}_{\alpha} = 0 \qquad (3.5.14)$$

$$j\omega m_{\alpha} N_{\alpha}^{(0)} \mathbf{v}_{\alpha} = j\gamma T_{\alpha} N_{\alpha}' \mathbf{k} + Z_{\alpha} e N_{\alpha}^{(0)} \mathbf{E}$$
(3.5.15)

Inspection of these equations shows that both \mathbf{v}_{α} and \mathbf{E} can be decomposed into a component perpendicular to \mathbf{k} and a component parallel to \mathbf{k} . This means that for our present model the longitudinal waves and transverse waves can propagate independently. Later we shall see that this is no longer true if there is a steady magnetic field. The component of \mathbf{v}_{α} perpendicular to \mathbf{k} can be used to drive an expression for the transverse dielectric constant. Inspection of (3.5.14) also shows that for transverse waves there are no associated density fluctuations. Discussion of the transverse dielectric constant and transverse waves is carried out in a separate section. We, therefore, turn our attention to the longitudinal dielectric constant and longitudinal waves.

For the longitudinal case \mathbf{v}_{α} and \mathbf{E} are parallel to \mathbf{k} . We can eliminate N'_{α} in (3.5.14) and (3.5.15) to obtain

$$\mathbf{v}_{\alpha} = -\frac{1}{1-\gamma T_{\alpha}k^2/m_{\alpha}\omega^2} \frac{jZ_{\alpha}e}{m_{\alpha}\omega} \mathbf{E}$$

The electric polarization density can be obtained by using the above relation.

$$\mathbf{P} = \sum_{\alpha} \left(Z_{\alpha} N_{\alpha}^{(0)} e/j \omega \right) \mathbf{v}_{\alpha}$$
$$= -\sum_{\alpha} \frac{\omega_{p\alpha}}{\omega^2 - \gamma T_{\alpha} k^2 / m_{\alpha}} \varepsilon_0 \mathbf{E}$$
(3.5.16)

where the angular plasma frequency of the α th species, $\omega_{p_{\alpha}}$, is defined by the relation

$$\omega_{p\alpha}^2 = N_{\alpha}^{(0)} Z_{\alpha}^2 e^2 / m_{\alpha} \varepsilon_0 \tag{3.5.17}$$

We note that the plasma frequency squared is proportional to number den-

sity and inversely proportional to the mass. Because of the light mass, the electron plasma frequency is usually much larger than the ion plasma frequency.

The factor multiplying $\varepsilon_0 \mathbf{E}$ on the right-hand side of (3.5.16) is by definition, the longitudinal electric susceptibility of the medium, i.e.,

$$\chi_{\parallel} = -\sum_{\alpha} \frac{\omega_{p\alpha}^2}{\omega^2 - (\gamma T_{\alpha} k^2 / m_{\alpha})}$$
(3.5.18)

The longitudinal dielectric constant is $\varepsilon_0(1 + \chi_{\parallel})$, or

$$\varepsilon_{\parallel}(\mathbf{k},\omega) = \varepsilon_0 \bigg[1 - \sum_{\alpha} \frac{\omega_{p\alpha}^2}{\omega^2 - (\gamma T_{\alpha} k^2 / m_{\alpha})} \bigg]$$
(3.5.19)

The dielectric constant (3.5.19) is both ω -dependent and k-dependent. This means that the medium is frequency dispersive as well as spatially dispersive although the medium is isotropic. We have discussed in Chapter 2 that the dispersion relation for longitudinal waves can be obtained by letting $\varepsilon_{\parallel} = 0$. Setting (3.5.19) to zero, we obtain the dispersion relation

$$1 - \sum_{\alpha} \frac{\omega_{p\alpha}^2}{\omega^2 - (\gamma T_{\alpha} k^2 / m_{\alpha})} = 0 \qquad (3.5.20)$$

We note that in the limit of zero temperature, (3.5.20) reduces effectively to $\omega^2 = \omega_{pe}^2$ which is no longer k-dependent. This simple case has been discussed in Section 3. Taking thermal effects into account makes ω kdependent. Inspection of (3.5.20) shows that there are as many modes as there are species of particles. For simplicity let us concern ourselves with the case of a two-component plasma in which there are electrons and singly charged neutralizing positive ions. The dispersion relation in this case is given by

$$1 - \frac{\omega_{pe}^2}{\omega^2 - (\gamma T_e k^2 / m_e)} - \frac{\omega_{pi}^2}{\omega^2 - (\gamma T_i k^2 / m_i)} = 0 \qquad (3.5.21)$$

The dispersion relation (3.5.21) is a biquadratic equation in ω and can be solved easily. But the resulting expressions are rather complex. The expression can be simplified if we note that there is at least two orders of magnitude difference between ionic mass and electronic mass. For the electronic branch we may assume that m_i is a large parameter. This is equivalent to assuming that ions form a neutralizing background and are inmobile. In this case the last term on the left-hand side of (3.5.21) can be ignored and we obtain the dispersion relation

$$\omega^2 = \omega_{pe}^2 + \gamma T_e k^2 / m_e \tag{3.5.22}$$

As before in (3.5.7) we may define a thermal velocity for electrons through

$$\gamma T_e = m_e v_{Te}^2 \tag{3.5.23}$$

The dispersion relation (3.5.22) can then be written in alternate forms

$$\omega^2 = \omega_{pe}^2 + \nu_{Te}^2 k^2$$
$$= \omega_{pe}^2 (1 + \gamma \lambda_D^2 k^2) \qquad (3.5.24)$$

The relation (3.5.22) or (3.5.24) with $\gamma = 3$ is sometimes known as Vlasov dispersion or Bohm and Gross dispersion. The dispersion relation (3.5.24) can be reexpressed in terms of refractive index, giving

$$n^{2} = \frac{c^{2}}{v_{Te}^{2}} \left(1 - \frac{\omega_{pe}^{2}}{\omega^{2}} \right)$$
(3.5.25)

which shows that electron plasma waves are attenuated at $\omega < \omega_{pe}$. We shall prove later strong damping occurs when $k\lambda_D \approx 1$ which, from (3.5.24), gives the upper propagation limit as $\omega \simeq 2\omega_{pe}$. Therefore, electronic plasma waves can propagate only within a narrow frequency range ω_{pe} to $2\omega_{pe}$.

The group velocity can be obtained by differentiating (3.5.24) with respect to k, yielding

$$v_g = \partial \omega / \partial k = v_{Te}^2 / v_p \tag{3.5.26}$$

where $v_p = \omega/k$ is the phase velocity. From (3.5.25) the product of the group velocity and phase velocity is related to the thermal velocity rather simply,

$$v_p v_g = v_{Te}^2$$
 (3.5.27)

The average coherent kinetic energy flux is given by (see 2.10.22)

$$|\langle \mathbf{S}^{\mathbf{i}}
angle| = -rac{1}{4}\omega E^2 \,\partial arepsilon/\partial k$$

We differentiate (3.5.19) with respect to k and take into account only the electronic contribution. Then we substitute the dispersion relation (3.5.24) into the resulting expression to obtain

$$\left|\left<\mathbf{S}^{1}\right>\right| = \omega E^{2} \varepsilon_{0} k v_{Te}^{2} / 2\omega_{pe}^{2}.$$
(3.5.28)

The internal energy in a wave packet is given by (see 2.10.20)

$$\langle W \rangle = \frac{1}{4} E^2 \partial(\omega \varepsilon) / \partial \omega$$

Again by differentiation and use of the dispersion relation the energy density can be computed to be

$$\langle W \rangle = \frac{1}{2} E^2 \varepsilon_0 \omega^2 / \omega_{pe}^2 \tag{3.5.29}$$

The energy velocity is the ratio of $|\langle S^1 \rangle|$ to $\langle W \rangle$ and from (3.5.28) and (3.5.29) it is

$$|\langle \mathbf{S}^{1} \rangle | \langle W \rangle = k v_{Te}^{2} / \omega = v_{g}$$
(3.5.30)

We have proven the equality of energy velocity to group velocity from very general considerations in Chapter 2. Equation (3.5.30) merely demonstrates that this is so for a specific model. We should note that for our longitudinal waves the transport of energy is carried entirely by streaming motions of charged particles.

The ionic branch of our dispersion relation for the two-component plasma can be obtained by letting $m_e \rightarrow 0$ in (3.5.21). This is equivalent to ignoring the electron inertia in its equation of motion. The dispersion relation (3.5.21) simplifies to

$$\omega^{2} = \frac{\gamma T_{i}k^{2}}{m_{i}} + \frac{\omega_{pi}^{2}}{1 + (\omega_{pe}^{2}m_{e}/\gamma T_{e}k^{2})}$$
(3.5.31)

If the wavelength of interest is much larger than the Debye length, (3.5.31) can be further simplified to

$$\omega^{2} = (\gamma k^{2}/m_{i})(T_{i} + T_{e})$$
(3.5.32)

For this case there is no dispersive effect since both phase velocity and group velocity are equal to $[\gamma(T_i + T_e)/m_i]^{1/2}$. These waves are sometimes known as ion sound waves. Using a distribution function approach, it is found that these waves are heavily damped unless the electron temperature is three times or larger than the ion temperature (Stix, 1964).

3.6 Plasma Density Fluctuations

We have seen that plasmas as fluids can sustain wave motions and these plasma waves have associated density fluctuations. It is then of interest to compute the spectrum of these fluctuations. As discussed earlier, the fluid 3. Waves in Fluid Plasma

approach is not entirely satisfactory, but since it affords simplicity and physical insights, we shall use it here. The derivation makes use of the theorem of equipartition of energy which is strictly valid for a plasma in thermal equilibrium and we shall assume this to be the case in this section.

Consider a large cube of side 2L of a plasma. Due to thermal agitations there are set up standing waves of many modes. For simplicity we assume that these standing waves are repeated outside of the cube so that Fourier series representation can be used. For the *i*th mode the spatial variation is given by the wave number $k_i = i\pi/L$, i = 1, 2, 3, ... The corresponding time variation is given by ω_i which must be computed by the dispersion relation; i.e., $\omega_i = \omega(\mathbf{k}_i)$. The standing wave for density fluctuations of α th species corresponding to the *i*th mode can be written as

$$N_{\alpha}^{\prime(i)} = A_1(\mathbf{k}_i) \cos \mathbf{k}_i \cdot \mathbf{r} \cos \omega_i t + A_2(\mathbf{k}_i) \cos \mathbf{k}_i \cdot \mathbf{r} \sin \omega_i t + A_3(\mathbf{k}_i) \sin \mathbf{k}_i \cdot \mathbf{r} \cos \omega_i t + A_4(\mathbf{k}_i) \sin \mathbf{k}_i \cdot \mathbf{r} \sin \omega_i t \quad (3.6.1)$$

Since the plasma is in thermal equilibrium, it must be statistically homogeneous in space and stationary in time. Further, as N' is the fluctuating part of the density it must have vanishing mean. The necessary and sufficient conditions for (3.6.1) to satisfy all these requirements are

$$\langle A_m(\mathbf{k}_i) \rangle = 0, \quad m = 1, 2, 3, 4, \text{ all } i$$

and

$$\langle A_m(\mathbf{k}_i)A_n(\mathbf{k}_j)\rangle = \delta_{mn} \,\delta_{ij}\langle A^2(\mathbf{k}_i)\rangle, \qquad m, n = 1, 2, 3, 4, \text{ all } i, j \quad (3.6.2)$$

where angular brackets are used to denote statistical average. The total density fluctuation for α th species is obtained by summing up all the modes,

$$N_{\alpha}'(\mathbf{r}, t) = \sum_{i} \left[A_{1}(\mathbf{k}_{i}) \cos \mathbf{k}_{i} \cdot \mathbf{r} \cos \omega_{i} t + A_{2}(\mathbf{k}_{i}) \cos \mathbf{k}_{i} \cdot \mathbf{r} \sin \omega_{i} t + A_{3}(\mathbf{k}_{i}) \sin \mathbf{k}_{i} \cdot \mathbf{r} \cos \omega_{i} t + A_{4}(\mathbf{k}_{i}) \sin \mathbf{k}_{i} \cdot \mathbf{r} \sin \omega_{i} t \right]$$
(3.6.3)

Making use of (3.6.2), the correlation function of the density fluctuation can be computed to be

$$B_{N}(\boldsymbol{\xi}, \tau) = \langle N_{\alpha}'(\mathbf{r} + \boldsymbol{\xi}, t + \tau) N_{\alpha}'(\mathbf{r}, t) \rangle$$

= $\sum_{i} \langle A^{2}(\mathbf{k}_{i}) \rangle \cos \mathbf{k}_{i} \cdot \boldsymbol{\xi} \cos \omega_{i} \tau$ (3.6.4)

The spectral density of density fluctuations is just the Fourier transform of the correlation function,

$$S_N(\mathbf{k},\omega) = \int B_N(\boldsymbol{\xi},\tau) e^{-j(\omega\tau - \mathbf{k}\cdot\boldsymbol{\xi})} d\boldsymbol{\xi} d\tau \qquad (3.6.5)$$

3.6 Plasma Density Fluctuations

Substituting (3.6.4) into (3.6.5), we find the spectral density

$$S_N(\mathbf{k},\omega) = 4\pi^4 \sum_i \langle A^2(\mathbf{k}_i) \rangle \,\delta(\omega \pm \omega_i) \,\delta(\mathbf{k} \pm \mathbf{k}_i) \qquad (3.6.6)$$

In the interest of compactness the product of two δ -functions in (3.6.6) actually represents sum of four terms with all combination of signs. The wave number of the *i*th mode is $k_i = i\pi/L$, $i = 1, 2, \ldots$ The change in wave number between the adjacent mode is $\Delta k_i = \pi/L$. For a large L, Δk_i is small. We may therefore approximate the sum (3.6.6) by an integral as done in the following.

$$S_{N}(\mathbf{k},\omega) = 4\pi^{4}(L/\pi)^{3} \int \langle A^{2}(\mathbf{k}_{i}) \rangle \,\delta(\omega \pm \omega_{i}) \,\delta(\mathbf{k} \pm \mathbf{k}_{i}) \,d\mathbf{k}_{i}$$
$$= 8\pi L^{3} \langle A^{2}(\mathbf{k}) \rangle \left[\delta(\omega + \omega(\mathbf{k})) + \delta(\omega - \omega(\mathbf{k}))\right] \qquad (3.6.7)$$

The average "power" in the density fluctuation $\langle A^2(\mathbf{k}) \rangle$ in (3.6.7) is an unknown quantity and it must be related to the thermodynamic parameters of the plasma system. Since the system is in thermodynamic equilibrium, the theorem of equipartition of energy tells us that the mean energy associated with each mode is T. (Note that we express the temperature T in energy units.) The energy associated with the wave has been derived and is given by the quantity in the square bracket of (3.5.13). However, in order to compute it we must specify our plasma and the region of interest.

For simplicity we again assume that we have a two-component plasma with electrons and singly-charged positive ions both of density N_0 . We shall be concerned first with fluctuations arising from electron plasma waves. For this case, ions just form a neutralizing background and do not provide density fluctuations while electron densities do fluctuate. If the electron density fluctuation has the form (3.6.1) for the *i*th mode, the associated velocity and electric field can be obtained by (3.5.8) and (3.5.10). They are

$$v_{e}^{(i)} = (\omega_i/k_iN_0)[A_1(\mathbf{k}_i)\sin\mathbf{k}_i\cdot\mathbf{r}\sin\omega_i t - A_2(\mathbf{k}_i)\sin\mathbf{k}_i\cdot\mathbf{r}\cos\omega_i t - A_3(\mathbf{k}_i)\cos\mathbf{k}_i\cdot\mathbf{r}\sin\omega_i t + A_4(\mathbf{k}_i)\cos\mathbf{k}_i\cdot\mathbf{r}\cos\omega_i t]$$
(3.6.8)

$$E^{(i)} = (e/\varepsilon_0 k_i) [-A_1(\mathbf{k}_i) \sin \mathbf{k}_i \cdot \mathbf{r} \cos \omega_i t - A_2(\mathbf{k}_i) \sin \mathbf{k}_i \cdot \mathbf{r} \sin \omega_i t + A_3(\mathbf{k}_i) \cos \mathbf{k}_i \cdot \mathbf{r} \cos \omega_i t + A_4(\mathbf{k}_i) \cos \mathbf{k}_i \cdot \mathbf{r} \sin \omega_i t]$$
(3.6.9)

Making use of conditions (3.6.2), the mean square values of velocity and electric field can be obtained.

$$\langle (v_e^{(i)})^2 \rangle = (\omega_i^2 / k_i^2 N_0^2) \langle A^2(\mathbf{k}_i) \rangle$$
(3.6.10)

$$\langle (E^{(i)})^2 \rangle = (e^2 / \varepsilon_0^2 k_i^2) \langle A^2(\mathbf{k}_i) \rangle \tag{3.6.11}$$

The mean energy density associated with density fluctuations of the *i*th mode due to electron plasma waves is

$$U^{(i)} = \frac{\gamma T}{2N_0} \langle (N_e')^2 \rangle + \frac{N_0}{2} m_e \langle (v_e^{(i)})^2 \rangle + \frac{1}{2} \varepsilon_0 \langle (E^{(i)})^2 \rangle \quad (3.6.12)$$

The total energy in a volume $(2L)^3$ corresponding to the *i*th electron plasma waves is then

$$W^{(i)} = \frac{(2L)^3 \langle A^2(\mathbf{k}_i) \rangle}{2N_0} \left(\gamma T + \frac{m_e \omega_i^2}{k_i^2} + \frac{N_0 e^2}{\varepsilon_0 k_i^2} \right)$$
(3.6.13)

According to the theorem of equipartition, (3.6.13) is also equal to T, i.e.,

$$\frac{(2L)^{3}\langle A^{2}(\mathbf{k}_{i})\rangle}{N_{0}} \frac{m\omega_{i}^{2}}{k_{i}^{2}} = T$$
(3.6.14)

where the dispersion relation (3.5.22) has been used. The expression (3.6.14) relates the average "power" of density fluctuation to thermodynamic parameters of the system,

$$\langle A^2(\mathbf{k}) \rangle = \frac{N_0 k^2 T}{8L^3 m_e \omega^2} = \frac{N_0 T}{8L^3} \frac{k^2}{m_e (\omega_{pe}^2 + \gamma T k^2 / m_e)}$$
 (3.6.15)

Substituting (3.6.15) into (3.6.7) we immediately get the spectral density for fluctuation in electron density caused by electron plasma waves as

$$S_N(\mathbf{k},\omega) = \frac{\pi N_0 T k^2}{m_e(\omega_{pe}^2 + \gamma T k^2/m_c)} \left[\delta(\omega - \omega(\mathbf{k})) + \delta(\omega + \omega(\mathbf{k}))\right] \quad (3.6.16)$$

The result (3.6.16) indicates that for a given **k** the spectra are given by two lines at ω satisfying the dispersion relation (3.5.22). More accurate theory derived from kinetic considerations shows that there is some spread in the spectrum although rather sharp.

Similar considerations can also be applied in deriving density fluctuations for the ionic branch. In this case the approximation is $m_e \rightarrow 0$. Since both electrons and ions have velocities and density fluctuations, we must sum up all contributions to the energy density. Again making use of equipartition of energy, the electron density fluctuations corresponding to ionic plasma waves have a spectral density given by

$$S_N(\mathbf{k},\omega) = \frac{\pi N_0 \omega_{p_e}^4}{(\omega_{p_e}^2 + \gamma T k^2 / m_e)(2\omega_{p_e}^2 + \gamma T k^2 / m_e)} \left[\delta(\omega - \omega(\mathbf{k})) + \delta(\omega + \omega(\mathbf{k}))\right]$$
(3.6.17)

where $\omega(\mathbf{k})$ is given by the dispersion relation for ionic plasma waves in (3.5.31). We should note that because of the difference in oscillation frequency the contribution of density fluctuation spectrum from electron plasma waves given by (3.6.16) is near the electron plasma frequency while that from ion plasma waves given by (3.6.17) is near the ion plasma frequency.

3.7 Two-Stream Instability

In the derivation of plasma waves in Section 6 we have assumed that the equilibrium plasma fluids are not in motion. This can be seen in (3.5.6) where the unperturbed velocity is assumed to vanish. Now we wish to consider the possibility that the equilibrium plasma fluids may be in motion. Physically, this case is of interest because strong interaction is expected if the fluid is moving with a velocity nearly that of the wave. As a matter of fact if conditions are right, perturbations may grow in time. When this condition occurs, the medium is said to be unstable. The instability mechanism can be explained by the charge bunching mechanism. Consider the density variation N_e' as shown in Fig. 3.7-1 at some instant. There is an excess of electrons in the positive portion of the wave (points 1 to 3 and 5 to 7 in the figure) and an excess of positive ions in the negative portion of the wave (points 3 to 5, etc.). If electrons are streaming in the z-direction with a velocity nearly equal to the phase velocity of the wave, strong interaction between electrons and the wave through the electric potential is expected. The electrons see a retarding potential from 1 to 2, accelerating potential from 2 to 4, retarding potential from 4 to 6, etc. As a result they reach their minimum velocity at points 2, 6, etc., and maximum velocity at points 4, etc. This has the effect of making electrons spend



Fig. 3.7-1. An example of electron density perturbation showing possibilities for growth.

more of their time in regions of negative charge excess, making them more negative. Similarly the regions of positive charge excess is made more positive. Therefore, the initial perturbation may grow with time, causing instability.

To approach this problem mathematically, we may go back to our basic equations (3.5.1) through (3.5.5) and replace the velocity assumption in (3.5.6) by

$$\mathbf{v}_{\alpha} = \mathbf{v}_{\alpha}^{(0)} + \mathbf{v}_{\alpha}^{\prime} \tag{3.7.1}$$

and repeat the manipulations carried out in Section 5. However, the problem can be more directly approached by noting that for nonrelativistic motions the main effect is that of a Doppler shift in frequency. The contribution of α th species to the longitudinal susceptibility in a frame in which the α th fluid is stationary is given by (3.5.18) as

$$\chi_{\parallel \alpha} = -\omega_{p\alpha}^2 / \omega^2 \tag{3.7.2}$$

where for the moment we have ignored the thermal effect. Suppose in the laboratory frame the α th fluid is moving with velocity $\mathbf{v}_{\mathbf{x}}^{(0)}$ and the appropriate susceptibility is (3.7.2) with the Doppler shift taken into account. Therefore, the longitudinal susceptibility in the laboratory frame is

$$\chi_{\parallel\alpha} = -\omega_{p\alpha}^2 / (\omega - \mathbf{k} \cdot \mathbf{v}_{\alpha}^{(0)})^2 \qquad (3.7.3)$$

The corresponding dielectric constant is then

$$\varepsilon_{\parallel}(\mathbf{k},\omega) = \varepsilon_0 \left(1 - \sum_{\alpha} \frac{\omega_{p_{\alpha}}^2}{(\omega - \mathbf{k} \cdot \mathbf{v}_{\alpha}^{(0)})^2} \right)$$
(3.7.4)

The dispersion relation for longitudinal waves is obtained by setting $\varepsilon_{\parallel}(\mathbf{k}, \omega) = 0$. For simplicity we shall assume that all beams are traveling in the same direction and study the case with **k** parallel to the beam direction. The dispersion relation can be written in the form

$$k^{2} = \sum_{\alpha} \omega_{px}^{2} / (\omega/k - \nu_{\alpha}^{(0)})^{2} \equiv F(\omega/k)$$
(3.7.5)

where we have denoted the right-hand side by $F(\omega/k)$. A typical term in the sum of $F(\omega/k)$ has a behavior sketched in Fig. 3.7-2. The value of the term decays toward zero continuously as $\omega/k \to \pm \infty$ and the value approaches to infinity as the phase velocity approaches to the beam velocity. A plot of $F(\omega/k)$ for three beams is shown in Fig. 3.7-3. The dispersion

104



Fig. 3.7-2. Diagram showing dependence of $\omega_{pa}^2/(\omega/k - v_a^{(0)})^2$ as a function of ω/k .

relation (3.7.5) has roots in ω twice as many as number of beams. Therefore, for the three beam example shown in Fig. 3.7-3 we would expect six roots. These roots are given by $F(\omega/k) = k^2$. We consider the following two cases:



Fig. 3.7-3. Diagram showing dependence of $F(\omega/k)$ as a function of ω/k for a threebeam system.

(i) Stability. For sufficiently large values of k, e.g., k larger than k_2 such as $k = k_1$ in Fig. 3.7-3, the line intersects $F(\omega/k)$ at six points, indicating there are six real roots in ω . As $k \to \infty$, the roots $\omega/k \to \mathbf{v}_{\alpha}^{(0)}$. The space-time behavior of perturbations becomes $e^{j(\omega t - kz)} \to e^{jk(\mathbf{v}_{\alpha}^{(0)}t-z)}$ which shows that perturbations travel with the beam. This is the regime where all beams behave individually and like free beams.

(ii) Instability. As the value of k decreases and reaches k_2 in Fig. 3.7-3, two real roots coalesce. For k smaller than k_2 such as k_3 there are only four real roots, and the two remaining roots must be complex. For $k = k_4$ there are two real roots and four complex roots. Since the dispersion relation (3.7.5) has real coefficients, these complex roots must form complex conjugate pairs. The time behavior $e^{j\omega t} = e^{(j\omega_r t - \omega_t t)}$ shows that one of each complex conjugate pair of roots must give rise to exponential growth and the system is said to be unstable. Note that instability is caused by interactions among beams and hence represents a collective effect.

3. Waves in Fluid Plasma

In the following let us consider the example of one streaming beam through a stationary beam. This can be the case of a streaming electron beam through stationary ions or of one streaming electron beam through another electron beam at rest with ions considered as neutralizing background. Anyway, the dispersion relation for this example takes the form

$$1 = \frac{\omega_{p_1}^2}{\omega^2} + \frac{\omega_{p_2}^2}{(\omega - k\nu^{(0)})^2} \equiv G(\omega, k)$$
(3.7.6)

The function $G(\omega, k)$ defined by (3.7.6) has a behavior represented by the curve of Fig. 3.7-4. As seen the curve reaches a minimum $G(\omega_0, k)$ when $\omega = \omega_0$. According to the foregoing discussion, the plasma system under consideration is stable if $G(\omega_0, k) < 1$, marginally stable if $G(\omega_0, k) = 1$, and unstable if $G(\omega_0, k) > 1$. Therefore, it is crucial to find the value of $G(\omega_0, k)$. To find the threshold of instability, we set $(\partial G/\partial \omega)_{\omega_0} = 0$ and obtain

$$\frac{\omega_{p_1}^2}{\omega_0^{3}} + \frac{\omega_{p_2}^2}{(\omega_0 - k\nu^{(0)})^3} = 0$$



Fig. 3.7-4. Dependence of $G(\omega, k)$ on ω . See Eq. (3.7.6).

The real root for ω_0 can be found easily by solving the above equation,

$$\omega_{0} = \frac{\omega_{p_{1}}^{2/3} k v^{(0)}}{\omega_{p_{1}}^{2/3} + \omega_{p_{2}}^{2/3}}$$
(3.7.7)

This is the value of ω at which $G(\omega, k)$ is a minimum. Substituting (3.7.7) into (3.7.6), the resulting expression can be simplified to

$$G(\omega_0, k) = \frac{(\omega_{p_1}^{2/3} + \omega_{p_2}^{2/3})^3}{(k\nu^{(0)})^2}$$
(3.7.8)

The condition for marginal instability is that (3.7.8) has a value 1 or

$$kv^{(0)} = (\omega_{p1}^{2/3} + \omega_{p2}^{2/3})^{3/2}$$
(3.7.9)

106

which divides the region of instability from stability. In $kv^{(0)}$ plane (3.7.9) represents a hyperbolic curve shown in Fig. 3.7-5. The condition for stability $G(\omega_0, k) < 1$ corresponds to the region above the hyperbolic curve, and the condition for instability $G(\omega_0, k) > 1$ corresponds to the region below the hyperbolic curve. These regions are clearly marked in Fig. 3.7-5. We

Fig. 3.7-5. Region of stability and instability in $k - \nu^{(0)}$ plane.

remark that these results are obtained by ignoring the thermal effects and certainly need corrections when the beam velocity $v^{(0)}$ is close to the thermal velocity. Therefore, let us now include the thermal effects in our consideration.

The longitudinal dielectric constant in a stationary plasma with thermal effects taken into account by using fluid equations has been derived in (3.5.19). When these plasma fluids are streaming with beam velocities $\mathbf{v}_{\alpha}^{(0)}$, the resulting dielectric constant is just (3.5.19) with a Doppler shift correction, i.e.,

$$\varepsilon_{\parallel}(k,\omega) = \varepsilon_0 \bigg[1 - \sum_{\alpha} \frac{\omega_{p\alpha}^2}{(\omega - \mathbf{k} \cdot \mathbf{v}_{\alpha}^{(0)})^2 - v_{T\alpha}^2 k^2} \bigg]$$
(3.7.10)

By setting (3.7.10) to zero, we obtain the dispersion relation

$$k^{2} = \sum_{\alpha} \frac{\omega_{p\alpha}^{2}}{(\omega/k - \nu_{\alpha}^{(0)})^{2} - \nu_{T\alpha}^{2}} \equiv F(\omega/k)$$
(3.7.11)

in place of (3.7.5). In (3.7.11), for simplicity, we again let all beams travel in the same direction as **k**. Comparing (3.7.11) with (3.7.5) we see that the number of roots are unchanged but the thermal effects are important when the phase velocity ω/k is near the beam velocity $v_{\alpha}^{(0)}$. Instead of Fig. 3.7-2 the contribution of each term in $F(\omega/k)$ has a behavior sketched in Fig. 3.7-6.





As an example let us return to the example considered earlier, remembering that in this example one beam is stationary and one beam is streaming with velocity $v^{(0)}$. Instead of (3.7.6) the dispersion relation is now,

$$k^{2} = \frac{\omega_{p_{1}}^{2}}{(\omega/k)^{2} - v_{T_{1}}^{2}} + \frac{\omega_{p_{2}}^{2}}{(\omega/k - v^{(0)})^{2} - v_{T_{2}}^{2}} \equiv F(\omega/k) \quad (3.7.12)$$

Let us consider the effect of reducing the beam velocity $v^{(0)}$ from a very large value to a very small value through the thermal velocity v_T . When $v^{(0)} \gg v_{T\alpha}$, $F(\omega/k)$ has the behavior shown in Fig. 3.7-7. As far as the consideration



Fig. 3.7-7. $F(\omega/k)$ as a function of ω/k for a large beam velocity.

for instability is concerned, the thermal correction for this case is not large. For a given value of $k^2 = k_1^2$, there are four intersections with $F(\omega/k)$, indicating four real roots. The plasma system is stable for this value of k_1 . As the value of v_0 is reduced, the infinites are brought closer together as shown in Fig. 3.7-8. The minimum of the curve C marked in Fig. 3.7-7 is raised until it is above k_1^2 as shown in Fig. 3.7-8 in which case there are two real roots and a pair of complex conjugate roots. The plasma is then



Fig. 3.7-8. $F(\omega/k)$ as a function of ω/k for a moderate beam velocity.

unstable when perturbed by perturbations of wave number k_1 . As ν_0 is reduced further, the situation becomes that given by Fig. 3.7-9. There are again four real roots and the plasma is stable. Therefore, the thermal motion tends to stabilize the plasma for beam velocity near the thermal velocity.



Fig. 3.7-9. $F(\omega/k)$ as a function of ω/k for a small beam velocity.

3.8 Interaction of Charged Particles with Longitudinal Waves

We have seen in Section 5 that there is an electric field associated with these longitudinal plasma waves. It is therefore possible for plasma particles to interact with plasma waves. The interaction may be strong if there are appreciable particles with thermal velocities approximating the phase velocity of the wave. This is particularly evident in Section 7 where two stream instability was discussed. In this section we shall discuss this interaction problem and derive the Landau damping formula by using the onedimensional equation of motion. In the absence of collisions and electric field a thermal electron will move according to

$$z = z_0 + v_0 t \tag{3.8.1}$$

where z_0 is its initial position and v_0 its initial velocity. Let us imagine that we have an ensemble of such noninteracting electrons in a neutralizing ion background. Consequently both z_0 and v_0 have certain probability distributions. In a uniform plasma the initial position z_0 must then be uniformly distributed. For convenience we shall assume a velocity distribution $f(v_0)$.

Now let us assume that there exists an electric field

$$E(z, t) = E_0 \cos(\omega t - kz) \tag{3.8.2}$$

excited by a plasma wave and turned on for $t \ge 0$. The presence of the electric field (3.8.2) will modify the motion of electrons. The resulting equation of motion under the combined influence of thermal motion and the electric field is nonlinear and is difficult to solve. We approach it by a perturbation technique by treating the amplitude of the electric field E_0 as a small expansion parameter. The position of an electron is then given by

$$z = z_0 + v_0 t + z_1 + z_2 + \cdots$$
 (3.8.3)

where $v_0 t$ is the displacement induced by thermal motion and $z_1 + z_2 + \cdots$ the displacement induced by the electric field (3.8.2) with z_1 proportional to E_0, z_2 proportional to E_0^2 , etc. Our initial conditions $z = z_0$ at t = 0and $\dot{z} = v_0$ at t = 0 become

$$z_1 = z_2 = \cdots = 0,$$
 at $t = 0$
 $\dot{z}_1 = \dot{z}_2 = \cdots = 0,$ at $t = 0$
(3.8.4)

The equation of motion for an electron is

$$m\ddot{z} = -eE_0\cos(\omega t - kz) \tag{3.8.5}$$

Substitute (3.8.3) into (3.8.5) and expand to the second order in E_0 . The result is

$$m(\ddot{z}_1 + \ddot{z}_2) = -eE_0 \cos[(\omega - k\nu_0)t - kz_0] - eE_0kz_1 \sin[(\omega - k\nu_0)t - kz_0] + O(E_0^3)$$
(3.8.6)

Equation (3.8.6) can be solved by equating terms of equal power in E_0 .

3.8 Interaction of Particles with Longitudinal Waves

For terms proportional to E_0 , we obtain from (3.8.6)

$$m\ddot{z}_{1} = -eE_{0}\cos[(\omega - k\nu_{0})t - kz_{0}]$$
(3.8.7)

Integrating and applying the initial condition (3.8.4)

$$\dot{z}_1 = -\frac{eE_0}{m(\omega - k\nu_0)} \left\{ \sin \left[(\omega - k\nu_0)t - kz_0 \right] + \sin kz_0 \right\} \quad (3.8.8)$$

Integrating again

$$z_{1} = -\frac{eE_{0}}{m(\omega - kv_{0})} \left\{ \frac{-\cos\left[(\omega - kv_{0})t - kz_{0}\right] + \cos kz_{0}}{\omega - kv_{0}} + t\sin kz_{0} \right\}$$
(3.8.9)

For terms proportional to E_0^2 in (3.8.6), we obtain

$$\ddot{z}_2 = -(eE_0kz_1/m)\sin[(\omega - kv_0)t - kz_0]$$
(3.8.10)

where z_1 has been found and given by (3.8.9). The time rate of increase of kinetic energy for each electron is given by

$$\frac{d}{dt} (\text{K.E.}) = \frac{d}{dt} \left[\frac{1}{2} m(\dot{z})^2 \right]$$

= $m \dot{z} \ddot{z} = m(v_0 + \dot{z}_1 + \cdots)(\ddot{z}_1 + \ddot{z}_2 + \cdots)$
= $m v_0 \ddot{z}_1 + m \dot{z}_1 \ddot{z}_1 + m v_0 \ddot{z}_2 + O(E_0^3)$ (3.8.11)

We substitute (3.8.8), (3.8.9), and (3.8.10) into (3.8.11) and take the ensemble average over the initial position. Remember that z_0 is uniformly distributed and hence terms periodic in z_0 vanishes on averaging.

$$\left\langle \frac{d}{dt} (\text{K.E.}) \right\rangle_{z_0} = \frac{e^2 E_0^2}{2m(\omega - k\nu_0)} \sin(\omega - k\nu_0)t \\ + \frac{e^2 E_0^2 k \nu_0}{2m(\omega - k\nu_0)} \left[\frac{\sin(\omega - k\nu_0)t}{\omega - k\nu_0} - t \cos(\omega - k\nu_0)t \right] \\ = \frac{e^2 E_0^2}{2m} \left[\frac{\omega \sin(\omega - k\nu_0)t}{(\omega - k\nu_0)^2} - \frac{k\nu_0 t}{\omega - k\nu_0} \cos(\omega - k\nu_0)t \right]$$
(3.8.12)

We wish now to average (3.8.12) over the initial velocity distribution v_0 . We note that by means of l'Hopital's rule that the right-hand side of (3.8.12) is actually well behaved at $\omega - kv_0 = 0$. Hence in integrating over the velocity space we may take the principal part of the sum, which is just the sum of the principal parts. Let the velocity distribution be so normalized that

$$\int_{-\infty}^{\infty} f(v_0) \, dv_0 = 1 \tag{3.8.13}$$

We proceed to find the velocity average of terms in the square bracket of (3.8.12). For large t, $\sin(\omega - kv_0)t$ and $\cos(\omega - kv_0)t$ are rapidly varying functions of v_0 or more conveniently of $\omega - kv_0$. The integration of any slowly varying function multiplied by these trigonometric functions has contribution coming mainly from the point $\omega/k = v_0$. Therefore, we expand $f(v_0)$.

$$f(v_0) = f(\omega/k) - [(\omega - kv_0)/k]f'(\omega/k) + \cdots$$
 (3.8.14)

The average of the first term in square bracket of (3.8.12) is

$$P\int_{-\infty}^{\infty} \frac{\omega \sin(\omega - kv_0)t}{(\omega - kv_0)^2} f(v_0) dv_0 = \omega f\left(\frac{\omega}{k}\right) P \int_{-\infty}^{\infty} \frac{\sin(\omega - kv_0)t}{(\omega - kv_0)^2} dv_0$$
$$- \frac{\omega}{k} f'\left(\frac{\omega}{k}\right) P \int_{-\infty}^{\infty} \frac{\sin(\omega - kv_0)t}{\omega - kv_0} dv_0 + \cdots$$
$$\simeq - \frac{\omega\pi}{k |k|} f'\left(\frac{\omega}{k}\right)$$
(3.8.15)

where P in front of the integral sign stands for principal part and $f'(\omega/k)$ the derivative of the distribution evaluated at $v_0 = \omega/k$. The average of the second term in the square brackets of (3.8.12) can be approached in the same manner. To the same accuracy as (2.8.15) it vanishes on averaging. Hence the rate of increase of kinetic energy per particle becomes

$$\left\langle \frac{d}{dt} (\text{K.E.}) \right\rangle_{z_0, v_0} = -\frac{\omega \pi e^2 E_0^2}{2mk \mid k \mid} f'\left(\frac{\omega}{k}\right)$$
 (3.8.16)

Note that even though there is no collision, electrons gain energy from the wave if the slope of the velocity distribution is negative at the phase velocity of the wave and that electrons lose energy to the wave if the slope is positive at the phase velocity of the wave. The dissipation of energy per unit volume from the wave to the particles is just (3.8.16) multiplied by the electron density N, i.e.,

$$\delta Q = -\frac{N\omega\pi e^2 E_0^2}{2mk \mid k \mid} f'\left(\frac{\omega}{k}\right)$$
(3.8.17)

The energy density in electron plasma waves has been found in (3.5.29) as

$$U = \frac{1}{2} \varepsilon_0 E_0^2 \omega^2 / \omega_{pe}^2$$
 (3.8.18)

The change in frequency in the dispersion relation due to dissipation (3.8.17) can be computed through $\delta \omega = j \, \delta Q/2U$ as shown in Chapter 2. Carrying out the ratio, we obtain

$$\delta\omega = -j \frac{\pi \omega_{p_e}^4}{2k \mid k \mid \omega} f'\left(\frac{\omega}{k}\right)$$
(3.8.19)

The above formula indicates that by taking particle-wave interaction into account the modification in the dispersion relation is the appearance of the imaginary part of ω . If the imaginary part of ω is positive, the wave is damped exponentially with time. This occurs when $f'(\omega/k)$ is negative or when there are more particles traveling at a velocity slightly slower than the phase velocity of the wave than faster. On the other hand when there are more particles traveling at a velocity faster than the phase velocity of the wave gains energy from the particles and its amplitude grows exponentially with time. The formula (3.8.19) is often referred to as the formula for Landau damping although it differs slightly from the original formula.

3.9 Excitation of Fields by a Test Particle

It was shown in Section 5 that there exist longitudinal waves in a plasma. The origin of these waves was not discussed although it was shown that such waves exist in association with thermal fluctuations in Section 6. In this section we shall consider the problem of exciting such waves by a test particle of charge Ze moving through the plasma at a constant velocity v_0 . We have seen in Section 4 that a stationary test particle is surrounded by a polarization cloud which acts to screen its interaction with other particles outside of the Debye sphere. If the test particle is in motion, the shape of the polarization cloud is expected to be modified. We note from Section 3.5 that plasma waves have phase velocity of the order of thermal velocity. The test particle can therefore have a velocity larger than the phase velocity of the wave. When this is the case Čerenkov radiation is possible.

The response of a plasma to a test charge is completely contained in the dielectric constant. However, in the derivation of the dielectric constant, certain approximations may have been made. For example, the formula

(3.5.19) was derived by using fluid theory and hence is valid only for wavelengths much larger than Debye length. This should be kept in mind as (3.5.19) is used in the following.

At present we are interested in the excitation of only the longitudinal part of the fields. These fields are described by the electrostatic equations

$$\nabla \cdot \mathbf{D}(\mathbf{r}, t) = \varrho_t(\mathbf{r}, t) \tag{3.9.1}$$

and

$$\varepsilon_0 \nabla \cdot \mathbf{E}(\mathbf{r}, t) = \varrho_t(\mathbf{r}, t) + \varrho_p(\mathbf{r}, t)$$
(3.9.2)

Here ϱ_p is the polarization charge induced by the test charge ϱ_t . For the longitudinal field the electric field is derivable from the gradient of a potential function

$$\mathbf{E}(\mathbf{r},t) = -\nabla V(\mathbf{r},t) \tag{3.9.3}$$

Substitute (3.9.3) into (3.9.1) and (3.9.2), respectively, and take the Fourier transform with respect to **r** and *t* on the resulting equations, giving

$$k^{2}\varepsilon_{0}K_{\parallel}(\mathbf{k},\omega)V(\mathbf{k},\omega) = \varrho_{t}(\mathbf{k},\omega) \qquad (3.9.4)$$

$$k^{2}\varepsilon_{0}V(\mathbf{k},\omega) = \varrho_{t}(\mathbf{k},\omega) + \varrho_{p}(\mathbf{k},\omega) \qquad (3.9.5)$$

Take the ratio of these two equations and solve for the polarization charge,

$$\varrho_p(\mathbf{k},\omega) = \varrho_l(\mathbf{k},\omega) \left[\frac{1}{K_{\parallel}(\mathbf{k},\omega)} - 1 \right]$$
(3.9.6)

For a test particle traveling at a constant velocity \mathbf{v}_0 and with charge Ze,

$$\varrho_t(\mathbf{r}, t) = Ze \ \delta(\mathbf{r} - \mathbf{v}_0 t) \tag{3.9.7}$$

where we have assumed that the particle is at $\mathbf{r} = 0$ when t = 0. The corresponding test charge density in the transformed domain is given by

$$\varrho_t(\mathbf{k},\omega) = 2\pi Z e \ \delta(\omega - \mathbf{k} \cdot \mathbf{v}_0) \tag{3.9.8}$$

Substituting (3.9.8) into (3.9.6), we obtain

$$\varrho_p(\mathbf{k},\omega) = 2\pi Z e \ \delta(\omega - \mathbf{k} \cdot \mathbf{v}_0) \left[\frac{1}{-K_{\parallel}(\mathbf{k},\omega)} - 1 \right]$$
(3.9.9)

The induced polarization density due to passage of a test charge moving at a constant velocity \mathbf{v}_0 is obtained by an inverse transformation of (3.9.9). The integration in ω is helped by the presence of δ -function in (3.9.9),

giving

$$\varrho_p(\mathbf{r},t) = \frac{Ze}{(2\pi)^3} \int \left[\frac{1}{K_{\parallel}(\mathbf{k},\mathbf{k}\cdot\mathbf{v}_0)} - 1\right] e^{j\mathbf{k}\cdot(\mathbf{v}_0t-\mathbf{r})} d^3k \quad (3.9.10)$$

Note that the relative dielectric constant $K_{\rm B}$ in (3.9.10) is evaluated at $\omega = \mathbf{k} \cdot \mathbf{v}_0$. The use of (3.9.4) and (3.9.5) can yield other physical quantities of interest, such as the potential and electric field. We give them in the following for later reference.

$$V(\mathbf{r},t) = \frac{Ze}{(2\pi)^3 \varepsilon_0} \int \frac{1}{k^2 K_{\parallel}(\mathbf{k},\mathbf{k}\cdot\mathbf{v}_0)} e^{j\mathbf{k}\cdot(\mathbf{v}_0t-\mathbf{r})} d^3k \qquad (3.9.11)$$

$$\mathbf{E}(\mathbf{r},t) = \frac{jZe}{(2\pi)^3\varepsilon_0} \int \frac{\mathbf{k}}{k^2 K_{\parallel}(\mathbf{k},\mathbf{k}\cdot\mathbf{v}_0)} e^{j\mathbf{k}\cdot(\mathbf{v}_0t-\mathbf{r})} d^3k \qquad (3.9.12)$$

We note that the electric field in (3.9.12) may be obtained directly by using (2.13.6) or the dyadic Green's function (2.14.30). Equations (3.9.10), (3.9.11), and (3.9.12) show that the response of the plasma depends on the dielectric constant of the medium. The expression given by (3.5.19) is valid only for small k (long wavelength). For large k, comparable to Debye wave number, plasma waves in a thermal equilibrium plasma are damped through particle-wave interactions as discussed in Section 8. Therefore, we may introduce an arbitrary cutoff at $k = k_D$ for the upper limit if the integral is to diverge.

The approach used in arriving at (3.9.10), (3.9.11), and (3.9.12) make use of the dielectric constant. It is equally valid if we start with the fluid equations and Maxwell's equations discussed in Section 5 with the appropriate inclusion of sources to take into account the test charge. After the equation is linearized about the perturbed quantities, the induced density perturbation is found to satisfy the inhomogeneous Klein-Gordon equation. This equation can be solved by the transform method. The integrals involved in the inversion are exactly identical to those considered here [see (3.9.14) below] and the results are also identical.

In the following we shall treat two cases separately, depending on the magnitude of the test particle velocity v_0 smaller or larger than the thermal velocity v_T . For convenience the test charge is assumed to be moving along the z-axis for both cases. The relative dielectric constant is given by

$$K_{\parallel}(\mathbf{k},\omega) = 1 - \omega_p^2 / (\omega^2 - k^2 \nu_T^2)$$
(3.9.13)

where contributions from ions have been ignored.

Let us first treat the slow test particle case in which $v_0 < v_T$. Substituting

(3.9.13) into (3.9.10), the polarization charge density is found to be

$$\varrho_{p}(\mathbf{r},t) = -\frac{Zek_{D}^{2}}{(2\pi)^{3}} \int \frac{1}{k_{x}^{2} + k_{y}^{2} + k_{z}^{2}(1-\beta^{2}) + k_{D}^{2}} e^{j\mathbf{k}\cdot(\mathbf{v}_{0}t-\mathbf{r})} d^{3}k$$
(3.9.14)

where $\beta = v_0/v_T$ and is less than 1 for the present case and $k_D^2 = \omega_p^2/v_T^2$. Let us make the following simultaneous substitutions in (3.9.14),

$$k_{x}' = k_{x}, \qquad k_{y}' = k_{y}, \qquad k_{z}' = (1 - \beta^{2})^{1/2} k_{z}$$

$$x' = x, \qquad y' = y, \qquad z' = (z - v_{0}t)/(1 - \beta^{2})^{1/2}$$
(3.9.15)

We obtain

$$\varrho_p(\mathbf{r},t) = -\frac{Zek_D^2}{(2\pi)^3(1-\beta^2)^{1/2}} \int \frac{1}{k_D^2 + k'^2} e^{-j\mathbf{k}'\cdot\mathbf{r}'} d^3k' \quad (3.9.16)$$

We note that the induced polarization charge is spherically symmetric in the new coordinates. We note also that (3.9.16) is identical in form to the transform of (3.4.4) for which a solution has been given in (3.4.8). Therefore, we can immediately write down the expression. Since Fourier inversion can be carried out rather easily, we shall proceed to evaluate the integral for the purpose of illustrating the technique. (Also, see Chapter 2, Section 14.) Rewrite (3.9.16) by using spherical polar coordinates in k-space with polar axis aligned with \mathbf{r}' . The integrations with respect to azimuthal angle ϕ with limits 0 and 2π and with respect to polar angle θ with limits 0 and π can be carried out easily.

$$\varrho_{p}(\mathbf{r},t) = -\frac{Zek_{D}^{2}}{(2\pi)^{3}(1-\beta^{2})^{1/2}} \int \frac{e^{-jk'r'\cos\theta}}{k'^{2}+k_{D}^{2}} k'^{2}\sin\theta \, dk' \, d\theta \, d\phi$$
$$= -\frac{Zek_{D}^{2}}{(2\pi)^{2}(1-\beta^{2})^{1/2}jr'} \int_{0}^{\infty} \frac{k'}{k'^{2}+k_{D}^{2}} \left(e^{jk'r'}-e^{-jk'r'}\right) dk' \quad (3.9.17)$$

The integrand in (3.9.17) is even in k' and therefore can be written as

$$\varrho_p(\mathbf{r},t) = -\frac{Zek_D^2}{(2\pi)^2 jr'(1-\beta^2)^{1/2}} \int_{-\infty}^{\infty} \frac{k'}{k'^2+k_D^2} e^{jk'r'} dk'$$

which can be integrated by contour integration. For ρ_p to remain bounded as $r' \to \infty$, we close the contour in the upper half-plane and pick up the contribution from the simple pole at $k' = jk_D$, yielding

$$\varrho_p(\mathbf{r},t) = -\frac{Zek_D^2}{4\pi r'(1-\beta^2)^{1/2}} e^{-k_D r'}$$
(3.9.18)

We note that in the transformed coordinates (3.9.18) is identical to the Debye screen cloud discussed in Section 4. When we transform it back to the original coordinates, we obtain the result

$$\varrho_{p}(\mathbf{r},t) = -\frac{Zek_{D}^{2}}{4\pi(1-\beta^{2})^{1/2}} \frac{\exp\{-k_{D}[x^{2}+y^{2}+(z-\nu_{0}t)^{2}/(1-\beta^{2})]^{1/2}\}}{[x^{2}+y^{2}+(z-\nu_{0}t)^{2}/(1-\beta^{2})]^{1/2}}$$
(3.9.19)

The induced polarization charge (3.9.19) is in the form of a screening cloud that comoves with the test particle. From the point of view of fluid theory the screening comes about because of Coulomb forces which displace the fluid until the electrostatic force is balanced by the pressure force. The screening cloud has an oblate spheroidal form centered at the test charge and being compressed in the direction of motion by the ratio 1 to $(1 - \beta^2)^{1/2}$. For the case of a stationary test charge, (3.9.19) should reduce to the results discussed in Section 4. These results are comparable if we set $\gamma = 1$.

As we have seen, for $v_0 < v_T$ the perturbed fields are evanescent. However, when the test particle is suprathermal, viz., $v_0 > v_T$, radiation fields are expected due to Čerenkov radiation. This is the case we wish to consider at present. The starting point is still (3.9.14) which is now written as

$$\varrho_p(\mathbf{r}, t) = -\frac{Zek_D^2}{(2\pi)^3} \int \frac{1}{D(\mathbf{k}, k_z v_0)} e^{-j\mathbf{k}\cdot\mathbf{r}'} d^3k \qquad (3.9.20)$$

where

$$D(\mathbf{k},\omega) = k^2 + k_D^2 - \omega^2 / v_T^2$$
(3.9.21)

$$\mathbf{r}' = \mathbf{r} - \mathbf{v}_0 t \tag{3.9.22}$$

The inversion in (3.9.20) can be carried out exactly, but it is long and tedious. In the following this integral will be evaluated asymptotically by using the method given in Appendix B. It so happens in this case that the asymptotic expression is exact. The asymptotic evaluation of (3.9.20) produces the following

$$\varrho_p = -\frac{Zek_D^2}{2\pi r'} \sum \frac{C}{|\nabla_k D(\mathbf{k}, k_2 \nu_0)| \sqrt{|K|}} e^{-j\mathbf{k}\cdot\mathbf{r}'} \qquad (3.9.23)$$

where $\mathbf{r}' = |\mathbf{r}'|$ and K is the Gaussian curvature of the dispersion surface. The summation is carried out over all values of **k** which has a corresponding group velocity along \mathbf{r}' .

We note from (3.9.21) that $D(\mathbf{k}, \omega) = 0$ is the dispersion relation (3.5.25) for electron plasma waves. The relation

$$D(\mathbf{k}, k_z v_0) = k_D^2 + k_x^2 + k_y^2 - k_z^2 (\beta^2 - 1) = 0 \qquad (3.9.24)$$

is just the Doppler shifted dispersion relation as seen by an observer moving with the particle velocity v_0 along z-axis. The effect of motion is to distort the otherwise spherical dispersion surface into the surface of hyperbolid of two sheets when $\beta > 1$. The surface is axially symmetric with respect to k_z -axis and its intersection with $k_y = 0$ plane is shown in Fig. 3.9-1. Since



Fig. 3.9-1. The dispersion surface in a moving plasma as given by (3.9.24).

the group velocity is normal to the dispersion surface, it is now pointing in a direction different from that of \mathbf{k} . The normal to the dispersion surface is in the same direction as

$$\nabla_k D(\mathbf{k}, k_z v_0) = 2[\hat{x}k_x + \hat{y}k_y - \hat{z}k_z(\beta^2 - 1)]$$
(3.9.25)

The group velocity is either parallel or antiparallel to the vector given by (3.9.25). A decision on the choice can be made by noting that $\mathbf{v}_g = \nabla_k \omega$ and therefore, \mathbf{v}_g must be in the direction of increasing ω . In applying this method to the problem at hand, we need to examine the dispersion surface with a small increase in ω , that is the following surface.

$$D(\mathbf{k}, k_{z}v_{0} + \delta) = k^{2} + k_{D}^{2} - (k_{z}v_{0} + \delta)^{2}/v_{T}^{2} = 0 \qquad (3.9.26)$$

where δ is a small positive parameter. The dispersion surface (3.9.26) is approximately the same as that given by (3.9.24) except that both surface branches are shifted downward slightly. Consequently, the group velocity for both branches of the surface has a negative z-component.

According to the asymptotic theory, the observer at \mathbf{r}' sees those waves with those wave numbers whose corresponding group velocities all point toward the observer at \mathbf{r}' . Suppose now a certain observer is at \mathbf{r}' , which is parallel to \mathbf{v}_{g1} with a corresponding wave vector \mathbf{k}_1 (See Fig. 3.9-1). Because of the symmetry of the dispersion surface and the foregoing discussion in regard to the correct choice of normal to the dispersion surface as the direction of \mathbf{v}_g , the observer also sees a wave with wave vector $\mathbf{k}_2 = -\mathbf{k}_1$ since its corresponding group velocity \mathbf{v}_{g2} is parallel to \mathbf{v}_{g1} as shown in Fig. 3.9-1. Therefore, for the dispersion surface the radiaton field at any point \mathbf{r}' has contributions coming from either two points on the surface or none at all. An examination of Fig. 3.9-1 shows there is field only when the observer is inside a cone which is identified as the Čerenkov cone later on in this section.

Let the observer be at $\mathbf{r}' = (x', y, z')$. Based on these discussions and (3.9.25), we must have

$$x': y': z' = k_x: k_y: -k_z(\beta^2 - 1)$$
(3.9.27)

Making use of the Doppler shifted dispersion relation (3.9.24), Eq. (3.9.27) can be solved to produce

$$\mathbf{k} = \pm \frac{k_D}{(z'^2/(\beta^2 - 1) - x'^2 - y'^2)^{1/2}} \left[\hat{x}x' + \hat{y}y' - \hat{z}z'/(\beta^2 - 1) \right] \quad (3.9.28)$$

where the upper sign applies to the upper branch of the dispersion surface of Fig. 3.9-1 and the lower sign to the lower branch. The Gaussian curvature is found to be

$$K = k_D^2 (\beta^2 - 1) / [k_x^2 + k_y^2 + k_z^2 (\beta^2 - 1)^2]^2$$
(3.9.29)

for both branches, where the Doppler shifted dispersion relation has been used. We note that the Gaussian curvature is positive, the constant C of (3.9.23) is 1 if the surface is convex to the direction of $\nabla_k D(\mathbf{k}, k_z v_0)$ and is -1 if convex to the opposite direction. For our case, C = 1 for both branches. Putting (3.9.25), (3.9.28), and (3.9.29) in (3.9.23) and using (3.9.28) to reexpress **k** in terms of **r'**, we obtain

$$\varrho_{p}(\mathbf{r},t) = -\frac{Zek_{D}^{2}}{2\pi r'} \cdot \frac{1}{|\nabla_{k} D| \sqrt{|K|}} (e^{-j\mathbf{k}\cdot\mathbf{r}'} + e^{j\mathbf{k}\cdot\mathbf{r}'})$$
$$= -\frac{Zek_{D}^{2}}{2\pi(\beta^{2}-1)^{1/2}} \frac{\cos k_{D}[z'^{2}/(\beta^{2}-1) - x'^{2} - y'^{2}]^{1/2}}{[z'^{2}/(\beta^{2}-1) - x'^{2} - y'^{2}]^{1/2}}$$

When transformed back to the original coordinates, the induced polarization

density is therefore given by

$$\varrho_{p}(\mathbf{r},t) = -\frac{Zek_{D}^{2}}{2\pi(\beta^{2}-1)^{1/2}} \frac{\cos k_{D}[(z-\nu_{0}t)^{2}/(\beta^{2}-1)-x^{2}-y^{2}]^{1/2}}{[(z-\nu_{0}t)^{2}/(\beta^{2}-1)-x^{2}-y^{2}]^{1/2}} \times u\left(-\frac{z-\nu_{0}t}{(\beta^{2}-1)^{1/2}}-(x^{2}+y^{2})^{1/2}\right)$$
(3.9.30)

where the unit step function u comes from the following considerations. The dispersion surface (3.9.24) has asymptotes which make an angle θ with the horizontal axis so that

$$\tan \theta = 1/(\beta^2 - 1)^{1/2} \tag{3.9.31}$$

These asymptotes in $k_x k_z$ — plane are shown by dotted lines in Fig. 3.9-1. The normals to these asymptotes form a cone with an apex angle equal to θ . Within the cone a polarization charge density is induced; outside of the cone the charge density is unperturbed. The cone of spherical cross section with an apex angle θ serves as the shock boundary and is sometimes referred to as the Čerenkov cone. The coordinates of the cone are given by

$$\tan \theta = -\frac{(x'^2 + y'^2)^{1/2}}{z'} = -\frac{(x^2 + y^2)^{1/2}}{z - v_0 t}$$
(3.9.32)

Eliminating tan θ between (3.9.31) and (3.9.32), the Čerenkov cone is found to be given by

$$\frac{z-v_0t}{(\beta^2-1)^{1/2}}+(x^2+y^2)^{1/2}=0$$

A point (x, y, z) inside the cone must satisfy the inequality

$$-(z - v_0 t) < [(\beta^2 - 1)(x^2 + y^2)]^{1/2}$$
(3.9.33)

which clearly shows that the argument of the unit step in (3.9.30) is correct.

From (3.9.31) the apex angle θ also satisfies the following relation

$$\sin\theta = 1/\beta = \nu_T/\nu_0 \tag{3.9.34}$$

The complimentary angle of θ is then given by $\theta_c = \cos^{-1}(\nu_T/\nu_0)$ and is referred to as the Čerenkov angle. The field vanishes outside of the wellknown Čerenkov cone with the apex at the moving particle and the apex angle θ . The faster the velocity of the test particle ν_0 , the smaller is the apex angle θ . Far behind the Čerenkov cone and nearly on the axis, (3.9.30) can be approximated by

$$\varrho_p(\mathbf{r},t) = \frac{-Zek_D^2}{2\pi(z-v_0t)} \cos\frac{k_D(z-v_0t)}{(\beta^2-1)^{1/2}}, \qquad z \ll v_0t \qquad (3.9.35)$$

Note that the field decays as inverse distance. The field given by (3.9.35) is the radiation field brought about by the electrostatic interaction between the test particle and the fluid plasma particles. This radiation is indicative of energy transfer from test particle to plasma waves which then propagate away. Unless there is external force to maintain the velocity, the test particle must slow down and be stopped eventually. It is therefore of interest to consider the energy loss through radiation per unit distance of travel of the test particle. This energy loss is known as the stopping power.

For a test particle given by (3.9.7) moving with a velocity v_0 , the corresponding current density is

$$\mathbf{J}(\mathbf{r}, t) = Ze \ \delta(\mathbf{r} - \mathbf{v}_0 t) \mathbf{v}_0 \tag{3.9.36}$$

The power dissipated per unit volume by the test particle is just $\mathbf{J}(\mathbf{r}, t) \cdot \mathbf{E}_p(\mathbf{r}, t)$ which when integrated over the volume gives the total power dissipation. The use of (3.9.36) reduces the volume integral to

$$dW/dt = Ze\mathbf{v}_0 \cdot \mathbf{E}_p(\mathbf{v}_0 t, t) \tag{3.9.37}$$

The electric field \mathbf{E}_p in the power dissipation expression (3.9.37) is the field generated by polarization charges and is evaluated at the position of the test charge. Due to passage of the test charge, a polarization field is induced and it reacts back on the test particle. The induced field can be obtained by substracting the self-field from the total electric field given by (3.9.12). The self-field is the electric field that exists in free space and can thus be obtained from (3.9.12) by simply setting K_{\parallel} to 1. These remarks enable us to write explicitly the expression for the power dissipation as

$$\frac{dW}{dt} = \frac{j(Ze)^2 v_0}{(2\pi)^2 \varepsilon_0} \int_0^\infty k_\perp dk_\perp \int_{-\infty}^\infty dk_z \frac{k_z}{k_z^2 + k_\perp^2} \left[\frac{1}{K_{\parallel}(\mathbf{k}, k_z v_0)} - 1 \right] (3.9.38)$$

where we have assumed, as before, that the test particle is moving along zaxis. Also the volume integral in k-space is written in cylindrical coordinates and it is assumed that K_{\parallel} does not depend on the azimuthal angle. As discussed in Section 5 of Chapter 2, $K_{\parallel}(\mathbf{k}, \omega) \rightarrow 1$ as $\omega \rightarrow \infty$. Hence the integrand in (3.9.38) vanishes as $k_z \rightarrow \infty$ and we may close the integral with a large semicircle. The question of interest here is whether the large semicircle is above or below the real axis. In Section 5 of Chapter 2 we have

found that $K(\mathbf{k}, \omega)$ must be regular in the lower half ω -plane if the medium is stable. The argument makes use of the principal of causality with the electric field intensity viewed as the excitation and the electric displacement viewed as the response of the system. If we switch their roles and view the electric displacement as the excitation and the electric field intensity as the response, we must similarly conclude that $1/K_{\mu}(\mathbf{k}, \omega)$ must be regular in the lower half ω -plane in a stable and causal medium. As a matter of fact the transform technique used here cannot be used to define functions in the upper ω -plane because the integral diverges and they can be defined in the upper ω -plane only as analytic continuation from the lower half ω -plane. In the integral (3.9.38) the argument ω of K_{\parallel} is replaced by $k_2 v_0$. If the medium is not spatially dispersive, the foregoing remarks indicate that $1/K_{\parallel}(k_z v_0)$ must be regular in the lower half k_z -plane. Of course there may be poles in the upper half k_z -plane, but they are outside of the contour and shall make no contribution to the integral. In the limit of the lossless case the poles of $1/K_{\parallel}(k_z v_0)$ move down from the upper k_z -plane to the real axis. The contour integral along the real k_z -axis must be indented downward at these poles. These poles are still outside the contour and hence still make no contribution to the integral. For a spatially dispersive medium we can no longer make such general statement about the regularity of $1/K_{\parallel}(\mathbf{k}, k_z v_0)$ in the lower half k_z -plane because K_{\parallel} depends on k_z through its dependence on **k** and on ω (in this case since we set $\omega = k_z v_0$). However we shall assume that this is the case and it is certainly so for the longitudinal dielectric constant of the plasma given by (3.9.13) which reduces to

$$K_{\parallel}(\mathbf{k}, k_z \nu_0) = 1 - \frac{\omega_p^2}{k_z^2 (\nu_0^2 - \nu_T^2) - k_{\perp}^2 \nu_T^2}$$
(3.9.39)

Because of the assumed regularity of K_{\parallel} in the lower half k_z -plane and the indentation of the integral in the lossless case, the only contribution to the contour integral is from the pole at $k_z = -jk_{\perp}$ in (3.9.38). Carry out the contour integration with respect to k_z , yielding

$$\frac{dW}{dt} = \frac{(Ze)^2 \omega_p^2}{4\pi \varepsilon_0 v_0} \int_0^\infty \frac{k_\perp dk_\perp}{k_\perp^2 + \omega_p^2 / v_0^2}$$
(3.9.40)⁺

^t For the lossless case, the k_z integration of (3.9.38) can also be carried out by writing it as the sum of a principle part of the integral and the contributions from the small semicircle indentations at the two poles of $1/K_{\parallel}(\mathbf{k}, k_z \nu_0)$ on the real k_z -axis. The principle part integration vanishes because the integrand is an odd function of k_z . Therefore the only contribution is from the two poles. The result is the same as (3.9.40). Here we can see that the energy loss of the test particle is due to the excitation of the plasma waves, corresponding to $K_{\parallel} = 0$.

Problems

where (3.9.39) has been used in simplifying the expression. The integral (3.9.40) diverges logarithmically at the upper limit. We recall from the discussion in the beginning of this section that the dielectric constant expression such as (3.9.34) is applicable only for k less than k_D . Therefore, we arbitrarily introduce k_D as the upper limit in the integral (3.9.40). This has the effect of including the collective interactions between the test particle and the wave but close collisions that have individual particle behavior are ignored. The time rate of loss of energy by the test charge is then

$$\frac{dW}{dt} = \frac{(Ze)^2 \omega_p^2}{8\pi\epsilon_0 v_0} \log\left(\frac{k_D^2 v_0^2}{\omega_p^2} + 1\right)$$
(3.9.41)

We note that k_D enters only logarithmically and hence its exact choice is not very critical. The stopping power is defined as the energy loss by the test particle per unit distance of travel and is just (3.9.41) divided by v_0 , i.e.,

$$\frac{dW}{ds} = \frac{(Ze)^2 \omega_p^2}{8\pi \varepsilon_0 v_0^2} \log\left(\frac{k_D^2 v_0^2}{\omega_p^2} + 1\right)$$
(3.9.42)

The stopping power (3.9.42) can also be derived by integrating over the surface of a large sphere of radiation energy per unit distance of travel of the test particle (see problem at the end of the chapter). Therefore, all of the energy computed in (3.9.42) is transferred from the test particle to the wave and radiated away. We note that the slower suprathermal test charge is more efficient in producing plasma waves than the faster one because of the approximate inverse square dependence on v_0 in (3.9.42).

In the above discussion, the effect of ions in the plasma is neglected. If in (3.9.39), we include the terms corresponding to the effect of ions, similar computation can be carried out. One finds that in addition to the energy loss due to the excitation of electron plasma waves, there will be contribution from the excitation of ion plasma waves. But this contribution is small compared to the electron contribution simply because of the large mass of the ions.

Problems

1. Consider a sphere of radius R containing electrons with uniform density N_e and singly charged ions with uniform density N_i not necessarily equal to N_e .

3. Waves in Fluid Plasma

(a) Find the expression of the electrostatic potential at the surface of the sphere, assuming the potential at $r = \infty$ is zero.

(b) Solve for the potential inside the sphere. What is the potential at r = 0?

(c) Take R = 100 km, a dimension certainly small as compared with the whole of the ionosphere. Take $N_e = 10^{12}/\text{m}^3$, a typical value for the F region of the ionosphere. Calculate the surface potential for a number of percent departures of ion density.

2. It has been found that the natural frequency of oscillation of the plasma parallel slab is just the plasma frequency. This is somewhat accidental since the natural frequency of oscillation depends not only on the density but also on the geometry of the problem. Consider an infinite circular cylindrical column with radius *a* and uniform electron density N_e and singly charged positive density N_i . When unperturbed, the cylindrical column is electrically neutral with $N_e = N_i$. Displace all electrons by a small identical distance ξ in a direction normal to the axis. Find the natural frequency of oscillation [L. Tonks, *Phys. Rev.* 37, 1458 (1931); 38, 1219 (1932); N. Herlofson, *Ark. Fys.* 3, 247-297, (1951)].

3. This problem is concerned with distribution of charged particles in a uniform gravitational field in a highly idealized manner. When in thermodynamic equilibrium the electrons and positive ions are distributed according to the Boltzmann distribution

$$N_e = N_{e0} e^{(-m_e gz + eV)/T}$$
$$N_i = N_{i0} e^{(-m_i gz + eV)/T}$$

where g is the constant gravitational acceleration, the coordinate z is pointed vertically upward and the temperature of electron and ion are identical, and both are T expressed in energy units. The electrical potential V satisfies the Poisson's equation

$$d^2 V/dz^2 = -e(N_i - N_e)/\varepsilon_0$$

(a) Assume strict charge neutrality. Show that the electric field in the medium must be uniform and has a value $(m_i - m_e)g/2e$. This electric field arises from charges at the surface of the boundaries. Show that both electron and ion densities are distributed exponentially with a scale height $H = 2T/(m_e + m_i)g$, i.e., they are distributed as if the plasma has the average mass of electron and ion.

124

Problems

(b) In general the charge neutrality condition may not be tenable in some regions of the atmosphere. The problem is prescribed by some associated boundary conditions. However, due to the nonlinear character of the equations, they become very difficult to solve exactly. We must then resort to asymptotic methods. Let

$$E_0 = (m_i - m_e)g/2e$$

be the electric field when there is strict charge neutrality. Define ϕ as the potential departure (with a normalizing factor) from the case considered in (a), i.e.,

$$\phi = (e/T)[V + E_0(z - z_0)]$$

where z_0 is the height at which electrons and ions would have equal density if they were not influenced by the presence of electric potential, i.e.,

$$z_0 = [T/(m_i - m_e)g] \ln(n_{i0}/N_{e0})$$

The Poisson's equation may be transformed into the following nonlinear differential equation

$$d^2\phi/d\xi^2 = a^2 e^{-\xi} \sin h\phi$$

with the following substitutions:

$$N_{0} = N_{e0}e^{-m_{e}gz_{0}/T} = N_{i0}e^{-m_{i}gz_{0}/T}$$

$$\xi = z/H$$

$$\lambda_{D}^{2} = \varepsilon_{0}T/2N_{0}e^{2}$$

$$\mu = (m_{i} - m_{e})/(m_{i} + m_{e})$$

$$p = 2\ln(H/\lambda_{D})$$

$$\ln a^{2} = (z_{0}/H + p)$$

Let the associated boundary conditions be V(0) = 0, $E(0) = E(\infty) = 0$. Then they transform to

$$\phi(0) = \mu(p - \ln a^2)$$

 $d\phi/d\xi = \mu$ at $\xi = 0$ and ∞

The nonlinear equation given above is now appropriate for the discussion of asymptotic solutions in the limit $p \gg 1$.

3. Waves in Fluid Plasma

(i) In the region $\phi \ll 1$, we have the approximate solution

$$\phi \sim (-\mu \lambda_D/H) e^{-z/\lambda_D}$$

and the distribution found in (a) is approximately valid. Show this.

(ii) In the region $\phi \gg 1$, the electrons and ions are distributed according to their respective scale heights $H_e = T/m_e g$ and $H_i = T/m_i g$. Show this.

(iii) Discuss the solution in the transition region [(J. E. Allen, S. E. Segre, and A. Turrin, *Nuovo Cimento* 31, 402–413 (1964)].

4. Generally there are two classes of problems of interest. The first class of problems is the forced oscillation case in which the angular frequency of oscillation ω is given and is thus real while the wave vector **k** may be complex showing absorption of the wave. The second class of problems is the initial value problems in which an oscillation is set up with a definite wavelength and hence **k** is given and real while the angular frequency ω is allowed to take complex values showing temporal damping of the oscillation. Show that in a slightly lossy isotropic medium the negative imaginary part of **k** for a given real ω is related to the imaginary part of ω for a real **k** by the relation

$$\omega^{\prime\prime} = \mathbf{v}_{\mathbf{a}} \cdot \mathbf{k}^{\prime\prime}$$

where v_q is the group velocity.

5. Suppose the electron-neutral collisions in the form of a frictional force cannot be ignored entirely in the consideration of electron plasma waves as done in Section 5. Derive the complex longitudinal dielectric constant for this case. The appearance of the imaginary part of the dielectric constant shows that plasma waves will be damped in time if initiated. Find the shift in the frequency and the damping time constant in the limit of small collisions.

6. Derive Eq. (3.7.4) in Chapter 3 from the fluid equations as carried out in Section 5 of Chapter 3.

7. Show that the average force per particle in a uniform plasma with initial distribution function $f(v_0)$ due to a longitudinal wave $E_0 \cos(\omega t - kz)$ is

$$F = \left\langle m \, \frac{dv}{dt} \right\rangle_{z_0, v_0} = \frac{-\pi e^2 E_0^2}{2mk} f'\left(\frac{\omega}{k}\right)$$

[T. H. Stix, Bull. Amer. Phys. Soc. [2], 5, 530 (1960)].

126

Problems

8. For a Maxwellian distribution

$$f(v_0) = \left(\frac{m}{2\pi T}\right)^{1/2} e^{-mv_0^2/2T}$$

compute the Landau damping rate.

9. For a two-component plasma, electrons and ions, show that the dispersion relation for longitudinal waves can be expressed as

$$n^{4} - [(1 - \omega_{pe}^{2}/\omega^{2})/\delta_{e}^{2} + (1 - \omega_{pi}^{2}/\omega^{2})/\delta_{i}^{2}]n^{2} + [1 - (\omega_{pe}^{2} + \omega_{pi}^{2})/\omega^{2}]/\delta_{e}^{2}\delta_{i}^{2} = 0$$

where *n* is the refractive index and $\delta_e^2 = v_{Te}^2/c^2$, $\delta_i^2 = v_{Ti}^2/c^2$, *c* is the velocity of light.

For $\omega_{pi} \gg \omega_{pi}$, sketch the solution n^2 as a function of ω .

10. Flow of a charged plasma in a metallic conductor of variable cross section leads to the appearance of the so-called configuration emf. Consider the flow of steady current through a conductor whose cross section decreases abruptly at some point. Show that associated with the cross-section change there must be a jump in electric potential given by $(I^2m/2Ne^3)(S_2^{-2} - S_1^{-2})$ where I is the total current, S_2 and S_1 the cross-sectional areas of the conductor. Note that in a conductor the charged carriers are conduction electrons which have constant density [configuration emf is discussed by A. A. Vedenov, Sov. Phys. Usp. 7, 809-822 (1965); M. Chester, Phys. Rev. A 133, 907 (1964)].

11. The interaction of a solid-state plasma particle with lattice vibrations may lead to modification of velocity of sound. The propagation of sound waves induces plasma motion which in turn modifies deformation force and changes the velocity of sound. Let $\boldsymbol{\xi}$ be the deformation, s the velocity of sound, M and N the mass and density of lattice atoms, q_{α} the interaction constants between the lattice and α th kind of plasma particles, and n_{α} the plasma particle density. Then the d'Alambert equation for the deformation is

$$(\partial^2 \boldsymbol{\xi}/\partial t^2) - s^2 \nabla^2 \boldsymbol{\xi} = -(1/MN_{\alpha})q_{\alpha} \nabla n_{\alpha}$$

The equation of motion of the α th kind plasma particles is coupled back through

$$m d\mathbf{v}/dt = q \nabla^2 \mathbf{\xi} - (1/n) \nabla p + Z_{\alpha} e \mathbf{E} - m \mathbf{v}_{\alpha} \mathbf{v}$$

Suppose the plasma is electrically neutral and composed of singly charged

3. Waves in Fluid Plasma

particles of electrons and holes at temperatures T_1 and T_2 (in energy units). Assume inertia and frictional forces are negligible and the unperturbed plasma is homogeneous. Show that the modified velocity of sound is s' and satisfies

$$s^{\prime 2} = s^2 + \frac{(q_1 + q_2)^2}{T_1 + T_2} \frac{n}{NM}$$

[A. A. Vedenov, Solid state plasma, Sov. Phys. Usp. 1, 809-822 (1965); G. Weinreich, Phys. Rev. 104, 321 (1956)].

12. The response of the plasma to a test charge can be treated by using fluid equations (3.5.1) through (3.5.4) and the Poisson's equation. Assume that the plasma is composed of electrons and neutralizing background ions. Into this plasma a test charge of density ϱ_t is introduced. Show that the perturbed electron density must satisfy the linearized equation

$$\left(\nabla^2 - \frac{\partial^2}{\nu_T^2 \partial t^2} - \frac{1}{\gamma \lambda_D^2}\right) N_e' = -\varrho_t / \gamma e \lambda_D^2.$$

An equation of this form is known as the inhomogeneous Klein-Gordon equation in the quantum field theory. Show also that this equation is consistent with (3.9.14) by taking the Fourier transform of the equation. The test charge can be assumed to be moving along the z-axis with a constant velocity as was done in (3.9.7) [M. H. Cohen, Radiation in a plasma I, Čerenkov effect. *Phys. Rev.* **123**, 711-721 (1966)].

13. Assume the electron distribution function satisfies the simplified Boltzmann equation

$$\partial f/\partial t + \mathbf{v} \cdot \nabla f + \mathbf{v} \cdot \nabla v f = -(f - f_0)/\tau$$

where f_0 is the equilibrium distribution function and is independent of position and time. Use the perturbation approach and let

$$f(\mathbf{r}, \mathbf{v}, t) = f_0(\mathbf{v}) + f'(\mathbf{r}, \mathbf{v}, t)$$

with $f' \ll f_0$.

(a) Assume a complete degeneracy so that

$$\nabla_{\boldsymbol{v}} f_0 = -(\mathbf{v}/\boldsymbol{v}) \,\,\delta(|\mathbf{v}-\mathbf{v}_0|).$$

Find the longitudinal dielectric constant.

References

(b) Find the self-energy of the charged particle moving at a constant velocity \mathbf{v}_0 in this degenerate gas. Introduce a lower cutoff at the Debye wave number to prevent the integral from diverging.

(c) Find the stopping power of this charged particle [J. Lindhard, Kgl. Dan. Vidensk. Selsk. Mat. Fys. Medd. 28, No. 8 (1954)].

References

- ALLIS, W. P., BUCHSBAUM, S. J., and BERS, A. (1963). "Waves in Anisotropic Plasma." MIT Press, Cambridge, Massachusetts.
- BERNSTEIN, I. B., and TREHAN, S. K. (1960). Plasma oscillations I. Nucl. Fusion 1, 3.
- CHANG, D. B. (1964). Landau damping and related phenomena. Phys. Fluids 1, 1980-1986.
- COHEN, M. H. (1961). Radiation in a plasma, Part 1. Čerenkov effect, *Phys. Rev.* 123, 711-721.
- DENISSE, J. F., and DELCROIX, J. L. (1961). "Théorie des Ondes dans les Plasmas." Dunod, Paris.
- JACKSON, J. D. (1960). Longitudinal plasma oscillations. J. Nucl. Energy Part C 1, 171.
- KLIMONTOVICH, YU. L. (1967). "The Statistical Theory of Non-Equilibrium Processes in a Plasma." MIT Press, Cambridge, Massachusetts.
- LANDAU, L. D. (1946). On the vibrations of the electronic plasma. J. Phys. (U.S.S.R.) 10, 25.
- LIU, C. H. (1970). Long wavelength ion-acoustic waves in a magneto-plasma in a gravitational field. J. Plasma Phys. 4, 617-627.
- MONTGOMERY, D. C., and TIDMAN, D. A. (1964). "Plasma Kinetic Theory." McGraw-Hill, New York.
- ROSENBLUTH, M. N., and ROSTOKER, N. (1962). Electromagnetic wave scattering in a nonequilibrium plasma. *Phys. Fluids* 5, 776.
- SITENKO, A. G. (1967). "Electromagnetic Fluctuations in Plasma." Academic Press, New York.
- So, K. C., and YEH, K. C. (1968). Generalized stochastic equations. Proc. Phys. Soc. London (Gen.) [2], 1, 447-454.
- STIX, T. H. (1964). "The Theory of Plasma Waves." McGraw-Hill, New York.
4. Waves in Fluid Plasma with a Steady Magnetic Field

Chapter 3 was concerned exclusively with longitudinal waves in a plasma in the absence of a steady magnetic field. But this is not the only possible wave that can exist in the medium. Even in the absence of a steady magnetic field, we may have transverse electromagnetic waves. When there is an external magnetic field, the waves are in general neither strictly longitudinal nor strictly transverse. We shall find that the medium is extremely rich in sustaining many types of wave motions. Because the plasma state is described by a large number of parameters and the parameters may vary over a wide range, it has been found useful, convenient, and almost essential to define a parameter space. The refractive index surfaces are classified in the parameter space. This is helpful in visualizing the transformation of refractive index surfaces as a given parameter varies. Such a classification is first done on cold plasmas and later extended to the warm plasma case.

4.1 Transverse Dielectric Constant and Index of Refraction

The fluid equations (3.5.1) through (3.5.4) used to describe longitudinal plasma waves can still be used to describe transverse waves, but now the particle velocity is perpendicular to the propagation vector. The equation of continuity immediately deduces that the associated particle density is not perturbed (see Eq. 3.5.14). Consequently, the pressure force in the equation of motion does not come in (see Eq. 3.5.15) and we may speak of force on an "average" electron.

The equation of motion for an "average" electron is then

$$m\mathbf{\ddot{\xi}} = -e\mathbf{E} \tag{4.1.1}$$

where ξ is the displacement of the electron from its equilibrium position due to the presence of electric field and *m* the mass of an electron. Here the Lorentz force, $e\xi \times B$, has been ignored because we are assuming in this section the absence of external magnetic field and because we are linearizing the equation of motion. For sinusoidal forced oscillations of dependence exp $j\omega t$, the displacement can be solved in terms of the electric field,

$$\boldsymbol{\xi} = \boldsymbol{e}\mathbf{E}/\omega^2 \boldsymbol{m} \tag{4.1.2}$$

The dipole moment of a dipole arrangement of charges is given by (2.3.27). In a volume of charge density N, the resulting electric polarization density is

$$\mathbf{P} = -Ne\boldsymbol{\xi} = -(\omega_p^2/\omega^2)\varepsilon_0 \mathbf{E} = \chi \varepsilon_0 \mathbf{E}$$
(4.1.3)

where ω_p is the angular plasma frequency given by (3.3.4); i.e.,

$$\omega_p^2 = N e^2 / m \varepsilon_0 \tag{4.1.4}$$

In the ionospheric literature it is conventional to adopt the notation X for the normalized plasma frequency squared,

$$X = \omega_p^2 / \omega^2 \tag{4.1.5}$$

In this notation, the electric susceptibility is

$$\chi = -X \tag{4.1.6}$$

Here we have ignored entirely the ionic contribution to the susceptibility because of their large mass. The corresponding expressions for the dielectric constant and the refractive index are

$$\varepsilon(\omega) = \varepsilon_0(1 - X) = \varepsilon_0(1 - \omega_p^2/\omega^2) \tag{4.1.7}$$

$$n = k/(\omega/c) = (1 - X)^{1/2} = (1 - \omega_p^2/\omega^2)^{1/2}$$
(4.1.8)

Here, the dielectric constant is real and ω -dependent but not k-dependent. Equation (4.1.8) shows that for $\omega \gg \omega_p$ the propagation is nearly unaffected by the presence of the medium since the medium can not respond to such a high frequency. As ω is decreased, the polarization current density, $\dot{\mathbf{P}} = -j\omega Ne\boldsymbol{\xi} = -j\omega_p^2 \varepsilon_0 \mathbf{E}/\omega$, is increased. Since the polarization current is 180° out of phase from the displacement current, the total current is effectively reduced. This shows the frequency dispersive nature of the medium. When $\omega = \omega_p$, the polarization current and the displacement current have equal magnitude and the resulting total current is zero. When $\omega < \omega_p$, the total current is negative corresponding to the region of imaginary refractive index. This is the region of evanescence where the wave is attenuated exponentially. The division of propagation and evanescence occurs at $\omega = \omega_p$ at which n = 0. For convenience, the condition of vanishing refractive index shall be called the cutoff condition.

From (4.1.8), we can obtain the expressions for the phase velocity and group velocity as

$$v_p = \omega/k = c/(1 - (\omega_p/\omega)^2)^{1/2} = c/n$$
 (4.1.9)

and

$$v_g = d\omega/dk = c(1 - (\omega_p/\omega)^2)^{1/2} = cn$$
 (4.1.10)

It is interesting to note that the product of the phase velocity and the group velocity is equal to the square of the velocity of light in free space, i.e.,

$$v_p v_q = c^2 \tag{4.1.11}$$

In the ionospheric literature a group refractive index is occasionally used. It is defined by

$$n_g = c/\nu_g \tag{4.1.12}$$

Substitution of (4.1.10) in (4.1.12) immediately gives

$$n_g = 1/n$$
 (4.1.13)

The time-averaged energy in a wave packet can be computed by using (4.1.7) and (2.10.20).

$$\langle W \rangle = \frac{1}{4} |E|^2 \, \partial(\omega \varepsilon) / \partial\omega + \frac{1}{4} \mu_0 |H|^2 = \frac{1}{4} (\varepsilon_0 |E|^2 + \mu_0 |H|^2) + \frac{1}{4} (\omega_p^2 / \omega^2) \varepsilon_0 |E|^2$$
(4.1.14)

The first term of Eq. (4.1.14) represents the energy stored in the corresponding free space electromagnetic fields. The second term represents the additional energy stored in the material medium and in the present case it is actually the kinetic energy of the electrons. The time-averaged kinetic energy per unit volume is given by

$$\langle T \rangle = \frac{1}{4} N m \dot{\mathbf{\xi}} \cdot \dot{\mathbf{\xi}}^* = \frac{1}{4} (\omega_p^2 / \omega^2) \varepsilon_0 |E|^2 \qquad (4.1.15)$$

where Eq. (4.1.2) has been used. We see that the average kinetic energy density (4.1.15) is just the last term of Eq. (4.1.14).

In writing the equation of motion (4.1.1), the collisional effects are ignored entirely. A rigorous consideration of the collisional dynamics is outside the scope of this book. We shall only describe it by using a very simple model. Collisions between two hard spheres of different masses will result in a transfer of momentum from the light sphere (electron) to the heavy sphere (molecule) of an amount

$$\Delta(m\mathbf{v}) = m\mathbf{v}(1 - \cos \chi_s) \tag{4.1.16}$$

where χ_s is the scattering angle through which the electron is deflected. (Note: $\chi =$ electric susceptibility and $\chi_s =$ scattering angle.) Average (4.1.16) over all scattering angles and assume that all scattering angles are equally probable; we obtain

$$\langle \Delta(m\mathbf{v}) \rangle = m\mathbf{v} \tag{4.1.17}$$

Therefore, on the average each collision results in a transfer of electron momentum equal to its initial momentum. We shall assume that such collisions take place v times per second. The net rate of transfer of momentum from electrons to molecules is then mvv. We note that this collisional force, like frictional force, is proportional to velocity. In some literature this force is referred to as the Langevin force. The collisional frequency is related to the particle density, collisional cross section, and thermal velocity through

$$\boldsymbol{\nu} = N \sigma \boldsymbol{\nu}_T \tag{4.1.18}$$

In a weakly ionized gas, collisions occur mainly between electrons and neutral particles. In this case the particle density in (4.1.18) should be neutral particle density and the cross section the neutral particle cross section. For ionospheric studies we may adopt the electron neutral collisional frequency as [Nicolet, 1959]

$$v = 3.3 \times 10^{-16} \sqrt{T} [N(O_2) + N(N_2) + 2N(O)] \text{ sec}^{-1}$$
 (4.1.19)

where $N(O_2)$, $N(N_2)$ and N(O) are number densities in m^{-3} , respectively,

of O_2 , N_2 , and O in the atmosphere. The factor 2 is used to account for the experimental observation that the scattering cross section of electrons by atomic oxygen is 32 a.u., about twice the accepted cross section for air.

The effect of collisions is to introduce an additional frictionlike term in the equation of motion (4.1.1). For forced oscillations, it becomes

$$-\omega^2 m(1-j\nu/\omega)\mathbf{\xi} = -e\mathbf{E} \qquad (4.1.20)$$

Comparison of (4.1.20) with (4.1.2) indicates that the collisional effect can be taken into account simply by replacing m in the collision-free case by $m(1 - j\nu/\omega)$. This is equivalent to replacing ω_p^2 or X by $\omega_p^2/(1 - j\nu/\omega)$ or $X/(1 - j\nu/\omega)$, respectively. Such replacements in (4.1.7) and (4.1.8) give us the dielectric constant and refractive index, now with collisions taken into account, as

$$\varepsilon(\omega) = \varepsilon_0 \left(1 - \frac{\omega_p^2 / \omega^2}{1 - j\nu / \omega} \right) = \varepsilon_0 (1 - X/U)$$
(4.1.21)

$$n = (1 - X/U)^{1/2}$$
(4.1.22)

where

$$U = 1 - j\nu/\omega \tag{4.1.23}$$

We note from (4.1.21) and (4.1.22) that both the dielectric constant and the refractive index are in general complex. For a real frequency ω , the real part and the negative imaginary part of the dielectric constant are, respectively,

$$\varepsilon'(\omega) = \varepsilon_0 \left(1 - \frac{\omega_p^2 / \omega^2}{1 + \nu^2 / \omega^2} \right) \tag{4.1.24}$$

$$\varepsilon^{\prime\prime}(\omega) = \varepsilon_0 \left(\frac{\omega_p^2 \nu / \omega^3}{1 + \nu^2 / \omega^2} \right) \tag{4.1.25}$$

These formulas show that the $\varepsilon'(\omega)$ is even and $\varepsilon''(\omega)$ is odd in (real) ω , in agreement with (2.5.1). Further, we can show that they satisfy the Kramers-Kronig relations (2.5.15) (see problem at the end of the chapter). The presence of the simple pole at $\omega = 0$ merely indicates that the medium has a dc conductivity given by (2.5.10)

$$\sigma(0) = j \lim_{\omega \to 0} \omega \varepsilon(\omega) = N e^2 / m \nu \qquad (4.1.26)$$

For a real frequency ω , the real part and the negative imaginary part of the refractive index can be obtained from Eq. (4.1.22). They are given

by the following relations:

$$2(n')^{2} = 1 - (X/UU^{*}) + [1 - X(2 - X)/UU^{*}]^{1/2} \quad (4.1.27a)$$

$$2(n'')^2 = 1 + (X/UU^*) + [1 - X(2 - X)/UU^*]^{1/2} \quad (4.1.27b)$$

These expressions are rather involved for hand computational purposes. Graphical methods do exist if only moderate accuracy is needed (Deschamps and Weeks, 1962).

It should be mentioned that for a time it was not certain whether the Lorentz polarization term should be included. The inclusion of the Lorentz term has very large effect if a propagation technique is used to measure the electron density (see problem at the end of this chapter). However, the refined theory [Darwin (1943); its modern version is given by Theimer and Taylor (1961), Kadomtsev (1958)] indicates that the effective electric field in a gaseous plasma with free electrons is just equal to the applied electric field and thus the Lorentz polarization term should not be included.

4.2 Reflection of a Plane Transient Wave from the Plasma Half-Space

We have seen in Section 3.3 that a plasma has a natural frequency of oscillation at the plasma frequency ω_p . Therefore, it is reasonable to expect that in some way this resonance effect will be revealed in transient behavior. If so, it can be very useful in diagnostic applications to determine the electron density in a plasma.

Let us consider a plane wave incident normally from free space on a plasma half-space. Part of the incident energy is reflected and the remaining part transmitted, but both signals are affected by the dispersive properties of the medium. We shall postpone our discussion on transmitted signal till the next section and only treat the reflected signal in this section.

The reflection of an incident sinusoidal plane wave by a plane boundary is described by Fresnel formulas. For the case of normal incidence the reflection coefficient reduces to

$$R(\omega) = \frac{\{\omega - (\omega^2 - \omega_p^2)^{1/2}\}}{\{\omega + (\omega^2 - \omega_p^2)^{1/2}\}}$$
(4.2.1)

If the incident wave is $E_i(t)$ with Fourier transform $E_i(\omega)$, the reflected wave can be found by inverting $R(\omega)E_i(\omega)$. Since the reflected wave is propagating in free space which is not dispersive, we shall assume that both the incident and reflected waves are found at the interface. If these waves are at some nonvanishing distance from the interface, we need only to take appropriate time delays into account.

To be specific, let us assume that the incident wave (at the interface) is an impulse whose transform is unity, i.e.,

$$E_i(t) = \delta(t), \quad E_i(\omega) = 1$$
 (4.2.2)

The reflected wave (also at the interface) is then just the Fourier inverse of

$$E_r(\omega) = R(\omega)E_i(\omega) = \omega_p^2/[\omega + (\omega^2 - \omega_p^2)^{1/2}]^2$$
(4.2.3)

This expression has the exact inverse, giving

$$E_{r}(t) = -(2/t)J_{2}(\omega_{p}t)u(t)$$
(4.2.4)

where u(t) is a unit step and J_2 is the Bessel function of the first kind of order two. For a time large compared with the inverse plasma frequency, we may use the asymptotic expression of the Bessel function for large argument. It reduces (4.2.4) to

$$E_r(t) = (8/\pi\omega_p t^3)^{1/2} \cos(\omega_p t - \pi/4), \qquad \omega_p t \gg 1$$
(4.2.5)

We see that the reflected wave tends to oscillate with an angular frequency equal to plasma frequency. The time behavior of (4.2.4) can be found in Fig. 4.2-1. Analogous to system analysis we may speak of (4.2.4) as the impulse response of the system. As is well known in system analysis, the



Fig. 4.2-1. Time behavior of the reflected wave on a lossless plasma half-space for an impulse incident wave. [After Wait (1969).]

response of a general input can be obtained by convolving the impulse response with the input, i.e.,

$$E_r(t) = -\int_0^t (2/\tau) J_2(\omega_p \tau) E_i(t-\tau) d\tau \qquad (4.2.6)$$

For example, if the input wave is a unit step, $E_i(t) = u(t)$, the reflected wave is then

$$E_{r}(t) = \left[-\int_{0}^{t} (2/\tau) J_{2}(\omega_{p}\tau) d\tau \right] u(t)$$

= $[(2/\omega_{p}t) J_{1}(\omega_{p}t) - 1] u(t)$ (4.2.7)

where the integration formula,

$$\int x^{-p+1} J_p(x) \, dx = -x^{-p+1} J_{p-1}(x)$$

with p = 2, has been used. We note again that the reflected wave has a tendency to oscillate at ω_p . As $\omega_p t$ approaches infinity, the reflected wave given by (4.2.7) tends to -1 asymptotically, showing that, for these long times, the plasma is behaving like a perfect conductor.

As another example, it can be shown that if the incident field is a turnedon sinusoidal wave

$$E_i(t) = u(t)\sin\omega_0 t \tag{4.2.8}$$

the reflected wave is found to be (see Problem 5 at the end of chapter)

$$E_r(t) = -2\sum_{n=0}^{\infty} (-1)^n a_n J_{2n+3}(\omega_p t) u(t)$$
(4.2.9)

where

$$a_n = \cos(2n+1)\theta$$
, $\cos \theta = \omega_0/\omega_p$ if $\omega_0 \le \omega_p$

and

 $a_n = \cosh(2n+1)\theta$, $\cosh \theta = \omega_0/\omega_p$ if $\omega_0 \ge \omega_p$ (4.2.10)

These results can be generalized to the case of square-wave-modulated sinusoids by applying the principle of superposition.

4.3 Signal Propagation in Lossless, Isotropic Plasma

The problem of propagation of a time-dependent signal in a dispersive medium interested Sommerfeld and many others [see Brillouin (1960)]. For the case of lossless plasma, an exact solution expressed as a series can be obtained by using the method similar to that used in the last section. However, the series is found to converge extremely slowly if the observer is far away from the interface (Knop, 1964). Therefore, we shall use the method of saddle point to find the asymptotic behavior (Haskell and Case, 1967).

Let us now consider a uniform, lossless, isotropic plasma in an infinite half-space for z > 0. An electromagnetic signal E(z, t) is assumed to propagate in the plasma in the z-direction. From the discussion in Chapter 2 we can write the signal in terms of a superposition of plane waves in the form of an inverse Fourier transform

$$E(z, t) = \frac{1}{2\pi} \int_C E(0, \omega) e^{j\omega[t - n(\omega)z/c]} d\omega \qquad (4.3.1)$$

where $E(0, \omega)$ is the Fourier transform of the signal at the plane z = 0and $n(\omega)$ is the refractive index. For the lossless, isotropic plasma, $n(\omega)$ is given by (4.1.8) as

$$n(\omega) = (1 - \omega_p^2 / \omega^2)^{1/2}$$
(4.3.2)

where ω_p is the electron plasma frequency. Since only high frequency waves will be considered, the effect of ions can be neglected. The contour C in (4.3.1) is chosen such that it is parallel to the real ω -axis and below all singularities of the integrand. Let the applied signal at z = 0 be a turned-on sinusoidal wave of frequency ω_0 ,

$$E(0, t) = u(t)\sin\omega_0 t \tag{4.3.3}$$

which has the Fourier transform $E(0, \omega) = \omega_0/(\omega_0^2 - \omega^2)$. The applied signal at z = 0 given by (4.3.3) has a frequency spectrum which is peaked at $\omega = \omega_0$. Since the medium is dispersive, each frequency component of this signal will propagate with a different velocity in the plasma; hence there will be distortion to the signal. In the following, we shall study this transient response of the signal.

Introducing the following dimensionless variables

$$\tau = \omega_0 t, \qquad \zeta = \omega_0 z/c = k_0 z, \qquad X_0 = \omega_p^2 / \omega_0^2 P = X_0^{1/2} = \omega_p / \omega_0, \qquad \xi = \omega / \omega_0 = \xi_1 + j\xi_2$$
(4.3.4)

(4.3.1) can then be written in the form

$$E(\zeta,\tau) = \frac{1}{2\pi} \int_{C_{\xi}} \frac{1}{1-\xi^2} e^{j\xi[\tau-\zeta(1-X_0/\xi^2)^{1/2}]} d\xi \qquad (4.3.5)$$



Fig. 4.3-1. Locations of poles and zeros and path of integration for the integral given by (4.3.5).

where C_{ξ} is the contour shown in Fig. 4.3-1. The integrand has two poles at $\xi = \pm 1$ and two branch points at $\xi = \pm P$. A branch cut is drawn between +P and -P.

We note first that by closing the contour with a large semicircle in the lower half ξ -plane, we obtain from (4.3.5) $E(\zeta, \tau) = 0$ for $\tau - \zeta = \omega_0(t - z/c)$ < 0. This agrees with the principle of causality. At any position z in the plasma, no signal arrives prior to the time t = z/c which is the time it takes for the signal to travel from the origin to z in free space. Physically this can be explained by the fact that the dispersive property of the plasma is due to the induced motion of charged particles. Therefore, prior to the arrival of the signal, the medium is electromagnetically void like a vacuum. The very first portion of the signal always sees a vacuum in front, hence it always travels with the vacuum speed c.

For t - z/c > 0, the integral (4.3.5) has to be investigated in detail. First, we consider the so-called Sommerfeld solution (Brillouin, 1960). We deform the original path C_{ξ} in Fig. 4.3-1 into a semicircle of very large radius R in the lower half ξ -plane plus the segments of the real ξ -axis as shown in Fig. 4.3-2.

On the real ξ -axis, the integrand of (4.3.5) goes to zero as $1/\xi^2$, while on the semicircular path of radius R in the upper half ξ -plane it vanishes exponentially as $R \to \infty$ for $\tau - \zeta = \omega_0(t - z/c) > 0$. Therefore we can add to the original path a path shown by the dotted line in Fig. 4.3-2 without changing the value of the integral. The original path is then replaced by a circle of very large radius R, and $E(\zeta, \tau)$ is expressed by

$$E(\zeta,\tau) = \frac{1}{2\pi} \oint_{R} \frac{1}{1-\xi^2} e^{j\xi[\tau-\zeta(1-X_0/\xi^2)^{1/2}]} d\xi \qquad (4.3.6)$$



Fig. 4.3-2. Deformed path of integration for Eq. (4.3.5).

Since on this path $|\xi| \ge 1$, the integrand can be approximated by expanding the exponent in Taylor's series and keeping only terms up to the first order, and by neglecting the unity in the denominator. We have then

$$E(\zeta, \tau) \cong \frac{-1}{2\pi} \oint_{R} (1/\xi)^2 e^{j\xi[\tau-\zeta+\zeta X_0/2\xi^2]} d\xi \qquad (4.3.7)$$

Let

$$\xi = Re^{j\theta} \tag{4.3.8}$$

on the circle of radius R; then (4.3.7) becomes

$$E(\zeta,\tau) \cong \frac{-j}{2\pi} \int_0^{2\pi} \frac{1}{R} e^{j[R(\tau-\zeta)e^{j\theta} + (\zeta X_0/2R)e^{-j\theta}] - j\theta} d\theta \qquad (4.3.9)$$

This integral can be put into a standard form if we require

$$[\zeta X_0/2(\tau-\zeta)]^{1/2} = R \tag{4.3.10}$$

This is a special case whose corresponding physical condition shall be discussed shortly. Substitution of the special condition (4.3.10) in (4.3.9) gives

$$E(\zeta, \tau) \cong \frac{-j}{2\pi} \left(\frac{2(\tau - \zeta)}{\zeta X_0} \right)^{1/2} \int_0^{2\pi} e^{j2(\zeta X_0(\tau - \zeta)/2)^{1/2} \cos \theta} e^{-j\theta} \, d\theta$$
$$= \left(\frac{2(\tau - \zeta)}{\zeta X_0} \right)^{1/2} J_1 \{ 2[\zeta X_0(\tau - \zeta)/2]^{1/2} \}$$
(4.3.11)

where $J_1(x)$ is the Bessel function of order one.

140

If we define

$$a = 2X_0\zeta(\tau - \zeta) = 2\omega_p^2(t - z/c)z/c$$
(4.3.12)

then the Sommerfeld solution can be written as

$$E(\zeta,\tau) \cong (1/X_0\zeta)\sqrt{a} J_1(\sqrt{a})u(a) \tag{4.3.13}$$

where u(a) is the unit step function.

We note that to write the Sommerfeld solution in the form of (4.3.13), (4.3.10) must be satisfied. Since $R \gg 1$, we can deduce the range of validity of this solution from (4.3.10). The solution is valid for

$$\zeta X_0/2 \gg (\tau - \zeta) \tag{4.3.14}$$

At a fixed point z, if ζ is large, this solution works well right after the arrival of the signal, but becomes inapplicable as t increases. This portion of the transient signal is called "the precursors" by Sommerfeld (1914).

To study the signal after the arrival of the "precursor," let us go back to (4.3.5). The integral can be rewritten as the sum of two integrals,

$$E(\zeta, \tau) = (1/4\pi)[I_+ + I_-] \tag{4.3.15}$$

where

$$I_{\pm} = \int_{C_{\xi}} e^{\xi f(\xi)} / (1 \pm \xi) \, d\xi \tag{4.3.16}$$

with

$$f(\xi) = j\zeta[\tau/\zeta - (1 - X_0/\xi^2)^{1/2}]$$

= f_1(\xi) + jf_2(\xi) (4.3.17)

We note that although $f(\xi)$ depends on ζ , the condition that $\tau - \zeta > 0$ indicates that τ/ζ can be taken as a time parameter which is equal to one at the time of arrival of the signal and increases with time thereafter. We can therefore consider $f(\xi)$ to be independent of ζ for the purpose in the present computation. Equation (4.3.16) is of the standard form discussed in Appendix A. It can be evaluated by the method of saddle point integration.

The saddle points of $f(\xi)$ can be found by setting $f'(\xi) = 0$ and are given by

$$\xi_0 = \pm (\tau/\zeta) P / [(\tau/\zeta)^2 - 1]^{1/2}$$
(4.3.18)

Thus we have two saddle points lying symmetrically on the real axis with

141

respect to the origin. They are functions of time. At $\tau/\zeta = 1$, they are at $\pm \infty$. As time increases, they move toward the origin and will approach the branch points $\pm P$ for very large time. The saddle points will pass through the poles of the integrand at $\xi = \pm 1$ at the time $\tau = \tau_g = \zeta/(1 - P^2)^{1/2}$. We shall see that this time corresponds to the time of arrival of the main signal.

At the saddle points, the real and imaginary parts of the function $f(\xi)$ are given by

$$f_1(\xi_0) = 0$$

$$f_2(\xi_0) = \pm [(\tau/\zeta)^2 - 1]^{1/2}P$$
(4.3.19)

The lines of steepest descent (or ascent) passing through saddle points are given by

$$f_2(\xi) = f_2(\xi_0) \tag{4.3.20}$$

In the vicinity of the saddle points, the steepest descent path can be easily found to be

$$\theta_s = n\pi/2 \pm \alpha/2, \quad n = 1, 3$$
 (4.3.21)

where $\theta_s = \arg(\xi - \xi_0)$ and $\alpha = \arg f''(\xi_0)$. These lines are shown in Fig. 4.3-3 where the corresponding valleys of $f_1(\xi)$ in the vicinity of the saddle points are shown by hatched regions.

Several cases must be considered separately.

(i) $|\xi_0| = (\tau/\zeta)P/[(\tau/\zeta)^2 - 1]^{1/2} \gg 1$. The saddle points are far away from the poles at $\xi = \pm 1$. For this case, the original path of integration C_{ξ} can be deformed into SDP's as shown by dashed lines in Fig. 4.3-3.



Fig. 4.3-3. Lines of steepest descent (or ascent) in complex ξ -plane.

The integrals I_{\pm} can then be evaluated by the method of steepest descent as discussed in Appendix A. We have

$$I_{+} \sim \left(\frac{2\pi P}{\zeta}\right)^{1/2} \frac{1}{A^{3/2}} \left[\frac{e^{-j\zeta PA - j\pi/4}}{1 - \tau P/\zeta A} + \frac{e^{j\zeta PA + j\pi/4}}{1 + \tau P/\zeta A}\right]$$

$$I_{-} \sim \left(\frac{2\pi P}{\zeta}\right)^{1/2} \frac{1}{A^{3/2}} \left[\frac{e^{-j\zeta PA - j\pi/4}}{1 + \tau P/\zeta A} + \frac{e^{j\zeta PA + j\pi/4}}{1 - \tau P/\zeta A}\right]$$
(4.3.22)

and

$$E(\zeta,\tau) \sim \left(\frac{2P}{\pi\zeta}\right)^{1/2} \frac{A^{1/2}}{A^2 - \tau^2 P^2/\zeta^2} \cos(\zeta PA - \pi/4) \qquad (4.3.23)$$

where

$$A = [(\tau/\zeta)^2 - 1]^{1/2}$$
(4.3.24)

The amplitude of $E(\zeta, \tau)$ increases with increasing τ/ζ until $A \simeq \tau P/\zeta$, or $\tau/\zeta \simeq 1/(1 - P^2)^{1/2}$, the time at which the saddle points move close to the poles. At this time, the expression is no longer valid, since when the saddle points are in the neighborhood of the poles, the method of saddle point integration used above is no longer applicable; it requires the use of a modified saddle point method by Van der Waerden (1950). This modified method is discussed in Appendix A.2.

The range of validity for (4.3.23) is given by

$$\left| \zeta (1 - \tau P / \zeta A)^2 A^3 / P \right| \gg 1 \tag{4.3.25}$$

(ii) $|\xi_0| \sim 1$, Arrival of the Main Signal. This corresponds to $\tau \to \tau_g$; the saddle points are in the vicinity of the poles. In evaluating I_+ , the contribution from the saddle point in the right half-plane still is the same as in (i) while that from the saddle point in the left half-plane must be treated separately. Let us divide the integral I_+ into two parts: I_+^+ and I_+^- , where I_+^+ represents the contribution of the integral along that portion of the path C_{ξ} which is in the right half ξ -plane while I_+^- is the contribution from the remaining portion of the path. In the neighborhood of the saddle point $\xi_0^- = -(\tau/\zeta)P/[(\tau/\zeta)^2 - 1]^{1/2}$, the integral I_+^- along the SDP can be written as

$$I_{+}^{-} \cong e^{\xi f(\xi_{0}^{-})} \int_{-e}^{+e} \frac{e^{\frac{1}{2} \xi f''(\xi_{0}^{-})(\xi - \xi_{0}^{-})^{2}}}{1 + \xi} d\xi$$
$$= e^{\xi f(\xi_{0}^{-})} \int_{-e}^{+e} \frac{e^{-a\varrho^{2}}}{-j(1 + \xi_{0}^{-}) e^{j\alpha/2} + \varrho} d\varrho \qquad (4.3.26)$$

where $a = \zeta f''(\xi_0^-)/2$, $\alpha = \arg f''(\xi_0^-)$, and (4.3.21) has been used.

For $\zeta \gg 1$, the limits of integration can be extended to $\pm \infty$ without introducing significant error. Therefore, we have

$$I_{+}^{-} \cong e^{\xi f(\xi_{0}^{-})} \int_{-\infty}^{+\infty} \frac{e^{-a\varrho^{2}}}{\varrho - \beta} d\varrho$$

$$\beta = j(1 + \xi_{0}^{-})e^{j\alpha/2} \qquad (4.3.27)$$

with

$$I_{+}^{-} \cong e^{\zeta f(\xi_{0}^{-})} \int_{-\infty}^{+\infty} \frac{(\varrho + \beta)}{\varrho^{2} - \beta^{2}} e^{-a\varrho^{2}} d\varrho$$

= $2\beta e^{\zeta f(\xi_{0}^{-})} \int_{0}^{\infty} \frac{e^{-a\varrho^{2}}}{\varrho^{2} - \beta^{2}} d\varrho$
= $\beta e^{\zeta f(\xi_{0}^{-})} \int_{0}^{\infty} t^{-1/2} \left(\frac{1}{t - \beta^{2}}\right) e^{-at} dt$ (4.3.28)

where the transform $t = \varrho^2$ has been used.

Equation (4.3.28) is of the standard form discussed in Appendix A.2 and the result of the integration is [see Eq. (A.2.7)]

$$I_{+}^{-} \sim -j\pi e^{\zeta f(\xi_{0}^{-})} e^{-a\beta^{2}} \operatorname{erfc} (-a\beta^{2})^{1/2}$$
(4.3.29)

where

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_{x}^{\infty} e^{-y^2} dy$$
 (4.3.30)

is the complementary error function.

In terms of the original variables, (4.3.29) can be written as

$$I_{+}^{-} \sim -j\pi \exp\{-jA\zeta P[1 + (A^{2}/2P^{2})(1 - \tau P/\zeta A)^{2}]\} \times \operatorname{erfc} [-j(A^{3}\zeta/2P)(1 - \tau P/\zeta A)^{2}]^{1/2}$$
(4.3.31)

where A is given by (4.3.24)

Equation (4.3.31) is the asymptotic expression for I_{+}^{-} when the saddle point is near the pole. We note that it is finite when $\tau P/\zeta A = 1$, corresponding to $|\xi_0| = 1$.

We can now write the asymptotic expression for the integral I_+ . For $\tau \leq \tau_g$ so that the saddle point has not yet crossed the pole, the integral is equal to the contribution from the saddle point ξ_0^+ plus that from ξ_0^- .

From (4.3.22) and (4.3.31) we have

$$I_{+} \sim \left(\frac{2\pi P}{\zeta}\right)^{1/2} \frac{1}{A^{3/2}} \frac{\exp[j(\zeta PA + \pi/4)]}{1 + \tau P/\zeta A}$$

-j\pi \exp\{-j\zeta PA[1 + A^{2}(1 - \tau P/\zeta A)^{2}/2P^{2}]\}
\times \exp\{crfc[-j\zeta A^{3}(1 - \tau P/\zeta A)^{2}/2P]^{1/2} for \tau \le \tau_{g} (4.3.32)\}

For $\tau \ge \tau_g$, the saddle point ξ_0^- has already crossed the pole at -1. Therefore, in deforming the contour to SDP, the contribution from the pole must



Fig. 4.3-4. Deformed path for $\tau \geq \tau_g$.

be included (Fig. 4.3-4). This contribution is easily calculated by the residue theorem. We have

$$I_{+} \sim (I_{+})_{\tau \leq \tau_{g}} + 2\pi j \exp\{-j\zeta[\tau/\zeta - (1 - X_{0})^{1/2}]\} \quad \text{for} \quad \tau \geq \tau_{g} \quad (4.3.33)$$

 I_{-} can be computed in exactly the same manner. We obtain

$$I_{-} \sim \left(\frac{2\pi P}{\zeta}\right)^{1/2} \frac{1}{A^{3/2}} \frac{\exp[-j(\zeta PA + \pi/4)]}{1 + \tau P/\zeta A}$$

+ $j\pi \exp\{j\zeta PA[1 + A^{2}(1 - \tau P/\zeta A)^{2}/2P^{2}]\}$
 $\times \operatorname{erfc}[j\zeta A^{3}(1 - \tau P/\zeta A)^{2}/2P]^{1/2} \quad \text{for} \quad \tau \leq \tau_{g} \quad (4.3.34)$

and

$$I_{-} \sim (I_{-})_{\tau \leq \tau_{g}} - 2\pi j \exp\{j\zeta[\tau/\zeta - (1 - X_{0})^{1/2}]\} \quad \text{for} \quad \tau \geq \tau_{g} \quad (4.3.35)$$

Substituting Eqs. (4.3.32)-(4.3.35) into (4.3.15), we have for the signal

$$E(\zeta, \tau) \sim \frac{P}{2\pi\zeta} \frac{1}{A^{3/2}(1+\tau P/\zeta A)} \cos(\alpha PA + \pi/4) + \frac{1}{4} \{-j \exp\{-j\zeta PA[1+A^2(1-\tau P/\zeta A)^2/2P^2]\} \times \operatorname{erfc}\left[\frac{\sqrt{\pi}}{2}(1-j)\left(\frac{\zeta}{-P\pi}\right)^{1/2}A^{3/2}(1-\tau P/\zeta A)\right] + \operatorname{c.c.}\}, \quad \tau \leq \tau_g$$
(4.3.36)

and

$$E(\zeta, \tau) \sim [E(\zeta, \tau)]_{\tau \le \tau_g} + \sin[\tau - (1 - X_0)^{1/2} \zeta], \quad \tau \ge \tau_g \quad (4.3.37)$$

where c.c. in (4.3.36) represents the complex conjugate of the first term in the bracket.

Equations (4.3.36) and (4.3.37) are the asymptotic expressions for the transient signal in the time range $\tau \sim \tau_g$. For large ζ , the first term in (4.3.36) is very small and can be neglected. The quantity $1 - \tau P/\zeta A$ vanishes for $\tau = \tau_g$ and is very small for $\tau \simeq \tau_g$. Therefore the term in the bracket of (4.3.36) can be approximated using Taylor's expansions for the exponential and erfc functions. The amplitude of this term is approximately $\frac{1}{2}$. Therefore as time increases, the signal is building up to about half its steady state value and then, when $\tau \ge \tau_g$, the main signal arrives, which is represented by the last term in (4.3.37). (See Fig. 4.3-5).

(iii) $|\xi_0| \ll 1$. In this region, the saddle points are again far away from the poles; hence the result of (4.3.23) is valid. In addition to that, the contribution from the poles must also be included. Therefore, we have

$$E(\zeta, \tau) \sim \left(\frac{2P}{\pi\zeta}\right)^{1/2} \frac{A^{1/2}}{A^2 - \tau^2 P^2/\zeta^2} \cos(\zeta PA - \pi/4) + \sin[\tau - (1 - X_0)^{1/2}\zeta]$$
(4.3.38)

The amplitude of the signal oscillates about the steady state value but as time increases, the first term in (4.3.36) diminishes and the amplitude approaches unity.

To summarize the above discussion, a numerical example is plotted in Fig. 4.3-5. This is a plot of the envelope of the transient response of a turned-on sinusoidal signal which has propagated in a lossless, isotropic plasma under the condition $\zeta = k_0 z = 10^4$ and $P = X_0^{1/2} = \omega_p / \omega_0 = 0.8$.

Prior to the time t = z/c ($\tau/\zeta = 1$), no signal will arrive at the position z. This is the consequence of the principle of causality. Right after t > z/c ($\tau/\zeta > 1$), the Sommerfeld solution is applicable. This solution is indicated by the region S in the figure. The amplitude of the signal is very small. As τ/ζ increases, solution (4.3.23) must be used and is indicated by the Region I in Fig. 4.3-5. Region II represents the main signal buildup solution of (4.3.36)



Fig. 4.3-5. Envelope of the transient response of a turned-on sinusoidal signal. [After Haskell and Case (1967).]

and (4.3.37). When $\tau = \tau_g$ the saddle points cross the poles and the amplitude increases to about half the steady state value. Region III in Fig. 4.3-5 shows a small oscillation of the amplitude about its final value after the arrival of the main signal. The oscillation decays as time increases.

The discussion in this section is for the special "turn-on" sinusoidal signal. The results can be used directly to study the response of a sinusoidal pulse of duration T. Since the pulse can be represented by

$$E(0, t) = u(t) \sin \omega_0 t - u(t - T) \sin \omega_0 (t - T)$$
 (4.3.39)

The results of this section can then be applied to the two terms of (4.3.39) separately.

Finally, it is noted that the discussion in this section can be generalized without much difficulty to include the effect of collisional loss in the plasma on the signal transient response. Since the collisional effect is proportional to ν/ω where ν is the collision frequency, additional distortion as well as damping of the signal is expected.

4.4 Gyrofrequency in the Ionosphere

As discussed in Section 3.3, a plasma has a tendency to oscillate at its plasma frequency. The presence of an external magnetic field introduces additional characteristic frequencies. It is known that a charge particle executes a spiral motion about the magnetic lines of force. For a particle of charge Ze and of mass m moving perpendicular to the magnetic field \mathbf{B}_0 with velocity v_{\perp} , the equation of motion is

$$mv_{\perp}^2/r = |Z|ev_{\perp}B_0 \tag{4.4.1}$$

where the centrifugal force on the left-hand side of the equation is balanced by the magnetic force of Lorentz on the right-hand side. The radius of gyration about the magnetic lines of force is

$$r = m v_{\perp} / |Z| e B_0 \quad m \tag{4.4.2}$$

The time required to complete a circular orbit about the magnetic field line is called the gyration period T and is given by $(2\pi r/\nu_{\perp}) = 2\pi m/|Z|eB_0$. The angular frequency of gyration can then be computed, i.e.,

$$\omega_B = 2\pi/T = |Z| eB_0/m \text{ rad/sec}$$
(4.4.3)

For convenience, we shall always use the angular gyrofrequency (4.4.3) as a positive number.

For ionospheric applications, the appropriate magnetic field is that of the earth. In many cases, the earth's field can be approximated by a dipole with south pole at 78.6°N, 70.1°W and with north pole at 78.6°S, 109.9°E. The magnetic moment of the earth's equivalent dipole is

$$M = 8.06 \times 10^{22} \quad \text{A-m}^2 \tag{4.4.4}$$

from which we can compute the magnetic intensity in the magnetic north direction X and in the vertically downward direction Z. For the dipole approximation, the magnetic field does not have a component in the magnetic east direction which has the symbol Y. We shall note that X, Y, Z are standard notations used in the study of terrestrial geomagnetism and they should not be confused with the X, Y, Z used in the theory of wave propagation in plasmas. Let Λ be the geomagnetic latitude, positive in the northern hemisphere, and r the distance from the center of the earth; the components of magnetic field intensity are given by

$$X = M \cos \Lambda / 4\pi r^3 \quad A/m$$

$$Z = 2M \sin \Lambda / 4\pi r^3 \quad A/m$$
(4.4.5)

The angle between the total field and the local horizon is called the magnetic dip I. In the dipole approximation, the dip is related to the geomagnetic latitude through

$$I = \tan^{-1}(Z/X) = \tan^{-1}(2 \tan \Lambda)$$
(4.4.6)

As expected, the magnetic dip is 0° at the magnetic equator and $\pm 90^{\circ}$, at the geomagnetic poles. We also note that *I* is positive in the northern hemisphere and negative in the southern hemisphere. The total magnetic field strength is denoted by *F* and from (4.4.5) we obtain

$$F = \frac{M}{4\pi r^3} (1 + 3\sin^2 \Lambda)^{1/2} \quad A/m$$
 (4.4.7)

The total field (4.4.7) decays like $1/r^3$ with r, characteristic for a dipole field. The field is strongest at the magnetic pole ($\Lambda = 90^{\circ}$) at which the field is twice as strong as the field at the magnetic equator ($\Lambda = 0^{\circ}$). Multiplying (4.4.7) by μ_0 we can compute the earth magnetic flux density on the surface of the earth (r = 6371 km) to be $3-6 \times 10^{-5}$ Wb/m². The corresponding gyrofrequency for electrons varies from 0.85 MHz at the equator to 1.7 MHz at the pole and that for protons from 460 Hz to 920 Hz. The electron gyrofrequency is right in the broadcast band and the propagation of broadcast radio signals is expected to be influenced by the earth magnetic field. The proton gyrofrequency is very small and the ions can be ignored except for propagation of lowest frequencies.

4.5 Dielectric Tensor of a Cold Magnetoplasma

We have seen in Section 4.1 that in an isotropic plasma there were no associated density perturbations and consequently no associated pressure perturbations with the propagation of electromagnetic waves. In other words, the pressure force in the equation of motion contributes only to longitudinal waves and not to transverse waves in our fluid model. This is no longer the case when there is a steady external magnetic field in the plasma. As we shall find in a later section, the pressure term introduces complexities and intricacies that may not be easily untangled at the first reading. Therefore, we choose to ignore the pressure term altogether in this section. The neglect of the pressure term is justified if the thermal velocity of the particle is small when compared with the phase velocity of the wave, viz., $v \ll v_p$. When this is the case, we speak of the plasma as being "cold."

The cold plasma model gives a satisfactory description except for waves with extremely slow phase velocity.

The equation of motion for an average particle of α th kind with mass m_{α} , charge $Z_{\alpha}e$ is

$$m_{\alpha} \ddot{\boldsymbol{\xi}}_{\alpha} = Z_{\alpha} \boldsymbol{e} (\mathbf{E} + \dot{\boldsymbol{\xi}}_{\alpha} \times \mathbf{B}_{0}) \qquad (4.5.1)$$

where \mathbf{B}_0 is the externally applied steady magnetic field and \mathbf{E} the electric field of the wave. For the sinusoidal case of dependence $e^{j\omega t}$ (4.5.1) can be rewritten in the form

$$\mathbf{E} = -(\omega^2 m_{\alpha}/Z_{\alpha} e) \boldsymbol{\xi}_{\alpha} - j \omega \boldsymbol{\xi}_{\alpha} \times \mathbf{B}_{0}$$
(4.5.2)

The electric polarization density due to α th kind of ions is related to its displacement from the equilibrium position by

$$\mathbf{P}_{\alpha} = N_{\alpha} Z_{\alpha} e \boldsymbol{\xi}_{\alpha} \tag{4.5.3}$$

where N_x is the density of α th kind of ions. It follows from (4.5.2) that the electric field and the polarization are related by

$$-\varepsilon_0 X_{\alpha} \mathbf{E} = (1 + j \mathbf{Y}_{\alpha} \times) \mathbf{P}_{\alpha}$$
(4.5.4)

Before we proceed further we need to clarify the notations used. We define angular plasma frequency and gyrofrequency by

$$\omega_{p_{\alpha}}^{2} = N_{\alpha}(Z_{\alpha}e)^{2}/m_{\alpha}\varepsilon_{0} \qquad (4.5.5)$$

$$\boldsymbol{\omega}_{B\alpha} = -(Z\alpha e/m_{\alpha})\mathbf{B}_{0} \tag{4.5.6}$$

Note that the plasma frequency is always a positive number while the gyrofrequency given by (4.5.6) is a vector. Charges of opposite signs are gyrating about the magnetic field in opposite directions. With the negative sign in (4.5.6) the orbit of charged particles of either sign will gyrate about ω_{Bx} in the right-hand sense. This is clearly demonstrated in Fig. 4.5-1. Again we define a set of normalized frequencies by

$$X_{\alpha} = \omega_{p\alpha}^2 / \omega^2$$
, proportional to N_{α} (4.5.7)

$$\mathbf{Y}_{\alpha} = \boldsymbol{\omega}_{B\alpha} / \omega$$
, proportional to B_0 (4.5.8)

To help to remind ourselves that (4.5.7) is the ratio of two frequencies squared and (4.5.8) is the ratio of two frequencies, it is useful to remember that X_{α} is proportional to N_{α} while Y_{α} is proportional to B_0 .



Equation (4.5.4) can be put in the form

$$\varepsilon_0 \mathbf{E} = \mathbf{\chi}_{\alpha}^{-1} \cdot \mathbf{P}_{\alpha} \tag{4.5.9}$$

with the inverse susceptibility tensor of the α th kind of charged particles expressed as

$$\mathbf{X}_{\alpha}^{-1} = -\frac{1}{X_{\alpha}} \begin{bmatrix} 1 & -jY_{\alpha z} & jY_{\alpha y} \\ jY_{\alpha z} & 1 & -jY_{\alpha x} \\ -jY_{\alpha y} & jY_{\alpha x} & 1 \end{bmatrix}$$
(4.5.10)

The determinant of (4.5.10) is found to be $(1 - Y_{\alpha}^2)/X_{\alpha}^3$. The inversion of (4.5.10) gives the susceptibility tensor as

$$\mathbf{X}_{\alpha} = -\frac{X_{\alpha}}{1 - Y_{\alpha}^{2}} \times \begin{bmatrix} 1 - Y_{\alpha x}^{2} & -Y_{\alpha x}Y_{\alpha y} + jY_{\alpha z} & -Y_{x x}Y_{\alpha z} - jY_{\alpha y} \\ -Y_{\alpha x}Y_{\alpha y} - jY_{x z} & 1 - Y_{x y}^{2} & -Y_{\alpha y}Y_{x z} + jY_{\alpha x} \\ -Y_{\alpha x}Y_{\alpha z} + jY_{\alpha y} & -Y_{\alpha y}Y_{\alpha z} - jY_{\alpha x} & 1 - Y_{\alpha z}^{2} \end{bmatrix}$$
(4.5.11)

The total electric polarization density due to all kinds of particles is just the sum of partial polarization densities, i.e.,

$$\mathbf{P} = \sum_{\alpha} \mathbf{P}_{\alpha} = \varepsilon_0 \sum_{\alpha} \mathbf{X}_{\alpha} \cdot \mathbf{E} = \varepsilon_0 \mathbf{X} \cdot \mathbf{E}$$
(4.5.12)

The total electric susceptibility due to all kinds of charged particles is therefore

$$\mathbf{X} = \sum_{\alpha} \mathbf{X}_{\alpha} \tag{4.5.13}$$

with \mathbf{X}_{α} given by (4.5.11). The dielectric tensor of the medium follows from (4.5.13) and is

$$\boldsymbol{\varepsilon} = \varepsilon_0 (\mathbf{I} + \mathbf{\chi}) = \varepsilon_0 \mathbf{K} \tag{4.5.14}$$



where I is a unit tensor. We note that the susceptibility tensor (4.5.11) and, consequently, also the dielectric tensor are Hermitian showing that the medium is lossless. The Onsager relation as discussed in Section 2.6 is also satisfied. In the limit of vanishing magnetic field, all Y's $\rightarrow 0$ and thus $\chi_{\alpha} \rightarrow -X_{\alpha}$. Correspondingly, the dielectric constant becomes isotropic and reduces to

$$\varepsilon = \varepsilon_0 (1 - \sum_{\alpha} X_{\alpha}), \quad \mathbf{B}_0 \to 0$$
 (4.5.15)

This is just the dielectric constant (4.1.7) derived earlier with contributions from all kinds of charged particles taken into account. Since ion plasma frequencies are negligibly small when compared with electron plasma frequency, (4.5.15) reduces to (4.1.7) for all practical purposes.

The dielectric tensor takes a simpler form when B_0 is along the positive z-coordinate. In this case the nine elements of K are given by

$$K_{xx} = K_{yy} = 1 - \sum_{\alpha} X_{\alpha} / (1 - Y_{\alpha}^{2})$$

$$K_{zz} = 1 - \sum_{\alpha} X_{\alpha}$$

$$K_{xy} = -K_{yx} = -j \sum_{\alpha} X_{\alpha} Y_{\alpha} / (1 - Y_{\alpha}^{2})$$

$$K_{xz} = K_{yz} = K_{zx} = K_{zy} = 0$$

$$B_{0} // \hat{z} - axis \qquad (4.5.16)$$

Here **K** not only satisfies the Onsager relation but is also rotationally symmetric about z-axis. A word of caution about the sign of Y_{α} for the element K_{xy} and K_{yx} in (4.5.16) is in order. As defined by (4.5.8) and (4.5.6) Y_{α} is positive for negatively charged particles and negative for positively charged particles. It is helpful to remember that a negatively charged particle has a gyration vector parallel to the magnetic field and hence Y_{α} is positively charged particle and hence Y_{α} is negative. An occasional reference to Fig. 4.5-1 may be useful.

In the literature of plasma theory several other tensors are used. One is the mobility tensor which for α th kind of particles is defined by the relation

$$\dot{\boldsymbol{\xi}}_{\alpha} = \boldsymbol{\mu}_{\alpha} \cdot \mathbf{E} \tag{4.5.17}$$

Comparison of (4.5.17) with (4.5.12) shows that the mobility tensor and the susceptibility tensor are related by

$$\boldsymbol{\mu}_{\alpha} = \frac{j\omega\varepsilon_{0}}{N_{\alpha}Z_{\alpha}e} \, \boldsymbol{\chi}_{\alpha} \tag{4.5.18}$$

For the lossless and cold magnetoplasma considered here, the mobility tensor is found to be

$$\boldsymbol{\mu}_{\alpha} = \frac{Z_{\alpha}e}{j\omega m_{\alpha}(1-Y_{\alpha}^{2})} \begin{bmatrix} 1 & jY_{\alpha} & 0 \\ -jY_{\alpha} & 1 & 0 \\ 0 & 0 & 1-Y_{\alpha}^{2} \end{bmatrix}, \quad \boldsymbol{B}_{0} // \hat{z} \text{-axis} \quad (4.5.19)$$

where the steady magnetic field is assumed to be in the z-direction.

Another tensor often used is the conductivity tensor. It can be related to the dielectric tensor by considering the total current on the right-hand side of the Maxwell equation

$$\nabla \mathbf{X} \mathbf{H} = \mathbf{J} + \dot{\mathbf{D}}$$

For the sinusoidal case with $e^{j\omega t}$ dependence the effect of the medium can be taken into account in the form of a polarization current as done here and we obtain in this case $j\omega \mathbf{\epsilon} \cdot \mathbf{E}$ for the right-hand side. However, the contribution to the total current may be equally treated by including a conduction current, and the right-hand side becomes $\boldsymbol{\sigma} \cdot \mathbf{E} + j\omega \varepsilon_0 \mathbf{E}$. Both methods give identical results and therefore

$$\boldsymbol{\sigma} = j\omega(\boldsymbol{\varepsilon} - \varepsilon_0 \mathbf{I})$$
$$= j\omega\varepsilon_0 \mathbf{X}$$
(4.5.20)

We note that for a lossless case \mathbf{x} is Hermitian while $\boldsymbol{\sigma}$ is not.

So far in this section we have ignored entirely the collisional effect. Since collision is a loss process its inclusion will make the dielectric tensor non-Hermitian. We shall now turn to this question in the next section.

4.6 Effect of Collisional Loss and DC Conductivity

The collisional effect can be taken into account simply if it is frictionlike. Let v_{α} be the effective collisional frequency for momentum transfer of the α th kind of ions with neutral particles. The equation of motion becomes

$$m_{\alpha} \ddot{\mathbf{\xi}}_{\alpha} = Z_{\alpha} e(\mathbf{E} + \dot{\mathbf{\xi}}_{\alpha} \times \mathbf{B}_{0}) - m_{\alpha} v_{\alpha} \dot{\mathbf{\xi}}_{\alpha} \qquad (4.6.1)$$

Comparing this equation with (4.5.1) we conclude that the frictional loss can be taken into account if we replace m_{α} for the lossless case by $m_{\alpha}(1 - jv_{\alpha}/\omega)$ as we have done in Section 4.1 for the isotropic case. Equivalently if we replace X_{α} in the lossless case by X_{α}/U_{α} and Y_{α} by Y_{α}/U_{α} simultaneously, we obtain the case with collisional loss taken into account. Here U_{α} is defined by

$$U_{\alpha} = 1 - j \nu_{\alpha} / \omega \qquad (4.6.2)$$

For example, for the case in which \mathbf{B}_0 is oriented along z-axis, such simultaneous substitutions in (4.5.11) give the susceptibility tensor

$$\mathbf{\chi}_{\alpha} = -\frac{X_{\alpha}}{U_{\alpha}^{2} - Y_{\alpha}^{2}} \begin{bmatrix} U_{\alpha} & jY_{\alpha} & 0 \\ -jY_{\alpha} & U_{\alpha} & 0 \\ 0 & 0 & (U_{\alpha}^{2} - Y_{\alpha}^{2})/U_{\alpha} \end{bmatrix}, \quad \mathbf{B}_{0} // \hat{z} \text{-axis}$$
(4.6.3)

Similarly, the elements of the relative dielectric tensor are obtained from (4.5.16) as

$$K_{xx} = K_{yy} = 1 - \sum_{\alpha} X_{\alpha} U_{\alpha} / (U_{\alpha}^{2} - Y_{\alpha}^{2})$$

$$K_{zz} = 1 - \sum_{\alpha} X_{\alpha} / U_{\alpha}$$

$$K_{xy} = -K_{yx} = -j \sum_{\alpha} X_{\alpha} Y_{\alpha} / (U_{\alpha}^{2} - Y_{\alpha}^{2})$$

$$K_{xz} = K_{yz} = K_{zx} = K_{zy} = 0$$

$$B_{0} // \hat{z} \text{-axis} \qquad (4.6.4)$$

These formulas reduce to the corresponding lossless case if we let $U_{\alpha} = 1$. We note that χ_{α} and **K** are not Hermitian. But they satisfy the Onsager relation and they are rotationally symmetric with respect to z-axis. Further, the elements (4.6.4) satisfy the Kramers-Kronig relation.

The susceptibility tensor given by (4.6.3) has a simple pole at the origin of the complex ω -plane. This indicates the presence of dc conductivity. The conductivity tensor is related to the susceptibility tensor by (4.5.20). In the limit of $\omega \rightarrow 0$ we obtain

$$\boldsymbol{\sigma}(\omega=0) = \begin{bmatrix} \sigma_P & \sigma_H & 0 \\ -\sigma_H & \sigma_P & 0 \\ 0 & 0 & \sigma_{\text{H}} \end{bmatrix}, \quad \mathbf{B}_0 // \hat{z} \text{-axis} \quad (4.6.5)$$

where

$$\sigma_{\rm H} = \sum_{\alpha} \varepsilon_0 \omega_{p\alpha}^2 / \nu_{\alpha}$$

$$\sigma_P = \sum_{\alpha} \varepsilon_0 \omega_{p\alpha}^2 \nu_{\alpha} / (\nu_{\alpha}^2 + \omega_{B\alpha}^2)$$

$$\sigma_H = \sum_{\alpha} \varepsilon_0 \omega_{p\alpha}^2 \omega_{B\alpha} / (\nu_{\alpha}^2 + \omega_{B\alpha}^2)$$
(4.6.6)

These conductivities are referred to as the parallel conductivity for σ_{\parallel} ,

154

Pederson conductivity for σ_P , and Hall conductivity for σ_H . The parallel conductivity is the conductivity for the isotropic case [compare with (4.1.26)]. The Pederson conductivity shows the effect of the magnetic field and its associated current, called the Pederson current, flows in the same direction as the electric field. The presence of Hall conductivity gives rise to a current flowing in a direction perpendicular to the electric field. These conductivities are quantities of central importance in the study of the dynamo theory and ionospheric currents (Maeda and Kato, 1966).

We note that the dc conductivity tensor in (4.6.5) may also be obtained from (4.6.4) by using the definition (2.5.10).

4.7 Longitudinal Oscillations

The electric field is parallel to the propagation vector for longitudinal oscillations. We deduce from Maxwell's equations, as we have done in Section 2.9, that $\mathbf{H} = 0$ and $\mathbf{D} = 0$, i.e.,

$$\varepsilon_0 \mathbf{K} \cdot \mathbf{E} = 0 \tag{4.7.1}$$

Eq. (4.7.1) has a nontrivial solution for **E** only when the determinant of the coefficient matrix vanishes, i.e., det $|\mathbf{K}| = 0$. Let us orient our coordinate axes so that $\mathbf{B}_0 // \hat{z}$ -axis. It follows that the elements given by (4.5.16) may be used. To be explicit, the longitudinal waves are obtained in a cold lossless magnetoplasma by requiring

$$\det |\mathbf{K}| = \begin{vmatrix} 1 - \sum_{\alpha} \frac{X_{\alpha}}{1 - Y_{\alpha}^{2}} & -j \sum_{\alpha} \frac{X_{\alpha}Y_{\alpha}}{1 - Y_{\alpha}^{2}} & 0\\ j \sum_{\alpha} \frac{X_{\alpha}Y_{\alpha}}{1 - Y_{\alpha}^{2}} & 1 - \sum_{\alpha} \frac{X_{\alpha}}{1 - Y_{\alpha}^{2}} & 0\\ 0 & 0 & 1 - \sum_{\alpha} X_{\alpha} \end{vmatrix} = 0,$$
(4.7.2)

where $\mathbf{B}_0 // \hat{z}$ -axis. For later convenience let us define three "relative" dielectric constants by

$$K_0 \equiv 1 - \sum_{\alpha} X_{\alpha} = K_{zz} \tag{4.7.3a}$$

$$K_{\rm I} \equiv 1 - \sum_{\alpha} \frac{X_{\alpha}}{1 - Y_{\alpha}^2} + \sum_{\alpha} \frac{X_{\alpha} Y_{\alpha}}{1 - Y_{\alpha}^2} = K_{xx} + jK_{xy} \qquad (4.7.3b)$$

$$K_{II} \equiv 1 - \sum_{\alpha} \frac{X_{\alpha}}{1 - Y_{\alpha}^{2}} - \sum_{\alpha} \frac{X_{\alpha}Y_{\alpha}}{1 - Y_{\alpha}^{2}} = K_{xx} - jK_{xy} \qquad (4.7.3c)$$

When the determinant is multiplied out the condition (4.7.2) reduces to

$$K_0 = 0$$
 (4.7.4a)

$$K_{\rm I} = 0$$
 (4.7.4b)

$$K_{\rm HI} = 0$$
 (4.7.4c)

For positive frequencies these three conditions give us three different modes. They are discussed in the following.

(i) $K_0 = 0$ Mode. We deduce from (4.7.1) that $E_x = E_y = 0, \quad E_z \neq 0$ (4.7.5)

for this mode. That is, the electric field is polarized in the same direction as the external magnetic field. As the induced motion is uninfluenced by the magnetic field, the resulting dispersion relation (4.7.4a) reduces to

$$\omega^2 = \sum_{\alpha} \omega_{p\alpha}^2 \approx \omega_{pe}^2 \tag{4.7.6}$$

which is identical to the isotropic case. Because of the heavy mass of ions when compared with electrons, the sum of the square of plasma frequencies can be replaced by the square of electron plasma frequency with negligible error. Properties of such waves have been discussed in Chapter 3 and shall not be repeated here.

(ii) $K_{I} = 0$ Mode. The ratios of three components of electric field are obtained from (4.7.1). If $1 - \sum_{\alpha} X_{\alpha} \neq 0$, then

$$E_{x}: E_{y}: E_{z}$$

$$= \begin{vmatrix} 1 - \sum_{\alpha} \frac{X_{\alpha}}{1 - Y_{\alpha}^{2}} & 0 \\ 0 & 1 - \sum_{\alpha} X_{\alpha} \end{vmatrix} : - \begin{vmatrix} j \sum_{\alpha} \frac{X_{\alpha} Y_{\alpha}}{1 - Y_{\alpha}^{2}} & 0 \\ 0 & 1 - \sum_{\alpha} X_{\alpha} \end{vmatrix}$$

$$: \begin{vmatrix} j \sum_{\alpha} \frac{X_{\alpha} Y_{\alpha}}{1 - Y_{\alpha}^{2}} & 1 - \sum_{\alpha} \frac{X_{\alpha}}{1 - Y_{\alpha}^{2}} \\ 0 & 0 \end{vmatrix}$$

$$= \left(1 - \sum_{\alpha} \frac{X_{\alpha}}{1 - Y_{\alpha}^{2}}\right) : -j \sum_{\alpha} \frac{X_{\alpha} Y_{\alpha}}{1 - Y_{\alpha}^{2}} : 0 \qquad (4.7.7)$$

For our mode, $K_{I} = 0$, it simplifies (4.7.7) to

$$E_x: E_y: E_z = -j: 1:0 \tag{4.7.8}$$

The electric field (4.7.8) is in the plane transverse to the magnetic field. E_x and E_y have equal magnitude and differ in phase by 90° with E_y leading. The electric field is therefore left circularly polarized in the plane transverse to B_0 . The dispersion relation $K_I = 0$ can be rearranged in several ways since

$$K_{I} = 1 - \sum_{\alpha} X_{\alpha} / (1 + Y_{\alpha})$$

= $1 - \sum_{\alpha} \omega_{p\alpha}^{2} / \omega(\omega + \omega_{B\alpha})$
= $1 + \sum_{\alpha} \omega_{p\alpha}^{2} / \omega_{B\alpha}(\omega_{B\alpha} + \omega) - \sum_{\alpha} \omega_{p\alpha}^{2} / \omega \omega_{B\alpha}$ (4.7.9)

In an electrically neutral plasma the last term vanishes and

$$\sum_{\alpha} \frac{\omega_{p\alpha}^2}{\omega \omega_{B\alpha}} = (\sum_{\alpha} N_{\alpha} Z_{\alpha} e) / \varepsilon_0 \omega B_0 = 0$$

Therefore, the dispersion relation simplifies to

$$1 + \sum_{\alpha} \frac{\omega_{p_{\alpha}}^2}{\omega_{B_{\alpha}}(\omega_{B_{\alpha}} + \omega)} = 0 \qquad (4.7.10)$$

We note here that the dispersion relation (4.7.10) is ω -dependent but not **k**-dependent. Consequently, the group velocity for the wave is zero. Any perturbation introduced locally will not propagate away via this mode. This comes about because the dielectric tensor used is that for a cold plasma.

The dispersion relation (4.7.10) yields as many roots for ω as the number of species of charged particles. However, some of the roots may be negative, corresponding to the mode to be discussed in (iii). In general the total number of modes in a cold magnetoplasma is equal to the number of species plus the mode considered in (i).

As an example, let us consider a two-component plasma with electrons and singly charged neutralizing ions. Remember that electron gyrofrequency is positive while that of ion is negative. The relation $(\omega_{pe}^2/\omega_{Be}) = -(\omega_{pi}^2/\omega_{Bi})$ can be used in (4.7.10) to give the dispersion relation

$$1 + \frac{\omega_{pe}^2}{\omega_{Be}} \left(\frac{1}{\omega_{Be} + \omega} - \frac{1}{\omega_{Bi} + \omega} \right) = 0$$
 (4.7.11)

When multiplied out, this is a quadratic equation in ω . If we assume $|\omega_{Be}| \gg |\omega_{Bi}|$, as is generally valid for gaseous plasmas, the positive root is found to be

$$\omega = -(\omega_{Be}/2) + [\omega_{Pe}^2 + (\omega_{Be}/2)^2]^{1/2}$$
(4.7.12)

In the limit of vanishing magnetic field this mode degenerates to the ordinary plasma oscillation considered in (i). When $\omega_{Be} \gg \omega_{pe}$, the oscillation occurs at $\omega = (\omega_{pe}^2/\omega_{Be})$, i.e., a small fraction of electron plasma frequency.

The effect of collisional loss can be taken into account by replacing ω_{Be} by $\omega_{Be}/(1 - j\nu_e/\omega)$ and ω_{pe}^2 by $\omega_{pe}^2/(1 - j\nu_e/\omega)$ in (4.7.12). For $\nu \ll \omega$, we may expand and find that the real part of ω is still given by (4.7.12) to the first order in ν_e/ω but now there exists an imaginary part of ω given by

$$\omega^{\prime\prime} = -(\omega_{Be}\nu_{e}/2\omega^{\prime}) + \omega_{pe}^{2} = (\omega_{Be}/2)^{2} \frac{\nu_{e}(\omega_{pe}^{2} + \omega_{Be}^{2}/2)}{2\omega(\omega_{pe}^{2} + \omega_{Be}^{2}/4)} \quad (4.7.13)$$

In (4.7.13) ω' is given by (4.7.12). The presence of the imaginary part of ω shows damping of the oscillation.

(iii) $K_{II} = 0$ Mode. Similar considerations as (ii) result in

$$E_x: E_y: E_z = j: 1:0 \tag{4.7.14}$$

The electric field is right circularly polarized in a plane transverse to B_0 . The dispersion relation in a neutral plasma is

$$1 + \sum_{\alpha} \frac{\omega_{p\alpha}^2}{\omega_{B\alpha}(\omega_{B\alpha} - \omega)} = 0 \qquad (4.7.15)$$

For the example of a two-component plasma considered in (ii), the dispersion relation yields one positive root,

$$\omega = (\omega_{Be}/2) + [\omega_{Pe}^2 + (\omega_{Be}/2)^2]^{1/2}$$
(4.7.16)

The oscillation frequency (4.7.16) is larger than that given by (4.7.12). When $\omega_{Be} \gg \omega_{pe}$, the oscillation occurs at approximately the electron gyro-frequency. The collisional effect can be considered similarly.

In case there are streaming motions of plasma particles along the magnetic field, it is possible that the plasma may become unstable. The effect of streaming motion will not change any of the polarization relations (4.7.5), (4.7.8), and (4.7.14), but the dispersion relations will have to be modified to take Doppler shift into account. These discussions have been made in Section 3.7 for the isotropic plasma. Since so far our dispersion relations are not yet k-dependent we shall delay our discussion until we have the appropriate relations in a later section.

4.8 Refractive Indices and Polarizations

One of the important properties of wave propagation in anisotropic media is the occurrence of characteristic waves or normal modes. A characteristic wave is a plane wave which can propagate in a given uniform medium without changing its wave polarization. When defined in this way, a wave with an arbitrary polarization is a characteristic wave in any isotropic medium without spatial dispersion. This is no longer true for the anisotropic case. As we have seen in Chapter 2, there exist two characteristic waves when the anisotropic medium is not spatially dispersive. These two waves have two generally distinct polarizations and refractive indices. In a homogeneous and unbounded medium, the two waves propagate independently in the linear limit. Mode conversion is therefore expected if either the medium is inhomogeneous, or there exist boundaries or the nonlinear effects become important.

Several forms for the expression of refractive index are in common use in the literature; each has the advantage over other forms in discussing certain properties of the wave. Some of these forms have been given in Sections 2.7 and 2.8 for general anisotropic media. The reader may wish to review these sections. We shall be more concerned with the cold magnetoplasmas here.

The starting point of the derivation is the set of Maxwell's equations. For plane waves with dependence $e^{j(\omega t - n\hat{s} \cdot \mathbf{r})}$, where \hat{s} is a unit vector in the direction of the propagation vector, Maxwell's equations reduce to

$$\mathbf{D} = -(n/c)\hat{s} \times \mathbf{H}$$

$$\mathbf{H} = (\varepsilon_0/\mu_0)^{1/2}n\hat{s} \times \mathbf{E}$$

$$\hat{s} \cdot \mathbf{D} = 0$$

$$\hat{s} \cdot \mathbf{B} = 0$$

(4.8.1)

Eliminating **H** from the first two equations of (4.8.1) and making use of the dielectric tensor, we obtain the wave equation

$$\mathbf{D} \cdot \mathbf{E} = \mathbf{0} \tag{4.8.2}$$

with

$$\mathbf{D} = k^2 \mathbf{I} - \mathbf{k} \mathbf{k} - k_0^2 \mathbf{K} \tag{4.8.3}$$

The symbol I in (4.8.3) stands for the unit dyad and $k_0 = \omega/c$. By comparing (4.8.3) with (2.7.8), **D** and **L** are found to be related through $\mathbf{D} = k_0^2 \mathbf{L}$.

The relative dielectric tensor, **K**, for the cold magnetoplasma has been derived and is given by (4.5.16). The set of homogeneous equations (4.8.2) has nontrivial solution if and only if **D** has a vanishing determinant. Let us orient the coordinate axes so that B_0 is along z-axis and **k** in the xz-plane with a polar angle θ as shown in Fig. 4.8-1. The vanishing of the determinant in this coordinate system is then

det
$$\mathbf{D} = \begin{vmatrix} n^2 \cos^2 \theta - K_{xx} & -K_{xy} & -n^2 \cos \theta \sin \theta \\ K_{xy} & n^2 - K_{xx} & 0 \\ -n^2 \cos \theta \sin \theta & 0 & n^2 \sin^2 \theta - K_{zz} \end{vmatrix} = 0$$
 (4.8.4)

where we have used the polar coordinates with $n = k/k_0$. The determinant (4.8.4) can be multiplied out and rearranged into the form

$$(K_{zz} - n^2)[K_{xx}(K_{xx} - n^2) + K_{xy}^2]\sin^2\theta + K_{zz}[(K_{xx} - n^2)^2 + K_{xy}^2]\cos^2\theta$$

= 0 (4.8.5)

The elements of the relative dielectric tensor are given by (4.5.16). They can be expressed in terms of three relative dielectric constants defined by (4.7.3) and their introduction simplifies our expression to

$$(n^{2} - K_{0})[n^{2}(K_{I} + K_{II})/2 - K_{I}K_{II}]\sin^{2}\theta + K_{0}(n^{2} - K_{I})(n^{2} - K_{II})\cos^{2}\theta$$

= 0 (4.8.6)

The form (4.8.6) was first used by Ästrom (1950) and it is especially convenient for studying the special cases of parallel propagation ($\theta = 0^{\circ}$) and perpendicular propagation ($\theta = 90^{\circ}$). These special cases shall be considered in a later section.

The equation for the refractive index may also be given in a different form. For a cold plasma, **K** is not **k**-dependent and consequently is neither a function of n nor θ . In this case (4.8.5) is a quadratic equation in n^2 , representing two modes of propagation. As done in Section 2.8, the biquadratic equation in n can be represented by

$$a_4n^4 + a_2n^2 + a_0 = 0 \tag{4.8.7}$$

with coefficients given by

$$a_{4} = K_{xx} \sin^{2} \theta + K_{zz} \cos^{2} \theta$$

$$a_{2} = -K_{xx} K_{zz} (1 + \cos^{2} \theta) - (K_{xx}^{2} + K_{xy}^{2}) \sin^{2} \theta \qquad (4.8.8)$$

$$a_{0} = \det |\mathbf{K}| = (K_{xx}^{2} + K_{xy}^{2}) K_{zz}$$

When compared with (2.8.12) we find that (4.8.8) is simpler because certain elements are zero in K. We note that the coefficients a_4 , a_2 , and a_0 are functions only of the polar angle θ , showing the axial symmetry about the magnetic field \mathbf{B}_0 . Further, the coefficients (4.8.8) are unchanged when θ is replaced by $180^\circ - \theta$. This shows that the refractive index has a plane symmetry about a plane perpendicular to the magnetic field \mathbf{B}_0 . The two solutions of (4.8.7) may be written in two forms:

$$n^{2} = \frac{-a_{2} \pm (a_{2}^{2} - 4a_{4}a_{0})^{1/2}}{2a_{4}}$$
$$= 1 - \frac{2(a_{4} + a_{2} + a_{0})}{2a_{4} + a_{2} \pm (a_{2}^{2} - 4a_{4}a_{0})^{1/2}}$$
(4.8.9)

In the ionospheric literature the second form of (4.8.9) is usually used. For the general multicomponent plasmas, (4.8.9) does not lead to any simple forms when expressed in terms of plasma parameters except for the special case of parallel and perpendicular propagation. If the frequency is so high that we need to take only electrons into account, (4.8.9) simplifies to the Appleton-Hartree equation which shall be discussed in a later





section. The manipulation of (4.8.8) for the general case from one form to another form is sometimes very laborious and requires considerable algebraic dexterity. For this reason (4.8.9) is seldom used except for numerical computations.

The characteristic polarizations for a general anisotropic medium have been discussed in Section 2.9. In the following, we obtain these polarizations for the magnetoplasma.

The state of polarization of a wave is described by the ratios of three component fields. For the coordinate system of Fig. 4.8-1, these ratios can be

obtained from (4.8.2) as ratios of determinants of certain cofactors of **D**, i.e.,

$$E_{x}: E_{y}: E_{z}$$

$$= \begin{vmatrix} n^{2} - K_{xx} & 0 \\ 0 & n^{2} \sin^{2} \theta - K_{zz} \end{vmatrix} : - \begin{vmatrix} K_{xy} & 0 \\ -n^{2} \cos \theta \sin \theta & n^{2} \sin^{2} \theta - K_{zz} \end{vmatrix}$$

$$: \begin{vmatrix} K_{xy} & n^{2} - K_{xx} \\ -n^{2} \cos \theta \sin \theta & 0 \end{vmatrix}$$

$$= (n^{2} - K_{xx})(n^{2} \sin^{2} \theta - K_{zz}) : - K_{xy}(n^{2} \sin^{2} \theta - K_{zz})$$

$$: (n^{2} - K_{xx}) n^{2} \cos \theta \sin \theta \qquad (4.8.10)$$

However, the polarization of the wave is commonly not discussed in the coordinates of Fig. 4.8-1 in which \mathbf{B}_0 is parallel to z-axis. The reason is simple. As seen from the last two equations of (4.8.1), the fields **B** and **D** are transverse fields even though **E** is not. In a coordinate system in which **k** is parallel to z-axis we expect **B** and **D** to be confined in the xy-plane. Therefore, let us rotate the coordinates shown in Fig. 4.8-1 about y-axis through an angle θ to obtain Fig. 4.8-2a. Subsequently, we rotate about z'-axis through 90° to obtain Fig. 4.8-2b. We wish now to express all field



Fig. 4.8-2. Coordinates in which the polarization of the wave is expressed.

components in the double primed coordinates. Note that the primed system Fig. 4.8-2a is the same as the one used in Section 2.9 while the double primed system is the one used more often in the ionosphere literature.

From the first two equations of (4.8.1) we obtain

$$\mathbf{D} = n^2 \varepsilon_0 (\mathbf{E} - \hat{s}\hat{s} \cdot \mathbf{E}) \tag{4.8.11}$$

When expressed in the double primed coordinates of Fig. 4.8-2b, (4.8.11)

becomes, in component form,

$$D_{x}^{''} = \varepsilon_{0} n^{2} E_{x}^{''} = \varepsilon_{0} E_{x}^{''} + P_{x}^{''}$$

$$D_{y}^{''} = \varepsilon_{0} n^{2} E_{y}^{''} = \varepsilon_{0} E_{y}^{''} + P_{y}^{''}$$

$$D_{z}^{''} = 0 = \varepsilon_{0} E_{z}^{''} + P_{z}^{''}$$
(4.8.12)

Express the electric field of a given characteristic mode by

$$\mathbf{E} = E_y''(\hat{x}''R_x'' + \hat{y}'' + \hat{z}''R_x''Q)$$
(4.8.13)

The state of polarization of the electric field is completely defined by the complex ratios R''_x and Q. Therefore

$$R_x'' = E_x''/E_y'' = D_x''/D_y'' = P_x''/P_y''$$
(4.8.14a)

$$Q = E_z''/E_x'' = P_z''(1 - n^2)/P_x''$$
(4.8.14b)

where (4.8.12) has been used. The state of polarization of the magnetic field can be related to that of the electric field by using the second equation of (4.8.1). Expressed in the double primed coordinates of Fig. 4.8-2,

$$\mathbf{H} = (\varepsilon_0/\mu_0)^{1/2} n(-\hat{x}'' E_y'' + \hat{y}'' E_x'')$$

which gives the relation

$$H_x''/H_y'' = -E_y''/E_x'' = -1/R_x''$$
(4.8.15)

The relation (4.8.15) is general and is applicable to the lossless as well as lossy case. It tells us that for a given characteristic wave the *H*-ellipse and the *E*-ellipse when projected on the wavefront are similar and both polarization ellipses are rotating in the same sense but their major axes are perpendicular. Actually, in a lossless medium *n* is real and hence from the second equation of (4.8.1) E and H must be perpendicular instantaneously. But if the medium is lossy, the instantaneous magnetic field and instantaneous electric field need not be perpendicular since *n* is complex.

The double primed coordinates of Fig. 4.8-2b is obtained from the coordinates of Fig. 4.8-1 through a coordinate transformation

$$\mathbf{T} = \begin{bmatrix} 0 & 1 & 0 \\ -\cos\theta & 0 & \sin\theta \\ \sin\theta & 0 & \cos\theta \end{bmatrix}$$
(4.8.16)

The dielectric tensor in the new coordinate system is related to that given by

(4.5.16), i.e.,

$$\mathbf{K}^{\prime\prime} = \mathbf{T} \cdot \mathbf{K} \cdot \mathbf{T}^{-1} \tag{4.8.17}$$

As the transformation (4.8.16) is linear and preserves orthogonality, its inverse \mathbf{T}^{-1} is equal to the transpose of \mathbf{T} . Carrying out the matrix products in (4.8.17), the elements of \mathbf{K}'' are expressed in terms of the elements (4.5.16) and the angle θ between the propagation vector and the steady magnetic field:

$$K_{xx}'' = K_{xx}$$

$$K_{xy}'' = -K_{yx}'' = K_{xy} \cos \theta$$

$$K_{xz}'' = -K_{zx}'' = -K_{xy} \sin \theta$$

$$K_{yy}'' = K_{xx} \cos^2 \theta + K_{zz} \sin^2 \theta$$

$$K_{yz}'' = K_{zy}'' = (-K_{xx} + K_{zz}) \sin \theta \cos \theta$$

$$K_{zz}'' = K_{xx} \sin^2 \theta + K_{zz} \cos^2 \theta$$
(4.8.18)

The wave equation (4.8.2) in the new coordinates is

$$\begin{bmatrix} n^{2} - K''_{xx} & -K''_{xy} & -K''_{xz} \\ K''_{xy} & n^{2} - K''_{yy} & -K''_{yz} \\ K''_{xz} & -K''_{yz} & -K''_{zz} \end{bmatrix} \begin{bmatrix} E''_{x} \\ E''_{y} \\ E''_{z'} \end{bmatrix} = 0$$
(4.8.19)

By a procedure similar to that used in Section 2.9, we obtain equations for the polarization ratio as

$$R_{x}^{\prime\prime2} + \frac{K_{xx}^{\prime\prime} - K_{yy}^{\prime\prime} + (K_{xz}^{\prime\prime2} + K_{yz}^{\prime\prime2})/K_{zz}^{\prime\prime}}{K_{xy}^{\prime\prime} - K_{xx}^{\prime\prime}K_{yz}^{\prime\prime}/K_{zz}^{\prime\prime}} R_{x}^{\prime\prime} + 1 = 0 \qquad (4.8.20)$$

$$Q = (K_{xz}^{\prime\prime} R_x^{\prime\prime} - K_{yz}^{\prime\prime}) / K_{zz}^{\prime\prime} R_x^{\prime\prime}$$
(4.8.21)

These relations can also be reexpressed in terms of K in the coordinates of Fig. 4.8-1 by using (4.8.18). The results are

$$R_{x}^{\prime\prime 2} + \left[(K_{xx}^2 - K_{xx}K_{zz} + K_{xy}^2) \sin^2 \theta / K_{xy}K_{zz} \cos \theta \right] R_x^{\prime\prime} + 1 = 0 \quad (4.8.22)$$

$$Q = \frac{(-K_{xy}R''_{x} + K_{xx}\cos\theta - K_{zz}\cos\theta)\sin\theta}{(K_{xx}\sin^{2}\theta + K_{zz}\cos^{2}\theta)R''_{x}}$$
(4.8.23)

We note that the two roots of (4.8.22) satisfy the relations

$$R_{x_1}^{\prime\prime} R_{x_2}^{\prime\prime} = 1 \tag{4.8.24a}$$

and

$$R_{x_1}^{\prime\prime} + R_{x_2}^{\prime\prime} = -(K_{xx}^2 - K_{xx}K_{zz} + K_{xy}^2)\sin^2\theta/K_{xy}K_{zz}\cos\theta \qquad (4.8.24b)$$

164

Therefore, the two characteristic polarization *E*-ellipses when projected on the wavefront must be mirror images about a line making a 45° angle with respect to the x''-axis and they are counterrotating. If the medium is lossless, K_{xx} , K_{zz} are purely real and K_{xy} is purely imaginary, making the righthand side of (4.8.24b) and also R''_{x_1} and R''_{x_2} purely imaginary. Thus the polarization ellipses have major axes aligned either along the x''- or y''-axis in a lossless medium. For the special case of propagation parallel to \mathbf{B}_0 (i.e., $\theta = 0^\circ$), the middle coefficient of (4.8.22) and Q both vanish; both characteristic waves are purely transverse and circularly polarized. For the special case of perpendicular propagation (i.e., $\theta = 90^\circ$), the middle coefficient of (4.8.22) approaches to infinity and the two roots of (4.8.22) are zero which corresponds to a wave linearly polarized along the y''-axis and infinity which corresponds to a wave polarized in the x''y''-plane. These special cases are discussed more extensively in the next few sections.

4.9 Propagation Parallel to Steady Magnetic Field

The first special case we shall discuss is that when the propagation vector is parallel to \mathbf{B}_0 . Setting $\theta = 0^\circ$ in (4.8.6), one of the following conditions must be satisfied,

$$K_0 = 1 - \sum_{\alpha} X_{\alpha} = 0 \tag{4.9.1}$$

$$n_L^2 = K_I = K_{xx} + jK_{xy}$$

= $1 - \sum_{\alpha} \frac{X_{\alpha}}{1 + Y_{\alpha}} = 1 + \sum_{\alpha} \frac{\omega_{p\alpha}^2}{\omega_{B\alpha}(\omega_{B\alpha} + \omega)}$ (4.9.2a)

or

$$n_R^2 = K_{II} = K_{xx} - jK_{xy}$$

= $1 - \sum_{\alpha} \frac{X_{\alpha}}{1 - Y_{\alpha}} = 1 + \sum_{\alpha} \frac{\omega_{p\alpha}^2}{\omega_{B\alpha}(\omega_{B\alpha} - \omega)}$ (4.9.2b)

The first equation (4.9.1) corresponds to longitudinal oscillations and has been discussed in Section 4.7. The remaining two equations, (4.9.2a) and (4.9.2b), are the desired dispersion relations. They give the refractive index for each characteristic mode of the medium. The corresponding polarizations are given by the ratios of (4.8.10) which reduces, in the present case, to

$$E_x: E_y: E_z = -K_{zz}(n^2 - K_{xx}): K_{zz}K_{xy}: 0$$
(4.9.3)
If $K_{xy} \neq 0$, substitution of (4.9.2a) and (4.9.2b) in (4.9.3) gives

$$E_x: E_y: E_z = \mp j: 1:0$$
 (4.9.4)

Both characteristic waves are therefore transverse and circularly polarized. The upper sign of (4.9.4) is associated with (4.9.2a) and corresponds to a left-handed polarization while the lower sign of (4.9.4) is associated with (4.9.2b) and corresponds to a right-handed polarization. The subscripts L and R in (4.9.2a) and (4.9.2b) are used to denote left and right polarization, respectively. If $K_{xy} = 0$, we then deduce from (4.9.2) and (4.9.3) the relations

$$n_1^2 = n_2^2 = K_{xx}$$

$$E_x : E_y : E_z = 0:0:0$$
 if $K_{xy} = 0$ (4.9.5)

i.e., both refractive indices are the same and the polarization of the wave is indeterminant. For this rather special case, we may deduce from (4.8.6) that one of the two modes actually becomes isotropic. The frequency at which $K_{xy} = 0$ is called a cross-over frequency. A wave propagating in a slowly varying medium can change its sense of polarization when the wave frequency becomes equal to the cross-over frequency. This interesting observation has been used as a diagnostic tool in measuring relative abundance of heavy ions in ionospheric research. We shall have more to say about its use and implication in Section 4.11.

Comparison of refractive indices (4.9.2a) and (4.9.2b) with (4.1.8) for the isotropic case shows a marked difference in behavior. While the wave is propagating above the plasma frequency but not below in the isotropic case, the effect of a steady magnetic field opens up a series of "windows" in frequency bands in which propagation is possible. The number of windows for a left circularly polarized wave is in general equal to $1 + N_+$ where N_+ is the number of species of positively charged particles, and the number of windows for a right circularly polarized wave is equal to $1 + N_-$ where N_- is the number of species of negatively charged particles. Examination of (4.9.2a) and (4.9.2b) shows that propagation is possible even in the limit of $\omega \to 0$. When $\omega \ll |\omega_{B_A}|$ all α 's, both refractive indices are equal and given by

$$n_L^2 = n_R^2 = 1 + \sum_{\alpha} \left(\omega_{p\alpha}^2 / \omega_{B\alpha}^2 \right) = 1 + \left(\varrho / \varepsilon_0 B_0^2 \right)$$
(4.9.6)

where the mass density is given by

$$\varrho = \sum_{\alpha} N_{\alpha} m_{\alpha} \tag{4.9.7}$$

The meaning of the second term on the right-hand side of (4.9.6) can be clarified by using some concepts used in magnetohydrodynamics. The study of magnetohydrodynamics is concerned with the behavior of a conducting fluid in an electromagnetic field. When the fluid is infinitely conducting, the magnetic field lines are found frozen into the fluid, i.e., fluid particles comove with magnetic field lines. A magnetic tube may therefore be thought of as possessing a mass of linear density equal to the mass density of the fluid. Analogous to the case of a vibrating string, when the magnetic field is perturbed, it tends to oscillate with a phase velocity equal to (tension/density)^{1/2}. The tension associated with the magnetic field is B_0^2/μ_0 with an associated hydrostatic pressure $B_0^2/2\mu_0$ as given by the Maxwell's stress tensor. Putting the tension and density in the expression, we obtain the Alfvén velocity as

$$\nu_A = \left(\frac{\text{tension}}{\text{density}}\right)^{1/2} = \left(\frac{B_0^2}{\mu_0 \varrho}\right)^{1/2}$$
(4.9.8)

The use of Alfvén velocity reduces (4.9.6) to

$$n_L^2 = n_R^2 = 1 + c^2 / v_A^2, \quad \omega \ll |\omega_{B\alpha}|$$
 (4.9.9)

In most cases the Alfvén velocity is much smaller than the velocity of light in free space and thus the unity in (4.9.9) can be ignored. Actually the term unity has its origin in the displacement current and the term c^2/v_A^2 in the polarization current. For slow time variations the displacement current can be ignored as is usually done in magnetohydrodynamics. The ratio c^2/v_A^2 can also be reinterpreted as

$$\frac{c^2}{v_A^2} = \frac{1}{2} \left(\frac{\varrho c^2}{B_0 H_0/2} \right) = \frac{1}{2} \frac{\text{rest energy}}{\text{magnetic energy}}$$
(4.9.10)

Except in extremely tenuous plasmas and/or exceptionally strong magnetic fields, the rest energy is usually much larger than the magnetic energy. All these equivalent interpretations allow us to write (4.9.9) in most cases, as

$$n_L^2 = n_R^2 = c^2 / v_A^2 v_p = v_A \qquad \omega \ll |\omega_{B\alpha}|$$
(4.9.11)

The phase velocity of the low frequency waves is then just equal to the Alfvén velocity.

As frequency increases a series of resonances occur at gyrofrequencies of each species of ions. The total number of resonant frequencies is equal to the number of species of charged particles. These resonances are called gyroresonances or cyclotron resonances. The gyrofrequency of positively charged particles is negative and hence contributes to poles of n_L^2 , corresponding to a left-hand polarized wave. Remember from Fig. 4.5-1 that the positively charged particles themselves are also rotating in a left-handed sense. Therefore, at resonance positively charged particles see a constant electric field in their own coordinate frame and considerable acceleration and transfer of energy between the wave and the resonant particles may result. The situation is not unlike that of Landau damping discussed in Section 3.8. Similarly at the gyrofrequency of negatively charged particles, a right circularly polarized wave interacts and exchanges energy with negative resonant particles. The study of resonant interaction requires the use of distribution functions. The resulting damping of the wave is called the cyclotron damping.

In a gaseous plasma, electrons are the lightest charged particles and hence have the highest gyrofrequency, being 1836 times higher than that of protons, the next lightest charged particles. It is therefore possible to find a frequency range in which ionic contributions can be ignored all together. In a twocomponent plasma consisting of electrons and singly charged neutralizing positive ions, the ionic effect can be neglected if

$$\left|\frac{X_e}{1\pm Y_e}\right| \gg \left|\frac{X_i}{1\pm Y_i}\right| \tag{4.9.12}$$

Let us assume that the required condition is

$$|Y_i| \ll 1$$
 or $\omega \gg |\omega_{Bi}|$ (4.9.13)

which must be checked for consistency as done in the following. The use of the inequality (4.9.13) reduces (4.9.12) to

$$\left| \frac{m_i}{1 \pm Y_e} \right| \gg m_e$$

which is equivalent to (4.9.13). Therefore, in a two-component plasma the ionic effects can be ignored if the frequency is much larger than the ionic gyrofrequency. Similar conclusion can be reached in a general multicomponent plasma as long as the electron concentration is of the same order as the ionic concentration. When this is true, the dispersion relations (4.9.2a) and (4.9.2b) reduce, respectively, to

$$n_L^2 = 1 - X/(1 + Y) \tag{4.9.14a}$$

 $\omega \gg |\omega_{Pi}|$

$$n_R^2 = 1 - X/(1 - Y)$$
 (4.9.14b)

where the subscripts e on X and Y are omitted for simplicity because only electronic quantities remain and thus there is no danger for confusion. These refractive index expressions indicate that in the high frequency region there is only one resonance at $\omega = \omega_B$ for the right circularly polarized wave and there is a cutoff frequency for each of the two modes at

$$\omega = \pm (\omega_B/2) + (\omega_p^2 + (\omega_B/2)^2)^{1/2}$$
(4.9.15)

The upper sign applies to the left circular wave and the lower sign to the right circular wave. These expressions are identical to those given by (4.7.12) and (4.7.16) as the cutoff condition is identical to the condition for longitudinal oscillations.

When the wave frequency is much larger than the electronic gyrofrequency and the electronic plasma frequency the refractive index of both waves approaches the free space value of 1.

To illustrate the behavior of the refractive index dependence on wave frequency a concrete example is considered. Figure 4.9-1 shows n^2 behavior for a three-component plasma of electrons neutralized by 60% protons and 40% singly charged oxygen ions. The refractive index is very large and equal for both characteristic waves in the extremely low frequency region. As frequency increases, the left circularly polarized wave goes through resonances at the oxygen gyrofrequency $|\omega_{B1}|$ and the proton gyrofre-



Fig. 4.9-1. Behavior of n^2 as a function of f when propagating parallel to the steady magnetic field in a three-component plasma of electrons neutralized by 60% protons and 40% atomic oxygen ions.

quency $|\omega_{B_2}|$ while the right circularly polarized wave has a resonance at the electron gyrofrequency ω_{Be} . There are two cutoffs at which the refractive index vanishes for the left circular wave and one cutoff for the right circular wave. These resonances and cutoffs divide the frequency spectrum into a series of windows within which propagation is possible. Inspection of Fig. 4.9-1 shows that there are three windows for the left circular wave and two for the right circular wave. When the frequency is very large the refractive index of both waves approaches to 1. Figure 4.9-1 also shows that there is a cross-over frequency ω_{co} at which both refractive indices have the same value.

4.10 Faraday Effect

It was discovered by Faraday in 1845 that certain substances become optically "active" when placed in a magnetic field parallel to the direction of propagation of the light wave. By optically active is meant that the plane of polarization undergoes a rotation that is found to be proportional to the strength of the magnetic field and the distance traversed. This phenomenon of the rotation of the polarization is known as the Faraday effect or Faraday rotation. In a magnetoplasma a similar phenomenon also occurs and is also known as the Faraday effect.

It is known that the resultant of two equiamplitude, oppositely rotating, circularly polarized waves propagating in the same direction is a linearly polarized wave. The plane of polarization depends on the phase relationship between the two circular waves. As shown in Section 4.9, both characteristic waves are circularly polarized when propagating parallel to the steady magnetic field. Let us assume that both waves are propagating in the z-direction with equal amplitude and at z = 0 they are in phase. The resultant is then a linearly polarized wave with an amplitude twice of that of each circular wave and with the plane of polarization oriented as shown on the top of Fig. 4.10-1. Next we examine what happens as the wave propagates a small distance Δz . We have seen in Section 4.9 that the refractive indices for characteristic waves are in general different. This means that these waves will be no longer in phase even if they are initially at z = 0. The phase shift for the left circular wave is $-k_0 n_L \Delta z$ and that for the right circular wave is $-k_0 n_R \Delta z$. The resultant of these two equiamplitude circular waves is still linearly polarized but the plane of polarization has been rotated through an angle $\Delta \Omega$ given by

$$\Delta \Omega = (k_0/2)(n_L - n_R) \,\Delta z \tag{4.10.1}$$



Fig. 4.10-1. Illustration of the rotation of the plane of polarization of the resultant wave as it propagates from z = 0 to $z = \Delta z$ along the magnetic field.

where a positive sign indicates a rotation in the right-handed sense. In a homogeneous medium we may integrate (4.10.1) to obtain the total angle of polarization rotation for a wave that has traveled a distance z,

$$\Omega = (k_0/2)(n_L - n_R)z \quad \text{rad}$$
 (4.10.2)

For concreteness we have shown in Fig. 4.10-1 the case in which $n_L > n_R$ so that the left circular wave undergoes a larger phase shift than the right circular wave. This results in a polarization rotation that twists like a right-handed screw as the wave propagates. The opposite case $n_L < n_R$ is also possible and the twist is in the left-hand sense. The regions in the frequency spectrum in which the respective case applies can be found from plots such as Fig. 4.9-1.

The proof of Faraday rotation can also be approached mathematically. The equiamplitude left and right circularly polarized waves are given by

$$\mathbf{E}_{L} = E_{0}[\hat{x}\cos(\omega t - k_{0}n_{L}z) + \hat{y}\cos(\omega t - k_{0}n_{L}z + \pi/2)] \quad (4.10.3a)$$

$$\mathbf{E}_{R} = E_{0}[\hat{x}\cos(\omega t - k_{0}n_{R}z) + \hat{y}\cos(\omega t - k_{0}n_{R}z - \pi/2)] \quad (4.10.3b)$$

where the waves are assumed to be in phase at z = 0. The resultant electric field is obtained by adding E_L and E_R , giving

$$\begin{split} \mathbf{E} &= \mathbf{E}_{L} + \mathbf{E}_{R} \\ &= 2E_{0}\{\hat{x}\cos[k_{0}(n_{R} - n_{L})z/2] + \hat{y}\cos[k_{0}(n_{R} - n_{L})z/2 + \pi/2]\} \\ &\times \cos[\omega t - k_{0}(n_{L} + n_{R})z/2] \end{split}$$
(4.10.4)

The expression (4.10.4) shows that the x- and y-components of the wave are in time phase and thus linearly polarized. The resultant electric vector makes an angle Ω with respect to the x-axis and

$$\Omega = \tan^{-1} \frac{E_y}{E_x} = \tan^{-1} \frac{\cos[k_0(n_R - n_L)z/2 + \pi/2]}{\cos[k_0(n_R - n_L)z/2]}$$

or

$$\Omega = -(k_0/2)(n_R - n_L)z \quad \text{rad}$$
 (4.10.5)

in an agreement with (4.10.2). The absolute value of the resultant wave is, from (4.10.4),

$$|E| = (E_x^2 + E_y^2)^{1/2}$$

= 2E₀ cos[$\omega t - k_0(n_R + n_L)z/2$] (4.10.6)

which shows a phase shift in an equivalent medium with the average refractive index $(n_R + n_L)/2$.

There are several interesting properties connected with Faraday effect that ought to be elucidated. First is that the polarization rotation is cumulative. Therefore, even in the case of a very weak steady magnetic field the twist of the electric vector may become appreciable provided the wave has traveled a sufficiently long distance. Next we note that when the wave travels parallel to \mathbf{B}_0 , the twist of the electric vector is right-handed if $n_R < n_L$ and left-handed if $n_R > n_L$. On the other hand, if the wave travels antiparallel to \mathbf{B}_0 , the subscripts L and R in (4.9.2) must be interchanged. This results in a reversal of twist on reversing the direction of propagation. For a wave making a round trip in the medium the rotation on the return trip is in the same direction as that in the first trip, making the total rotation twice that of a one-way trip.

Because of the sensitivity of Faraday rotation on the presence of electrons, the technique has been used in ionospheric research. Most experiments choose a frequency much higher than the electron plasma frequency and electron gyrofrequency so that we may approximate (4.9.14) by a binomial expansion,

$$n_{L,R} = [1 - X/(1 \pm Y)]^{1/2} \cong 1 - X(1 \mp Y)/2 \qquad (4.10.7)$$

Substituting (4.10.7), we obtain the Faraday rotation

$$\Omega = (k_0 X Y/2)z$$

= $\omega_p^2 \omega_B z/2c\omega^2$ rad (4.10.8)

The rotation is proportional to electron density, magnetic field intensity, and distance of travel and is inversely proportional to the square of frequency. The positive sign indicates a right-handed twist. Numerically, (4.10.8) reduces to

$$\Omega = 2.97 \times 10^{-2} N H_0 z / f^2 \quad \text{rad} \tag{4.10.9}$$

where all quantities are expressed in the mks system. We note from the approximate expression (4.10.7) that the average refractive index is just 1 - X/2 which is equal to the refractive index of the medium in the absence of the steady magnetic field. Therefore, we may describe propagation of a linearly polarized wave as a right-handed twist in polarization given by (4.10.8) combined with a phase shift equal to the equivalent isotropic case.

4.11 Electron and Ion Whistlers

We have seen that one of the effects of a steady magnetic field in the plasma is that it opens up a series of windows in the frequency spectrum, in which one or both characteristic waves may propagate. These waves, especially in low frequency region, may be very slow and dispersive. This means that the frequency components of an initial impulse may spread out in arrival time after propagating through the medium. When such a signal is detected by a radio it produces a whistling tone and therefore is called a whistler.

For frequencies much larger than all ionic gyrofrequencies, the effect of ions can be ignored when compared with electrons. The refractive index of a right circularly polarized wave is given by

$$n_R^2 = 1 - X/(1 - Y)$$

= 1 + f_p^2/f(f_B - f) (4.11.1)

The group velocity of the wave is given by

$$v_{a} = d\omega/dk = c/(n + fdn/df) \qquad (4.11.2)$$

Substitute (4.11.1) in (4.11.2) and carry out the differentiation. The following expression is obtained.

$$\nu_g = \frac{cn}{1 + f_p^2 f_B / 2f(f_B - f)^2}$$
(4.11.3)

As is usually the case in ionospheric applications, the wave frequency is

much lower than the plasma frequency so that the inequalities $f_p^2 \gg f f_B$, $f_p \gg f$ are valid. These inequalities reduce the expression (4.11.3) for group velocity to

$$v_g = \frac{2c}{f_p f_B} \left[f(f_B - f)^3 \right]^{1/2}$$
(4.11.4)

The corresponding phase velocity is

$$v_p = (c/f_p)[f(f_B - f)]^{1/2}$$
(4.11.5)

The ratio of group velocity to phase velocity can be found to be

$$v_g/v_p = 2(1 - f/f_B) \tag{4.11.6}$$

which is a linear function of f/f_B . In the frequency range $0 < f < f_B/2$, the group velocity is greater than the phase velocity; and in the frequency range $f_B/2 < f < f_B$, the phase velocity is greater. It is interesting to note that the group velocity expression (4.11.4) has a maximum when $f = f_B/4$ at which the group velocity is given by

$$\nu_{g\max} = \frac{3\sqrt{3}}{8} \frac{f_B}{f_p} c$$
 (4.11.7)

In a uniform plasma the signal at a frequency of one-fourth the gyrofrequency suffers a minimum time delay. Therefore, if the transmitting source is an impulse, the received spectrum as a function of time delay has the form shown in Fig. 4.11-1. The frequency that gives the minimum delay is called the nose frequency. In a uniform plasma, the nose frequency is equal to $f_B/4$.

The expression (4.11.4) simplifies to

$$v_g \simeq (2c/f_p)(ff_B)^{1/2}$$
 (4.11.8)



Fig. 4.11-1. Sketch showing group delay of a nose whistler.

when $f \ll f_B$. The corresponding time delay is

$$T = \int_{\text{path}} \frac{dp}{v_g} = \frac{1}{2c\sqrt{f}} \int_{\text{path}} \frac{f_p}{f_B^{1/2}} dp$$
(4.11.9)

which predicts a $1/\sqrt{f}$ dependence.

The impulses that generate naturally occurring whistlers are ordinary lightning discharges. The history of the discovery of whistlers is very interesting. This history and the methods used by ionospheric researchers in obtaining electron density information in the magnetosphere can be found in the comprehensive book of Helliwell (1965).

As the wave frequency is decreased the ions may become important so that instead of (4.11.1) we must go back to (4.9.2a) and (4.9.2b) (also, see discussion in Section 4.13). Of particular interest is the occurrence of cross-over frequencies. At the cross-over frequency the wave polarizations of both characteristic modes are indeterminant. A characteristic wave propagating in a slowly varying medium may couple energy into the second characteristic wave when the wave frequency becomes equal to the cross-



Fig. 4.11-2. Electron and proton whistlers observed on a satellite at 2950 km. The calculated fractional abundance of H^+ is 0.67 and O^+ is 0.33. [After McEwen and Barrington (1968). By permission of North Holland Publishing Company.]

over frequency. For example, in the ionosphere it has been observed that a right-handed circularly polarized electron whistler may couple part of the energy into the left-handed circularly polarized ion whistler (Gurnett *et al.*, 1965). An example is shown in Fig. 4.11-2. The condition for crossover is $K_{xy} = 0$ or

$$\sum_{\alpha} \frac{X_{\alpha} Y_{\alpha}}{1 - Y_{\alpha}^{2}} = 0$$
 (4.11.10)

Let us consider a three-component plasma of electrons and two kinds of singly charged positive ions of fractional ionization A_1 and A_2 with $A_1 + A_2 = 1$. Since the cross-over frequency is near the ion gyrofrequency, we may let $Y_e \gg 1$ and (4.11.10) reduces to

$$\frac{1}{Y_e^2} = \frac{A_1}{Y_e^2 - (m_{i_1}/m_e)^2} + \frac{A_2}{Y_e^2 - (m_{i_2}/m_e)^2}$$

Multiplying out all the factors, we obtain the cross-over frequency

$$f_{\rm co}^2 = f_{Bi_2}^2 A_1 + f_{Bi_1}^2 A_2 \tag{4.11.11}$$

It is clear from (4.11.11) that the measurement of cross-over frequency can determine the fractional abundance of positive ions.

If there are more kinds of positive ions, a cross-over frequency occurs between each ionic resonance. For example, in a three-ion plasma, there are two cross-over frequencies. Measurements of these cross-over frequencies can give the fractional abundance of all positive ions in the plasma.

4.12 Propagation Perpendicular to Steady Magnetic Field

The second special case we shall discuss is that when the propagation vector is perpendicular to \mathbf{B}_{0} . Setting $\theta = \pi/2$ reduces (4.8.6) to

$$n_0^2 = K_0 = 1 - \sum_{\alpha} X_{\alpha}$$
(4.12.1)
$$n_x^2 = \frac{2K_{\rm I}K_{\rm II}}{K_{\rm I} + K_{\rm II}} = \frac{[1 - \sum_{\alpha} X_{\alpha}/(1 - Y_{\alpha})][1 - \sum_{\alpha} X_{\alpha}/(1 + Y_{\alpha})]}{1 - \sum_{\alpha} X_{\alpha}/(1 - Y_{\alpha}^2)}$$
(4.12.2)

The use of (4.8.10) to find the polarization corresponding to the wave with refractive index n_0 leads to indeterminacy and hence we must go back to (4.8.2). In the coordinates of Fig. 4.8-1 with $\theta = \pi/2$, we find that the electric field vector is entirely along z-axis, i.e., parallel to \mathbf{B}_0 . The wave is purely transverse. Since the electric field is parallel to \mathbf{B}_0 , the induced motion is not affected by the presence of a steady magnetic field. Consequently, the refractive index expression (4.12.1) is identical to that derived in the absence of a steady magnetic. Therefore, this characteristic wave is called the ordinary wave and a subscript 0 is used to designate its refractive index. The remaining wave with the refractive index given by (4.12.2) is definitely affected by the steady magnetic field and is called the extraordinary wave. We use a subscript x to denote it. The corresponding polarization is given by

$$E_x: E_y: E_z = -j(K_{\rm I} - K_{\rm II}): (K_{\rm I} + K_{\rm II}): 0 \qquad (4.12.3)$$

which can be obtained from (4.8.10). The coordinate system is that given by Fig. 4.8-1 with $\theta = \pi/2$, i.e., **k** along the x-axis, **B**₀ along the z-axis. The extraordinary wave (4.12.2) is thus in general not transverse. It is polarized elliptically in a plane perpendicular to **B**₀. It is interesting to note that the wave is purely longitudinal when $K_{\rm I} + K_{\rm II} = 0$ for which the resonance for n_x occurs. As $\omega \to 0$, n_0^2 is large and negative, showing the evanescent nature of the ordinary wave. The refractive index of the extraordinary wave reduces to, as $\omega \to 0$,

$$n_x^2 = 1 + c^2 / v_A^2 \tag{4.12.4}$$

where v_A is the Alfvén velocity.

The ordinary wave has just one cutoff at $\omega = (\sum_{\alpha} \omega_{\alpha}^2)^{1/2}$ while the extraordinary wave has all the cutoffs of n_L and n_R discussed in Section 4.9. The ordinary wave has a single resonance at $\omega = 0$. The resonant frequencies of the extraordinary wave are given by the condition $K_{\rm I} + K_{\rm II} = 0$ or

$$1 + \sum_{\alpha} \frac{\omega_{p_{\alpha}}^{2}}{\omega_{B_{\alpha}}^{2} - \omega^{2}} = 0$$
 (4.12.5)

which shows that there are as many positive resonant frequencies as number of species of charged particles.

For concreteness let us consider a two-component plasma of electrons neutralized by singly charged positive ions. The resonance frequencies are given by the equation

$$1 + \frac{\omega_{p_e}^2}{\omega_{B_e}^2 - \omega^2} + \frac{\omega_{p_i}^2}{\omega_{B_i}^2 - \omega^2} = 0$$
 (4.12.6)

This equation gives two resonant frequencies one of which is near the electron characteristic frequencies. For the high resonant frequency we may assume $\omega^2 \gg \omega_{Bi}^2$, ω_{pi}^2 and thus (4.12.6) immediately yields a solution

$$\omega_{\rm uh}^2 = \omega_{Be}^2 + \omega_{pe}^2 \tag{4.12.7}$$

This frequency is known as the upper hybrid resonant frequency and, as seen from (4.12.7), is independent of the ionic gyrofrequency and ionic

plasma frequency. Because of the smallness of the electron mass, the electrons contribute almost all the polarizability of the medium at the upper hybrid frequency. The remaining resonant frequency can be obtained by multiplying out (4.12.6) and solving it by assuming the smallness of the electron mass, obtaining

$$\omega_{\rm lh}^2 = |\omega_{Bi}| \omega_{Be} (|\omega_{Bi}| \omega_{Be} + \omega_{pe}^2) / (\omega_{Be}^2 + \omega_{pe}^2)$$
(4.12.8)

This frequency is known as the lower hybrid resonant frequency. In the low density limit the lower hybrid approaches the ionic gyrofrequency. In the high density limit, the lower hybrid becomes the geometric mean of the electron and ionic gyrofrequency,

$$\omega_{\rm lh}^2 = |\omega_{Bi}| \omega_{Be}, \qquad \omega_{pe}^2 \gg \omega_{Be}^2 \tag{4.12.9}$$

At the lower hybrid resonance frequency, $E_y = E_z = 0$ as seen from (4.12.3), i.e., the electric field vector is parallel to **k**. The corresponding polarization density is given by $\mathbf{P}_{\alpha} = e_0 \mathbf{\chi}_{\alpha} \cdot \mathbf{E}$ with susceptibility tensor given by (4.5.11). Since at the lower hybrid, $X_i = X_e(m_e/m_i)$, $Y_i = -(m_e/m_i)^{1/2}$, $Y_e = (m_i/m_e)^{1/2}$, the polarization densities of electrons and ions are, respectively,

$$\mathbf{P}_{e} = \varepsilon_{0} X_{e} (m_{e}/m_{i}) [\hat{x} - j\hat{y}(m_{i}/m_{e})^{1/2}] E_{x}
\mathbf{P}_{i} = -\varepsilon_{0} X_{e} (m_{e}/m_{i}) [\hat{x} + j\hat{y}(m_{e}/m_{i})^{1/2}] E_{x}$$
(4.12.10)

We note that in the direction of the electric field, the displacements of the electron cloud and ionic cloud are in phase at the lower hybrid resonance so as to contribute vanishing total polarization density. In a direction perpendicular to both the electric field and the steady magnetic field, electron and ion clouds are oscillating out of phase, giving rise to space charge effect.

The behavior of n^2 as a function of frequency in a three-component plasma identical to that of Fig. 4.9-1 is depicted in Fig. 4.12-1. It shows the presence of one upper hybrid and two lower hybrids. At the lowest hybrid resonance the electrons remain relatively motionless while the oscillations of two ion clouds perpendicular to the steady magnetic field is 180° out of phase with each other (see problem at the end of this chapter).

It is interesting to mention the topside sounding results in which some of these resonances have been observed. An ionospheric sounder, or ionosonde for short, is essentially a radar which measures the time delay of returned echos at either a fixed frequency or as the frequency is swept over a wide frequency band. Ground based sounders throughout the world have been probing the ionosphere since 1926. More recently the sounder has been placed on the satellite to sound the ionosphere from above. The purpose is



Fig. 4.12-1. Behavior of n^2 as a function of f for the extraordinary wave when propagating perpendicular to \mathbf{B}_0 . The plasma is identical to that of Fig. 4.9-1. The resonance occurs at the upper hybrid frequency f_{uh} and two lower hybrid frequencies f_{lh_1} and f_{lh_2} .

to obtain an ionization profile above the peak of the layer. The data of sweptfrequency sounders are usually presented in the form of time delay or apparent range as a function of frequency. Such data are commonly called ionograms. Immediately after the launch of the satellite-borne sounder, it was observed that many ionograms show vertical spikes at some discrete frequencies. At these frequencies the energy transmitted by the sounder is stored in the plasma in the form of a stationary disturbance or as waves with group velocity comparable to that of the satellite. These spikes have been observed at electron gyrofrequency and its harmonics, plasma frequency, upper and lower hybrid frequencies, and many other cutoffs and resonances. Note that the upper hybrid resonance gives a direct measurement of electron density when the electron gyrofrequency is known. Some of the satellite results have been summarized by Chapman and Warren (1968).

4.13 Hydromagnetic Waves-Low Frequency Approximation

In the low frequency limit, i.e., $\omega \ll |\omega_{B\alpha}|$, $\omega_{p\alpha}$ all α , we may expand K_{I} and K_{II} given by (4.7.3b) and (4.7.3c) to obtain

$$K_{\rm I} = 1 + \sum_{\alpha} \frac{\omega_{p\alpha}^2}{\omega_{B\alpha}(\omega_{B\alpha} + \omega)} = 1 + \frac{c^2}{v_A{}^2} - \omega \sum_{\alpha} \frac{\omega_{p\alpha}^2}{\omega_{B\alpha}^2} + O(\omega^2) \quad (4.13.1a)$$

$$K_{\rm II} = 1 + \sum_{\alpha} \frac{\omega_{p\alpha}^2}{\omega_{B\alpha}(\omega_{B\alpha} - \omega)} = 1 + \frac{c^2}{\nu_A^2} + \omega \sum_{\alpha} \frac{\omega_{p\alpha}^2}{\omega_{B\alpha}^3} + O(\omega^2) \quad (4.13.1b)$$

The harmonic mean of K_{I} and K_{II} accurate to the first order in ω is therefore just

$$\frac{2K_{\rm I}K_{\rm II}}{K_{\rm I}+K_{\rm II}} = 1 + c^2/\nu_A^2 + O(\omega^2)$$
(4.13.2)

When these relations are substituted, the dispersion relation (4.8.6) reduces to

$$[n^{2} - (1 + c^{2}/\nu_{A}^{2})] \times \left\{ (n^{2} - K_{0}) \sin^{2}\theta + \frac{K_{0}}{1 + c^{2}/\nu_{A}^{2}} [n^{2} - (1 + c^{2}/\nu_{A}^{2})] \cos^{2}\theta \right\} = 0$$
(4.13.3)

which yields two roots for refractive index. They are given by

$$n_c^2 = 1 + c^2 / v_A^2 \tag{4.13.4}$$

and

$$n_t^2 = \frac{K_0}{\sin^2 \theta + K_0 \cos^2 \theta / (1 + c^2 / \nu_A^2)}$$
(4.13.5)

We note that the mode with refractive index given by n_c is isotropic since it does not depend on the angle θ . The mode corresponding to n_t is anisotropic. The resonance angle at which the refractive index approaches infinity is given by the relation

$$\tan \theta_r = \left(\frac{-K_0}{1+c^2/\nu_A^2}\right)^{1/2} \approx \frac{\omega_{pe}\nu_A}{\omega c}$$
(4.13.6)

For $\theta \ll \theta_r$, we may approximate (4.13.5) by

$$n_t^2 \approx (1 + c^2/v_A^2)/\cos^2\theta$$
 (4.13.7)

In most cases θ_r departs only slightly from $\pi/2$; the approximate expression (4.13.7) is valid for nearly all angles except for a small cone near the exact perpendicular condition. The refractive index surface given by (4.13.7) is a plane surface perpendicular to \mathbf{B}_0 as sketched in Fig. 4.13-1. The group



Fig. 4.13-1. Sketch showing $n_t(\theta)$. The refractive index surface is obtained by revolving the solid line about \mathbf{B}_0 . The symmetric lower portion of the curve is not shown.

180

velocity is known to be perpendicular to the refractive index. Waves with the refractive index terminated on the plane surface portion of the refractive index surface have energy all channeled to the same direction, namely, parallel to B_0 .

The characteristic wave polarization of these waves is given by the ratios of (4.8.10) in the coordinates of Fig. 4.8-1. In the low frequency approximation

$$K_{xx} = 1 + c^2 / v_A^2, \qquad |K_{xy}| = |j\omega \Sigma \omega_{p\alpha}^2 / \omega_{B\alpha}^3| \ll K_{xx}$$

For $\theta \neq 0$, we deduce

$$E_x: E_y: E_z \approx \begin{cases} 0: E_y: 0 & \text{corresponding to mode } n_c \\ E_x: 0: 0 & \text{corresponding to mode } n_t \end{cases}$$
(4.13.8)

Hence the electric field for these two modes are approximately linearly polarized for θ not near 0°. The induced fluid velocity is related to the electric field by the equation of motion (4.5.1) which in the low frequency approximation reduces to

$$\mathbf{E} + \mathbf{v}_{\alpha} \times \mathbf{B}_0 = 0 \tag{4.13.9}$$

With the electric field given by (4.13.8), the fluid velocity, the electric field, and the steady magnetic field are mutually orthogonal in the present approximation. The velocity corresponding to mode n_e is

$$\mathbf{v}_{\alpha} \propto \hat{x} \, e^{j(\omega t - k_x x - k_z x)} \tag{4.13.10}$$

while that corresponding to mode n_t is

$$\mathbf{v}_{\alpha} \propto \hat{\mathbf{y}} e^{j(\omega t - k_x x - k_z z)} \tag{4.13.11}$$

The velocity given by (4.13.10) is compressional and the mode is referred to as a compressional mode which accounts for the subscript c on the refractive index. The velocity given by (4.13.11) has a zero divergence and the corresponding mode is referred to as the torsional or shear mode. A subscript t is used to denote it. We note that the compressional mode is isotropic and it reduces to the extraordinary wave when $\theta = \pi/2$. The anisotropic mode (4.13.5) becomes the nonpropagation ordinary wave when $\theta = \pi/2$. When $\theta = 0$, both modes are circularly polarized. The transition from circular polarization to linear polarization as θ departs from 0 can be investigated by using a better approximation.

The dispersion relations (4.13.4) and (4.13.5) are essentially zero frequency limits of the general dispersion relation. As the frequency is raised the

approximations (4.13.1a) and (4.13.1b) are no longer valid especially when the frequency is approaching the ionic gyrofrequency. In the neighborhood of ion gyrofrequency, ions are expected to play an important role. Hence these waves are called ion whistlers or ion cyclotron waves. In a moderately dense magnetoplasma in which the electron plasma frequency and electron gyrofrequency have the same order of magnitude, K_0 is expected to be very large for ion whistlers because of the large ion-to-electron mass ratio, namely,

$$|K_0| = |1 - \sum_{\alpha} X_{\alpha}| \simeq X_e \gg 1$$

We may also assume $|K_0| \gg |K_I|$, $|K_{II}|$ if the frequency is not exactly equal to the ion gyrofrequency. These approximations reduce the dispersion relation (4.8.6) to

$$n^4 \cos^2 \theta - n^2 (1 + \cos^2 \theta) (K_{\rm I} + K_{\rm II})/2 + K_{\rm I} K_{\rm II} = 0 \qquad (4.13.12)$$

if θ is not too close to $\pi/2$. In the neighborhood of any ionic gyrofrequency $|K_{I}|$ or $|K_{II}|$ may be very large, depending on the sign of the charge. For definiteness, let us assume the ion is positive and hence $|K_{I}| \gg |K_{II}|$. The approximate roots of (4.13.12) are then

$$n_c^2 = 2K_{\rm II}/(1 + \cos^2\theta) \tag{4.13.13}$$

$$n_{iw}^2 = K_{\rm I}(1 + \cos^2\theta)/2\cos^2\theta \qquad (4.13.14)$$

These refractive indices are dependent on $K_{\rm I}$ and $K_{\rm II}$ whose properties have been discussed in Section 4.9 and on θ . The refractive index n_e given by (4.13.13) is for the compressional wave in the neighborhood of an ion gyrofrequency of positive ion. It is now anisotropic by its dependence on θ . The compressional wave becomes a wave with right-handed circular polarization when $\theta = 0$. The wave satisfying the dispersion relation (4.13.14) is called the ion whistler wave or ion cyclotron wave. In the neighborhood of the gyrofrequency of the *i*th ion, we may approximate $K_{\rm I}$ by

$$K_{\rm I} \approx \frac{\omega_{pi}^2}{\omega_{Bi}(\omega_{Bi}+\omega)}$$

which has a pole at $\omega = -\omega_{Bi}$. The ion whistler is polarized circularly with left-hand sense near the gyrofrequency of a positive ion when $\theta = 0$.

Similar expressions can be obtained for frequencies in the neighborhood of a negative ion gyrofrequency. The refractive index expressions are still given by (4.13.3) and (4.13.14) except that K_{II} and K_{II} must be interchanged. We note now K_{II} has a pole at the gyrofrequency. In the limit of $\theta = 0$, both waves are still circularly polarized, but the compressional wave is left-handed and ion whistler right-handed.

The angular dependence of the refractive index of ion whistler as given by (4.13.14) constrains wave energy to propagate along a small cone centered about the steady magnetic field. It can be shown (see problem at the end of this chapter) that the angle ψ between the group ray and **B**₀ is given by

$$\tan \psi = \sin \theta \cos^3 \theta / (1 + \cos^4 \theta) \tag{4.13.15}$$

The largest value of ψ is 12.3°. Hence, the ray path of ion whistlers near the ion gyrofrequency is not more than 12.3° from along the steady magnetic field.

When the frequency is above all the ion gyrofrequency no simplification of the dispersion relation (4.8.6) or its equivalent is possible since all the ion terms as well as the electron term contribute to the dispersion relation. However, in the frequency range $\omega \gg |\omega_{Bi}|$ of all ions, only electrons contribute to the polarizability of the medium. When this is the case we shall refer to it as the high frequency approximation. The dispersion relation in the high frequency approximation is called the Appleton-Hartree formula which is discussed in the next section.

4.14 Appleton-Hartree Formula—High Frequency Approximation

When the frequency is much larger than all ionic gyrofrequencies, induced ionic motions are negligible because of their mass. Ions can therefore be viewed as forming a stationary neutralizing background. Only electrons contribute to the polarizability of the medium, i.e., the relative dielectric tensor with elements given by (4.5.16) has only the electron component. The wave propagation in such a medium can be studied, in principle, by simplifying the dispersion relations (4.8.6) or (4.8.7) or (4.8.9) and the wave polarization relations (4.8.10) or (4.8.22) and (4.8.23). But, in practice, the algebra involved to put the results in commonly used forms is very tedious (see problem at the end of this chapter). We choose to start our derivation from the beginning with the two Maxwell's curl equations. The derivation is very much simplified if the inverse susceptibility tensor is made use of. Maxwell's curl equations for plane waves with dependence $e^{j\omega(t-n\hat{s}\cdot \mathbf{r}_lc)}$ are

$$\mathbf{D} = -(n/c)\hat{s} \times \mathbf{H} \tag{4.14.1}$$

$$\mathbf{H} = (\varepsilon_0/\mu_0)^{1/2} \, n \hat{s} \times \mathbf{E} \tag{4.14.2}$$

The vector \mathfrak{F} is a unit vector in the direction of the propagation vector. When (4.14.2) is substituted into (4.14.1), we obtain

$$\mathbf{D} = n^2 \varepsilon_0 [\mathbf{E} - \hat{\mathbf{s}} (\hat{\mathbf{s}} \cdot \mathbf{E})] \tag{4.14.3}$$

Orient the coordinates so that $\beta // z$ -axis and B_0 in the *yz*-plane as shown in Fig. 4.14-1. This is identical to the doubled primed coordinates of Fig.



4.8-2b. Here, for simplicity, we shall denote our coordinates without primes. In this coordinate system, we can write (4.14.3) in component form

$$D_x = \varepsilon_0 n^2 E_x = \varepsilon_0 E_x + P_x \tag{4.14.4a}$$

$$D_y = \varepsilon_0 n^2 E_y = \varepsilon_0 E_y + P_y \tag{4.14.4b}$$

$$D_z = 0 \qquad = \varepsilon_0 E_z + P_z \qquad (4.14.4c)$$

The first two equations of (4.14.4) and (4.14.2) yield

$$R = E_x/E_y = P_x/P_y = D_x/D_y = -H_y/H_x$$
(4.14.5)

This is a relation obtained before [see (4.8.14a) and (4.8.15)]. Its implications on the state of polarization have been discussed in Section 4.8 and shall not be repeated here. The electric field is also related to the polarization density by

$$\varepsilon_0 \mathbf{E} = \mathbf{\chi}_e^{-1} \cdot \mathbf{P}_e = \mathbf{\chi}^{-1} \cdot \mathbf{P} \tag{4.14.6}$$

where the last equality comes from the fact that only electrons contribute

184

to the polarizability in the high frequency approximation. Let us denote the elements of the inverse susceptibility tensor as Γ_{ij} , i.e., $\chi^{-1} = [\Gamma_{ij}]$. In the coordinates of Fig. 4.14-1, these elements are, from (4.5.10),

$$\Gamma_{xx} = \Gamma_{yy} = \Gamma_{zz} = -1/X$$

$$\Gamma_{xy} = -\Gamma_{yx} = jY_z/X$$

$$\Gamma_{zz} = -\Gamma_{zx} = -jY_y/X$$

$$\Gamma_{yz} = \Gamma_{zy} = 0$$
(4.14.7)

Since all quantities are referring to electrons, the subscript e is ignored on X and Y as done in (4.14.7) and elsewhere in this section. The symmetry of χ^{-1} as indicated by (4.14.7) can be used to good advantage to simplify our derivation. In component form, (4.14.6) can be written as

$$\varepsilon_0 E_x = \Gamma_{xx} P_x + \Gamma_{xy} P_y + \Gamma_{xz} P_z \qquad (4.14.8a)$$

$$\varepsilon_0 E_y = -\Gamma_{xy} P_x + \Gamma_{xx} P_y \tag{4.14.8b}$$

$$\varepsilon_0 E_z = -\Gamma_{xz} P_x + \Gamma_{xx} P_z \tag{4.14.8c}$$

From (4.14.4c) and (4.14.8c) we eliminate E_z to obtain

$$P_z = \frac{\Gamma_{xz}}{1 + \Gamma_{xx}} P_x \tag{4.14.9}$$

Substitute P_z as given by (4.14.9) into (4.14.8a) and make use of (4.14.5); we can then rewrite (4.14.8a) and (4.14.8b), respectively, as

$$\varepsilon_0 E_x = [\Gamma_{xx} + \Gamma_{xz}^2/(1 + \Gamma_{xx})]RP_y + \Gamma_{xy}P_y \qquad (4.14.10a)$$

$$\varepsilon_0 E_y = -\Gamma_{xy} R P_y + \Gamma_{xx} P_y \tag{4.14.10b}$$

Take the ratio of above two equations and again make use of (4.14.5). The resulting equation is a quadratic equation in R and it simplifies to

$$R^{2} + \frac{\Gamma_{xz}^{2}}{\Gamma_{xy}(1 + \Gamma_{xx})} R + 1 = 0 \qquad (4.14.11)$$

This equation has two roots, R_1 and R_2 . The two roots are related by

$$R_1 R_2 = 1 \tag{4.14.12}$$

$$R_1 + R_2 = -\Gamma_{xz}^2 / \Gamma_{xy} (1 + \Gamma_{xx})$$

= $j Y_y^2 / Y_z (1 - X)$ (4.14.13)

where the elements (4.14.7) have been substituted. The polarization ellipses of the two characteristic waves in a plane transverse to \hat{s} are similar, counterrotating, and with the major axis aligned with either the x-axis or y-axis. These results are identical to those discussed in Section 4.8. The roots of (4.14.11)can be written down easily:

$$R = \frac{j}{Y_{z}} \left[\frac{Y_{y}^{2}}{2(1-X)} \mp \left(\frac{Y_{y}^{4}}{4(1-X)^{2}} + Y_{z}^{2} \right)^{1/2} \right]$$

= $\frac{j}{\cos \theta} \left[\frac{Y \sin^{2} \theta}{2(1-X)} \mp \left(\frac{Y^{2} \sin^{4} \theta}{4(1-X)^{2}} + \cos^{2} \theta \right)^{1/2} \right]$ (4.14.14)

We note that in the literature there is always confusion in determining the sign of Y_z , depending on the conventions used. In the present case, Y is always positive in the second form of (4.14.14) as it refers to electrons.

In the magnetoplasma there is, in general, a nonvanishing longitudinal electric field component. Hence in addition to R, we need a second complex ratio to express the full state of wave polarization. This ratio is defined by $Q = E_z/E_x$ and is obtained by first rewriting (4.14.4a) as

$$\varepsilon_0(n^2-1)E_x = P_x = \frac{1+\Gamma_{xx}}{\Gamma_{xz}}P_z = -\frac{\varepsilon_0(1+\Gamma_{xx})}{\Gamma_{xz}}E_z$$
 (4.14.15)

where the second equality and third equality come from (4.14.9) and (4.14.4c), respectively. The equality of the first and last expressions of (4.14.15) gives

$$Q = E_z / E_x = \Gamma_{xz} (1 - n^2) / (1 + \Gamma_{xx})$$

= $j Y \sin \theta (1 - n^2) / (1 - X)$ (4.14.16)

This ratio is expressed in terms of the refractive index.

Now let us derive the Appleton-Hartree formula. From (4.14.4b) we obtain

$$n^2 = 1 + P_y / \varepsilon_0 E_y \tag{4.14.17}$$

The ratio $P_y/\varepsilon_0 E_y$ is given by (4.14.10b), and when it is substituted into (4.14.17) we get

$$n^{2} = 1 + \frac{1}{\Gamma_{xx} - \Gamma_{xy}R}$$
$$= 1 - \frac{X}{1 + jYR\cos\theta}$$
(4.14.18)

The refractive index (4.14.8) is still expressed in terms of R. The expression

is useful in determining the corresponding n and R for a given characteristic wave. When the polarization ratio (4.14.14) is substituted into (4.14.18) we obtain the Appleton-Hartree formula

$$n^{2} = 1 - \frac{X}{1 - \frac{Y^{2} \sin^{2} \theta}{2(1 - X)} \pm \left(\frac{Y^{4} \sin^{4} \theta}{4(1 - X)^{2}} + Y^{2} \cos^{2} \theta\right)^{1/2}}$$
(4.14.19)

The upper sign of (4.14.14) and the upper sign of (4.14.19) refer to the same characteristic wave, similarly for the lower signs. In our convention Y is positive. It is also convenient to take the positive value of the square root in both (4.14.14) and (4.14.19).

All expressions derived so far in this section are for the lossless case. If collisions cannot be ignored modifications are needed. When the collision is frictionlike with an effective collision frequency ν for electrons, simultaneous substitutions of X by X/U and Y by Y/U in (4.14.14), (4.14.16), and (4.14.19) give us the desired results. Here U is given by

$$U = 1 - j\nu/\omega \qquad (4.14.20)$$

For completeness we list the expressions in the following.

$$R = \frac{j}{\cos\theta} \left[\frac{Y\sin^2\theta}{2(U-X)} \mp \left(\frac{Y^2\sin^4\theta}{4(U-X)^2} + \cos^2\theta \right)^{1/2} \right]$$
(4.14.21)

$$Q = jY \sin \theta (1 - n^2) / (U - X)$$
(4.14.22)

$$n^{2} = 1 - \frac{X}{U - \frac{Y^{2} \sin^{2} \theta}{2(U - X)} \pm \left(\frac{Y^{4} \sin^{4} \theta}{4(U - X)^{2}} + Y^{2} \cos^{2} \theta\right)^{1/2}}$$
(4.14.23)

Since U is complex, all above quantities are in general complex. The computations of these quantities are very tedious but very essential in understanding the behavior of waves. We shall study some simple cases in the next section.

The field components when referred to, for example, the amplitude E_{0x} in the coordinates of Fig. 4.14-1, can be found by using (4.14.4), (4.14.5), and (4.14.16). They are given by

$$\mathbf{E} = E_{0x}(\hat{x} + \hat{y}/R + \hat{z}Q)$$

$$\mathbf{D} = \varepsilon_0 n^2 E_{0x}(\hat{x} + \hat{y}/R)$$

$$\mathbf{H} = (nE_{0x}/c\mu_0)(-\hat{x}/R + \hat{y})$$

$$\mathbf{P} = -\varepsilon_0 E_{0x}[\hat{x}(1 - n^2) + \hat{y}(1 - n^2)/R + \hat{z}Q]$$
(4.14.24)

187

where the multiplying factor $e^{j(\omega t - nz/c)}$ common to all has not been written out explicitly and E_{0x} is the amplitude of E_x . These expressions are applicable even when the medium is lossy since they are expressed in terms of R, Q, and n. If the medium is lossless, R and Q are imaginary and n is real. If the medium is lossy, all are complex in general.

4.15 Some Properties of the Appleton-Hartree Formula

The understanding of Appleton-Hartree formula and its associated polarization relations are very crucial to studies of wave propagation in the ionosphere. There exist in the literature many papers that discuss properties of these equations. We shall discuss some of these.

(i) Parallel Propagation. In this case the wave vector and the steady magnetic field are parallel and thus $\theta = 0$. Both waves are purely transverse as Q given by (4.14.16) vanishes. The refractive index (4.14.19) and the polarization relation (4.14.14) reduce to

$$\begin{array}{c} n^{2} = 1 - \frac{X}{1 \pm Y} \\ P = - \frac{T}{i} \end{array} \right\} \qquad \theta = 0 \qquad (4.15.1a) \\ \theta = 0 \qquad (4.15.1b) \end{array}$$

Both waves are circularly polarized; the wave corresponding to the upper sign is left-handed and the lower sign wave is right-handed. The square of refractive index depends linearly on X which is proportional to electron density. The cutoffs at which n = 0 occur at X = 1 + Y and 1 - Y. These properties are shown in Fig. 4.15-1. We see from Fig. 4.15.1b that when the frequency is less than the electronic gyrofrequency, the refractive index



Fig. 4.15-1. Square of refractive index as a function of $X, \theta = 0$. (a) Y < 1 or $f > f_B$. (b) Y > 1 or $f < f_B$.

of the right circular mode stays real for any positive X. This mode is just the electron whistler mode discussed in Section 4.11 and it can propagate in a plasma with any density.

When the wave vector and the steady magnetic field are antiparallel, i.e., $\theta = \pi$, the wave with the refractive index given by the upper sign of (4.15.1a) becomes right-handed and the lower sign becomes left-handed. That is,

$$R = \pm j, \quad \theta = \pi \tag{4.15.1c}$$

where the right-handed wave corresponding to the upper sign has the refractive index given by the upper sign of (4.15.1a) and similarly the lower sign of (4.15.1c) corresponds to the lower sign of (4.15.1a).

When the frequency is lowered to the ion gyrofrequency we must include ionic contributions. The effect of ions has been discussed in Section 4.9.

(ii) Perpendicular Propagation. When $\theta = \pi/2$, the refractive index (4.14.19) and the polarization relations (4.14.14) and (4.14.16) reduce to

$$n_0^2 = 1 - X$$
, $R_0 = 0$, $Q_0 = jXY/(1 - X)$ (4.15.2a)

and

$$n_x^2 = 1 - \frac{X(1-X)}{1-X-Y^2}, \quad R_x = \infty, \quad Q_x = jXY/(1-X-Y^2)$$
(4.15.2b)

The formulas (4.15.2a) are referred to a wave which is polarized linearly along \mathbf{B}_0 (see Fig. 4.14-1 with $\theta = \pi/2$) and whose refractive index is not influenced by the steady magnetic field. This wave is called the ordinary wave. The extraordinary wave given by (4.15.2b) is polarized in a plane perpendicular to \mathbf{B}_0 and in general is not a transverse wave like the ordinary wave. The cutoffs occur at X = 1 for the ordinary wave and at $X = 1 \pm Y$ for the extraordinary wave. The ordinary wave has no resonance and the extraordinary wave has a resonance at $X = 1 - Y^2$ when Y < 1 but not when Y > 1. These properties are evident in Fig. 4.15-2.

(iii) General Case. For a general angle θ , we must use the full Appleton-Hartree formula (4.14.19). Typical curves for a given θ are shown in Fig. 4.15-3. Dotted boundaries show regions in which the refractive index curves for any general θ must lie. It is obvious that curves for the special cases of parallel and perpendicular propagation serve as boundaries. As expected, curves for any θ pass through zero at the same three points.



Fig. 4.15-2. Square of refractive index as a function of $X, \theta = \pi/2$. (a) Y < 1 or $f > f_B$; (b) Y > 1 or $f < f_B$.



Fig. 4.15-3. Square of refractive index as a function of Y, arbitrary θ . (a) Y < 1, $f > f_B$, plotted for $Y = \frac{1}{2}$; (b) Y > 1, $f < f_B$, plotted for Y = 2. [After Ratcliffe (1959), "The Magnetoionic Theory and its Application to the Ionosphere," Cambridge Univ. Press. Reproduced by permission.]

This is because the cutoff condition is independent of θ . The three cutoffs are given by

(i)
$$X = 1 - Y$$
, $\omega = (\omega_B/2) + [\omega_p^2 + (\omega_B/2)^2]^{1/2}$, $f_p^2 = f^2 - ff_B$
(4.15.3a)

(ii)
$$X = 1$$
, $\omega = \omega_p$, $f_p^2 = f^2$
(4.15.3b)

(iii)
$$X = 1 + Y$$
, $\omega = -(\omega_B/2) + [\omega_p^2 + (\omega_B/2)^2]^{1/2}$, $f_p^2 = f^2 + ff_B$
(4.15.3c)

We note that the conditions of cutoff in the second form of (4.15.3a-c) are just the dispersion relations of longitudinal oscillations (4.7.16), (4.7.6), and (4.7.12), respectively. The third form of (4.15.3) is sometimes useful in ionospheric investigations. The single resonance condition occurs when

$$X = (1 - Y^2)/(1 - Y^2 \cos^2 \theta)$$
(4.15.4)

which depends on θ .

We note that the general case is much more complicated than the special case of perpendicular or parallel propagation. Under certain conditions the propagation can be described as quasi-perpendicular or quasi-parallel depending on the magnitude of the first term in relation to the second term of the two terms under the square root sign of (4.14.19). If it is smaller, the propagation is quasi-parallel. These conditions can be restated mathematically as

$$\cos^2 \theta \ll \frac{f^2 f_B^2 \sin^4 \theta}{4(f^2 - f_p^2)^2}$$
, quasi-perpendicular condition (4.15.5)

$$\cos^2 \theta \gg \frac{f^2 f_B^2 \sin^4 \theta}{4(f^2 - f_p^2)^2}$$
, quasi-parallel condition (4.15.6)

Roughly, the quasi-perpendicular condition is expected to be valid for a large range of θ when f is close to f_p or when both f and f_p are small when compared with f_B . On the other hand, the quasi-parallel condition is expected to be valid for a large range of θ when f is larger than both f_B and f_p or when f_p is larger than both f and f_B . Under these special conditions, the refractive index and polarization expressions can be simplified.

Quasi-perpendicular condition:

$$n_0^2 = 1 - X,$$
 $R_0 = 0$ (4.15.7a)

$$n_x^2 = 1 - \frac{X(1-X)}{1-X-Y^2\sin^2\theta}$$
, $R_x = jY\sin^2\theta/(1-X)\cos\theta$, large (4.15.7b)

Quasi-parallel condition:

$$n^2 = 1 - \frac{X}{1 \pm Y \cos \theta}, \quad R = \mp j |\cos \theta| / \cos \theta \quad (4.15.8)$$

In the transverse plane, the quasi-perpendicular waves are still approximately linearly polarized, the quasi-parallel waves are still circularly polarized. With the exception of the ordinary wave, the refractive indices (4.15.7b) and (4.15.8) have important modifications when compared with their corresponding expressions for $\theta = \pi/2$ or $\theta = 0$. These modifications are important in studying several phenomena. For example, the propagation of electron whistlers discussed in Section 4.11 can be studied in terms of the refractive index

$$n_R^2 = 1 - X/(1 - Y\cos\theta)$$
 (4.15.9)

when the quasi-parallel condition (4.15.8) is met. This formula has the interesting property that the group ray must be within 19.5° of the direction of the magnetic field (see problem at the end of chapter). Like ion whistlers, there is therefore considerable channeling of energy along the steady magnetic field. As another example, the Faraday rotation discussed in Section 4.10 must be modified by introducing the factor $\cos \theta$ in (4.10.8) when the propagation is quasi-parallel. The total twist of the electric vector of a linearly polarized wave in a uniform magnetoplasma is now

$$\Omega = \omega_p^2 \omega_B z \cos \theta / 2c \omega^2 \quad \text{rad} \tag{4.15.10}$$

Let us now turn our attention to the Poynting vector. The time-averaged Poynting vector shows the direction of energy flow and is given by

$$\langle \mathbf{S} \rangle = \frac{1}{2} \operatorname{Re}(\mathbf{E} \times \mathbf{H}^*)$$
 (4.15.11)

In the coordinates of Fig. 4.14-1, we have $H_z = 0$. The three components of the average Poynting vector therefore simplifies to, for the lossless case,

$$\langle S_x \rangle = -\frac{1}{2} \operatorname{Re}(E_z H_y^*) = -(E_{0x} E_{0x}^*/2c\mu_0) \operatorname{Re} Qn^* = 0 \langle S_y \rangle = \frac{1}{2} \operatorname{Re}(E_z H_x^*) = -(E_{0x} E_{0x}^*/2c\mu_0) \operatorname{Re}(Qn^*/R^*) = -E_{0x} E_{0x}^* Qn/2c\mu_0 R^* \langle S_z \rangle = \frac{1}{2} \operatorname{Re}(E_x H_y^* - E_y H_x^*) = (E_{0x} E_{0x}^*/2c\mu_0) \operatorname{Re} n^*(1 + 1/RR^*) = (E_{0x} E_{0x}^* n/2c\mu_0)(1 + 1/RR^*)$$
(4.15.12)

where the field components given by (4.14.24) have been used. In the lossless

192

case the x-component of the Poynting vector vanishes. Therefore, the propagation of waves in a lossless magnetoplasma is accompanied by a lateral deviation of energy in the magnetic meridian and by an oscillation of energy perpendicular to the magnetic meridian. Let the angle between the wave vector and Poynting vector be α . The lateral deviation of energy is then given by

$$\tan \alpha = -\langle S_y \rangle / \langle S_z \rangle = \frac{QR}{1 + RR^*}$$
(4.15.13)

The quantity $\tan \alpha$ is the lateral deviation of energy in the magnetic meridian plane away from **k** per unit distance of propagation along **k**. The reason for introducing the minus sign in (4.15.13) is that we wish to define α as positive if **k** and **B**₀ are on the same side of the group velocity. This definition is identical to that used in Section 2.12 in which

$$\tan \alpha = -\frac{1}{n} \frac{\partial n}{\partial \theta}$$
(4.15.14)

is obtained. It can be shown that (4.15.14) is equal to (4.15.13) by differentiating the Appleton-Hartree formula (see Problem 19 at the end of the chapter). The algebra is rather involved.

When the medium is lossy, n, R, and Q are all in general complex. The Poynting vector is given by the next to the last expression of (4.15.12) times the amplitude attenuation factor $e^{-2(\operatorname{Im} n/c)z}$. All three components of the Poynting vector in general exist. The energy can therefore deviate out of the magnetic meridian plane as well as in the plane.

4.16 Cutoffs and Resonances in Parameter Space

We have seen that a characteristic wave in a cold anisotropic medium must have a specific polarization and a specific wave vector. The extremity of the wave vector must be on the dispersion surface

$$\det \mathbf{D}(\omega, \mathbf{k}) = 0 \tag{4.16.1}$$

where **D** is given by (4.8.3). In spherical coordinates the surface (4.16.1) for a given ω is a plot of k versus the azimuthal and polar angles of **k**. When the plot is made for the refractive index $n = k/k_0$, the surface is referred to as the refractive index surface or simply index surface. As pointed out in Chapter 2 the index surface not only depicts the magnitude of the refractive

194

index in a given direction including the possibilities of cutoffs and resonances, but more. For a given refractive index *n* terminated on the index surface, the normal to the index surface at the terminal is parallel to the group velocity vector corresponding to the given n. Further, the Gaussian curvature of the index surface plays an important role in radiation problems. It is thus of interest to find the nature of these surfaces when parameters of the medium are varied. In this connection it has been found convenient to define a parameter space by using X_e and Y_e^2 as coordinate axes. A point in the parameter space corresponds to a definite plasma whose parameters are specified by the location of the point. As the point moves about in the space the set of plasma parameters are changing continuously in some fashion. Since both X_e and Y_e^2 are positive, only the first quadrant of the space is physically meaningful. For a fixed frequency, X_e is proportional to electron density and Y_e^2 proportional to B_0^2 . For a fixed plasma, the change in frequency is equivalent to a radial motion on the parameter space. The infinite frequency is at the origin. As the frequency is lowered from infinity, the parameters (X_e, Y_e^2) move out radially. Therefore, the high frequency approximation in which only electrons contribute to the polarizability of the medium is expected to apply for the portion of the parameter space near the origin. As the frequency is lowered further, the ions begin to participate. The ions are expected to be important for the portion of the parameter space far away from the origin. The low frequency approximation of hydromagnetic waves occupies the space very far from the origin. We shall demonstrate these points with concrete examples in the following.

The cutoff condition (i.e., $\mathbf{k} = 0$) is the vanishing of det **K** and is independent of the direction of **k**. It is also the condition for longitudinal oscillations discussed in Section 4.7. The condition is given by

$$K_0 = 0, \quad K_{\rm I} = 0, \quad \text{or} \quad K_{\rm II} = 0$$
 (4.16.2)

In the high frequency approximation in which only electrons need to be taken into account, (4.16.2) reduces to

$$X = 1$$
, $X = 1 + Y$, or $X = 1 - Y$ (4.16.3)

respectively. These curves are shown in Fig. 4.16-1 as dotted lines. When the frequency is reduced enough the ions may begin to contribute. For concreteness let us consider a three-component plasma of electrons neutralized by 60% protons and 40% singly charged atomic oxygen ions. The portion of the parameter space in which ions are important is very far away from the origin. It is therefore convenient to use logarithmic scale



Fig. 4.16-1. Contours of θ_r in the parameter space for an electronic plasma are shown as solid lines. Dotted lines show conditions of cutoff.



Fig. 4.16-2. Regions in which resonances may occur are shown shaded. The plasma is composed of electrons neutralized by 60% protons and 40% atomic oxygen ions. Dotted lines show conditions of cutoff.

for both X_e and Y_e^2 . The cutoffs given by (4.16.2) are shown as dotted lines in Fig. 4.16-2.

The condition for resonance is $n \to \infty$ which requires, as seen in (4.8.7), $a_4 = 0$. Setting a_4 equal to zero in (4.8.8), it leads to an expression for the resonance angle θ_r ,

$$\tan^2 \theta_r = -K_{zz}/K_{xx} = -(1 - \sum_{\alpha} X_{\alpha})/[1 - \sum_{\alpha} X_{\alpha}/(1 - Y_{\alpha}^2)] \qquad (4.16.4)$$

The high frequency approximation of (4.16.4) is

$$\tan^2 \theta_r = -\frac{(1-X)(1-Y^2)}{1-X-Y^2} \tag{4.16.5}$$

Since both X and Y^2 are positive, the real resonance angle given by (4.16.5) can occur only in two regions. They are

- (i) X < 1, $Y^2 < 1$, and $1 X < Y^2$
- (ii) X > 1 and $Y^2 > 1$.

196

In the parameter space the two regions in which resonance occurs can be found easily. Contours of θ_r have been computed by using (4.16.5). The results are shown in Fig. 4.16-1. The curves in the region $X, Y^2 > 1$ approach asymptotically to $X = \sec^2 \theta_r$ when Y^2 is very large. Similarly, when X is very large, the curves approach to $Y^2 = \sec^2 \theta_r$. The bounding curves of resonance regions are given by either $\theta_r = 0$ or $\theta_r = \pi/2$. In Fig. 4.16-1 we have

$$\theta_r = \pi/2 \quad \text{on} \quad 1 - X = Y^2
\theta_r = 0 \quad \text{on} \quad X = 1, \quad \text{or} \quad Y^2 = 1$$
(4.16.6)

For large values of X_e and Y_e , ion effects must be included. They introduce additional resonance regions. For the example considered earlier, regions in which resonance may occur are shown shaded in Fig. 4.16-2. The resonance angles on the bounding curves are

$$\theta_r = \pi/2 \quad \text{on} \quad K_I + K_{II} = 2K_{xx} = 0$$

 $\theta_r = 0 \quad \text{on} \quad K_0 = 0 \quad \text{or} \quad Y_{\alpha}^2 = 1$
(4.16.7)

They are marked on the curves of Fig. 4.16-2.

4.17 Index Circle and Index Surface

Systematic classification of index surfaces was first carried out by Clemmow and Mullaly (1955) and Allis [see Allis *et al.* (1962)]. Hence, plots of index surfaces on the parameter space are called CMA diagrams by Stix (1962).⁺ The construction of the index surface can be done graphically as shown by Deschamps (1965). The method also assists in visualizing the transition of topological genera of the index surface in the parameter space. The method makes use of the index circle. Therefore, we shall first discuss the construction of index circle.

An index circle is a circle of unit diameter, on which scales for the refractive index are calibrated. Since the medium is lossless, n is either purely real or purely imaginary. The real n going from 0 to ∞ are calibrated on the right-hand half of the circle and the negative imaginary n going from 0 to ∞ on the left-hand half. At the point n = 0 a tangent is drawn with linear scale in n^2 . The refractive index scale on the circle is then obtained by projecting any point on the tangent line to a point on the circle with the top point of the circle (i.e., $n = \infty$ point) as the center of projection. The whole process is shown in Fig. 4.17-1. For illustration, a line is drawn be-



Fig. 4.17-1. Graduation of index circle.

tween the point $0.25 = (0.5)^2$ on the tangent and the point $n = \infty$ on the circle. The line intersects the circle at n = 0.5. We note that the points for two reciprocal values of n are on the same vertical as the points 0.5 and 2. A fully calibrated index circle is shown in Fig. 4.17-2.

The index circle can be used to construct index surfaces. The method is based on the special form of the dispersion relation (4.8.6) or

$$K_0(n^2 - K_{\rm I})(n^2 - K_{\rm II})\cos^2\theta + K_{xx}(n^2 - K_0)(n^2 - K_x)\sin^2\theta = 0 \quad (4.17.1)$$

⁺ Stix actually plotted the phase velocity of the wave on the CMA diagrams.



Fig. 4.17-2. Index circle. [Courtesy of G. A. Deschamps.]

where $K_{xx} = (K_{I} + K_{II})/2$, $K_x = K_{I}K_{II}/K_{xx}$. Equation (4.17.1) can also be written in the form

$$K_0(n^2 - K_{\rm I})(n^2 - K_{\rm II}) + (K_{xx} - K_0)n^2(n^2 - K_{\infty})\sin^2\theta = 0 \quad (4.17.2)$$

where

$$K_{\infty} = (K_{\rm I}K_{\rm II} - K_{xx}K_0)/(K_{xx} - K_0) \qquad (4.17.3)$$

There are two basic rules that must be followed in order to construct the index surface. The two rules are:

(1) For any given $\theta = \arcsin \sqrt{t}$, the two roots of (4.17.1) for *n* can be located on the index circle. A line L_t joining these two points can be drawn. The first rule is that all the lines L_t for arbitrary real values of $t = \sin^2 \theta$ must pass through a fixed point *J*. This applies for real values of θ for which $t \le 1$ as well as for imaginary values of θ for which t > 1.

(2) A special line that can be drawn is L_{∞} . This line is drawn for $t = \infty$. As seen from (4.17.2), the line L_{∞} must pass through n = 0 and $n = \sqrt{K_{\infty}}$. The rule (1) must also apply to L_{∞} so that it passes through the point J. Draw a line parallel to L_{∞} . The second rule is that the intersection of L_t with this parallel line describes a linear scale in $t = \sin^2 \theta$.

These two rules can be used to construct the index surface. The method is as follows. Mark on the index circle $n_L = \sqrt{K_I}$ and $n_R = \sqrt{K_{II}}$. The line joining these two points is L_0 . Similarly, draw a line L_1 through the points $n_0 = \sqrt{K_0}$ and $n_x = \sqrt{K_x}$. The lines L_0 and L_1 meet at a fixed point J. The line L_∞ is obtained by drawing through the point n = 0 and J. The line L_∞ intersects the index circle at $n = \sqrt{K_\infty}$. A scale linear in $\sin^2 \theta$ is placed parallel to L_∞ with points t = 0 (or $\theta = 0$) and t = 1 (or $\theta = \pi/2$) on lines L_0 and L_1 , respectively. The desired values of refractive index for any given t (or θ) can be found by joining the value t on the scale and J and then reading off the intersected values of the index circle with the line L_t . As t varies, the line L_t sweeps through the shaded region as shown in Fig. 4.17-3. For convenience the scale linear in $t = \sin^2 \theta$ can be recali-



Fig. 4.17-3. The use of index circle in constructing the index surface. The index surface is axially symmetric with respect to B_0 . Because of the symmetry of the dispersion surface only one quarter of its meridional section is given. The solid curves show the real refractive index and dotted curve shows the negative imaginary refractive index. [After Deschamps (1965).]

brated in terms of θ as shown in Fig. 4.17-2. The example of Fig. 4.17-3 shows one closed sheet of the index surface and one open sheet with the resonance phenomenon.

In the high frequency approximation for which only electron terms remain, (4.17.3) reduces to

$$K_{\infty} = 1 \tag{4.17.4}$$

The line L_{∞} in this case is a fixed line passing through the points n = 0

and n = 1 on the index circle. The line L_{∞} is a useful line because the sweep of L_t for any real angle θ from L_0 to L_1 must not contain L_{∞} . The index circle representation of dispersion properties of electronic plasmas in the parameter space is shown in Fig. 4.17-4. The resulting meridional sections of the index surface at different points in the parameter space are shown in



Fig. 4.17-4. Index-circle representation of dispersion properties in the parameter space. The dot on the circle indicates the position of n_0 . [After Deschamps (1965).]

Fig. 4.17-5. The circle n = 1 is shown merely for reference purposes. The solid curve is used when the refractive index is real and dotted curve when it is negative imaginary. The steady magnetic field is in the vertical direction. To explain certain properties of the medium and the wave that propagates in the medium we divide the parameter space into eight regions by using bounding curves corresponding to cutoffs and resonances at $\theta_r = 0$ and $\pi/2$. These eight regions are numbered in Fig. 4.17-6. In discussing these regions, it is useful to refer to Figs. 4.17-4 through 4.17-6.

Region I. This is the region in which the frequency is high, the electron density low, the magnetic field weak. Both characteristic waves can propagate in all directions. Near the origin of the parameter space, both refractive indices are close to 1. As the cutoff X = 1 - Y is approached, both refrac-

200



Fig. 4.17-5. The refractive index surfaces in the parameter space. The unit circles are drawn to indicate the scale of the index surfaces. [Courtesy of G. A. Deschamps.]

tive indices are reduced from 1, showing phase velocities larger than the free space velocity of light. The refractive index corresponding to the extraordinary wave is reduced at a faster rate than the ordinary wave and at the cutoff X = 1 - Y the ordinary wave can still propagate but the extraordinary wave has zero refractive index and stops propagating.

Region II. Only the ordinary wave can propagate in this region. The extraordinary wave is evanescent. The bounding curve between Regions I and II is the cutoff condition X = 1 - Y. In passing through the curve, the refractive index surface of the extraordinary wave becomes imaginary and is "destroyed" and therefore the transition is called a destructive transition for the extraordinary wave. The index surface for the ordinary wave deforms continuously from Region I to Region II. The transition for the ordinary wave is an intact transition.

Region III. Both characteristic waves can propagate, but the extraordinary wave has a resonance angle within which it can not propagate.
202 4. Waves in Fluid Plasma with a Steady Magnetic Field

The resonance angle decreases as the gyroresonance or plasma cutoff is approached. The transition at $X = 1 - Y^2$ between Regions II and III is a destructive transition for the extraordinary wave and intact transition for the ordinary wave.

Region IV. Only the extraordinary wave can propagate in this region. In going through the bounding surface X = 1, the resonance angle is reduced to zero so that the extraordinary refractive index surface is now closed. Because of this change in the surface, the transition from Region III to Region IV is called a reshaping transition for the extraordinary wave. The ordinary wave goes through a destructive transition on crossing X = 1.

Region V. The value of the extraordinary refractive index in Region IV is continuously decreased as the cutoff X = 1 + Y is approached. On crossing X = 1 + Y the extraordinary wave no longer propagates. In Region V both characteristic waves are evanescent. The ordinary wave has the larger attenuation than the ordinary wave.

Region VI. In going from Region III to VI, the resonance angle of the extraordinary is reduced to zero at Y = 1. Therefore, the extraordinary wave has now a closed index surface. It is a reshaping transition for the extraordinary wave at Y = 1. The transition for the ordinary wave is intact.

Region VII. On crossing the bounding curve X = 1 from Region VI to Region VII, the refractive index for the ordinary wave disappears at the origin but reappears at infinity. The shape of the ordinary index surface is



Fig. 4.17-6. The eight regions in the parameter space for an electronic plasma.

now changed; only waves within the resonance cone can propagate. The transition is a reshaping transition for the ordinary wave. The index surface of the extraordinary wave is intact. We note that the transition at Y = 1 from Region VII to Region IV is destructive for the ordinary wave and intact for the ordinary wave.

Region VIII. Only ordinary waves propagate in this region. The transition at X = 1 + Y from Region VII to Region VIII is intact for the ordinary wave and destructive for the extraordinary wave. The transition at Y = 1from Regions VIII to V is destructive for the ordinary wave.

In summary, we see that both characteristic waves can propagate in Regions I, III, VI, and VII of Fig. 4.17-6. In Regions II and VIII only



Fig. 4.17-7. Parameter space for a plasma containing electrons neutralized by 60% protons and 40% atomic oxygen ions. Representative index surfaces in each region are shown on the top. The right and left circular polarizations when propagating along the steady magnetic field are marked by *R* and *L*, respectively. Similarly, the ordinary (*O*) and extraordinary (*X*) indices are marked when propagating perpendicular to magnetic field.

ordinary waves can propagate, and in Region IV only extraordinary waves can propagate. Both waves do not propagate in Region V. In going from one region to another region, three types of transitions on the index surface may occur for a given characteristic wave. They are intact transition, reshaping transition, and destructive transition.

Complications arise for large values of X_e and Y_e due to presence of ions. The ions may give rise to hybrid resonances, ion gyroresonances, additional cutoffs, and a phenomenon called cross-overs. The cross-over occurs when $K_{xy} = 0$ as discussed in Sections 4.9 and 4.11. In passing through the cross-over condition the index surfaces maintain their topological genera except that the right and left circular polarizations are interchanged. Such a transition is called a cross-over transition. Typical index surfaces in all regions of an example plasma are shown in Figure 4.17-7. A cross-over transition occurs in going from Region Xa and Region VIIb. The propagation characteristics in each region and the type of transition in going from one region to the next can also be found by studying Fig. 4.17-7.

4.18 Dielectric Tensor of a Warm Magnetoplasma

We have so far in this chapter ignored the effects associated with plasma temperatures. The resulting theory is applicable to the cold plasma, i.e., a plasma in which the thermal velocity of plasma particles is negligibly small when compared with the phase velocity of the wave. This condition is violated in two cases: (i) The plasma waves discussed in Chapter 3 have velocities of the order of the thermal velocities. (ii) Near resonance the phase velocity of the electromagnetic wave is very small. To extend the cold plasma model, a pressure term in the equation of motion can be included. The plasma is still treated with a fluid model, commonly referred to as the warm plasma model because effects such as Landau damping arising from the velocity distribution of particles are not taken into account.

The fluid equations are essentially a set of conservation equations. They have been used in Section 3.5 to study the electron and ion plasma waves. The equations are generalized here by inclusion of a Lorentz force term due to the presence of a steady magnetic field. The set of equations starts with the equation of continuity for the α th species of particles

$$\partial N_{\alpha}/\partial t + \operatorname{div}(N_{\alpha}\mathbf{v}_{\alpha}) = 0 \tag{4.18.1}$$

Then we have the equation of motion

$$m_{\alpha}N_{\alpha}D\mathbf{v}_{\alpha}/Dt = -\operatorname{grad} p_{\alpha} + Z_{\alpha}eN_{\alpha}(\mathbf{E} + \mathbf{v}_{\alpha} \times \mathbf{B}_{0}) \qquad (4.18.2)$$

the ideal gas law

$$p_{\alpha} = N_{\alpha} T_{\alpha} \tag{4.18.3}$$

and the equation of state

$$D(p_{\alpha}N_{\alpha}^{-\gamma})/Dt = 0$$
 (4.18.4)

The symbol $D/Dt = \partial/\partial t + \mathbf{v}_{\alpha} \cdot \mathbf{g}$ rad is just the convective derivative. The temperature T_{α} is expressed in energy units. As explained in Section 3.5, the adiabatic condition with $\gamma = 3$ is predicted by the more exact calculations for high frequency electron plasma waves. It is not expected to be valid for low frequency waves such as ion acoustic waves in which the electron thermal velocity may exceed the phase velocity of the wave. In this latter case we would expect any variation in temperature to be thermalized by the speedy electrons. Therefore, an isothermal condition with $\gamma = 1$ may exist for low frequency waves. For our purposes here, we shall not be too concerned with the exact value of γ and let it take up whatever values demanded by the more exact theory. Experimentally, propagation studies have been carried out in the laboratory and γ found to be close to 1 for ion sound waves in several plasmas (Alexeff and Jones, 1965).

Let the homogeneous plasma be perturbed so that

$$N_{\alpha} = N_{\alpha}^{(0)} + N_{\alpha}', \qquad p_{\alpha} = p_{\alpha}^{(0)} + p_{\alpha}'$$

$$\mathbf{v}_{\alpha} = \mathbf{0} + \mathbf{v}_{\alpha}, \qquad \mathbf{E} = \mathbf{0} + \mathbf{E}$$
(4.18.5)

The first terms on the right-hand side of (4.18.5) denote the unperturbed quantities and they are constant in a homogeneous plasma. The second terms on the right-hand side of (4.18.5) denote perturbations. The implication here is that under equilibrium the plasma is not in motion and is electrically neutral. For small perturbations, we may linearize all equations of concern. The linearization of the equation of state (4.18.4) with the help of the ideal gas law gives

$$p_{\alpha}' = \gamma T_{\alpha} N_{\alpha}' \tag{4.18.6}$$

Consequently, the linearized equations of continuity and motion are, respectively,

$$\partial N_{\alpha}'/\partial t + N_{\alpha}^{(0)} \operatorname{div} \mathbf{v}_{\alpha} = 0 \tag{4.18.7}$$

$$m_{\alpha}N_{\alpha}\partial \mathbf{v}_{\alpha}/\partial t = -\gamma T_{\alpha} \operatorname{grad} N_{\alpha}' + N_{\alpha}^{(0)}Z_{\alpha}e(\mathbf{E} + \mathbf{v}_{\alpha} \times \mathbf{B}_{0}) \qquad (4.18.8)$$

For plane waves with dependence $e^{j(\omega t - \mathbf{k} \cdot \mathbf{r})}$, the linear equations (4.18.7)

205

and (4.18.8) reduce to

$$\omega N_{\alpha}' - N_{\alpha}^{(0)} \mathbf{k} \cdot \mathbf{v}_{\alpha} = 0 \tag{4.18.9}$$

$$j\omega m_{\alpha} N_{\alpha} \mathbf{v}_{\alpha} = j\gamma T_{\alpha} N_{\alpha}' \mathbf{k} + N_{\alpha}^{(0)} Z_{\alpha} e(\mathbf{E} + \mathbf{v}_{\alpha} \times \mathbf{B}_{0}) \qquad (4.18.10)$$

The electric polarization density is given by

$$\mathbf{P}_{\alpha} = N_{\alpha}^{(0)} Z_{\alpha} e \boldsymbol{\xi}_{\alpha} = N_{\alpha}^{(0)} Z_{\alpha} e \mathbf{v}_{\alpha} / j \omega \qquad (4.18.11)$$

Eliminating N_{α}' from (4.18.9) and (4.18.10) and expressing \mathbf{v}_{α} in terms of the polarization density through the use of (4.18.11), we may obtain

$$\varepsilon_0 \mathbf{E} = -(1/X_{\alpha})[\mathbf{P}_{\alpha} + j\mathbf{Y}_{\alpha} \times \mathbf{P}_{\alpha} - \delta_{\alpha}\mathbf{nn} \cdot \mathbf{P}_{\alpha}] \qquad (4.18.12)$$

where by definition

$$X_{\alpha} = \omega_{p\alpha}^{2}/\omega^{2} = N_{\alpha}^{(0)}(Z_{\alpha}e)^{2}/m_{\alpha}\varepsilon_{0}\omega^{2}$$

$$Y_{\alpha} = \omega_{B\alpha}/\omega = -Z_{\alpha}eB_{0}/m_{\alpha}\omega$$

$$\delta_{\alpha} = \gamma T_{\alpha}/m_{\alpha}c^{2} = \nu_{T\alpha}^{2}/c^{2}$$

$$\mathbf{n} = \mathbf{k}/k_{0} = \mathbf{k}c/\omega$$
(4.18.13)

Since this is a nonrelativistic theory we implicitly assume $\delta \ll 1$. Equation (4.18.12) is of the form

$$\varepsilon_0 \mathbf{E} = \mathbf{\chi}_{\alpha}^{-1} \cdot \mathbf{P}_{\alpha} \tag{4.18.14}$$

where the inverse susceptibility can be identified. Without loss of generality, choose a coordinate system in which \mathbf{B}_0 is along the z-axis and **n** is in the xz-plane such as Fig. 4.8-1. That is, in this coordinate system $\mathbf{B}_0 = \hat{z}B_0$ and $\mathbf{n} = \hat{x}n \sin \theta + \hat{z}n \cos \theta$ where θ is the angle between **n** and \mathbf{B}_0 . The inverse susceptibility tensor is then

$$\mathbf{\chi}_{\alpha}^{-1} = -\frac{1}{X_{\alpha}} \cdot \begin{bmatrix} 1 - n^2 \,\delta_{\alpha} \sin^2\theta & -jY_{\alpha} & -n^2 \,\delta_{\alpha} \sin\theta \cos\theta \\ jY_{\alpha} & 1 & 0 \\ -n^2 \,\delta_{\alpha} \sin\theta \cos\theta & 0 & 1 - n^2 \,\delta_{\alpha} \cos^2\theta \end{bmatrix}$$
(4.18.15)

The susceptibility tensor is obtained from (4.18.15) by a matrix inversion and is found to be

$$\mathbf{X}_{\alpha} = -\frac{X_{\alpha}}{1 - Y_{\alpha}^{2} - n^{2} \,\delta_{\alpha} (1 - Y_{\alpha}^{2} \cos^{2} \theta)} \\ \times \begin{bmatrix} 1 - n^{2} \,\delta_{\alpha} \cos^{2} \theta & jY_{\alpha} (1 - n^{2} \,\delta_{\alpha} \cos^{2} \theta) & n^{2} \,\delta_{\alpha} \sin \theta \cos \theta \\ -jY_{\alpha} (1 - n^{2} \,\delta_{\alpha} \cos^{2} \theta) & 1 - n^{2} \,\delta_{\alpha} & -jY_{\alpha} n^{2} \,\delta_{\alpha} \sin \theta \cos \theta \\ n^{2} \,\delta_{\alpha} \sin \theta \cos \theta & jY_{\alpha} n^{2} \,\delta_{\alpha} \sin \theta \cos \theta & 1 - Y_{\alpha}^{2} - n^{2} \,\delta_{\alpha} \sin^{2} \theta \end{bmatrix}$$

$$(4.18.16)$$

206

The dielectric tensor for a warm plasma is then just

$$\boldsymbol{\varepsilon} = \varepsilon_0 (\mathbf{I} + \sum_{\alpha} \mathbf{X}_{\alpha}) = \varepsilon_0 \mathbf{K}$$
(4.18.17)

In the limit $n^2 \delta_{\alpha} \rightarrow 0$, the dielectric tensor (4.18.17) reduces to the cold plasma expression of (4.5.14) as expected. But this reduction is not possible near a resonance for which $n \rightarrow \infty$, even though δ_{α} is very small. The dielectric tensor (4.18.17) is ω -dependent as well as k-dependent through its dependence on **n**. The medium is therefore both temporally and spatially dispersive. The form of the dielectric tensor is identical to that of (2.6.10) which was derived from the Onsager relation. The propagation of plane waves is governed by the wave equation

$$\mathbf{D} \cdot \mathbf{E} = \mathbf{0} \tag{4.18.18}$$

with

$$\mathbf{D} = k^2 \mathbf{I} - \mathbf{k} \mathbf{k} - k_0^2 \mathbf{K}$$
(4.18.19)

The system of equations (4.18.18) has a unique nontrivial solution (outside of a multiplying constant) only if

$$\det \mathbf{D} = 0 \tag{4.18.20}$$

The dispersion relation (4.18.20) is an algebraic equation of order three in n^2 . The finding of these roots does not present any difficulties in principle. However, because of the large number of parameters the systematic study of the behavior of these roots when the parameters are varied is more tedious than the cold plasma case. We shall only discuss the high frequency case and one special example of ion-acoustic waves in the following two sections.

4.19 Warm Plasma Correction to the High Frequency Waves

In the high frequency approximation only electrons contribute to the dielectric constant of (4.18.17). The dispersion relation (4.18.20) reduces to, after long and tedious algebra,

$$(n^{2} - K_{0})\{(\delta/X)(K_{s} - 1)n^{4} + [K_{s} - (\delta/X)(K_{1}K_{11} - K_{s})]n^{2} - K_{1}K_{11}\}\sin^{2}\theta + (n^{2} - K_{1})(n^{2} - K_{11})[(\delta/X)(K_{0} - 1)n^{2} + K_{0}]\cos^{2}\theta = 0$$
(4.19.1)

$$K_{\rm I} = 1 - X/(1 + Y),$$
 $K_{\rm II} = 1 - X/(1 - Y)$
 $K_0 = 1 - X,$ $K_s = (K_{\rm I} + K_{\rm II})/2 = 1 - X/(1 - Y^2)$ (4.19.2)

With the help of (4.19.2) the dispersion relation (4.19.1) can also be written in the form

$$\delta(1 - Y^{2}\cos^{2}\theta)n^{6}$$
+ [(-1 + X + Y^{2} - XY^{2}\cos^{2}\theta) + 2\delta(-1 + X + Y^{2}\cos^{2}\theta)]n^{4}
+ [2 - 4X + 2X^{2} - 2Y^{2} + XY^{2}(1 + \cos^{2}\theta)
+ $\delta(1 - 2X + X^{2} - Y^{2}\cos^{2}\theta)]n^{2} + (1 - X)[Y^{2} - (1 - X)^{2}] = 0$
(4.19.3)

The form (4.19.1) is useful for studying special cases of $\theta = 0$ or $\theta = \pi/2$. When $\theta = 0$, the propagation vector is parallel to the steady magnetic field. In this case, (4.19.1) reduces to

$$n^{2} = \begin{cases} K_{\rm I} \\ K_{\rm II} \\ (1 - X)/\delta \end{cases}$$
 when $\theta = 0$ (4.19.4)

The first two modes are the characteristic electromagnetic waves discussed in Section 4.9 and the last mode is just the electron plasma waves of Section 3.5. We see that all three modes can propagate independently.

When the propagation vector is perpendicular to the steady magnetic field, i.e., $\theta = \pi/2$, the dispersion relation (4.19.1) reduces to

$$n^2 = 1 - X \tag{4.19.5a}$$

$$\delta n^4 + [(X-1)(1+\delta) + Y^2]n^2 + (1-X)^2 - Y^2 = 0 \quad (4.19.5b)$$

The refractive index given by (4.19.5a) is just that for the ordinary wave. But the extraordinary wave and the plasma wave are now coupled as clearly shown by (4.19.5b). The biquadratic equation (4.19.5b) has the form

$$an^4 + bn^2 + c = 0 \tag{4.19.6}$$

with $a = \delta \ll 1$. Since the discriminant is positive,

$$b^2 - 4ac = [(X - 1)(1 - \delta) + Y^2]^2 + 4 \,\delta XY > 0$$

208

The wave either propagates $(n^2 > 0)$ or attenuates $(n^2 < 0)$. The roots of (4.19.6) are given by

$$n^{2} = \frac{-b \pm (b^{2} - 4ac)^{1/2}}{2a}, \quad b > 0 \quad (4.19.7)$$

We shall explain the reason for giving the condition b > 0. If the inequality $b^2 \gg 4ac$ holds, the solutions (4.19.7) can be approximated as follows.

$$n^{2} = \begin{cases} -(c/b)(1 + ac/b + \cdots) \\ -(b/a) + (c/b) + (ac^{2}/b^{3}) + \cdots \end{cases}$$
$$= \begin{cases} 1 - \frac{(X + \delta)(1 - X)}{(1 - X)(1 + \delta) - Y^{2}} + O(\delta) & \text{extraordinary} \\ (1 - X - Y^{2})/\delta + XY^{2}/(1 - X - Y^{2}) + O(\delta) & \text{plasma (4.19.8b)} \end{cases}$$

In the limit $\delta \rightarrow 0$, (4.19.8a) reduces to that for the extraordinary wave in the cold plasma theory. Therefore, (4.19.8a) is defined as the expression for the extraordinary wave and (4.19.8b) as the expression for the plasma wave. When defined in this way, the upper sign in (4.19.7) corresponds to the extraordinary wave and the lower sign in (4.19.7) corresponds to the plasma wave only when b is positive. When b is negative the signs must be switched, i.e.,

$$n^{2} = \frac{|b| \mp (|b|^{2} - 4ac)^{1/2}}{2a}, \quad b < 0 \quad (4.19.9)$$

in which the upper sign refers to the extraordinary wave and the lower sign the plasma wave. The switch of expressions from (4.19.7) to (4.19.9) occurs when b passes from positive values through zero to negative values. Let us therefore examine the neighborhood of b = 0. We note that the condition b = 0 corresponds to

$$(1 - X)(1 + \delta) = Y^2 \tag{4.19.10}$$

which shows

$$1 - X > 0$$

Since $\delta \ll 1$, (4.19.10) is near $1 - X = Y^2$ which is just the upper hybrid resonance of the cold plasma and is shown in Fig. 4.16-1. When b = 0, we have

$$c = (1 - X)^2 - Y^2 = -((1 - X)X + \delta) < 0$$

and the refractive indices at this point are given by

$$n^2 = \pm (-c/\delta)^{1/2} = \pm [(1-X)(X+\delta)/\delta]^{1/2}, \quad b = 0$$
 (4.19.11)

The behavior of n^2 as a function of b is depicted in Fig. 4.19-1. As seen from this figure coupling from the extraordinary to the plasma wave occurs when $1 - X - Y^2 = \delta(X - 1) \simeq 0$ which can be called the quasi-resonance condition. This coupling of the extraordinary wave with the plasma wave is expected to modify the dispersion surfaces in Region III of the parameter space given by Fig. 4.17-6. Mathematically, n^2 varies smoothly with b. The switch in designation occurs at b = 0 because we wish to identify the modes by using limiting expressions (4.19.8).



Fig. 4.19-1. Coupling of the extraordinary wave with the plasma wave at $1 - X - Y^2 = \delta(X - 1)$ and $\theta = \pi/2$.

When θ is arbitrary, we must go back to (4.19.1) or (4.19.3). The following several properties can be deduced from (4.19.3). The only resonance occurs when

$$\cos \theta_{pr} = 1/Y$$
 or $\tan^2 \theta_{pr} = Y^2 - 1$ (4.19.12)

which is for the plasma wave. When $-1 + X + Y^2 - XY^2 \cos^2 \theta \neq 0$ or when

$$\tan^2\theta \neq -\frac{(1-X)(1-Y^2)}{1-X-Y^2}$$
(4.19.13)

the refractive index for the plasma wave which has the largest index is given by

$$n^{2} = (1 - X - Y^{2} + XY^{2}\cos^{2}\theta)/\delta(1 - Y^{2}\cos^{2}\theta) \qquad (4.19.14)$$

We note that the right-hand side of (4.19.13) is just $\tan^2 \theta_r$ predicted by the cold plasma theory as indicated (4.16.5). Therefore, the plasma expression (4.19.14) is valid when θ is not equal to the resonance angle of the cold plasma theory. In the parameter space of Fig. 4.17-6, only Regions III, VII, and VIII have resonances. The coupling of the plasma wave to the electromagnetic waves is expected to occur in these regions at the cold plasma resonance angle.

With the foregoing discussions we can go back to the parameter space and show the warm plasma modifications. Referring to Fig. 4.17-6 we make the following comments.

Regions I and *II*. The electromagnetic waves have refractive indices less than 1 while those of plasma waves are much larger than 1. Hence, there is very little coupling between them. In Fig. 4.19-2, only the index surface for the plasma wave is shown.

Region III. The extraordinary wave of the cold plasma theory has a resonance at θ_r given by (4.16.5), and at θ_r the coupling to plasma waves is expected from the warm plasma theory. The warm plasma resonance is given by (4.19.12). Since in this region $Y^2 < 1$, the surface must be closed as shown in Fig. 4.19-2.



Fig. 4.19-2. Warm plasma additions and corrections of the dispersion surfaces in the parameter space. Surfaces unaffected are not shown. Coupling from the electromagnetic wave to the plasma occurs at the resonance angle θ_r . The resonance angle θ_{pr} is that for the plasma wave.

212 4. Waves in Fluid Plasma with a Steady Magnetic Field

Regions IV and V. The plasma wave attenuates as indicated by (4.19.14).

Region VI. Except near bounding curves of this region, the refractive indices of both electromagnetic characteristic waves are finite according to the cold plasma theory. There is therefore very little coupling with the plasma wave except near Y = 1. The refractive index for the plasma wave is (4.19.14) and has a resonance given by (4.19.12) as shown in Fig. 4.19-2.

Regions VII and VIII. The ordinary wave has a resonance and according to the cold plasma theory the resonance angle is [see (4.16.5)]

$$\tan^{-1}((Y^2-1)(X-1)/(X+Y^2-1))^{1/2} < \theta_{pr}$$
 (4.19.15)

Therefore, the ordinary wave is coupled to the plasma wave which has a larger resonance angle. The effect can be seen in Fig. 4.19-2.

4.20 Plasma Waves and Two-Stream Instabilities

The warm plasma corrections including the effects of ions are rather complicated and they also depend on the composition of the plasma. We shall not discuss the general case, but only concern ourselves with the plasma waves. The refractive index for the plasma wave is very large. When the dispersion relation (4.18.20) is written in the form

$$a_4n^4 + a_2n^2 + a_0 = 0 \tag{4.20.1}$$

the approximate dispersion relation for the plasma wave is given by setting (See Section 2.9)

 $a_{4} = 0$

or

$$K_{xx}\sin^2\theta + 2K_{xz}\cos\theta\sin\theta + K_{zz}\cos^2\theta = 0 \qquad (4.20.2)$$

The expression (4.20.2) for a_4 has been found in (2.8.12). Because this is a warm plasma theory, the elements of the dielectric tensor are *n*-dependent as well as ω -dependent. Substituting these elements as given by (4.18.17) into (4.20.2), we obtain the following dispersion relation for plasma waves

$$1 - \sum_{\alpha} \frac{X_{\alpha}(1 - Y_{\alpha}^{2} \cos^{2}\theta)}{1 - Y_{\alpha}^{2} - n^{2} \,\delta_{\alpha}(1 - Y_{\alpha}^{2} \cos^{2}\theta)} = 0 \qquad (4.20.3)$$

In the following, for simplicity, let us assume that the plasma is composed of electrons and singly charged positive ions.

When $\theta = 0$, the dispersion relation (4.20.3) reduces to

$$1 - \frac{X_e}{1 - n^2 \,\delta_e} - \frac{X_i}{1 - n^2 \,\delta_i} = 0 \tag{4.20.4}$$

The dispersion relation (4.20.4) is identical to (3.5.21) from which we deduced expressions for electron plasma waves and ion sound waves. The reader is referred to Section 3.5 for details.

When $\theta = \pi/2$, (4.20.3) reduces to

$$1 - \frac{X_e}{1 - Y_e^2 - n^2 \,\delta_e} - \frac{X_i}{1 - Y_i^2 - n^2 \,\delta_i} = 0 \qquad (4.20.5)$$

The dispersion relation (4.20.5) is a biquadratic equation in n^2 and can be solved easily. Because of the large ion-to-electron mass ratio, we can get approximate expressions by letting $m_i \rightarrow \infty$ for electron plasma waves and $m_e \rightarrow 0$ ionic sound waves. The refractive indices in these limits are

$$n^2 = (1 - X_e - Y_e^2)/\delta_e$$
 for electron plasma waves (4.20.6)

and

$$n^{2} = -\frac{Y_{e}^{2}(1 - Y_{i}^{2} - X_{i}) + X_{e}}{\delta_{i}(Y_{e}^{2} + X_{e})} \quad \text{for ionic sound waves} \qquad (4.20.7)$$

We note that the electron plasma wave has a cutoff at the upper hybrid resonance and the ionic sound wave has a cutoff at the lower hybrid resonance. The hybrid resonances were discussed in Section 4.12.

In ionospheric applications, one problem of special interest is the excitation of ion sound waves. We have already seen in Section 3.7 that the energy in the streaming motion may be fed to the growth of plasma waves. These waves have associated density perturbations. In radio science these density fluctuations are called irregularities because they scatter radio waves and they also cause radio signals to scintillate. These irregularities have been observed throughout the ionosphere for all latitudes and longitudes. But one type of irregularities which occur at a height 100 km near the equator in daytime has been convincingly proved experimentally as caused by two stream instabilities.

Near the equator a strong current known as the equatorial electroject flows in a height range of 5 to 10 km centered about a height of 105 km and within $2-3^{\circ}$ in latitude from the magnetic dip equator. The current in the electroject is driven by the electric field which is believed to be generated by the tidal motion of the earth's atmosphere. Because of the large ion mass we may assume ions are stationary. The electroject current is then entirely caused by drifting electrons. Let the electron streaming velocity be $\mathbf{v}^{(0)}$. In the nonrelativistic theory, the effect of motion is a Doppler shift in ω and invariance in \mathbf{k} when transformed from the rest frame to the moving frame of reference with respect to the medium. The electrons see a wave whose frequency is Doppler shifted to $\omega - \mathbf{k} \cdot \mathbf{v}^{(0)}$. At the height of interest, typical parameters of interest are of the following orders.

$$v_{en} \sim 10^5, \quad \omega_{Be} \sim 10^7, \quad \omega_{pe} \sim 10^7 \\ v_{in} \sim 10^4, \quad \omega_{Bi} \sim 10^2, \quad \omega_{pi} \sim 10^5$$

$$(4.20.8)$$

As seen the collisional effects are not negligible. The inclusion of frictionlike collisional effects in the equation of motion (4.18.2) is equivalent to replacing X_{α} , Y_{α} , and δ_{α} in the collisionless theory by X_{α}/U_{α} , Y_{α}/U_{α} , and $\delta_{\alpha}/U_{\alpha}$, respectively. The inclusion of collisions and a Doppler shift for electrons but not for ions and the neglect of ω_{Bi} in comparison with v_{in} and ω modify the dispersion relation (4.20.3) which can be put in the following form.

$$1 + \frac{\omega_{pe}^{2}}{j\omega_{D}(j\omega_{D} + v_{en}) \cdot \frac{\omega_{Be}^{2} + (j\omega_{D} + v_{en})^{2}}{\omega_{Be}^{2} \cos^{2}\theta + (j\omega_{D} + v_{en})^{2}} + k^{2}v_{Te}^{2}} + \frac{\omega_{pi}^{2}}{j\omega(j\omega + v_{in}) + k^{2}v_{Ti}^{2}} = 0$$
(4.20.9)

where ω_D is the Doppler shifted frequency, i.e., $\omega_D = \omega - \mathbf{k} \cdot \mathbf{v}^{(0)}$. In general ω is complex for a real k in (4.20.9). Remember that the assumed time dependence is $e^{j\omega t}$. When ω has a positive imaginary part the wave is damped exponentially with time. But when ω has a negative imaginary part, the wave will grow in time. The transition between damping and growth occurs when ω is real for real **k**. This condition is known as the condition for marginal stability. In the following, we shall find this condition.

The parameters of interest have the magnitudes given by (4.20.8). We may assume $\omega_D^2 \ll v_{en}^2 \ll \omega_{Be}^2$ but for generality we allow θ to take any values. The factor that appears in (4.20.9) may be then approximated.

$$\frac{(j\omega_{D} + v_{en})[\omega_{Be}^{2} + (j\omega_{D} + v_{en})^{2}]}{\omega_{Be}^{2}\cos^{2}\theta + (j\omega_{D} + v_{en})^{2}} = \frac{\omega_{Be}^{2}v_{en}}{\omega_{Be}^{2}\cos^{2}\theta + v_{en}^{2}} + j\frac{\omega_{D}\omega_{Be}^{2}(\omega_{Be}^{2}\cos^{2}\theta - v_{en}^{2})}{(\omega_{Be}^{2}\cos^{2}\theta + v_{en}^{2})^{2}}$$

Problems

We substitute this approximation into (4.20.9), and the imaginary part of the equation at the marginal stability, for which ω and hence ω_D are real. The following equation results

$$\frac{\omega_D}{\omega} = -\left(\frac{\omega_{pe}^2 \nu_{in}}{\omega_{pi}^2 \nu_{en}}\cos\theta + \frac{\nu_{in}\nu_{en}}{\omega_{Bi}\omega_{Be}}\right)$$
(4.20.10)

Because of the presence of collisions the Doppler shifted frequency per frequency for marginal stability must be negative. Equation (4.20.10) can also be reexpressed as

$$\frac{\nu^{(0)}}{(\omega/k)} = 1 + (\nu_{in}\nu_{en}/|\omega_{Bi}|\omega_{Be}) + (\omega_{pe}^2\nu_{in}/\omega_{pi}^2\nu_{en})\cos^2\theta \quad (4.20.11)$$

which states that the marginal stability occurs only when the streaming velocity of electrons is larger than the phase velocity of the wave. Because of the occurrence of the large factor $\omega_{pe}^2/\omega_{pi}^2$, the required streaming velocity is extremely large unless the angle θ is very close to $\pi/2$. For moderate values of the streaming velocity, excitation of ion sound waves occurs only for θ near $\pi/2$. The phase velocity of the wave at the marginal stability can be found by substituting (4.20.10) back to the dispersion relation. After rearrangement, we obtain

$$(\omega/k)^{2} = 2\nu_{Ti}^{2}/[1 + (\omega_{pe}^{2}\nu_{in}^{2}/\omega_{pi}^{2}\nu_{en}^{2})\cos^{2}\theta - \nu_{in}^{2}/\omega_{pi}^{2} - \nu_{in}^{2}/\omega_{Be} |\omega_{Bi}|]$$
(4.20.12)

Roughly, the phase velocity is of the order of ion thermal velocity for $\theta = \pi/2$. As the angle θ departs from $\pi/2$, the phase velocity is reduced.

The fact that the instability is most easily achieved when k is perpendicular to \mathbf{B}_0 means that the irregularities associated with the longitudinal ion sound waves are magnetic field aligned. The experimental verification of these theoretical predictions has been quite successful [see Buneman (1963); Farley (1963); Cohen and Bowles (1963)].

Problems

1. In the absence of a steady magnetic field, the relative dielectric constant for a cold plasma is (4.1.14) which has the real part given by (4.1.17), the negative imaginary part by (4.1.18). Show by actually carrying out the improper integrals that they satisfy the Kramers-Kronig relations (2.5.15). 2. For some time there was a question whether the Lorentz polarization term should be included for the plasma case. The Lorentz polarization term comes about because the effective electric field at any point in the medium is the sum of the applied electric field and the electric field radiated by other dipoles in the medium. The inclusion of the Lorentz term is the addition on the right-hand side of the equation of motion (4.1.1) of a term $-e\mathbf{P}/3\varepsilon_0$. Show that if this is done, the refractive index must now satisfy

$$n^2 = 1 - 3\omega_p^2/(3\omega^2 + \omega_p^2)$$

which has a corresponding cutoff condition at $\omega_p^2 = 3\omega^2/2$. Discuss its consequence on measuring electron density (viz., ω_p) by using the propagation technique.

3. Let a uniform plane wave with electric field $\mathbf{E} = \hat{x}E_0 e^{j(\omega t - k_0 z)}$ strike normally on a slab of lossless isotropic plasma of thickness $\Delta z'$ at z'. Find the scattered electric and magnetic fields from such a slab.

4. Now let us assume that electrons are distributed uniformly throughout the space. Imagine that the space is made up of many thin slabs, each of which will scatter waves like that shown in the previous problem. We require that these infinitely many thin slabs scatter coherently in a self-consistent manner so that the sum of all these scattered waves just makes up the originally assumed plane wave. Show that the condition of self-consistency requires the refractive index to satisfy $n^2 = 1 - \omega_p^2/\omega$.

5. Consider the normal incidence of a turned-on sinusoidal wave at the lossless, isotropic plasma half-space of the form given by (4.2.8), i.e.,

$$E_i(t) = u(t) \sin \omega_0 t$$

Show that the reflected wave at the interface is given by (4.2.9). The following identity will be helpful

$$\sin(z\cos\theta) = 2\sum_{n=0}^{\infty} (-1)^n J_{2n+1}(z)\cos[(2n+1)\theta]$$

in putting $E_i(t)$ into a convenient form [C. M. Knop, Further comments on "The transient phenomenon in an isotropic plasma without collisional loss." *Proc. IEEE* 53, 751–752 (1965)].

6. Verify the expressions (4.3.22) and (4.3.23) by applying the method of steepest descent.

Problems

7. Show that the equation of magnetic lines of force is given by

$$r/\sin^2 \Lambda = \text{constant}$$

for a dipole field. The lines orthogonal to the magnetic lines of force are constant potential lines. (Note that if the small contribution to the earth field from ionospheric currents is ignored, the magnetic field exterior to earth is irrotational and we may define a scalar potential whose negative gradiant is the magnetic field.) Show that the constant potential lines are given by

$$r^2/\sin \Lambda = \text{constant}$$

8. Find the expression for the dc conductivity tensor from (4.6.5) by a transformation of coordinates so that \mathbf{B}_0 is in the *xz*-plane.

In the study of dynamo theory, it is sometimes assumed that the vertical current is zero, i.e., $i_z = 0$. This assumption comes from the vague inference that currents flow mainly in the *E* region of the ionosphere when conductivity is appreciable and any vertical current is inhibited by the polarization field. Show that if this is the case we then have

$$egin{aligned} & i_x = ar{\sigma}_{xx}E_x + ar{\sigma}_{xy}E_y \ & i_y = ar{\sigma}_{xy}E_x + ar{\sigma}_{yy}E_y \end{aligned}$$

Express the elements $\bar{\sigma}_{xx'}$, $\bar{\sigma}_{xy'}$, $\bar{\sigma}_{xy}$, and $\bar{\sigma}_{yy}$ in terms of σ_0 , σ_P , σ_H , and the polar angle of the magnetic field.

9. Flow of a neutral plasma in a semiconductor in the presence of an external magnetic field can lead to the appearance of the magnetic moment of the plasma. Consider a cylindrical sample of radius R in which radial diffusion takes place such that the neutrality condition is maintained under the steady-state condition but an azimuthal current may be produced by unequal azimuthal velocities of electrons and holes. Suppose the cylindrical sample is situated in an axial external magnetic field. Find the average axial magnetic moment density [A. A. Vedenov, Solid state plasma. Sov. Phys. Usp. 1, 809–822 (1965); A. R. Moore and J. O. Kessler, Phys. Rev. 132, 1494 (1963)].

10. Consider a multicomponent cold magnetoplasma in which \mathbf{B}_0 , \mathbf{k} , and $\mathbf{v}_{\alpha}^{(0)}$ are all in the same direction. Here $\mathbf{v}_{\alpha}^{(0)}$ is the streaming velocity of α th kind of particles. Find the dispersion relations for all three longitudinal modes. Consider the special example of electrons streaming through positive ions at rest. What is the condition for marginal instability?

218 4. Waves in Fluid Plasma with a Steady Magnetic Field

11. In Section 4.10 we discussed the Faraday effect for characteristic waves propagating parallel to the magnetic field. Suppose there are equiamplitude ordinary and extraordinary waves propagating across the steady magnetic field. What happens to the transverse wave polarization of the resultant along the direction of propagation? Draw polarization ellipses to illustrate their change.

12. In most of ionospheric experiments the total Faraday rotation from the transmitter to the receiver is not measured but only its time rate of change. Consider a transmitter (a beacon satellite) which moves parallel to the plane earth at a constant velocity v. The plasma density between the transmitter and ground may be assumed homogeneous in a constant steady magnetic field. Show that the time rate of change of Faraday rotation of a high frequency signal is constant and is given by

$$d\Omega/dt = 2.97 \times 10^{-2} N \nu H_{\text{path}}/f^2$$
 rad/sec

where H_{path} is the component of the steady magnetic field resolved in the direction parallel to the path of the satellite. In the high frequency approximation the refractive index is given by the quasi-parallel expression for almost all directions [S. A. Bowhill, The Faraday rotation rate of a satellite radio signal. J. Atmos. Terr. Phys. 13, 175 (1956)].

13. Consider a slab of plasma in a steady magnetic field as shown. Suppose that we displace all electrons by a small distance ξ_x from their equilibrium



positions. Show that the system will have a resonant frequency ω satisfying

$$\omega^2 = \omega_{pe}^2 + \omega_{Be}^2$$

This is just the upper hybrid resonance.

14. Consider resonances for the special case of perpendicular propagation in a three-component plasma consisting of electrons and two neutralizProblems

ing positive ion species with masses M_1 and M_2 ($M_2 > M_1$) and fractional concentrations A_1 and A_2 ($A_1 + A_2 = 1$).

(a) Show that in the limit of low densities, resonances occur at each of the gyroresonances. In the limit of high densities, the square of the angular resonant frequencies are

 ω_{pe}^2 , $\omega_{Be}(A_1 | \omega_{Bi1} | + A_2 | \omega_{Bi2} |)$

and

 $|\omega_{Bi1}\omega_{Bi2}| (A_1 | \omega_{Bi2} | + A_2 | \omega_{Bi1} |)/(A_1 | \omega_{Bi1} | + A_2 | \omega_{Bi2} |)$

Note that the third new resonance involves only the ion gyrofrequencies.

(b) At the new resonance, the oscillations of two ion clouds perpendicular to the steady magnetic field are 180° out of phase, while electrons remain relatively motionless. Show that this is the case [S. J. Buchsbaum, *Phys. Fluids* 3, 418-420 (1960)].

15. Consider a two-component plasma with electrons and neutralizing positive ions. In the high density limit we find, in Section 4.12, a resonance ω_0 at the geometric mean of electron and ion gyrofrequency when k is perpendicular to B_0 and when collisions are all ignored.

(a) Take collisional damping in a form of a frictional force into account. Find the shift in resonance frequency from ω_0 when ν/ω_0 is much smaller than 1 and much greater than 1.

(b) Neglecting shift due to damping, find the real part and the imaginary part of the refractive index near $\omega = \omega_0$ still for k perpendicular to \mathbf{B}_0 .

(c) Neglecting damping entirely, discuss the dependence of resonant frequency on the direction of propagation which departs from the exact perpendicular condition by a small angle [H. Schlitter and C. J. Ransom, *Ann. Phys.* (New York) **33**, 360–380 (1965)].

16. The dispersion relation for ion whistlers near the ion gyrofrequency is given by (4.13.14).

(a) Show that the angle ψ between the group ray and **B**₀ of an ion whistler is given by

$$\tan \psi = \sin \theta \cos^3 \theta / (1 + \cos^4 \theta)$$

(b) From (a) show that the largest value of ψ is 12.3°.

(c) Find the group velocity of an ion whistler propagating along \mathbf{B}_0 [D. A. Gurnett and S. D. Shawhan, Determination of hydrogen ion concentration, electron density, and proton gyrofrequency from the dispersion of proton whistlers. J. Geophys. Res. 71, 741-754 (1966)].

17. A special case which is sometimes of interest is the strong magnetic field case. When the steady magnetic field is strong, it is convenient to expand the expression in powers of $1/Y_{\alpha}$.

(a) Expand the polarization density \mathbf{P}_{α} in a cold magnetoplasma to the second order in $1/Y_{\alpha}$. Show that \mathbf{P}_{α} has contributions from three terms. The first term comes from the electric field induced motion along the steady magnetic field. The second term has the origin of $\mathbf{E} \times \mathbf{B}_0$ drift and does not contribute to the total polarization since the drift is the same for charged particles of different masses. The third term comes from the polarization drift. In the orbit theory the polarization drift is given by $(m_{\alpha}/Z_{\alpha}eB_0^2)\partial \mathbf{E}/\partial t$.

(b) Show that the dielectric constant accurate to the second order in 1/Y is diagonal and is given by

$$m{\epsilon} = arepsilon_0 egin{bmatrix} 1 + c^2 / {v_A}^2 & 0 & 0 \ 0 & 1 + c^2 / {v_A}^2 & 0 \ 0 & 0 & 1 - \sum\limits_lpha X_lpha \end{bmatrix} + Oigg(rac{1}{Y^3}igg)$$

(c) Find the refractive indices in this medium.

18. Show that (4.8.9) reduces to the Appleton-Hartree formula (4.15.19) in the high frequency approximation in which only electrons contribute to the polarizability of the medium.

19. Show by actually differentiating the Appleton-Hartree formula that (4.15.13) and (4.15.14) are identical.

20. The propagation of electron whistlers under the quasi-parallel condition is described by the refractive index (4.15.9). Show that the group ray must be within 19.5° of the direction of the magnetic field [L. R. O. Storey, An investigation of whistling atmospherics. *Phil. Trans. Roy. Soc. London* Ser. A 246, 113 (1953)].

21. Consider a wave incident normally on a stratified lossless magnetoplasma such as vertical sounding of an ionosphere. In this case the direction of the wave vector is known as the wave penetrates into the plasma. Hence knowing tan α would be sufficient to trace the group ray in the medium. Show that the reflecting ray at X = 1 is perpendicular to the steady magnetic field. Also find the direction of a ray at the point of reflection when X = 1 - Y and X = 1 + Y.

22. Consider a medium in which the constitutive relations are given by

$$\mathbf{D} = \varepsilon \mathbf{E} + \alpha \mathbf{H}, \qquad \mathbf{B} = \beta \mathbf{E} + \mu \mathbf{H}$$

Find the polarization and the dispersion relation of a characteristic wave that can propagate in the medium. In this medium it is also possible to have Faraday rotation. How does the Faraday rotation in this medium differ from that in the magnetoplasma? Especially for waves making a round trip through the medium?

23. The force acting on a moving charge by an electromagnetic wave can be written as

$$\mathbf{F}(\mathbf{r}, t) = e[\mathbf{E}(\mathbf{r}, t) + \mathbf{v}(\mathbf{r}, t) \times \mathbf{B}(\mathbf{r}, t)]$$

where **E** and **B** are the wave fields and **v** is the velocity of the charge. To the linear approximation, **v** can be taken as the unperturbed trajectory of the particle. The fields are plane waves of the form $e^{j(\omega t - \mathbf{k} \cdot \mathbf{r})}$.

(a) For the case without static magnetic field, the unperturbed trajectory of the charge particle is $\mathbf{r} = \mathbf{r}_0 + \mathbf{v}_0 t$, where \mathbf{r}_0 and \mathbf{v}_0 are the initial position and velocity of the charge, respectively. Find the condition such that the force acting on the charge remains constant in time. This is the resonance condition. Is it physically possible to have resonance under the present situation?

(b) With a static magnetic field \mathbf{B}_0 in the z-direction and the k vector in the xz-plane, find the resonance conditions. (The formula $e^{ia \sin x} = \sum_{n=-\infty}^{+\infty} J_n(a)e^{inx}$ will be found useful.)

References

- ALEXEFF, I., and JONES, W. D. (1965). Collisionless ion-wave propagation and the determination of the compression coefficient of plasma electrons. *Phys. Rev. Lett.* 15, 286–288.
- ALLIS, W. P., BUCHSBAUM, S. J., and BERS, A. (1962). "Waves in Plasmas." Wiley, New York.
- ÄSTROM, E. (1950). On waves in an ionized gas. Ark. Fys. 2, 443-457.
- BRILLOUIN, L. (1960). "Wave Propagation and Group Velocity." Academic Press, New York.
- BUDDEN, K. G. (1961). "Radio Waves in the Ionosphere." Cambridge Univ. Press, London and New York.
- BUNEMAN, O. (1963). Excitation of field aligned sound waves by electron streams. *Phys. Rev. Lett.* **10**, 285-287.
- CHAPMAN, J. H., and WARREN, E. S. (1968). Topside sounding of the Earth's ionosphere. Space Sci. Rev. 8, 846-865.
- CLEMMOW, P. C. and MULLALY, R. F. (1955). Dependence of the refractive index in magneto-ionic theory on the direction of the wave normal. Physics of the Ionosphere. *Rep. Phys. Soc. Conf., Cavendish Lab.*, 1955, p. 340.

- COHEN, R., and BOWLES, K. L. (1963). The association of plane-wave electron-density irregularities with the equatorial electroject, J. Geophys. Res. 68, 2503.
- DARWIN, C. G. (1943). The refractive index of an ionized medium. Proc. Roy. Soc. Ser. A 182, 152-166.
- DESCHAMPS, G. A. (1965). Angular dependence of the refractive index in the ionosphere. J. Res. Nat. Bur. Stand. Sect. D 69, 395-400.
- DESCHAMPS, G. A., and WEEKS, W. L. (1962). Charts for computing the refractive indexes of a magneto-ionic medium. *IRE Trans. Antennas Propagat.* AP-10, 305-317.
- FARLEY, D. T., JR. (1963). A plasma instability resulting in field-aligned irregularities in the ionosphere, J. Geophys. Res. 68, 6083.
- GURNETT, D. A., SHAWHAN, S. D., BRICE, N. M., and SMITH, R. L. (1965). Ion cyclotron whistlers, J. Geophys. Res. 70, 1665–1668.
- HASKELL, R. E., and CASE, C. T. (1967). Transient signal propagation in lossless, isotropic plasma, *IEEE Trans. Antennas Propagat.* AP-15, 458-464.
- HELLIWELL, R. A. (1965). "Whistlers and Related Ionospheric Phenomena," Stanford Univ. Press, Stanford, California.
- KADOMTSEV, B. B. (1958). On the effective field in a plasma, Sov. Phys. JEPT 6, 117.
- KNOP, C. M. (1964). Pulsed electromagnetic wave propagation in dispersive media. *IEEE Trans. Antennas Propagat.* AP-12, 494-496.
- KNOP, C. M. (1965). Further comments on "The transient phenomenon in an isotropic plasma without collisional loss." *Proc. IEEE* 53, 751-752.
- McEwen, D. J., and BARRINGTON, R. E. (1968). Ion composition below 3000 km derived from ion whistler observations. "Space Research VIII," pp. 396–404. North-Holland Publ. Amsterdam.
- MAEDA, K., and KATO, S. (1966). Electrodynamics of the ionosphere. Space Sci. Rev. 5, 57-79.
- NICOLET, M. (1959). Collisional frequency of electrons in the terrestrial atmosphere. *Phys. Fluids*, **2**, 95-99.
- RATCLIFFE, J. A. (1959). "The Magnetoionic Theory and Its Applications to the Ionosphere," Cambridge Univ. Press, London and New York.
- RAWER, K., and SUCHY, K. (1967). Radio observations of the ionosphere. In "Handbuch der Physik," (S. Flugge, ed.), Vol. XLIX/2, "Geophysics" 111, Part 11, pp. 1–537. Springer-Verlag, Berlin and New York.
- SOMMERFELD, A. (1914). About the propagation of light in dispersive media. Ann. Physik 44, 177-202.
- STIX, T. H. (1962). "Theory of Plasma Waves," McGraw-Hill, New York.
- THEIMER, O., and TAYLOR, L. S. (1961). On the index of refraction in the ionosphere. J. Geophys. Res. 66, 3157-3162.
- VAN DER WAERDEN, B. L. (1950). On the method of saddle points. Appl. Sci. Res. B 2, 33-45.
- WAIT, J. R. (1969). Reflection of a plane transient electromagnetic wave from a cold lossless plasma slab. *Radio Sci.* 4, 401-405.

5. Wave Propagation in Inhomogeneous Media

5.1 Introduction

In previous chapters our discussion on wave propagation in dispersive media in general and in plasmas in particular has been limited to homogeneous structures. In reality, however, the phenomenon of wave propagation in inhomogeneous media occurs quite often; in fact it occurs much more frequently than the phenomenon of propagation in homogeneous media. To name just a few examples, we list the following natural phenomena:

- (1) underwater propagation of acoustic waves;
- (2) acoustic-gravity waves in the ocean and the atmosphere;
- (3) multilayer optics;
- (4) seismic waves;
- (5) radio waves in the ionosphere.

Therefore the problem of wave propagation in inhomogeneous media is indeed a very important one both theoretically and experimentally. As is the case for most real physical problems, the wave propagation problems under the most general inhomogeneous conditions are hopelessly complicated and usually no meaningful solutions (analytic or numerical) can be obtained. Therefore the first task in treating these problems is to isolate different classes of problems and to study various limiting conditions of real physical situations. The hope is that by setting up different models, physical situations that are responsible for different behavior of the waves may be studied independently, and meaningful conclusions may be drawn from these investigations. In setting up these models, one has to be very careful, since in many cases oversimplifications may hide clues to solution of the real physical problem (Karbowiak, 1967).

One of the limiting conditions for wave propagation in inhomogeneous medium is the geometrical optics approximation, the limit of zero wavelength. In Sections 5.2 to 5.7, we shall discuss various aspects of this particular limiting case. A set of ray equations, which are suitable for numerical computations, will be derived for general inhomogeneous, anisotropic media.

In many real physical situations, the properties of the medium may be taken as varying in only one particular direction. This type of media is called stratified media. Sections 5.8–5.20 are devoted to problems of wave propagation in such a medium. The WKB approximation technique which is valid for high frequency waves is discussed in Sections 5.9–5.11 and is applied to the stratified isotropic media in Sections 5.12 and 5.13. Another approximation method which is valid for the other extreme of very low frequency waves is discussed in Sections 5.15 and 5.16, the signal propagation problem is considered. In Sections 5.17–5.20, stratified anisotropic media are discussed. Although we have the application of ionospheric propagation in mind, the discussions are kept in fairly general terms so that the techniques are applicable to other problems of a similar mathematical nature but of entirely different physical situations.

Since the problem of wave propagation in inhomogeneous media is already very complex, we shall not make the discussion even more difficult by introducing spatial dispersion. Throughout this chapter the relations between D and E fields will be assumed to be local.

Finally, we mention in passing that there is another set of problems belonging to the general problems of wave propagation in inhomogeneous media, namely, wave propagation in random media. For these media, the inhomogeneities vary randomly. This problem will be considered in the next chapter.

5.2 Foundations of Geometrical Optics—Isotropic Media

One of the most successful ways of treating the problem of wave propagation in an inhomogeneous medium is the method of geometrical optics. It is the branch of optics which is characterized by taking the limit of zero wavelength in investigating propagation of electromagnetic waves. Physically, this amounts to neglecting the diffractional effects. The historical development of geometrical optics goes back to the nineteenth century and its relation to wave optics is analogous to the relation between classical mechanics and quantum mechanics. Recent advances of the theory have generalized the method a great deal and linked it closely to the technique of asymptotic expansions [see for example, Kline and Kay (1965)]. In this section we shall derive the fundamental equations of geometrical optics from Maxwell's equations and indicate the limitations and implications of the results.

Let us consider a lossless, isotropic, spatially inhomogeneous medium. The medium is assumed to be frequency dispersive but with no spatial dispersion. For monochromatic harmonic time variation $e^{j\omega t}$, the Maxwell equations for a source free region become

$$\nabla \times \mathbf{E}(\mathbf{r},\omega) = -j\omega\mu(\mathbf{r},\omega)\mathbf{H}(\mathbf{r},\omega)$$
(5.2.1a)

$$\nabla \times \mathbf{H}(\mathbf{r},\omega) = j\omega\varepsilon(\mathbf{r},\omega)\mathbf{E}(\mathbf{r},\omega)$$
(5.2.1b)

$$\nabla \cdot [\varepsilon(\mathbf{r}, \omega) \mathbf{E}(\mathbf{r}, \omega)] = 0 \qquad (5.2.1c)$$

$$\nabla \cdot [\mu(\mathbf{r}, \omega) \mathbf{H}(\mathbf{r}, \omega)] = 0$$
 (5.2.1d)

where $\varepsilon(\mathbf{r}, \omega)$ and $\mu(\mathbf{r}, \omega)$ are functions of spatial coordinates indicating the inhomogeneous properties of the medium.

In the homogeneous medium, the plane wave solution of (5.2.1) is of the form

$$\mathbf{E} = \mathbf{e} e^{-jk_0 n\hat{s} \cdot \mathbf{r}}$$

$$\mathbf{H} = \mathbf{h} e^{-jk_0 n\hat{s} \cdot \mathbf{r}}$$
 (5.2.2)

where $k_0 = \omega/c$ and *n* is the refractive index defined by

$$n = c(\varepsilon \mu)^{1/2} \tag{5.2.3}$$

and \hat{s} is the direction of propagation. **e** and **h** are two constant vectors. For the inhomogeneous medium, we can still define the refractive index as in (5.2.3) and write the solution of (5.2.1) as

$$\mathbf{E} = \mathbf{e}(\mathbf{r})e^{-jk_0\psi(\mathbf{r})}$$

$$\mathbf{H} = \mathbf{h}(\mathbf{r})e^{-jk_0\psi(\mathbf{r})}$$
 (5.2.4)

where the function $\psi(\mathbf{r})$ is a real scalar function of position and $\mathbf{e}(\mathbf{r})$ and $\mathbf{h}(\mathbf{r})$ are vector functions of position and may be complex. Note that the ω dependence has been dropped to save unnecessary writing.

Substituting (5.2.4) into (5.2.1), we obtain

$$\nabla \psi \times \mathbf{e} - c\mu \mathbf{h} = (\nabla \times \mathbf{e})/(jk_0)$$
 (5.2.5)

$$\nabla \psi \times \mathbf{h} + c\varepsilon \mathbf{e} = (\nabla \times \mathbf{h})/(jk_0) \tag{5.2.6}$$

$$\mathbf{e} \cdot \nabla \psi = (\mathbf{e} \cdot \nabla \ln \varepsilon + \nabla \cdot \mathbf{e})/(jk_0) \tag{5.2.7}$$

$$\mathbf{h} \cdot \nabla \psi = (\mathbf{h} \cdot \nabla \ln \mu + \nabla \cdot \mathbf{h})/(jk_0)$$
 (5.2.8)

The r dependence of the functions has been omitted in writing (5.2.5)-(5.2.8). Up to this point, no approximation has been made. For geometrical optics, we are interested in the case where the wavelength is approaching zero so that k_0 is approaching infinity. We can therefore neglect the righthand side of (5.2.5)-(5.2.8) since they are of the order $(1/k_0)$. These equations then reduce to

$$\nabla \psi \times \mathbf{e} - c\mu \mathbf{h} = 0 \tag{5.2.5a}$$

$$\nabla \psi \times \mathbf{h} + c\varepsilon \mathbf{e} = 0 \tag{5.2.6a}$$

$$\mathbf{e} \cdot \nabla \psi = 0 \tag{5.2.7a}$$

$$\mathbf{h} \cdot \nabla \psi = 0 \tag{5.2.8a}$$

We note that (5.2.7a) and (5.2.8a) can be derived from (5.2.5a) and (5.2.6a), respectively. Therefore we shall only concentrate on the two equations (5.2.5a) and (5.2.6a). Substituting **h** from (5.2.5a) into (5.2.6a), we have

$$(1/c\mu)[\nabla \psi \times (\nabla \psi \times \mathbf{e})] + c\varepsilon \mathbf{e} = 0$$

or

$$(\mathbf{e} \cdot \nabla \psi)\nabla \psi - \mathbf{e}(\nabla \psi)^2 + c^2 \mu \varepsilon \mathbf{e} = 0 \qquad (5.2.9)$$

But the first term is zero from (5.2.7a). Since e is not identically zero, we obtain finally

$$(\nabla \psi)^2 = n^2 \tag{5.2.10}$$

where (5.2.3) has been used for the definition of n.

The function $\psi(\mathbf{r})$ is called the eikonal and (5.2.10) is called the eikonal equation. It is a fundamental equation in the discussion of geometrical optics.

From (2.10.21), the time-averaged Poynting vector under the present approximation can be written as

$$\langle \mathbf{S}^{(0)} \rangle = \frac{1}{4} [\mathbf{E} \times \mathbf{H}^* + \mathbf{E}^* \times \mathbf{H}] = \frac{1}{4} (\mathbf{e} \times \mathbf{h}^* + \mathbf{e}^* \times \mathbf{h}) \quad (5.2.11)$$

Using (5.2.5a), we can reduce (5.2.11) to

$$\langle \mathbf{S}^{(0)} \rangle = c(\varepsilon e^2 + \mu h^2) \, \nabla \psi / 4n^2 \tag{5.2.12}$$

Therefore the average power flows in the direction of $\nabla \psi$. Since $\nabla \psi$ is normal to the wavefront surfaces $\psi(\mathbf{r}) = \text{constant}$, we may now define the rays as the orthogonal trajectories to the geometrical wave-fronts $\psi(\mathbf{r}) = \text{constant}$. If $\mathbf{r}(s)$ denotes the position vector of a point P on a ray, considered as a function of the arc length s, along the ray path, then the unit tangential vector, \hat{t} , along the ray is defined by

$$\hat{t} = d\mathbf{r}/ds = \nabla \psi / |\nabla \psi| = \nabla \psi / n \qquad (5.2.13)$$

We note that according to our definition, the ray direction \hat{t} is the same as the wave normal direction. This is because the medium is assumed isotropic.

With the definition (5.2.13), (5.2.12) can be put in the form

$$\langle \mathbf{S}^0 \rangle = \frac{1}{4} (\varepsilon e^2 + \mu h^2) (c/n) \hat{t} = \langle W \rangle (c/n) \hat{t} \qquad (5.2.12a)$$

where $\langle W \rangle = (\epsilon e^2 + \mu h^2)/4$ is the average energy density for the monochromatic wave [see (2.10.20)]. c/n can be taken as the velocity at which the ray propagates. $\langle W \rangle (c/n)$ is defined as the intensity of the ray, *I*.

In a lossless medium without spatial dispersion, the average energy density $\langle W \rangle$ is independent of time. From (2.10.19), we have

 $\nabla \cdot \mathbf{S}^{(0)} = \mathbf{0}$ $\nabla \cdot (l\hat{t}) = \mathbf{0}$ (5.2.14)

or

Geometrically, the significance of (5.2.14) can be interpreted by using Fig. 5.2-1. Let us take a narrow tube formed by all the rays proceeding from an element dS_1 of a wave point surface $\psi = a$. These rays intercept any other wave front in an element dS_2 as shown. This tube is defined as a tube of rays.

Integrating (5.2.14) over the volume of the tube we find that the energy



Fig. 5.2-1. Tube of rays.

flux flows into and out of the tube only through dS_1 and dS_2 . This implies that the product I dS along the tube of rays remains constant. As dS increases along the tube, I must decrease and vice versa. This is the intensity law of geometric optics.

From (5.2.10) and (5.2.13), it is possible to eliminate $\nabla \psi$ and obtain an equation for the rays. Differentiating (5.2.13) with respect to s, we have

$$\frac{d}{ds}\left(n\frac{d\mathbf{r}}{ds}\right) = \frac{d}{ds}\left(\nabla\psi\right)$$
$$= \frac{d\mathbf{r}}{ds} \cdot \nabla\left(\nabla\psi\right)$$
$$= (1/n)\nabla\psi \cdot \nabla\left(\nabla\psi\right) \qquad \text{(by 5.2.13)}$$
$$= (1/2n)\nabla\left[\left(\nabla\psi\right)^2\right]$$
$$= (1/2n)\nabla(n^2) \qquad \text{(by 5.2.10)}$$

Therefore,

$$(d/ds)(n \, d\mathbf{r}/ds) = \nabla n \tag{5.2.15}$$

This is a vector equation for the ray. It is the basis for tracing rays in an isotropic, inhomogeneous medium. As an example, let us consider the rays in a homogeneous medium. Then (5.2.15) becomes

$$\frac{d^2r}{ds^2} = 0 \tag{5.2.16}$$

which has a solution

$$\mathbf{r} = \mathbf{a}\mathbf{s} + \mathbf{b} \tag{5.2.17}$$

where a and b are two constant vectors. The rays in a homogeneous, isotropic medium are straight lines, not a surprising result.

5.3 Amplitude Variation along the Ray

From (5.2.1), we can also derive a wave equation for the inhomogeneous medium. We have

$$\nabla^{2}\mathbf{E} + \omega^{2}\varepsilon\mu\mathbf{E} + (\nabla \ln \mu) \times (\nabla \times \mathbf{E}) + \nabla(\mathbf{E} \cdot \nabla \ln \varepsilon) = 0 \quad (5.3.1)$$

and

$$\nabla^{2}\mathbf{H} + \omega^{2}\varepsilon\mu\mathbf{H} + (\nabla\ln\varepsilon) \times (\nabla\times\mathbf{H}) + \nabla(\mathbf{H}\cdot\nabla\ln\mu) = 0 \quad (5.3.2)$$

228

These are wave equations for electric and magnetic fields in an inhomogeneous medium and will be studied in detail in later sections for some special cases.

If we substitute (5.2.4) into (5.3.1), we obtain the following equation:

$$[n^{2} - (\nabla \psi)^{2}]\mathbf{e} - \frac{1}{jk_{0}} \mathbf{L}(\mathbf{e}, \psi, n, \mu) + \frac{1}{(jk_{0})^{2}} \mathbf{M}(\mathbf{e}, \psi, n, \mu) = 0 \quad (5.3.3)$$

where

$$\mathbf{L}(\mathbf{e}, \psi, n, \mu) = (\nabla \psi \cdot \nabla \ln \mu - \nabla^2 \psi)\mathbf{e} - 2(\mathbf{e} \cdot \nabla \ln n)\nabla \psi - 2(\nabla \psi \cdot \nabla)\mathbf{e}$$
(5.3.4)

 $\mathbf{M}(\mathbf{e}, \psi, n, \mu) = (\nabla \times \mathbf{e}) \times \nabla \ln \mu - \nabla^2 \mathbf{e} - \nabla (\mathbf{e} \cdot \nabla \ln \varepsilon)$ (5.3.5)

A similar equation can be obtained from (5.3.2) for **h**. The geometrical optics approximation is obtained by neglecting terms of the order $(1/k_0)$ and $(1/k_0^2)$ in (5.3.3). We note that the eikonal equation $n^2 = (\nabla \psi)^2$ is recovered immediately.

If now we solve the eikonal equation for ψ and substitute it back into (5.3.3), we obtain an equation relating **e** and ψ . We can use this equation to derive the amplitude of the wave along the ray. The equation now becomes

$$\mathbf{L} - \frac{\mathbf{M}}{jk_0} = \mathbf{0} \tag{5.3.6}$$

Under the present approximation, the second term can be neglected, and we obtain L = 0 or

$$\frac{\partial \mathbf{e}}{\partial s} + \frac{1}{2} \left(\frac{1}{n} \nabla^2 \psi - \frac{\partial \ln \mu}{\partial s} \right) \mathbf{e} + (\mathbf{e} \cdot \nabla \ln n) \hat{\iota} = 0 \qquad (5.3.7)$$

where $\hat{t} = (\nabla \psi/n)$ given by (5.2.13) has been used.

Equation (5.3.7) is the transport equation for the variation of e along the ray. A similar equation can be derived for h which is

$$\frac{\partial \mathbf{h}}{\partial s} + \frac{1}{2} \left(\frac{1}{n} \nabla^2 \psi - \frac{\partial \ln \varepsilon}{\partial s} \right) \mathbf{h} + (\mathbf{h} \cdot \nabla \ln n) \hat{t} = 0 \qquad (5.3.8)$$

Next, we take the scalar product of (5.3.7) by e^* and add the resulting equation to its complex conjugate equation; we obtain

$$\frac{\partial}{\partial s} \left(\mathbf{e} \cdot \mathbf{e}^* \right) + \left(\frac{1}{n} \nabla^2 \psi - \frac{\partial \ln \mu}{\partial s} \right) \left(\mathbf{e} \cdot \mathbf{e}^* \right) = 0 \qquad (5.3.9)$$

where (5.2.7a) has been used.

Dividing (5.3.9) by $(\mathbf{e} \cdot \mathbf{e}^*)$ and combining the first and third term, we obtain

$$\frac{\partial}{\partial s} \left[\ln \left(\frac{\mathbf{e} \cdot \mathbf{e}^*}{\mu} \right) \right] = -\frac{1}{n} \nabla^2 \psi \qquad (5.3.10)$$

which yields a solution

$$\left(\frac{\mathbf{e}\cdot\mathbf{e}^*}{\mu}\right)\Big|_B = \left(\frac{\mathbf{e}\cdot\mathbf{e}^*}{\mu}\right)\Big|_A \exp\left\{-\int_A^B \left[(\nabla^2\psi)/n\right]ds\right\} \quad (5.3.11)$$

after it is integrated along the ray from a point A to a point B. Equation (5.3.11) gives the variation of the magnitude of the vector e along the ray.

The next logical step would be to find the variation of the direction of the vector \mathbf{e} . This actually will indicate the variation of the polarization of the wave along the ray. Let us consider the complex unit vector defined by

$$\hat{\mu} = \mathbf{e}/(\mathbf{e} \cdot \mathbf{e}^*)^{1/2} \tag{5.3.12}$$

Dividing (5.3.7) by $(\mathbf{e} \cdot \mathbf{e}^*)^{1/2}$ and using the definition (5.3.12), we obtain

$$\frac{\partial \hat{\mu}}{\partial s} + \frac{1}{2} \left[\frac{\partial \ln(\mathbf{e} \cdot \mathbf{e}^*)}{\partial s} + \frac{1}{n} \nabla^2 \psi - \frac{\partial \ln \mu}{\partial s} \right] \hat{\mu} + (\hat{\mu} \cdot \nabla \ln n) \hat{t} = 0$$

The second term vanishes on account of (5.3.9). Therefore we obtain the equation

$$\partial \hat{\mu} / \partial s = -(\hat{\mu} \cdot \nabla \ln n)\hat{t}$$
 (5.3.13)

which governs the variation of the polarization of the wave along the ray. We note that for the homogeneous medium, $\nabla \ln n = 0$, $\hat{\mu}$ remains a constant along the ray.

Let us now consider a particularly simple situation as an example. Let the medium be horizontally stratified and nonmagnetic so that n = n(z)and μ is a constant. For a vertically incident ray, the ray path will remain vertical. Therefore everything varies as a function of only z and hence ds = dz. Equation (5.3.11) now reduces to

$$(\mathbf{e} \cdot \mathbf{e}^*)_B = (\mathbf{e} \cdot \mathbf{e}^*)_A [n(A)/n(B)]$$
(5.3.14)

Therefore, we conclude from (5.3.14) that in a horizontally stratified medium, at any point along the vertically incident ray, the magnitude of the vector **e** is proportional to the inverse square root of the refractive index at that point, i.e.,

$$|\mathbf{e}| = C/\sqrt{n} \tag{5.3.15}$$

where C is a constant of proportionality.

230

For this particular case, the eikonal equation also yields a simple solution

$$\psi(z) - \psi(A) = \int_{A}^{z} n \, d\xi \qquad (5.3.16)$$

where z is an arbitrary point along the path. Combining (5.3.15) and (5.3.16), we see that under the geometrical optics approximation, the electric field can be written as

$$\mathbf{E}(z) = [\mathbf{E}_0/(n(z))^{1/2}] \exp\left[-jk_0 \int_A^z n(\xi) \, d\xi\right]$$
(5.3.17)

where E_0 is a constant vector. In later sections, we will see that (5.3.17) is also called the WKB solution of the wave equation.

The discussion of this example can also be extended to the case of an obliquely incident wave; i.e., the wave propagates in directions other than vertical.

We note that so far the computations are made by neglecting terms of the order $(1/k_0)$ and higher. An asymptotic series in ascending powers of $(1/k_0)$ can be obtained by considering the problem in more detail [see, for example, Kline and Kay (1965)].

Now, let us say a few words about the range of validity of the geometrical optics. Since the approximation is made by dropping the right-hand sides of (5.2.5)-(5.2.8), the method is valid only when the variations of the inhomogeneities of the medium and the changes in **e** and **h** are small in one wavelength. More exact conditions can be obtained by deriving higher order approximations to the wave equation.

5.4 Fermat's Principle

The fundamental equations in geometric optics can also be derived from the Fermat's principle of classical mechanics. For our purpose, the principle can be stated in the following manner.

The time t it takes for the ray to travel from a point A to a point B can be expressed as

$$t = \int_{A}^{B} ds/(c/n)$$
 (5.4.1)

where integration is along a certain ray path.

The distance a wave will travel in free space in this time period is called the optical path length L.

$$L = ct = \int_{A}^{B} n \, ds \tag{5.4.2}$$

Fermat's principle for our problem can be now stated as follows: When a ray travels from a point A to a point B, it travels along a path for which the optical path length L has a stationary value.

Mathematically, this principle can be stated as

$$\delta \int_{A}^{B} n \, ds = 0 \tag{5.4.3}$$

where the symbol δ represents the variation of the integral.

If we express the path C from A to B by a set of parametric equations

$$x = x(u), \quad y = y(u), \quad z = z(u)$$
 (5.4.4)

where u is some parameter along the path, then the element of the arc length along the path is given by

$$ds = [(dx)^{2} + (dy)^{2} + (dz)^{2}]^{1/2}$$

= $[(dx/du)^{2} + (dy/du)^{2} + (dz/du)^{2}]^{1/2} du$
= $(\dot{x}^{2} + \dot{y}^{2} + \dot{z}^{2})^{1/2} du$ (5.4.5)

where an overdot is used to denote differentiation with respect to u. In terms of the parameter u, (5.4.3) becomes

$$\delta \int_{A}^{B} n(x, y, z) \cdot (\dot{x}^{2} + \dot{y}^{2} + \dot{z}^{2})^{1/2} du = 0 \qquad (5.4.6)$$

From calculus of variation [see, for example, Courant and Hilbert (1953)], the Euler equations for the variation

$$\delta \int_{A}^{B} F(x, y, z, \dot{x}, \dot{y}, \dot{z}) \, du = 0 \tag{5.4.7}$$

can be derived in the following way. Let us suppose that there is some neighboring curve C' connecting A and B which is also defined by the same parameter u. The variation in $F(\mathbf{r}, \dot{\mathbf{r}})$ in passing from a point P on the path C to a corresponding point P' on C' can be written as

$$\delta F = \nabla F \cdot \delta \mathbf{r} + \nabla_{\dot{\mathbf{r}}} F \cdot \delta \dot{\mathbf{r}}$$
(5.4.8)

where

$$\nabla_{\mathbf{\dot{r}}} = (\partial/\partial \dot{x})\dot{x} + (\partial/\partial \dot{y})\dot{y} + (\partial/\partial \dot{z})\dot{z}$$

$$\delta \mathbf{r} = \delta x \, \dot{x} + \delta y \, \dot{y} + \delta z \, \dot{z}$$

$$\delta \dot{\mathbf{r}} = \delta \dot{x} \, \dot{x} + \delta \dot{y} \, \dot{y} + \delta \dot{z} \, \dot{z} \qquad (5.4.9)$$

The variation in (5.4.7) can be now written as

$$\delta \int_{A}^{B} F(\mathbf{r}, \dot{\mathbf{r}}) du = \int_{A}^{B} \delta F(\mathbf{r}, \dot{\mathbf{r}}) du$$
$$= \int_{A}^{B} \nabla F \cdot \delta \mathbf{r} \, du + \int_{A}^{B} \nabla_{\dot{\mathbf{r}}} F \cdot \delta \dot{\mathbf{r}} \, du$$
$$= \int_{A}^{B} \nabla F \cdot \delta \mathbf{r} \, du + \int_{A}^{B} \nabla_{\dot{\mathbf{r}}} F \cdot (d/du) (\delta \mathbf{r}) \, du \quad (5.4.10)$$

where $\delta \dot{\mathbf{r}} = \delta (d\mathbf{r}/du) = (d/du)(\delta \mathbf{r})$ has been used.

The second integral in (5.4.10) can be integrated by parts to yield

$$\int_{A}^{B} \nabla_{\mathbf{i}} F \cdot \frac{d}{du} \left(\delta \mathbf{r} \right) du = \left[\nabla_{\mathbf{i}} F \cdot \delta \mathbf{r} \right]_{A}^{B} - \int_{A}^{B} \frac{d}{du} \left(\nabla_{\mathbf{i}} F \right) \cdot \delta \mathbf{r} \, du \quad (5.4.11)$$

Remember that $\delta \mathbf{r}$ is the variation of \mathbf{r} from C to C'. Since the two curves have common fixed end points A and B, the variations $\delta \mathbf{r}$ must vanish. Therefore the boundary term in (5.4.11) is zero and (5.4.10) becomes

$$\delta \int_{A}^{B} F(\mathbf{r}, \dot{\mathbf{r}}) \, du = \int_{A}^{B} \left[\nabla F - (d/du) (\nabla_{\dot{\mathbf{r}}} F) \right] \cdot \delta \mathbf{r} \, du \qquad (5.4.12)$$

Substituting (5.4.12) into (5.4.7), we have

$$\int_{A}^{B} \left[\nabla F - (d/du) (\nabla_{\mathbf{r}} F) \right] \cdot \delta \mathbf{r} \, du = 0 \qquad (5.4.13)$$

Since $\delta \mathbf{r}$ is any arbitrary variation, (5.4.13) can be satisfied only when the intergrand vanishes or

$$(d/du)(\nabla_{\mathbf{t}}F) - \nabla F = 0 \tag{5.4.14}$$

This is the Euler equation for the variation integral (5.4.7).

For the present case, from (5.4.6), we have $F = n(\mathbf{r})(\dot{x}^2 + \dot{y}^2 + \dot{z}^2)^{1/2} = n(\mathbf{r}) |\dot{\mathbf{r}}|$, and the corresponding Euler equation now becomes

$$\frac{d}{du} \left[\frac{n(\mathbf{r})\dot{\mathbf{r}}}{|\dot{\mathbf{r}}|} \right] - |\dot{\mathbf{r}}| \nabla n(\mathbf{r}) = 0$$
(5.4.15)

If we choose du = ds such that $(\dot{x}^2 + \dot{y}^2 + \dot{z}^2)^{1/2} = |\dot{\mathbf{r}}| = 1$, (5.3.9) becomes

$$(d/ds)(n \, d\mathbf{r}/ds) = \nabla n \tag{5.4.16}$$

This is exactly the same equation for rays (5.2.15) derived from the eikonal equation in Section 2. Thus we see that the geometrical optics can be formulated from the Fermat's principle. In fact, for anisotropic media, it is more convenient to discuss the fundamentals of the geometrical optics starting from the Fermat's principle rather than from Maxwell's equations. The next section is devoted to this problem.

5.5 Ray Equations in Anisotropic Media

In Chapter 2 we have discussed the refractive indices for different modes in a homogeneous, anisotropic medium. In general, for a spatially nondispersive medium, the refractive index is a function of the direction of the wave normal. As a consequence, the ray and wave normal have different directions in general. The angle between the two directions, α , is given by (2.12.5). Just as in the case for the isotropic medium, we can extend the definition of the refractive index to the inhomogeneous, anisotropic medium. The refractive index n will now be a function of position. At each point, the wave normal direction is orthogonal to the wave front passing through that point and the ray direction is the direction of energy flow at that point. We wish to find, for a particular mode, the trajectory of a ray in such a medium.

Let the direction of the wave normal be $\hat{s} = (s_x, s_y, s_z)$ and that of the ray be $\hat{t} = (t_x, t_y, t_z)$. Both \hat{s} and \hat{t} are unit vectors. We define two vectors:

$$\boldsymbol{\sigma} = n\hat{\boldsymbol{s}} \tag{5.5.1a}$$

$$\boldsymbol{\xi} = \hat{t}/n \cos \alpha \tag{5.5.1b}$$

Equation (5.5.1b) can also be written as

$$\boldsymbol{\xi} = \hat{t}/n_r \tag{5.5.1c}$$

where $n_r = n \cos \alpha$ is the ray refractive index. The ray velocity is then defined as $v_r = c/n_r$. Following the discussion in Section 5.3, the Fermat's principle for the ray in an anisotropic medium is just

$$\delta \int_{A}^{B} n_{r} \, ds = 0 \tag{5.4.2}$$

Before we discuss this principle any further, let us consider first some important relations between the wave normals and rays. At any point $P(\mathbf{r})$, $n(\mathbf{r})$ is given. If we trace the endpoint of the vector $\mathbf{\sigma}$ for all possible directions of s, we obtain a refractive index surface about P (Section 2.12). Since $\mathbf{\sigma}$ depends on the direction s in an anisotropic medium, the surface in general is not spherical. In Fig. 5.5-1, we draw a portion of the cross section of this surface in the $(s_x s_z)$ -plane about $P(\mathbf{r})$.



Fig. 5.5-1. Construction of ray surface from refractive index surface.

From the discussion in Section 2.12, we know that for each wave normal direction \hat{s} , the corresponding ray direction is normal to the refractive index surface at that point. For example, in Fig. 5.5-1 for the point P_1 on the refractive index surface, we can obtain the corresponding ray direction by taking the normal to the refractive index surface at P_1 . If from the point P, we draw a perpendicular line to the plane which is tangent to the refractive index surface at P_1 , we obtain the line PP_1' which is in the ray direction. The angle between PP_1 and PP_1' is α by our definition. If we now take $PP_2 = 1/n \cos \alpha$, we obtain the vector $\boldsymbol{\xi} = PP_2 \hat{i}$. Corresponding to every point on the refractive index surface, we can construct the vector $\boldsymbol{\xi}$ in this manner. The end points of $\boldsymbol{\xi}$ then trace out a new surface. This is known as the ray surface. In Fig. 5.5-1, a portion of the ray surface is also drawn. Therefore at each point $P(\mathbf{r})$, we can construct the refractive index surface.

We note that the two vectors σ and ξ are reciprocal vectors since

$$\boldsymbol{\sigma} \cdot \boldsymbol{\xi} = \boldsymbol{\$} \cdot \boldsymbol{\hat{t}} / \cos \alpha = 1 \tag{5.5.3}$$

Therefore, the refractive index surface and ray surface are called reciprocal surfaces.

At any point $P(\mathbf{r})$, let us now define the ray surface by the relation

$$F(\mathbf{r}, \xi) = |\xi| n_r = 1$$
 (5.5.4)

This is the right surface, since it implies that the magnitude of the vector $\mathbf{\xi}$ is equal to $1/n \cos \alpha$ which is (5.5.1c), the ray surface.

Similarly we define the refractive index surface by

$$G(\mathbf{r}, \boldsymbol{\sigma}) = |\boldsymbol{\sigma}|/n = 1 \tag{5.5.5}$$

The reciprocal properties of these two surfaces will be shown in the following.

The ray refractive index n_r is a function of position as well as the direction of the ray. Since we are interested in spatially nondispersive media, n_r depends only on the direction of ξ but not on its magnitude. Similarly, the refractive index n is a function of position and direction of the wave normal *s* but not the magnitude of σ . We have

$$n_r(\mathbf{r}, \lambda \boldsymbol{\xi}) = n_r(\mathbf{r}, \boldsymbol{\xi})$$

$$n(\mathbf{r}, \lambda \boldsymbol{\sigma}) = n(\mathbf{r}, \boldsymbol{\sigma})$$
(5.5.6)

Mathematically, any function satisfying the relation (5.5.6) is a homogeneous function of degree zero. The function $F(\mathbf{r}, \boldsymbol{\xi})$, however, is a homogeneous function of degree one in the variable $\boldsymbol{\xi}$ since from (5.5.4) and (5.5.6), we have

$$F(\mathbf{r}, \lambda \boldsymbol{\xi}) = \lambda F(\mathbf{r}, \boldsymbol{\xi}) \tag{5.5.7}$$

From the theory of homogeneous functions (Goursat and Hedrick, 1959), we have the following theorem: For a function $\phi(\mathbf{x})$ such that

$$\phi(\lambda \mathbf{x}) = \lambda^m \phi(\mathbf{x}) \tag{5.5.8}$$

then the following relation holds:

$$m\phi(\mathbf{x}) = \mathbf{x} \cdot \nabla_x \phi(\mathbf{x}) \tag{5.5.9}$$

Applying this result to the fraction $F(\mathbf{r}, \boldsymbol{\xi})$, we have

$$F(\mathbf{r}, \boldsymbol{\xi}) = \boldsymbol{\xi} \cdot \nabla_{\boldsymbol{\xi}} F(\mathbf{r}, \boldsymbol{\xi})$$
(5.5.10)

Similarly, $G(\mathbf{r}, \boldsymbol{\sigma})$ is a homogeneous function of degree one for the variable $\boldsymbol{\sigma}$. Therefore, we have

$$G(\mathbf{r}, \boldsymbol{\sigma}) = \boldsymbol{\sigma} \cdot \boldsymbol{\nabla}_{\boldsymbol{\sigma}} G(\mathbf{r}, \boldsymbol{\sigma}) \tag{5.5.11}$$

Using the general theorem of homogeneous functions, it is possible to derive certain relations among the derivatives of the functions F and G

236

and the vectors $\boldsymbol{\sigma}$ and $\boldsymbol{\xi}$. Let us consider the refractive index surface G = 1. Referring to Fig. 5.5-1, the normal to the tangent plane passing through P_1 on this surface is given by $\nabla_{\sigma} G(\mathbf{r}, \boldsymbol{\sigma})$ at P_1 . But since this is also the direction of the ray, we can write

$$\boldsymbol{\xi} = \lambda_1 \nabla_{\boldsymbol{\sigma}} G(\mathbf{r}, \boldsymbol{\sigma}) \tag{5.5.12}$$

where λ_1 is a proportionality constant. Similarly, for the ray surface F = 1, we have

$$\boldsymbol{\sigma} = \lambda_2 \nabla_{\boldsymbol{\xi}} F(\mathbf{r}, \boldsymbol{\xi}) \tag{5.5.13}$$

where λ_2 is also a proportionality constant. Since σ and ξ are reciprocal vectors, we have

$$\boldsymbol{\sigma} \cdot \boldsymbol{\xi} = \lambda_1 \boldsymbol{\sigma} \cdot \boldsymbol{\nabla}_{\sigma} \boldsymbol{G} = 1$$

$$\boldsymbol{\xi} \cdot \boldsymbol{\sigma} = \lambda_2 \boldsymbol{\xi} \cdot \boldsymbol{\nabla}_{\xi} \boldsymbol{F} = 1$$
 (5.5.14)

But since G and F are homogeneous functions, (5.5.10) and (5.5.11) may be applied. We can then determine the two constants λ_1 and λ_2 from (5.5.14):

$$\lambda_1 G = 1$$
 or $\lambda_1 = 1/G$
 $\lambda_2 F = 1$ or $\lambda_2 = 1/F$
(5.5.15)

Therefore, we obtain the relations

$$\boldsymbol{\xi} = (1/G) \boldsymbol{\nabla}_{\sigma} G$$

$$\boldsymbol{\sigma} = (1/F) \boldsymbol{\nabla}_{\varepsilon} F$$
(5.5.16)

Equation (5.5.16) indicates the reciprocal property of the two surfaces $F(\mathbf{r}, \boldsymbol{\xi}) = 1$ and $G(\mathbf{r}, \boldsymbol{\sigma}) = 1$.

From (5.5.16) and (5.5.1), it is easy to prove the following relations:

$$F(\mathbf{r},\boldsymbol{\xi})G(\mathbf{r},\boldsymbol{\xi}) = F(\mathbf{r},\boldsymbol{\sigma})G(\mathbf{r},\boldsymbol{\sigma}) = 1 \qquad (5.5.17)$$

which is an alternative way of expressing the reciprocal relations of the two surfaces F = 1 and G = 1.

Up to this point, we have discussed the geometrical relationships between the wave normal vector and the ray as well as the reciprocal properties between the corresponding surfaces. Let us now go back to the Fermat's principle to discuss the equations of the ray. The mathematical statement of the problem is

$$\delta \int_{A}^{B} n_{\tau} \, ds = 0 \tag{5.5.18}$$


Fig. 5.5-2. Relation between the wave front, the wave normal, and the ray.

We first introduce the time parameter t which is a function of the path length s. In Fig. 5.5-2, we see that in a time dt the wavefront AA' has advanced a distance ds along the ray direction \hat{t} and the ray now is at P'. The corresponding distance along the wave normal direction \hat{s} is $dl=ds \cos \alpha$. But $dl = v_p dt$, where v_p is the phase velocity $v_p = c/n$. Therefore

$$ds = dl/\cos \alpha = c \, dt/n \cos \alpha = (c/n_r) \, dt \tag{5.5.19}$$

From (5.5.1c) we have

$$ds = c \mid \mathbf{\xi} \mid dt \tag{5.5.20}$$

Substituting (5.5.20) into (5.5.18), we obtain

$$\delta \int_{A}^{B} n_{\mathbf{r}} | \mathbf{\xi} | d\tau = \delta \int_{A}^{B} F(\mathbf{r}, \mathbf{\xi}) d\tau = 0 \qquad (5.5.21)$$

where $\tau = ct$.

The corresponding Euler's equation is

$$(d/d\tau)(\nabla_{\mathbf{t}}F) - \nabla F = 0 \tag{5.5.22}$$

where

$$\dot{\mathbf{r}} = \frac{d\mathbf{r}}{d\tau} = \frac{d\mathbf{r}}{ds} \frac{ds}{d\tau} = \hat{t} |\mathbf{\xi}| = \mathbf{\xi}$$
(5.5.23)

Equation (5.5.22) is the ray equation in an anisotropic medium. From the definition of F, we see that the ray refractive index n_r and the direction of the ray \hat{t} are involved in this form of the ray equation. It can be put into a more convenient form for computation. From (5.5.16) and (5.5.17), we have

 $\nabla_{\mathbf{t}}F = \nabla_{\boldsymbol{\xi}}F = F\boldsymbol{\sigma} = \boldsymbol{\sigma}/G$

Therefore (5.5.22) can be written as

$$(d/d\tau)(\mathbf{\sigma}/G) - \mathbf{\nabla}(1/G) = 0 \tag{5.5.24}$$

The first term can be expanded:

$$\frac{d}{d\tau} \frac{\mathbf{\sigma}}{G} = \frac{1}{G} \frac{d\mathbf{\sigma}}{d\tau} - \frac{\mathbf{\sigma}}{G^2} \frac{dG}{d\tau}$$

Since G = 1 along the ray, the total derivative $dG/d\tau$ must then vanish. However the partial derivatives with respect to **r**, ξ , and σ are not necessarily zero. Therefore (5.5.24) reduces to

$$d\mathbf{\sigma}/d\mathbf{r} = -\nabla G(\mathbf{r}, \mathbf{\sigma}) \tag{5.5.25}$$

Next, combining (5.5.16) and (5.5.23), we have

$$d\mathbf{r}/d\tau = \nabla_{\sigma} G(\mathbf{r}, \,\boldsymbol{\sigma}) \tag{5.5.26}$$

Note that the pair of equations (5.5.25) and (5.5.26) are equivalent to those given by (2.12.24) derived by using a kinematic approach. They are two first-order equations which describe the trajectory of the ray. They have the form of Hamilton's canonical equations in mechanics with σ playing the role of the generalized momenta. The integration of this set of equations determines the ray path in the parametric form $\mathbf{r} = \mathbf{r}(\tau)$.

While (5.5.22) is equivalent to the sets (5.5.25) and (5.5.26), the latter set is much more convenient for numerical computation, since only the phase refractive index $n(\mathbf{r}, \sigma)$ is involved, while F depends on the ray refractive index n_r .

For an isotropic medium such that $n = n_r$ and is independent of σ or ξ , (5.5.22) becomes

$$(d/d\tau)(\xi n/|\xi|) - |\xi| \nabla n = 0$$

But $\xi/|\xi| = \hat{t} = d\mathbf{r}/ds$ and $d\tau = c dt = ds/|\xi|$. Therefore we obtain

$$(d/ds)(n d\mathbf{r}/ds) = \nabla n$$

which is exactly the same as (5.4.16).

The Hamilton's equations for the ray, (5.5.25) and (5.5.26), were derived for Cartesian coordinates. It is possible, starting from (5.5.21), to derive the ray equations for any generalized coordinate system (Haselegrove, 1954). For example, in a spherical coordinate system a point on the ray is given by $P(r, \theta, \psi)$. The direction of the wave normal at the point P can be expressed by its components in the local spherical coordinate system as $\sigma(\sigma_r, \sigma_{\theta}, \sigma_{\phi})$ where the magnitude of σ is again n given by (5.5.1a). The Hamilton's equations for rays in the spherical coordinates can be written as (Haselegrove, 1957)

$$\frac{dr}{d\tau} = \frac{1}{n^2} \left(\sigma_r - n \frac{\partial n}{\partial \sigma_r} \right)$$

$$\frac{d\theta}{d\tau} = \frac{1}{rn^2} \left(\sigma_\theta - n \frac{\partial n}{\partial \sigma_\theta} \right)$$

$$\frac{d\phi}{d\tau} = \frac{1}{r \sin \theta n^2} \left(\sigma_\phi - n \frac{\partial n}{\partial \sigma_\phi} \right)$$

$$\frac{d\sigma_r}{d\tau} = \frac{1}{n} \frac{\partial n}{\partial r} + \sigma_\theta \frac{d\theta}{d\tau} + \sin \theta \sigma_\phi \frac{d\phi}{d\tau}$$

$$\frac{d\sigma_\phi}{d\tau} = \frac{1}{r} \left(\frac{1}{n} \frac{\partial n}{\partial \theta} - \sigma_\theta \frac{dr}{d\tau} + r \cos \theta \sigma_\phi \frac{d\phi}{d\tau} \right)$$

$$\frac{d\sigma_\phi}{d\tau} = \frac{1}{r \sin \theta} \left(\frac{1}{n} \frac{\partial n}{\partial \phi} - \sin \theta \sigma_\phi \frac{dr}{dt} - r \cos \theta \sigma_\phi \frac{d\theta}{d\tau} \right)$$

where the relation $G = |\sigma|/n = 1$ has been used.

This set of equations has been used to trace the rays in the ionosphere on high speed computers by various authors.

5.6 Effect of Boundary on the Ray and Generalized Snell's Law

Up to this point, we have assumed that the ray propagates in a continuous medium without any boundary. It is of interest to see what will happen to the rays when they cross the sharp boundary between two different media. From elementary optics, we known that reflection and refraction will occur when rays meet the sharp boundary. Our problem now is to determine the ray path when reflection or refraction occurs in general anisotropic media. The answer can be obtained from the Fermat's principle.

Suppose now two different anisotropic media M_1 and M_2 are separated by a boundary surface S. We want to find the ray path for a ray which starts from a point A in M_1 and ends at a point B in M_2 . The integral in the Fermat's principle of (5.5.2) can be written in a slightly different form:

$$\int_{A}^{B} n_{r} ds = \int_{A}^{B} \boldsymbol{\sigma} \cdot d\mathbf{r}, \quad \text{with} \quad G(\mathbf{r}, \boldsymbol{\sigma}) = 1 \quad (5.6.1)$$

The direction σ at any point **r** on the ray must be found from the constrain-

ing condition $G(\mathbf{r}, \boldsymbol{\sigma}) = 1$. We note that $G(\mathbf{r}, \boldsymbol{\sigma}) = |\boldsymbol{\sigma}|/n(\mathbf{r}, \boldsymbol{\sigma})$ is discontinuous at the boundary S since n is different for the two media.

The variation of the integral in (5.6.1) can be expressed as

$$\delta \int_{A}^{B} \boldsymbol{\sigma} \cdot d\mathbf{r} = \int_{A}^{B} \left(\delta \boldsymbol{\sigma} \cdot d\mathbf{r} + \boldsymbol{\sigma} \cdot \delta \, d\mathbf{r} \right)$$
(5.6.2)

The second integral in (5.6.2) may be integrated by parts to yield

$$\int_{A}^{B} \mathbf{\sigma} \cdot \delta \, d\mathbf{r} = \int_{A}^{B} \mathbf{\sigma} \cdot d \, \delta \mathbf{r}$$
$$= \mathbf{\sigma} \cdot \delta \mathbf{r} \Big]_{A}^{P-\epsilon} + \mathbf{\sigma} \cdot \delta \mathbf{r} \Big]_{P+\epsilon}^{B} - \int_{A}^{P-\epsilon} \delta \mathbf{r} \cdot d\mathbf{\sigma} - \int_{P+\epsilon}^{B} \delta \mathbf{r} \cdot d\mathbf{\sigma} \quad (5.6.3)$$

where P is a point on the surface S and ε is a small parameter to be set equal to zero in the limit. The reason that the integral is divided into parts is that $G(\mathbf{r}, \boldsymbol{\sigma})$ is different for the two media. At the end points A and B, $\delta \mathbf{r}$ vanishes. Therefore, combining (5.6.2) and (5.6.3), we obtain

$$\delta \int_{A}^{B} \mathbf{\sigma} \cdot d\mathbf{r} = -\mathbf{\sigma} \cdot \delta \mathbf{r} \Big]_{P-\varepsilon}^{P+\varepsilon} + \int_{A}^{P-\varepsilon} (\delta \mathbf{\sigma} \cdot d\mathbf{r} - \delta \mathbf{r} \cdot d\mathbf{\sigma}) \\ + \int_{P+\varepsilon}^{B} (\delta \mathbf{\sigma} \cdot d\mathbf{r} - \delta \mathbf{r} \cdot d\mathbf{\sigma})$$
(5.6.4)

Taking the limit $\varepsilon \to 0$ in (5.6.4) and substituting it into (5.6.1) after the Fermat's principle is applied, we obtain

$$-(\boldsymbol{\sigma}_2-\boldsymbol{\sigma}_1)\boldsymbol{\cdot}\boldsymbol{\delta}\mathbf{r}+\int_A^B\left(\boldsymbol{\delta}\boldsymbol{\sigma}\boldsymbol{\cdot}\boldsymbol{d}\mathbf{r}-\boldsymbol{\delta}\mathbf{r}\boldsymbol{\cdot}\boldsymbol{d}\boldsymbol{\sigma}\right)=0 \qquad (5.6.5)$$

$$\delta G = \nabla G \cdot \delta \mathbf{r} + \nabla_{\sigma} G \cdot \delta \mathbf{\sigma} = 0 \tag{5.6.6}$$

where σ_2 and σ_1 correspond to the wave normal directions at the boundary in M_2 and M_1 , respectively, and (5.6.6) is obtained by taking the variation of the equation G = 1.

Equation (5.6.5) must hold for all variations $\delta \mathbf{r}$ with $\delta \mathbf{\sigma}$ given by (5.6.6). If we now define the so-called natural boundary condition

$$(\boldsymbol{\sigma}_2 - \boldsymbol{\sigma}_1) \cdot \delta \mathbf{r} = 0 \tag{5.6.7}$$

then (5.6.5) and (5.6.6) yield

$$d\sigma/d\tau = -\nabla G(\mathbf{r}, \sigma), \qquad d\mathbf{r}/d\tau = \nabla_{\sigma} G(\mathbf{r}, \sigma)$$
 (5.6.8)

in both M_1 and M_2 . Thus the rays in M_1 and M_2 satisfy the Hamilton's equations separately. At the boundary, the direction of the ray is determined by the natural boundary condition. Since $\varepsilon \to 0$, $\delta \mathbf{r}$ is an arbitrary variation of \mathbf{r} on the boundary surface S, (5.6.7) states that the vector $\boldsymbol{\sigma}_2 - \boldsymbol{\sigma}_1$ is in the direction of the normal to the surface S. If we denote this normal direction by the unit vector \hat{n} , then (5.6.7) may be expressed as

$$\boldsymbol{\sigma}_2 - \boldsymbol{\sigma}_1 = \alpha \hat{n} \tag{5.6.9}$$

where α is a proportionality constant.

From (5.6.9) it follows that the components of σ tangential to the surface S are continuous across the boundary.

For a given incident ray in medium one, it is easy to get the vector σ_2 from (5.6.4). Then using the Hamilton equation $(d\mathbf{r}/d\tau) = \nabla_{\sigma}G_2$ for the second medium we can find $d\mathbf{r}/d\tau$. The unit vector along $d\mathbf{r}/d\tau$ is the ray direction \hat{t} for the refracted ray. To obtain the reflected ray, the function G_1 for the first medium should be used instead of G_2 . Therefore we can find both the refracted and reflected rays at the boundary surface.

This method of finding the refracted and reflected rays is best carried out graphically. For a point P on the boundary surface S, let us draw two refractive index surfaces $G_1 = 1$ and $G_2 = 1$ as shown in Fig. 5.6-1. The





direction of the normal to S, \hat{n} , is also drawn schematically in the same figure. Construct from P the vector σ_1 for the incident ray. The endpoint of σ_1 must lie on the surface $G_1 = 1$ by definition. Therefore we have $\overrightarrow{PA_1} = \sigma_1$ in Fig. 5.6-1. The vector σ_2 is found by constructing a line through A_1 parallel to \hat{n} and finding its intersection with the surface $G_2 = 1$, since the endpoint σ_2 must lie on $G_2 = 1$. In the figure, we see that there is no intersection with $G_2 = 1$, hence there is no refraction for this case. Instead, reflection can occur since the line intersects $G_1 = 1$ at A_2 . The vector $\overrightarrow{PA_2} = \sigma_2$ is the direction of the wave front normal for the reflected ray. The vector $\overrightarrow{A_2A_1}$ is equal to $\alpha \hat{n}$. To insure that reflection does occur, it must be shown that the reflected ray actually does emerge back in medium one. If this is the case, there is total reflection, since no refraction occurs. Otherwise neither reflection nor refraction occurs.

For the other case shown in Fig. 5.6-1, the incident ray has the vector $\sigma_1 = \overrightarrow{PB_1}$. The line from B_1 which is parallel to \hat{n} intersects $G_2 = 1$ at C_1 and C_2 and $G_1 = 1$ at B_2 . Hence we have two possible refracted rays and one reflected ray. The three possibilities must be considered separately to see if the refracted rays emerge in medium two and if the reflected ray emerges in medium one. We note that the ray direction can be obtained from the figure very easily. For example at point A_1 , the ray direction is the direction of the normal to the surface G = 1 at A_1 .

The above discussion on the ray equation and the boundary condition between two media can be easily generalized to the cases where more than two media are present.

In many practical cases where the medium is stratified, instead of solving the ray equation, it is possible to divide the medium into many layers and apply the graphical method described above to trace the ray in each layer. Poeverlein (1948) used the technique to trace the rays in the ionosphere. In the following, we shall discuss a simple example to indicate the procedure of this technique (Forsgren, 1951).

Let us consider the earth's ionosphere. The electron density is assumed to be a function of height only. The refractive indices are therefore functions of height also. In Fig. 5.6-2, cross sections of the refractive index surfaces for the ordinary mode are drawn for different values of X, each representing the index surface at certain level of the ionosphere. The inclination of the external magnetic field is also shown. The outmost curve is a circle of unit radius representing the free space below the ionosphere. A vertical line $\overrightarrow{AA'}$ is drawn a distance $\sin \theta_i$ away from the vertical axis where θ_i is the incidence angle for a ray incident on the ionosphere from below. The direction \overrightarrow{PA} is the direction of the wave normal as well as the direction of the ray since the medium is the free space at this point. At the point B, the magnitude $|\overrightarrow{PB}|$ of the vector \overrightarrow{PB} is equal to the refractive index n at a certain level in the ionosphere. From the figure, we have

$$|PB| \sin \not\triangleleft PBA' = \sin \theta_i$$

This is just Snell's law. Therefore, the direction of the wave normal at the level corresponding to point B is \overrightarrow{PB} . Similarly we have the directions of the



Fig. 5.6-2. Ray tracing in a stratified magnetoplasma. [After Forsgren (1961).]

wave normal at levels corresponding to points C, D, etc. At these points, the normals to the respective surfaces of refractive index are shown by small arrows. From our earlier discussion, we note that these are the directions of the rays at the respective levels. We see that energy propagates upward until the ray reaches the level corresponding to the point E at which the cross section of the refractive index surface is tangential to the vertical line AA'. At this point, the normal to the surface is in the horizontal direction. Beyond this level, the ray directions point downward as shown in the figure. The level corresponding to the point E is called the reflection level. Energy begins to be reflected downward at this level. Thus we have seen the graphical way of tracing the ray in an anisotropic medium. Better approximations can be obtained if more refractive index surfaces are used, corresponding to finer division of the stratified medium.

The above discussion is also closely related to the generalized Snell's law. If we substitute $G = |\sigma|/n$ into (5.5.24), we obtain

$$\frac{d}{d\tau}\left(n\frac{\mathbf{\sigma}}{|\mathbf{\sigma}|}\right) = \frac{1}{|\mathbf{\sigma}|}\nabla n \qquad (5.6.10)$$

If we define a vector $\boldsymbol{\eta}$ which is parallel to ∇n , then (5.6.10) yields

$$\frac{d}{d\tau} \left(\frac{n}{\mid \mathbf{\sigma} \mid} \mathbf{\sigma} \right) \, \mathbf{X} \, \mathbf{\eta} = 0$$

This is the generalized differential form of Snell's law. In particular, if the medium is stratified horizontally so that n does not depend on x or y coordinates, then η is in the vertical direction and (5.6.10) reduces to the expression

$$\frac{d}{dt}(n\sin\theta) = 0 \tag{5.6.11}$$

where θ is the angle the wave normal makes with the z-axis. Equation (5.6.11) states that along the ray path, the quantity $n \sin \theta$ at any point is a constant. This indeed is the ordinary form of Snell's law.

Thus far we have discussed the propagation of a monochromatic ray in an anisotropic medium. Since in general the medium is frequency dispersive, it is of interest to consider the propagation of a pulsed signal in the medium. From (2.12.6), the magnitude of the group velocity is given by

$$v_q = c/[\partial(n\omega)/\partial\omega] \cos \alpha$$
 (5.6.12)

The time it takes for the pulse to travel along the ray path from A to B is equal to

$$t = \int_{A}^{B} ds / \nu_g \tag{5.6.13}$$

Now if we define a "group path" P' as the distance traveled by the pulse in free space in the time period t, we have

$$P' = ct = \int_{A}^{B} \cos \alpha \, \partial(n\omega) / \partial \omega \, ds \qquad (5.6.14)$$

But

$$\partial(n\omega)/\partial\omega = n + \omega(\partial n/\partial\omega) = n + f(\partial n/\partial f)$$
 (5.6.15)

Therefore we can define a "group index of refraction" n' by

$$n' = n + f(\partial n/\partial f)$$
 (5.6.16)

and (5.6.14) becomes

$$P' = \int_{A}^{B} n' \cos \alpha \, ds \qquad (5.6.17)$$

But from (5.5.19), $ds = (c/n \cos \alpha) dt$. Thus (5.6.17) becomes

$$P' = c \int_{A}^{B} (n'/n) dt = c \int_{A}^{B} [1 + (f/n)(\partial n/\partial f)] dt \qquad (5.6.18)$$

Or, putting it in the form of a differential equation, we have

$$\frac{dP'}{dt} = \left[1 + \frac{f}{n} \frac{\partial n}{\partial f}\right]c \qquad (5.6.19)$$

The group path is a quantity which is relatively easy to measure in many experimental situations. It plays an important role in ionosphere studies since a knowledge of variations of the group path yields a great amount of information concerning the medium in which the ray propagates (see Sections 5.15 and 5.16).

5.7 Reflection and Transmission of Waves at Sharp Boundaries

In the previous sections we have discussed the geometric optics in a general inhomogeneous medium. The discussion is based on the approximation of a zero wavelength limit. This approximation breaks down in situations where the diffractional effects are no longer negligible, e.g., near a caustic. For these cases, one must go back to the exact wave equation. However, solutions of the exact wave equation for a general inhomogeneous medium are not obtainable in general. Only asymptotic solutions for short wavelength waves have been discussed as generalizations of geometric optics (Kline and Kay, 1965). Fortunately, in many natural situations in which wave propagation phenomena occur, the properties of the medium can be assumed to be constant throughout each plane perpendicular to certain fixed directions. This kind of media is called "stratified media." The theory of wave propagation in stratified media has been well developed. Many situations in applied physics, such as optics of multilayers and underwater acoustics as well as ionospheric propagation, can be approximated by assuming a stratification. In the remaining part of this chapter, we shall concentrate on the study of the particular type of inhomogeneous media.

The simplest kind of stratified medium is a system of two isotropic homogeneous media separated by a sharp plane boundary. Let us now consider wave propagation in such a system. Assuming the sharp boundary is at z = 0, the medium for z < 0 is characterized by ε_1 and μ_1 , the medium for z > 0 is characterized by ε_2 and μ_2 (Fig. 5.7-1).

246



In a homogeneous, isotropic medium, it can be shown easily from the Maxwell equations that waves with two different polarizations propagate independently in the medium. One is called the horizontal polarized wave for which $E_x = E_z = H_y = 0$, and the other is called the vertical polarized wave for which $H_x = H_z = E_y = 0$ (Fig. 5.7-1).

Let us first consider a horizontally polarized wave with unit amplitude in medium I incident upon medium II with an incident angle θ_i . Since the two media are homogeneous, the solutions of the wave equations (5.3.1) and (5.3.2) for the two media for horizontal polarization are easily obtained. They are

$$E_y = E_y^i + E_y^r$$

= exp[-jk_0n_1(x sin \theta_i + z cos \theta_i)]
+ R_i exp[-jk_0n_1(x sin \theta_i + z cos \theta_i)], z < 0 (5.7.1)

$$E_y = E_y^t = T_\perp \exp[-jk_0 n_2(x\sin\theta_t + z\cos\theta_t)], \quad z > 0 \quad (5.7.2)$$

$$H_x = H_x^i + H_x^r$$

$$= (\varepsilon_1/\mu_1)^{1/2} [-\cos\theta_i E_y^{i} + \cos\theta_r E_y^{r}], \qquad z < 0$$
(5.7.3)

$$H_x = H_x^t = (\varepsilon_2/\mu_2)^{1/2} [-\cos\theta_t E_y^t], \qquad z > 0$$
(5.7.4)

where $n_1 = c(\varepsilon_1 \mu_1)^{1/2}$ and $n_2 = c(\varepsilon_2 \mu_2)^{1/2}$; θ_r and θ_t indicate the directions of the reflected and transmitted wave normals, respectively; and R_{\perp} and T_{\perp} are the reflection and transmission coefficients, respectively. The subscript \perp indicates the horizontal polarization.

At the boundary z = 0, the tangential components of the electric and magnetic fields are continuous for all values of x. This can be true only if

$$n_1 \sin \theta_i = n_1 \sin \theta_r = n_2 \sin \theta_t \tag{5.7.5}$$



Therefore

$$\theta_r = \pi - \theta_i$$

$$n_1 \sin \theta_i = n_2 \sin \theta_t \qquad (5.7.6)$$

which is just the Snell's law.

Next, matching the fields at the boundary, we have

$$1 + R_{\perp} = T_{\perp}$$

$$(\varepsilon_1/\mu_1)^{1/2} \cos \theta_i (1 - R_{\perp}) = (\varepsilon_2/\mu_2)^{1/2} \cos \theta_t T_{\perp}$$
(5.7.7)

Solving for R_{\perp} and T_{\perp} , we obtain

$$R_{\perp} = \frac{(\varepsilon_1/\mu_1)^{1/2} \cos \theta_i - (\varepsilon_2/\mu_2)^{1/2} \cos \theta_t}{(\varepsilon_1/\mu_1)^{1/2} \cos \theta_i + (\varepsilon_2/\mu_2)^{1/2} \cos \theta_t}$$
(5.7.8)

and

$$T_{\perp} = \frac{2(\varepsilon_1/\mu_1)^{1/2}\cos\theta_i}{(\varepsilon_1/\mu_1)^{1/2}\cos\theta_i + (\varepsilon_2/\mu_2)^{1/2}\cos\theta_t}$$
(5.7.9)

For most cases, $\mu_1 = \mu_2 = \mu_0$, (5.7.8) and (5.7.9) can be written as

$$R_{\perp} = \frac{n_1 \cos \theta_i - n_2 \cos \theta_t}{n_1 \cos \theta_i + n_2 \cos \theta_t}$$
(5.7.10)

$$T_{\perp} = \frac{2n_1 \cos \theta_i}{n_1 \cos \theta_i + n_2 \cos \theta_t}$$
(5.7.11)

We note that $R_{\perp} = E_y^{t}/E_y^{i}$ given by (5.7.8) and $T_{\perp} = E_y^{t}/E_y^{i}$ given by (5.7.9) are referred to as the Fresnel formulas.

Similarly, for vertical polarization, the Fresnel formulas can be obtained by matching the boundary condition at z = 0 (see problem at the end of the chapter). We have

$$R_{\parallel} = \frac{H_y^r}{H_y^i} = \frac{n_2 \cos \theta_i - n_1 \cos \theta_t}{n_2 \cos \theta_i + n_1 \cos \theta_t}$$
(5.7.12)

$$T_{\parallel} = \frac{n_1 H_y^{\ i}}{n_2 H_y^{\ i}} = \frac{2n_1 \cos \theta_i}{n_2 \cos \theta_i + n_1 \cos \theta_t}$$
(5.7.13)

Note that the definitions for reflection and transmission coefficients for the vertically polarized wave are different from those for the horizontally polarized wave. Also, we note that because of the boundary condition, there is no coupling between the two polarizations at the boundary.

Suppose now that in the previous two media system, we let medium two

248

and



be anisotropic. This situation occurs quite often in practice. From the discussion in Chapter 2, we know that in general in an anisotropic medium without spatial dispersion two characteristic modes can propagate. The polarization of these two modes is in general elliptical. Therefore, a linearly polarized incident wave may become elliptically polarized after reflection in order to satisfy the boundary conditions at z = 0. Hence it is convenient to define four coefficients, ${}_{\perp}R_{\perp}$, ${}_{\perp}R_{\parallel}$, ${}_{\parallel}R_{\parallel}$, ${}_{\parallel}R_{\perp}$ to indicate the ratios of the various components of the reflected and incident electric fields. The first subscript denotes whether the incident electric vector is parallel (||) or perpendicular (\perp) to the plane of incidence, and the second subscript refers in the same way to the reflected electric vector. In exactly the same manner, we can define four transmission coefficients ${}_{\perp}T_{\perp}$, ${}_{\perp}T_{\parallel}$, ${}_{\parallel}T_{\parallel}$, ${}_{\parallel}T_{\parallel}$, ${}_{\parallel}T_{\parallel}$, ${}_{\parallel}T_{\parallel}$, ${}_{\parallel}T_{\perp}$.

As a sample illustration for dealing with the problem of reflection and transmission of waves of this nature, let us consider a system where medium one is the free space and medium two is a magnetoplasma, with the external magnetic field perpendicular to the boundary surface as shown in Fig. 5.7-2.

A linearly polarized wave in the x-direction in the free space is incident normally on the magnetoplasma from below. In the magnetoplasma, in general, the refracted wave will consist of two waves corresponding to the two characteristic modes in the medium. From (4.14.14), the polarization for the two modes are $\pm j$, respectively. Therefore, just above z = 0, the field components may be written as

$$E_{x} = E_{xa} + E_{xb}$$

$$E_{y} = -jE_{xa} + jE_{xb}$$

$$H_{y} = (\varepsilon_{0}/\mu_{0})^{1/2}(n_{a}E_{xa} + n_{b}E_{xb})$$

$$H_{x} = (\varepsilon_{0}/\mu_{0})^{1/2}(jn_{a}E_{xa} - jn_{b}E_{xb})$$
(5.7.14)

where

$$n_a^2 = 1 - X/(1 - Y)$$

$$n_b^2 = 1 - X/(1 + Y)$$
(5.7.15)

 E_{xa} and E_{xb} are, respectively, the x-components of the two modes.

Just below z = 0, the field components are

$$E_{x} = (1 + \|R_{\|})E_{x}^{(i)}$$

$$E_{y} = \|R_{\perp}E_{x}^{(i)}$$

$$H_{x} = (\epsilon_{0}/\mu_{0})^{1/2}\|R_{\perp}E_{x}^{(i)}$$

$$H_{y} = (\epsilon_{0}/\mu_{0})^{1/2}(1 - \|R_{\|})E_{x}^{(i)}$$
(5.7.16)

where $E_x^{(i)}$ is the incident wave and $\|R_\|$, $\|R_{\perp}\|$ are the two reflection coefficients.

Matching the four field components at z = 0, we obtain four equations for the four unknowns E_{xa} , E_{xb} , $||R_{||}$, and $||R_{\perp}$ in terms of $E_x^{(i)}$ and n_a , n_b . Solving them, we obtain

$$E_{xa} = \frac{E_x^{(i)}}{1+n_a}, \qquad E_{xb} = \frac{E_x^{(i)}}{1+n_b}$$

$$_{\parallel}R_{\parallel} = \frac{1}{2} \left[\frac{1-n_a}{1+n_a} + \frac{1-n_b}{1+n_b} \right], \qquad _{\parallel}R_{\perp} = j \left(\frac{1}{1+n_b} - \frac{1}{1+n_a} \right)^{(5.7.17)}$$

Thus we see the splitting of the incident wave $E_x^{(i)}$ into the two characteristic modes \mathbf{E}_a and \mathbf{E}_b of the magnetoplasma at the boundary. Also, the reflected wave consists of two polarizations, one $\|R_{\parallel}E_x^{(i)}$ in the x-direction and one $\|R_{\perp}E_x^{(i)}$ in the y-direction. The resultant reflected wave mode is in general elliptically polarized.

More general cases may be treated in the similar manner and results have been given by Budden (1961).

5.8 Wave Propagation in Stratified Media—Isotropic Case

We now turn to the more general case of stratified media, where the properties of the medium vary smoothly as functions of z. Let the medium be characterized by the permittivity and permeability

$$\varepsilon = \varepsilon(z), \qquad \mu = \mu(z)$$
 (5.8.1)

250

Consider now a plane, time-harmonic wave propagating through such a medium. Since the coefficients in the Maxwell's equations are functions of z only, the fields can be expressed as

$$\mathbf{E} = \mathbf{E}(z)e^{j[\omega t - k_x x - k_y y]}$$

$$\mathbf{H} = \mathbf{H}(z)e^{j[\omega t - k_x x - k_y y]}$$
(5.8.2)

respectively. Furthermore, without loss of generality, we can rotate the coordinate system about z-axis such that the wave normal vector is in the xz-plane, i.e., $k_y = 0$. Substituting (5.8.2) into Maxwell equations, we have

(a) $dE_y/dz = j\omega\mu H_x$ (d) $-dH_y/dz = j\omega\varepsilon E_x$

(b)
$$dE_x/dz + jk_xE_z = -j\omega\mu H_y$$
 (e) $dH_x/dz + jk_xH_z = j\omega\varepsilon E_y$ (5.8.3)

(c)
$$jk_x E_y = j\omega\mu H_z$$
 (f) $-jk_x H_y = j\omega\varepsilon E_z$

where the unknown fields **E** and **H** are functions of z only. As in the case of homogeneous medium, these equations can also be separated into two independent sets. One set involves (a), (c), and (e) of (5.8.3) for E_y , H_x , and H_z and is called the horizontally polarized wave (or transverse electric wave). The other set involves (b), (d), and (f) of (5.8.3) for E_x , E_z , and H_y and is called the vertically polarized wave (or transverse magnetic wave). Since Maxwell equations remain unchanged when **E** and **H** and simultaneously ε and $-\mu$ are interchanged, it is sufficient to study the horizontally polarized wave in detail in our general discussion. Any theorem relating to vertically polarized waves may be deduced from the corresponding result for the horizontally polarized wave by making the above-mentioned change.

From (a), (c), and (e) of (5.8.3), we obtain, for the horizontally polarized wave, the following equations:

$$dH_x/dz = j\omega(\varepsilon - k_x^2/\omega^2\mu)E_y$$
(5.8.4a)

$$dE_y/dz = j\omega\mu H_x \tag{5.8.4b}$$

$$k_x E_y - \omega \mu H_z = 0 \tag{5.8.4c}$$

We note here the close resemblance of (5.8.4a,b) to the transmission line equations. From (5.8.4) we can derive equations for H_x and E_y :

$$\frac{d^2 E_y}{dz^2} - \frac{d(\ln \mu)}{dz} \frac{dE_y}{dz} + (k_0^2 n^2 - k_x^2)E_y = 0 \qquad (5.8.5a)$$

and

$$\frac{d^2H_x}{dz} - \frac{d[\ln(\varepsilon - k_x^2/\omega^2\mu)]}{dz} \frac{dH_x}{dz} + (k_0^2n^2 - k_x^2)H_x = 0 \quad (5.8.5b)$$

where $k_0 = \omega/c$ and $n = c(\varepsilon \mu)^{1/2}$.

According to symmetry of the Maxwell equations, the equations for the vertically polarized wave can be obtained easily.

$$dE_x/dz = -j\omega(\mu - k_x^2/\omega^2\varepsilon)H_y \qquad (5.8.6a)$$

$$dH_y/dz = -j\omega\varepsilon E_x \tag{5.8.6b}$$

$$k_x H_y + \omega \varepsilon E_z = 0 \tag{5.8.6c}$$

and

$$\frac{d^2H_y}{dz^2} - \frac{d(\ln \varepsilon)}{dz} \frac{dH_y}{dz} + (k_0^2n^2 - k_x^2)H_y = 0 \quad (5.8.7a)$$

$$\frac{d^2 E_x}{dz^2} - \frac{d[\ln(\mu - k_x^2/\omega^2 \varepsilon)]}{dz} \frac{dE_x}{dz} + (k_0^2 n^2 - k_x^2)E_x = 0 \quad (5.8.7b)$$

Equations (5.8.4)-(5.8.7) form the basis of our discussion in the next few sections on wave propagation in stratified media.

5.9 The WKB Solution

For a given $\varepsilon(z)$ and $\mu(z)$, (5.8.5) and (5.8.7) can be transformed into the standard form

$$\frac{d^2u}{dz^2} + \frac{h^2q^2(z)u}{u} = 0 \tag{5.9.1}$$

where u is related to the components of the unknown fields E or H, q(z) is related to the refractive index of the stratified medium, and h^2 is related to k_0^2 , the free space wave number.

In general, the second-order ordinary differential equation (5.9.1) does not have closed form solutions in terms of known functions. Although series solutions can be obtained and numerical computation of the solution can be made, usually the slow convergence of the series prohibits these approaches. For the case where the parameter h is a large number, approximation techniques have been developed to treat this equation in order to yield simple solutions. One of the techniques is the WKB method which we are going to discuss in some detail in the following. But before we go

252

into the discussion of solving (5.9.1) analytically, let us consider a simple example which will reveal certain interesting aspects of the WKB technique. Let us assume that a stratified medium characterized by n(z) occupies the half-space z > 0. A horizontally polarized wave is incident normally from the free space z < 0. Assuming n(z) = 1 at z = 0, we wish to find the transmitted wave in the stratified region z > 0. Let us first replace the stratified region by a set of homogeneous layers $0 < z < z_1$, $z_1 < z < z_2$, $z_2 < z < z_3$, ..., etc., with the successive refractive indices n_1, n_2, n_3, \ldots as shown in Fig. 5.9-1. This discrete medium can be made to approach the



Fig. 5.9-1. Waves in multilayer medium.

continuous one by making the layers infinitely thin. For z < 0, the incident wave is given by

$$E_{v}^{i} = E_{0} e^{-jk_{0}z} \tag{5.9.2}$$

The reflected and transmitted waves at the boundary z = 0 are given, respectively, by

$$E_{y}^{r} = R = \frac{n_0 - n_1}{n_0 + n_1} E_0 e^{jk_0 z}$$
(5.9.3)

and

$$E_y^{\ t} = T_1 = \frac{2n_0}{n_0 + n_1} E_0 e^{-jk_0 n_1 z}$$
(5.9.4)

where the reflection coefficient (5.7.10) and the transmission coefficient (5.7.11) have been used.

The wave T_1 propagates upward to $z = z_1$. At this boundary, it will be split into a reflected wave R_2 and a transmitted wave T_2 . These waves are proportional to $e^{jk_0n_1z}$ and $e^{-jk_0n_2z}$, respectively. This procedure is repeated at each boundary. At any arbitrary level $z = z_s$, we have the relation

$$T_{s+1}(z_s)/T_s(z_s) = 2n_s/(n_s + n_{s+1})$$
(5.9.5)

The sequence of the transmitted waves T_1, T_2, T_3, \ldots may be termed the principal wave. They are the upward traveling waves due to the transmission at each level of the primary incident wave. There are other contributions to the overall up-going wave. For example, in layer Δz_1 , when R_2 reaches z = 0 it will be split into reflected and transmitted waves. The reflected wave of R_2 is actually up-going and contributes to the total up-going wave in layer Δz_1 . As can be seen from Fig. 5.9-1, this wave is produced after two successive reflections. Similiar argument shows that there are contributions to the up-going wave from waves that have undergone multiple reflections. At this point, however, let us just concentrate on the principal wave. From (5.9.5) and the fact that T_s is proportional to $e^{-jk_0n_s z}$ in the sth layer, it is easy to show that just below the level $z = z_n$ the principal wave has the form

$$E_0 \frac{2n_0}{n_0 + n_1} e^{-jk_0 n_1 \Delta z_1} \frac{2n_1}{n_1 + n_2} e^{-jk_0 n_2 \Delta z_2} \cdots \frac{2n_{n-1}}{n_{n-1} + n_n} e^{-jk_0 n_n \Delta z_n} \quad (5.9.6)$$

This expression can be put in the following form

$$T(z_n^{-}) = E_0 \exp\left\{-\sum_{s=0}^{n-1} \ln(1 + \Delta n_s/2n_s) - j \sum_{s=1}^n k_0 n_s \, \Delta z_s\right\} \quad (5.9.7)$$

where $\Delta n_s = n_{s+1} - n_s$ has been used.

Passing to the limit of $\Delta z_s \rightarrow 0$, the second sum in the exponential is transformed into the integral $-jk_0 \int_0^{z_n} n(s) ds$ and the first sum is transformed to

$$-\int_{s=0}^{s=z_n} dn_s/2n_s = -\frac{1}{2} \ln n(z_n)/n(0)$$
 (5.9.8)

Thus, by taking the limit of $\Delta z_s \rightarrow 0$, the principal wave T at z is given by

$$T(z) = E_0 \exp\left\{-\frac{1}{2} \ln\{n(z)/n(0)\} - jk_0 \int_0^z n(s) \, ds\right\}$$
$$= E_0 [n_0/n(z)]^{1/2} \exp\left[-jk_0 \int_0^z n(s) \, ds\right]$$
(5.9.9)

Comparing (5.9.9) with (5.3.17), we see that the principal wave is identical to the wave we obtained by applying the geometrical optics. This is called the WKB solution of the wave propagation problem we posed in Fig. 5.9-1. From the discussion above, we see that physically the WKB solution can be interpreted as the principal wave which does not include any multiple reflected waves. By taking into account the contributions from the multiple reflections, higher order terms in the approximate solution can be obtained (Bremmer, 1951). Although our discussion here is for normally incident horizontally polarized waves only, similar investigation can be carried out for waves propagating in arbitrary directions and also for vertically polarized waves.

5.10 The Matrix Method

Having in mind the physical picture of the WKB approximation discussed in the previous section, let us now turn to the analytic aspects of the technique. Historically, the mathematical technique known as the WKB method can be traced back to the nineteenth century [see Froman and Froman (1965) for a discussion on the historical development of the WKB method]. Modern developments of the theory started in 1915 (Gans, 1915; Jeffreys, 1923; Kramers, 1926). In 1926, Brillouin, Wentzel, and Kramers introduced the method in quantum mechanics to treat Schrodinger's wave equation. Since then, the name WKB method has been adopted by most authors (sometimes the name JWKB is also used to acknowledge the contribution by Jeffreys). Extensions of the theory have been developed in many respects and are closely related to the higher order geometrical optics. There are several different approaches in discussing this technique, each having its specific advantage. In what follows we shall adopt the so-called matrix method in our discussion.

The second-order differential equation (5.9.1) can be put into a system of two simultaneous first-order differential equations by defining a column vector \mathbf{x}

$$\mathbf{x}(z) = \begin{bmatrix} x_1(z) \\ x_2(z) \end{bmatrix} = \begin{bmatrix} u(z) \\ du(z)/dz \end{bmatrix}$$
(5.10.1)

Equation (5.9.1) can be written as

$$d\mathbf{x}/dz = \mathbf{A}(z) \cdot \mathbf{x} \tag{5.10.2}$$

where the matrix \mathbf{A} is given by

$$\mathbf{A}(z) = \begin{bmatrix} 0 & 1\\ -h^2 q^2(z) & 0 \end{bmatrix}$$
(5.10.3)

We now change the dependent variable \mathbf{x} by the transformation

$$\mathbf{x} = \mathbf{\Phi}(z) \cdot \mathbf{y} \tag{5.10.4}$$

where y is the new dependent vector and $\Phi(z)$ is the transformation matrix yet to be determined. Substituting (5.10.4) into (5.10.2), we have

$$d\mathbf{\Phi}/dz \cdot \mathbf{y} + \mathbf{\Phi} \cdot d\mathbf{y}/dz = \mathbf{A} \cdot \mathbf{\Phi} \cdot \mathbf{y}$$

Multiplying through by Φ^{-1} , provided Φ is nonsingular, we obtain

$$d\mathbf{y}/dz = (\mathbf{\Phi}^{-1} \cdot \mathbf{A} \cdot \mathbf{\Phi}) \cdot \mathbf{y} - \mathbf{\Phi}^{-1} \cdot d\mathbf{\Phi}/dz \cdot \mathbf{y}$$
(5.10.5)

The idea here is to introduce a transform Φ such that the matrix

$$\mathbf{D} = \mathbf{\Phi}^{-1} \cdot \mathbf{A} \cdot \mathbf{\Phi} \tag{5.10.6}$$

becomes diagonal. If this is achieved, (5.10.5) becomes two uncoupled equations when the second term on the right-hand side is neglected. Solutions can then be obtained by direct integration. From matrix theory, we know that the diagonalization of the matrix **A** is achieved by choosing **Φ** such that the columns in the matrix **Φ** are the eigenvectors of the eigenvalue problem

$$\mathbf{A} \cdot \mathbf{v} = \lambda \mathbf{v} \tag{5.10.7}$$

For A given by (5.10.3), the eigenvalues are obtained easily to be

$$\lambda_1 = -jhq, \qquad \lambda_2 = jhq \qquad (5.10.8)$$

The corresponding eigenvectors are

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -jhq \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ jhq \end{bmatrix}$$
(5.10.9)

Therefore the transform matrix has the form

$$\mathbf{\Phi} = \begin{bmatrix} 1 & 1 \\ -jhq & jhq \end{bmatrix}$$
(5.10.10)

256

Substituting (5.10.10) into (5.10.6), we obtain the diagonal matrix **D**:

$$\mathbf{D} = \mathbf{\Phi}^{-1} \cdot \mathbf{A} \cdot \mathbf{\Phi} = \begin{bmatrix} -jhq & 0\\ 0 & jqh \end{bmatrix}$$
(5.10.11)

where

$$\mathbf{\Phi}^{-1} = \frac{1}{2jhq} \begin{bmatrix} jhq & -1\\ jhq & 1 \end{bmatrix}$$
(5.10.12)

has been used.

The transformed equation (5.10.5) now has the form

$$\frac{d\mathbf{y}}{dz} = \frac{d}{dz} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} -jhq & 0 \\ 0 & jhq \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} - \frac{1}{4q^2} \frac{d(q^2)}{dz} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$
(5.10.13)

Equation (5.10.13) is still a set of two coupled equations. But if under the condition that the off-diagonal terms on the right-hand side of (5.10.13) can be neglected (the exact condition will be shown in the following), then we obtain two uncoupled equations for y_1 and y_2 :

$$\frac{dy_1}{dz} = \left[-jhq - \frac{1}{2q} \frac{dq}{dz}\right]y_1$$

$$\frac{dy_2}{dz} = \left[jhq - \frac{1}{2q} \frac{dq}{dz}\right]y_2$$
(5.10.14)

The solutions can be obtained immediately by integration:

$$y_1 = c_1 q^{-1/2} e^{-j\hbar \int^z q ds}$$
(5.10.15a)

$$y_2 = c_2 q^{-1/2} e^{j\hbar \int^z q ds}$$
(5.10.15b)

where c_1 and c_2 are constants and the lower limit of the integration depends on the phase reference level that one chooses. y_1 and y_2 may be termed the two independent characteristic waves in the stratified medium. Comparing (5.10.15a) with (5.9.9), we note that they are identical if $q^{1/2}$ and h are set equal to n and k_0 , respectively, and the phase reference level in (5.10.15) is chosen at z = 0. Thus y_1 represents the principal wave propagating in the positive z-direction. Similarly, y_2 represents the principal wave propagating in the direction of decreasing z. Substituting (5.10.15) into (5.10.4), using (5.10.10), we obtain the general solution for the original equation (5.9.1)

$$u(z) = x_1(z) = y_1(z) + y_2(z)$$
 (5.10.16)

 y_1 and y_2 are the WKB solutions of (5.9.1). If we form the Wronskian of y_1 and y_2 , we have

$$y_1 \, dy_2 / dz - y_2 \, dy_1 / dz = 2ihc_1 c_2 \tag{5.10.17}$$

provided the same lower limit is chosen for the integrals in both y_1 and y_2 . Therefore, the WKB solutions are independent solutions for the equation (5.9.1).

To see the exact condition for the validity of the WKB solution, let us assume that $c_1 = 1$ and $c_2 = 0$ in (5.10.15). This is legitimate since in the framework of WKB approximation y_1 and y_2 are uncoupled. Differentiating u(z) with respect to z twice, we have

$$\frac{d^2u}{dz^2} = q^{3/2} \left[\frac{5}{16} \frac{1}{q^6} \left(\frac{dq^2}{dz} \right)^2 - \frac{1}{4} \frac{1}{q^4} \left(\frac{d^2q^2}{dz^2} \right) - h^2 \right] e^{-j\hbar \int^z q ds}$$

Substituting this and (5.10.16) into (5.9.1), we have

$$\left\{\frac{q^{3/2}}{h^2}\left[\frac{5}{16} \ \frac{1}{q^6}\left(\frac{dq^2}{dz}\right)^2 - \frac{1}{4} \ \frac{1}{q^4}\left(\frac{d^2q^2}{dz^2}\right)\right] + q^{3/2} - q^{3/2}\right\}e^{-jh\int^z qds} = 0$$
(5.10.18)

This equation is approximately satisfied only if

$$\left|\frac{3}{4h^2} \frac{1}{q^3} \left(\frac{dq}{dz}\right)^2 - \frac{1}{2h^2 q^2} \frac{d^2 q}{dz^2}\right| \ll 1$$
 (5.10.19)

which shows that h must be large and q must not be too close to zero. Thus (5.10.19) gives the necessary condition for the WKB solution to be valid. Only when (5.10.19) is satisfied, can we write the general solution for (5.9.1) as the sum of the two independent solutions y_1 and y_2 . The error that one makes in using the WKB solution can be roughly estimated from (5.10.19).

From (5.10.13), an iterative procedure can be used to compute the higher order terms in the solution. Keeping all the terms in (5.10.13), we have

$$\frac{dy_1}{dz} + \left(jhq + \frac{1}{2q} \ \frac{dq}{dz}\right)y_1 = \frac{1}{2q} \ \frac{dq}{dz} y_2$$

$$\frac{dy_2}{dz} + \left(-jhq + \frac{1}{2q} \ \frac{dq}{dz}\right)y_2 = \frac{1}{2q} \ \frac{dq}{dz} y_1$$
(5.10.20)

The equations can be put into dimensionless form by letting z' = hz.

We have

$$\frac{dy_1}{dz'} + \left(jq^{1/2} + \frac{1}{2q} \ \frac{dq}{dz'}\right)y_1 = \frac{1}{2q} \ \frac{dq}{dz'} y_2$$

$$\frac{dy_2}{dz'} + \left(-jq^{1/2} + \frac{1}{2q} \ \frac{dq}{dz'}\right)y_2 = \frac{1}{2q} \ \frac{dq}{dz'} y_1$$
(5.10.21)

Let us now consider an example for which an up-going (increasing z') wave is incident vertically upon a stratified medium occupying the halfspace z' > 0 at z' = 0. q is assumed to be continuous at z' = 0 and q = 1for z < 0. This is the same problem we have studied in Section 5.9. The formal solutions for (5.10.21) can be obtained in a straightforward manner to yield

$$y_1(z') = y_{10}(z') + \int_0^{z'} \frac{1}{2q(s)} \frac{dq}{ds} [y_{10}(z')/y_{10}(s)]y_2(s) ds \quad (5.10.22a)$$

$$y_2(z') = y_{20}(z') + \int_{\infty}^{z'} \frac{1}{2q(s)} \frac{dq}{ds} [y_{20}(z')/y_{20}(s)]y_1(s) ds \quad (5.10.22b)$$

where

$$y_{10}(z') = c_1 q^{-1/2} e^{-j \int_0^{z'} q ds}$$

$$y_{20}(z') = c_2 q^{-1/2} e^{j \int_0^{z'} q ds}$$
(5.10.23)

The lower limit of integration in (5.10.22a) is obtained by using the condition that at z' = 0, the up-going wave is the incident wave y_{10} . The lower limit of integration in (5.10.22b) is obtained by using the condition that as $z' \rightarrow \infty$ there is no reflected (down-coming) wave. In addition, since the incident wave is upgoing, there is no original down-coming wave. Therefore c_2 in (5.10.23) must be zero.

Equation (5.10.21) is now formally transformed into a set of coupled integral equations by substituting (5.10.23) into (5.10.22). We have

$$y_{1}(z') = y_{10}(z') + \frac{1}{2[q(z)]^{1/2}} \int_{0}^{z'} \frac{1}{[q(s)]^{1/2}} \frac{dq(s)}{ds}$$

$$\times e^{j\int_{s}^{s} qd\tau} y_{2}(s) ds \qquad (5.10.24a)$$

$$y_{2}(z') = -\frac{1}{2[q(z')]^{1/2}} \int_{z'}^{+\infty} \frac{1}{[q(s)]^{1/2}} \frac{dq(s)}{ds}$$

$$\times e^{j\int_{s}^{s'} qd\tau} y_{1}(s) ds \qquad (5.10.24b)$$

This set of integral equations can be used as the basis for an interative procedure to find a higher order solution for the problem. For $|(1/q)dq/ds| \ll 1$, the zeroth-order solution is $y_{10}(z')$, up-going wave. There is no coupling

to the down-coming wave. It is the WKB solution or the principal wave discussed in the last section. The second-order solution can be obtained by substituting y_{10} into (5.10.24b):

$$y_{2}(z') = -\frac{1}{2[q(z')]^{1/2}} \int_{z'}^{\infty} \frac{1}{[q(s)]^{1/2}} \frac{dq(s)}{ds} \times e^{j\int_{z'}^{z'} qd\tau} y_{10}(s) \, ds \qquad (5.10.25)$$

This represents the contribution from partial reflections of the principal incident wave at each level along the path of the wave. The third-order solution for y_1 is obtained by substituting (5.10.25) into the integral of (5.10.24a). Each iterative step adds the contribution from multiple reflections to the total solution. This is exactly the mechanism we discussed in the last section in regard to the physical picture depicted in Fig. 5.9-1. It can be shown that the iterative procedure will generate a series in ascending powers of (1/k). The first term of this series is the WKB solution. Thus, we have an analytic procedure of obtaining the WKB solution and higher order approximations. However, there are still certain important topics remaining to be considered in the theory of the WKB technique. First, let us turn back to (5.10.19) to take a closer look at the condition of validity for WKB solutions. Obviously, in the neighborhood of q = 0, the solution is not valid. In this neighborhood, the off-diagonal coupling terms in (5.10.13) are no longer negligible; y_1 and y_2 are no longer independent. The points at which q = 0 are called transition points or reflection points, or turning points. The behavior of a WKB solution when it passes through a transition point is one of the most important topics in the development of the WKB technique. Furthermore, if we extend the domain for the solution to the complex z-plane, then from (5.9.1), the solution should be single valued in a domain where q(z) is free from singularities. But the WKB solutions y_1 and y_2 clearly are not single valued due to the presence of the roots of q in the expression. Therefore y_1 and y_2 obviously can only be valid in some restricted domain of the complex z-plane. In the following section, we shall consider these questions via an example.

5.11 The Stokes Phenomenon

To study the behavior of the solution to (5.9.1) near a turning point we may expand q about this point. In the neighborhood of the turning point we may discard all higher order terms except the linear term provided it

does not vanish, i.e., the turning point is of first order (see definition on order of turning point given in Section 5.13). This leads directly to consideration of the Airy equation (or the Stokes equation)

$$\frac{d^2u}{dz^2} - zu = 0 \tag{5.11.1}$$

The series solutions of (5.11.1) may be obtained by the standard method to yield

$$u(z) = a_0 \left[1 + \frac{z^3}{3 \cdot 2} + \frac{z^6}{6 \cdot 5 \cdot 3 \cdot 2} + \frac{z^9}{9 \cdot 8 \cdot 6 \cdot 5 \cdot 3 \cdot 2} + \cdots \right] + a_1 \left[z + \frac{z^4}{4 \cdot 3} + \frac{z^7}{10 \cdot 9 \cdot 7 \cdot 6 \cdot 4 \cdot 3} + \cdots \right]$$
(5.11.2)

where a_0 and a_1 are two arbitrary constants. The standard Airy functions are defined by setting (Jeffreys and Jeffreys, 1956)

$$a_0 = 3^{-2/3} / \Gamma(2/3), \quad a_1 = -3^{-1/3} / \Gamma(1/3) \quad \text{for } Ai(z)$$

$$a_0 = 3^{-1/6} / \Gamma(2/3), \quad a_1 = 3^{1/6} / \Gamma(1/3) \quad \text{for } Bi(z)$$
(5.11.3)

Ai(z) and Bi(z) are the two standard solutions for the Airy equation (5.11.1). The behavior of Ai and Bi is plotted in Fig. 5.11-1.



Fig. 5.11-1. The behavior of Airy functions.

For large values of z, instead of the series solution (5.11.2), WKB solutions are much more useful. The two WKB solutions that are valid for $|z| \gg 1$ are readily obtained from (5.10.15).

$$y_1 = z^{-1/4} e^{-2z^{3/2}/3} \tag{5.11.4a}$$

$$y_2 = z^{-1/4} e^{2z^{3/2}/3}$$
 (5.11.4b)

Let us now extend the solution into the complex z-plane. To do this, we have to know the behavior of y_1 and y_2 in various regions of the z-plane. Since z = 0 is a branch point for both y_1 and y_2 , we take a branch cut OF as shown in Fig. 5.11-2 in the complex z-plane such that y_1 and y_2 become single-valued functions in the proper manner. In addition we drawn the lines OB, OD, and OG which correspond to

$$\operatorname{Re}[z^{3/2}] = 0$$
 or $\arg z = \pi/3, \pi, -\pi/3$ (5.11.5)



Fig. 5.11-2. The Stokes and anti-Stokes Lines for Airy equation in the complex z-plane.

respectively. These are called anti-Stokes lines along which the magnitude of the exponential function in (5.11.4) is unity. In Regions 1 and 7 where $-\pi/3 < \arg z < \pi/3$, y_2 is much greater than y_1 for large values of |z|. Therefore in this region y_2 is termed dominant and denoted by y_{2d} while y_1 is a subdominant and denoted by y_{1s} . These notations are also used in the other regions to indicate a dominant or subdominant solution. Along an anti-Stokes line, it is not possible to distinguish which one of y_1 and y_2 is dominant. The lines *OA*, *OC*, and *OE* correspond to

$$\text{Im}[z^{3/2}] = 0$$
 or $\arg z = 0, \pm 2\pi/3$ (5.11.6)

respectively. They are the Stokes lines along which a subdominant solution has its minimum and a dominant solution has its maximum.

The branch cut from z = 0 may be drawn anywhere as long as proper care is taken to assign values to the functions. If a WKB solution y_{2d} is continued across the branch cut in the counterclockwise sense, just before it crosses the branch cut at arg $z = \delta$, we have $z = re^{j\delta}$ such that

$$y_{2d} = r^{-1/4} \exp[-j\delta/4] \exp[2r^{3/2} \exp(3j\delta/2)/3]$$
 (5.11.7)

In order that this solution is continuous across the branch cut it must be

changed to $-jy_{1d}$ since across the branch cut $z = re^{-j(2\pi-\delta)} = re^{-j2\pi}e^{j\delta}$ and

$$-jy_{1d} = -jr^{-1/4} \exp(j\pi/2) \exp(-j\delta/4) \exp[-2r^{3/2} \exp(3j\delta/2) \exp(-j3\pi)/3]$$

= $r^{-1/4} \exp(-j\delta/4) \exp[2r^{3/2} \exp(3j\delta/2)/3]$ (5.11.8)

which is exactly the same as (5.11.7). Therefore, we have the rules of continuing the two WKB solutions across the branch cut counterclockwise as follows

$$y_2 \rightarrow -jy_1, \quad y_1 \rightarrow -jy_2$$
 (5.11.9)

The property of dominancy or subdominancy of the solution is preserved in the process. If the branch cut is crossed clockwise, -j is replaced by j in (5.11.9). When a solution crosses the anti-Stokes line, dominant solution becomes subdominant and vice versa.

The WKB approximation for the most general solution of the Airy equation may be written as the linear combination of y_1 and y_2 given by (5.11.4). We have

$$u(z) = Ay_1(z) + By_2(z)$$

= $Az^{-1/4}e^{-2z^{3/2}/3} + Bz^{-1/4}e^{2z^{3/3}/3}$ (5.11.10)

where A and B are two arbitrary constants. In any region of the complex z-plane, except along anti-Stokes lines, one of the two WKB solutions in (5.11.10) is dominant and the other is subdominant. When the solution u(z) is traced about z = 0 in the complex z-plane, the dominancy and subdominancy of y_1 and y_2 will change according to the rules discussed above. In general, it can be shown that different values of A or B must be used in representing u(z) in different regions of the complex z-plane (Budden, 1961; Heading, 1962; Jeffreys and Jeffreys, 1956). This is known in the asymptotic theory as "the Stokes phenomenon of the discontinuity of the arbitrary constants." It is obvious that in a particular region, the change should occur on the coefficient of the term which is subdominant in this region so that the accuracy of the solution is not affected. The most logical place for this discontinuous change of subdominant coefficient to occur is on the Stokes line where the subdominant term is minimum. Stokes (1858) has shown that for maximum accuracy in the summation of an asymptotic series, the change of coefficients must take place on the Stokes lines. A formula may be written for this coefficient change when a solution is traced across a Stokes line in the counterclockwise sense. We write

new subdominant coefficient = old subdominant coefficient

 $+ \lambda \times \text{dominant coefficient}$ (5.11.11)

 λ is called the Stokes constant associated with the particular Stokes line. To determine the Stokes constants, let us trace a WKB solution around the origin. In Fig. 5.11-2, let us assume that a solution is given on the anti-Stokes line *OB* by the expression

$$u(z) = Ay_1(z) + By_2(z)$$

where the dominancy and subdominancy of y_1 and y_2 are not determined.

In Region 2, the solution is

$$Ay_{1d} + By_{2s} \tag{5.11.12}$$

From Regions 2 to 3, a Stokes line is crossed in a counterclockwise sense. Applying (5.11.11), we have in Region 3

$$Ay_{1d} + (B + \lambda_C A)y_{2s} \tag{5.11.13}$$

where λ_C is the Stokes constant for the Stokes line OC.

In Region 1, the solution is

$$Ay_{1s} + By_{2d} \tag{5.11.14}$$

In Region 7, we have

$$(A - \lambda_A B)y_{1s} + By_{2d} \tag{5.11.15}$$

where λ_A is the Stokes constant for the Stokes line OA; the negative sign is due to the fact that the Stokes line is crossed in the clockwise sense.

In Region 6, the solution is

$$(A - \lambda_A B)y_{1d} + By_{2s} \tag{5.11.16}$$

From Regions 6 to 5, the branch cut OF is crossed in clockwise sense. Applying (5.11.9), we have in Region 5

$$j(A - \lambda_A B)y_{2d} + jBy_{1s}$$
 (5.11.17)

In Region 4, the solution is

$$j(A + \lambda_A B)y_{2d} + j[B - \lambda_E (A - \lambda_A B)]y_{1s} \qquad (5.11.18)$$

where λ_E is the Stokes constant for the Stokes line OE.

When this solution is traced across the anti-Stokes line OD, it should agree with the solution in Region 3 given by (5.11.13). Therefore we have

$$A = j[B - \lambda_E(A - \lambda_A B)]$$
(5.11.19)

$$B + \lambda_C A = j(A - \lambda_A B) \tag{5.11.20}$$

These relations should be valid for arbitrary values of A and B. Therefore we may equate the coefficients of A and B in the two equations, respectively. From (5.11.19), we obtain

$$1 = -j\lambda_E, \quad 1 + \lambda_E\lambda_A = 0$$

From (5.11.20), we have

$$\lambda_C = j, \quad 1 = -j\lambda_A$$

Therefore the Stokes constants are given by

$$\lambda_A = \lambda_C = \lambda_E = j \tag{5.11.21}$$

With the Stokes constants given, we can trace any WKB solution of (5.11.1) in the complex z-plane by applying rules (5.11.9) and (5.11.11). For example, a solution y_1 in the Region 1 and 7 of Fig. 5.11-2 is subdominant. When it is traced counterclockwise to the anti-Stokes line, we have

Regions 7, 1:
$$y_{1s}$$

Region 2: y_{1d}
Region 3: $y_{1d} + jy_{2s}$

Therefore on OD, the solution is y_1 and jy_2 . The same result is obtained if the solution is traced clockwise. Thus we have the procedure of extending the WKB solution to the whole complex z-plane.

The asymptotic expressions for the two standard solutions Ai(z) and Bi(z) of the Airy equation can be related to the two WKB solutions y_1 and y_2 by applying the above procedure. From the behavior of Ai(z), it can be shown that the WKB approximation for Ai(z) is given by

$$Ai(z) \sim -\frac{1}{2\sqrt{\pi}} y_1(z)$$
 (5.11.22)

in Regions 1 and 7 of Fig. 5.11-2.

On the positive real axis where arg z = 0, y_1 is subdominant and decreases exponentially. On the negative real axis where arg $z = 180^{\circ}$, from the dis-

cussion above, we see that

$$Ai(-|z|) \sim \frac{1}{2\sqrt{\pi}} (y_1 + jy_2)$$

= $\pi^{-1/2} |z|^{-1/4} \cos(2|z|^{3/2}/z - \pi/4)$ (5.11.23)

We could have started from (5.11.23) and continued it around z = 0and obtained (5.11.22) for the right asymptotic behavior of Ai(z) along positive z-axis.

For Bi(z), however, we cannot start from the positive z-axis since Bi is unbounded for large positive z, corresponding to a dominant WKB solution. But in the limit of WKB approximation, there could also be a subdominant WKB term in this region and the constant multiplier in front of this subdominant term must be chosen in such a way that when continued around the origin, the WKB solution gives the correct asymptotic behavior of Bi(z). There is no way of determining this constant a priori; therefore we start from the negative z-axis. Along the negative z-axis, the WKB approximation for Bi(z) can be written as

Region 3:
$$Bi(z) \sim \frac{1}{2\sqrt{\pi}} [jy_{1d} + y_{2s}]$$

Region 2: $Bi(z) \sim \frac{1}{2\sqrt{\pi}} [jy_{1d} + 2y_{2s}]$
Region 1: $Bi(z) \sim \frac{1}{2\sqrt{\pi}} [jy_{1s} + 2y_{2d}]$
Region 7: $Bi(z) \sim \frac{1}{2\sqrt{\pi}} [-jy_{1s} + 2y_{2d}]$ (5.11.24)
Region 6: $Bi(z) \sim \frac{1}{2\sqrt{\pi}} [-jy_{1d} + 2y_{2s}]$
Region 5: $Bi(z) \sim \frac{1}{2\sqrt{\pi}} [y_{2d} + j2y_{1s}]$
Region 4: $Bi(z) \sim \frac{1}{2\sqrt{\pi}} [y_{2d} + jy_{1s}]$

Along the positive real axis, the subdominant term can be neglected as

compared to the dominant term. Therefore

$$Bi(|z|) \sim \frac{1}{\sqrt{\pi}} y_{2d} = \frac{1}{\sqrt{\pi}} |z|^{-1/2} e^{2|z|^{3/2}/3}$$
(5.11.25)

Bi is exponentially large along the positive real axis.

From (5.11.24) and (5.11.25) we can write

$$z^{-1/4}[je^{-2z^{3/2}/3} + e^{2z^{3/2}/3}] \to 2z^{-1/4}e^{2z^{3/2}/3}$$
(5.11.26)

to indicate the connection of the WKB solution on the negative real axis to that on the positive real axis. This is called the connection formula. We note that the arrow in (5.11.26) is one way indicating the fact that to go from +z to -z, a subdominant term should be added to (5.11.26), a fact we discussed above. Similarly, for Ai(z), the connection formula is given by

$$z^{-1/4}[e^{-2z^{3/2}/3} + je^{2z^{3/2}/3}] \leftrightarrow z^{-1/4}e^{-2z^{3/2}/3}$$
(5.11.27)

A two-way arrow is used here since the connections in both directions are permissible.

The asymptotic behavior of Ai and Bi on the real axis as given by (5.11.22), (5.11.23), (5.11.24), and (5.11.25) are evident in Fig. 5.11-2.

Summarizing the essentials of this section, we note that we have shown a procedure to trace the WKB solutions of the Airy equation about z = 0, the turning point in the complex z-plane, such that when a WKB solution is known in certain domains in the z-plane, the form of this WKB solution in other regions of the z-plane can be obtained. The detailed behavior of the solution in the neighborhood of the turning point can not be obtained from the WKB solution. However after it has passed through the turning point and reaches a region where the WKB solution is again applicable, the solution in this region is predicted by the connection formulas.

The discussion on the Stokes phenomenon and connection formulas can be extended to asymptotic solutions of differential equations of other types where q(z) is an arbitrary function of z with higher order turning points (see Section 5.13 for definition). Some such general cases will be discussed in the following sections.

5.12 An Example

We now apply the results derived in the last section to study an example. Let us consider a horizontally polarized wave incident from free space upon a plasma. The electron density of the plasma is assumed to vary linearly with z. From the discussion in Chapter 4, we have the refractive index for z > 0.

$$n^2 = 1 - \omega_{pe}^2 / \omega^2 = 1 - az, \quad \mu = \mu_0$$
 (5.12.1)

where the electron density is assumed to vary as αz and where $a = e^2 \alpha / m \varepsilon_0 \omega^2$ is a constant.

Substituting (5.12.1) into (5.6.5a), we have

$$\frac{d^{2}E_{y}}{dz^{2}} + (k_{0}^{2} - k_{x}^{2} - k_{0}^{2}az)E_{y} = 0, \quad z \ge 0$$

$$\frac{d^{2}E_{y}}{dz^{2}} + (k_{0}^{2} - k_{x}^{2})E_{y} = 0, \quad z \le 0$$
(5.12.2)

The generalized Snell's law gives $k_x = k_0 \sin \theta_i$ where θ_i is the incident angle of the wave. Therefore, (5.12.2) can be written as

$$\frac{d^{2}E_{y}/dz^{2} + k_{0}^{2}(\cos^{2}\theta_{i} - az)E_{y} = 0, \quad z \ge 0}{d^{2}E/dz^{2} + k_{0}^{2}\cos^{2}\theta_{i}E_{y} = 0, \quad z \le 0}$$
(5.12.3)

Comparing (5.12.3) with (5.9.1), we see that E_y , k_0 correspond to μ and h, respectively, and

$$q^{2}(z) = \begin{cases} \cos^{2}\theta_{i} - az, & z \ge 0\\ \cos^{2}\theta_{i}, & z \le 0 \end{cases}$$
(5.12.4)

The equation for $z \ge 0$ can be transformed into the standard Airy equation by the following transformation:

$$\xi = (k_0^2 a)^{1/3} (z - \cos^2 \theta_i / a) \tag{5.12.5}$$

We have

$$d^{2}E_{y}/d\xi^{2}-\xi E_{y}=0, \qquad \xi \geq -(k_{0}/a)^{2/3}\cos^{2}\theta_{i}$$
 (5.12.6)

Hence the general solution of E_y for $z \ge 0$ is a linear combination of $Ai(\xi)$ and $Bi(\xi)$. As z becomes very large, ξ also is large and positive. From our discussion in the previous section on the asymptotic behavior of Ai and Bi, we know that for ξ large and positive $Bi(\xi)$ is exponentially large while $Ai(\xi)$ is exponentially small. Physical argument shows that for $z \to \infty$ we can not have infinite amplitude for the wave. Therefore, in the expression for E_y , the term $Bi(\xi)$ should not be included. Thus, we write

$$E_y = TAi(\xi), \qquad z \ge 0 \tag{5.12.7a}$$

268

The corresponding horizontal magnetic field H_x can be obtained from the Maxwell equation (5.8.3*a*):

$$H_x = -j \frac{1}{\omega \mu} \frac{dE_y}{dz} = -j[k_0^2 a)^{1/3} / \omega \mu] TAi'(\xi), \qquad z \ge 0 \qquad (5.12.7b)$$

where $Ai'(\xi) = dAi(\xi)/d\xi$.

For $z \leq 0$, the solution for (5.12.3) is easily found to be

$$E_{y} = e^{-jk_{0}\cos\theta_{i}z} + Re^{jk_{0}\cos\theta_{i}z}, \quad z \le 0$$

$$H_{x} = (1/\omega\mu)(-k_{0}\cos\theta_{i})[e^{-jk_{0}\cos\theta_{i}z} - Re^{jk_{0}\cos\theta_{i}z}] \quad (5.12.8)$$

At the boundary z = 0, E_y and H_x must be continuous. Thus we obtain the relations

$$IAi(\xi_0) = 1 + R$$

$$jT(k_0^2 a)^{1/3} Ai(\xi_0) = k_0 \cos \theta_i (1 - R)$$
(5.12.9)

where

$$\xi_0 = \xi(z=0) = -(k_0/a)^{2/3} \cos^2 \theta_i$$
 (5.12.10)

Eliminating T in (5.12.9), we obtain the reflection coefficient R

$$R = \frac{\cos \theta_i Ai(\xi_0) - j(a/k_0)^{1/3} Ai'(\xi_0)}{\cos \theta_i Ai(\xi_0) + j(a/k_0)^{1/3} Ai'(\xi_0)}$$
(5.12.11)

The transmission coefficient T is obtained from (5.12.9)

$$T = \frac{2\cos\theta_i}{\cos\theta_i Ai(\xi_0) + j(a/k_0)^{1/3} Ai'(\xi_0)}$$
(5.12.12)

We note first that in the region $\xi \ge 0$, or $z \ge z_0 = \cos^2 \theta_i/a$, the waves are evanescent. No power propagates into the region $z \ge z_0$. This can be seen by forming the complex Poynting vector $\mathbf{E} \times \mathbf{H}^*$ for $z \ge z_0$ and noting that $\operatorname{Re}(\mathbf{E} \times \mathbf{H}^*) = 0$. Therefore, all the incident power is reflected. This is evident from the fact that |R| = 1 in (5.12.11), since the numerator and denominator are complex conjugates in the expression for R.

The expressions (5.12.11) and (5.12.12) are exact. If $|\xi_0| = (k_0/a)^{2/3} \cos^2 \theta_i$ $\gg 1$, then the WKB approximation for Ai [(5.11.23) for arg $\xi = \pi$)] can be used in (5.12.11). The WKB approximation for Ai'(ξ) is obtained by a direct differentiation of the expression for Ai. We have

$$Ai'(-|\xi_0|) \sim \pi^{-1/2} |\xi_0|^{1/4} \sin(2|\xi_0|^{3/2}/3 - \pi/4) \qquad (5.12.13)$$

Substituting (5.11.23) and (5.12.13) into (5.12.11) we obtain

$$R \sim \frac{\cos(2 \mid \xi_0 \mid^{3/2}/3 - \pi/4) - j \sin(2 \mid \xi_0 \mid^{3/2}/3 - \pi/4)}{\cos(2 \mid \xi_0 \mid^{3/2}/3 - \pi/4) + j \sin(2 \mid \xi_0 \mid^{3/2}/3 - \pi/4)}$$

= exp $j[\pi/2 - 4 \mid \xi_0 \mid^{3/2}/3]$
= $j \exp[-j4 \mid \xi_0 \mid^{3/2}/3]$
= $j \exp[-j4k_0 \cos^3 \theta_i/3a]$ (5.12.14)

The condition $|\xi_0| \gg 1$ implies that $\cos^2 \theta_i$ should not be too small and $|k_0/a| \gg 1$. This latter condition indicates that in order that the WKB approximation may be used, the electron density should vary slowly as compared to the wavelength of the wave. Or, in other words, in one wavelength, the electron density should not change much. Equation (5.12.14) indicates that the reflected wave has the same amplitude as the incident wave with a phase difference $\pi/2 - 4k_0 \cos^3 \theta_i/3a$. We obtain this WKB approximation for the reflection coefficient by applying the WKB approximation to the exact expression of R. In the next section, we shall show that it is possible to obtain the WKB approximation for the reflection coefficient directly without knowing its exact expression. The technique is based on the connection formula for the WKB solutions.

We note that in the exact formula for the reflection coefficient R[(5.12.11)] partial reflections from z = 0 to $z = z_0$ are included. In the approximate expression (5.12.14), however, when WKB approximation is used, these partial reflections are neglected.

5.13 Reflection Coefficients for Stratified Media—High Frequency Approximation

In the following, we shall discuss the technique of obtaining the WKB approximation for the reflection coefficient directly without knowing its exact expression.

As an illustrative example of the technique, let us consider the propagation of a horizontally polarized wave in a stratified plasma. This time, the electron density varies arbitrarily as a function of z. The wave equation can be written in general as

$$d^{2}E_{y}/dz^{2} + k_{0}^{2}q^{2}(z)E_{y} = 0$$
(5.13.1)

q(z) is assumed to possess the following properties:

(i) q(z) is real on the real z-axis and approaches the value 1 as $z \to -\infty$ along the real axis.

(ii) The point $z = z_0$ is a turning point of order unity for q(z). By that we mean $q^2(z)$ can be expanded about $z = z_0$ by

$$q^{2}(z) = a_{1}(z - z_{0}) + a_{1}(z - z_{0})^{2} + a_{3}(z - z_{0})^{3} + \cdots$$
 (5.13.2)

where $a_1 \neq 0$. [If $(z - z_0)^n$ is the first nonvanishing term in the expansion, the turning point is defined to be of order n.]

(iii) Consistent with the consideration of wave reflections the point $z = z_0$ is assumed to be the only turning point. Consequently we may assume $q^2(z) < 0$ for $z > z_0$, which requires $a_1 < 0$. Actually the technique to be discussed can be easily extended to cases of isolated turning points in the sense that between two neighboring turning points there exists a region in which WKB solutions are valid.

If k_0^2 is large, the two WKB solutions of (5.13.1) are given by

$$y_1 = q^{-1/2} e^{-jk_0 \int^z q ds} \tag{5.13.3}$$

$$y_2 = q^{-1/2} e^{jk_0 \int^z q ds} \tag{5.13.4}$$

where y_1 represents a wave propagating in the direction of increasing z (see Section 5.10) and y_2 a wave propagating in the direction of decreasing z. The branch of q is chosen such that when q < 0, q is negative imaginary. Therefore y_1 becomes a decaying wave as z increases. These solutions are not valid near the turning point $z = z_0$.

Just as in the discussion of the Airy equation, we can define the anti-Stokes lines radiating from the turning point $z = z_0$ by

$$\operatorname{Re}\left[j\int_{z_0}^z q\,ds\right] = 0 \tag{5.13.5}$$

and the Stokes lines by

$$Im\left[j\int_{z_0}^{z} q \, ds\right] = 0 \tag{5.13.6}$$

In the neighborhood of $z = z_0$, $q^2 = a_1(z - z_0)$, $a_1 < 0$; therefore the Stokes and anti-Stokes lines are the same as those shown in Fig. 5.11-2 for the Airy function. In Fig. 5.13-1, we draw the circle *b* about the turning point $z = z_0$. Within this circle, the linear approximation of *q* is assumed to be valid; hence the Stokes and anti-Stokes lines are drawn exactly the same way as in Fig. 5.11-2.



Fig. 5.13-1. The Stokes and anti-Stokes lines in complex z-plane.

The circle *a* is drawn to indicate the region in which the WKB solutions are not applicable. We assume that the region *b* (in which *q* can be assumed linear) is large enough such that at the boundary of the *b* circle the WKB solutions of (5.13.15) can be used. The exact requirements on *q* and k_0 for this assumption to hold may be obtained analytically but will not be considered here [see, for example, Buddin (1961); Heading (1962)].

Physically, the solution of (5.13.1) for the electric field should be an outgoing or decaying wave as $z \to +\infty$. Therefore, the WKB solution for (5.13.1) should be y_1 in Regions 1 and 7 and is subdominant. When this solution is traced about $z = z_0$ in the complex z-plane, the same technique discussed in the Section 5.11 can be used. We note that the Stokes constants derived in Section 5.11 are for WKB solutions with the "phase-reference" (the lower limit of the integral in WKB solution) taken at the turning point. Therefore, in tracing the solution, we should write, in different regions, the following expressions.

In Regions 1 and 7,

$$y_{1s} = [q^{-1/2}e^{-jk_0\int_0^z qds}]_s$$

= $e^{-jk_0\int_0^{z_0} qds}[q^{-1/2}e^{-jk_0\int_{z_0}^z qds}]_s$

The term in the bracket has the phase reference level at the turning point $z = z_0$. It is this term that we shall trace in the complex z-plane about z_0 . Following the rules explained in Section 5.11, we obtain in Region 2,

$$y_{1d} = e^{-jk_0 \int_0^{z_0} qds} [q^{-1/2} e^{-jk_0 \int_{z_0}^{z} qds}]_d$$

In Region 3,

$$y_{1d} + jy_{2s} = e^{-jk_0 \int_0^{z_0} qds} \{ [q^{-1/2} e^{-jk_0 \int_{z_0}^{z} qds}]_d + j [q^{-1/2} e^{jk_0 \int_{z_0}^{z} qds}]_s \}$$

Therefore on the anti-Stokes line OD, the WKB solution is

$$q^{-1/2}e^{-jk_0\int_0^z qds} + je^{-j2k_0\int_0^{z_0} qds}[q^{-1/2}e^{jk_0\int_0^z qds}]$$
(5.13.7)

Hence, on the real z-axis for $z < z_0$, the WKB solution of the wave equation (5.13.1) is given by (5.13.7). This expression satisfies the boundary condition that as $z \to \infty$ the WKB solution is an outgoing wave. The first term in (5.13.7) is the incident wave and the second term is the reflected wave. The reflection coefficient is given by the coefficient of this second term. Thus,

$$R = j e^{-j2k_0 \int_0^{z_0} q ds} \tag{5.13.8}$$

This is the WKB approximation for the reflection coefficient. This technique of computing the reflection coefficient is sometimes referred to as the "phase integral" method. The technique can be extended to cases where lossy media are considered. For those cases, q becomes complex. The phase integral method is still applicable provided a suitable contour path from 0 to z_0 in the complex s-plane is chosen. Details of this method can be found in Heading's monograph (1962).

As an example, let us use (5.13.8) to compute the approximate reflection coefficient for the case of plasma with linear electron density. q(z) is given by (5.12.4); therefore.

$$R = j \exp\left[-j2k_0 \int_0^{z_0} (\cos^2 \theta_i - as)^{1/2} ds\right]$$

= $j \exp[-j4k_0 \cos^3 \theta_i/3a]$ (5.13.9)

where $z_0 = \cos^2 \theta_i / a$ has been used. Equation (5.13.9) is identical with (5.12.14) which is obtained from the exact expression.

In general, we see from (5.9.22) that for lossless media such that $q^{1/2}$ is real for $z < z_0$, |R| is unity. The incident energy is totally reflected. Equation (5.13.8) may be used to compute the reflection coefficient for any arbitrary stratified medium, the refractive index of which has an isolated turning point of first order provided the wavelength is short as compared to the typical dimension characterizing the inhomogeneity of the medium. For media having turning points of higher order, the formula (5.13.8) may no longer be applied. To study these more complicated cases, the starting point is the investigation of the Stokes phenomena about the higher order turning points. Some results are given by Heading (1962).

Finally, we note that the above discussion for horizontally polarized waves may be easily extended to the case for vertically polarized waves.
5. Wave Propagation in Inhomogeneous Media

Physically, as mentioned in the previous section, the WKB approximate expression for the reflection coefficient R[(5.13.8)] is the result of neglecting all partial reflections in the medium. For high frequency waves, this is a a good approximation in general. However, for the case where no turning point exists, there is no total reflection and according to WKB approximation, the reflection coefficient is zero. In this case, partial reflections become important since they are the only contributions to the reflection coefficient. The iterative procedure discussed in Section 5.10 may be used to find the contributions from the partial reflections. Instead of treating the problem this way, however, an alternative technique is introduced in the following and explicit expression for the reflection coefficient will be found.

Let us consider again

$$d^{2}E_{y}/dz^{2} + k_{0}^{2}q^{2}(z)E_{y} = 0$$
(5.13.1)

where $q^2 > 0$ always and approaches one for z < 0, and some positive constant q_+^2 as $z \to +\infty$.

In view of the partial reflections, the boundary conditions for E_y at $z \rightarrow -\infty$ is

$$E_y \sim e^{-jk_0 z} + R e^{jk_0 z} \tag{5.13.10}$$

and at $z \to +\infty$ is

$$E_y \sim T e^{-jk_0 q_+ z}$$
 (5.13.11)

Now, define a new variable

$$\xi(z) = k_0 \int_0^z q(\tau) \, d\tau \tag{5.13.12}$$

and construct a Green's function

$$G(\xi, \xi') = \begin{cases} (j/2)e^{-j\xi+j\xi'}, & \xi' < \xi \le +\infty\\ (j/2)e^{-j\xi'+j\xi}, & -\infty \le \xi \le \xi' \end{cases}$$
(5.13.13)

which satisfies the equation

$$d^{2}G(\xi, \xi')/d\xi^{2} + G(\xi, \xi') = \delta(\xi - \xi')$$
 (5.13.14)

In the variable z, (5.13.13) becomes

$$G(z, z') = \begin{cases} (j/2)e^{-jk_0 \int_{z'}^{z} q(\tau)d\tau}, & z' < z \le +\infty \\ (j/2)e^{jk_0 \int_{z'}^{z} q(\tau)d\tau}, & -\infty \le z \le z' \end{cases}$$
(5.13.15)

and (5.3.14) becomes

$$\frac{d^2G(z,z')}{dz^2} + k_0^2 q^2 G(z,z') = k_0 q(z) \ \delta(z-z') + \frac{1}{q} \ \frac{dq}{dz} \ \frac{dG(z,z')}{dz}$$
(5.13.16)

Multiplying (5.13.1) by G(z, z') and (5.3.16) by $E_y(z)$ and subtracting one from the other, we obtain

$$G(z, z') \frac{d^2 E_y}{dz^2} - E_y(z) \frac{d^2 G(z, z')}{dz^2} = -k_0 q(z) E_y(z) \,\delta(z - z') - \frac{1}{q} \frac{dq}{dz} E_y(z) \frac{dG(z, z')}{dz} \quad (5.13.17)$$

Integrating (5.13.17) with respect to z from $-\infty$ to $+\infty$, we have

$$\left[G(z, z')\frac{dE_y}{dz} - E_y(z)\frac{dG(z, z')}{dz}\right]_{-\infty}^{+\infty}$$

= $-k_0q(z')E_y(z') - \int_{-\infty}^{+\infty}\frac{1}{q(z)}\frac{dq}{dz}E_y(z)\frac{dG(z, z')}{dz}dz$ (5.13.18)

Using the boundary conditions (5.13.10) and (5.13.11), we obtain from (5.13.18)

$$E_{y}(z') = \frac{1}{q(z')} e^{-jk_{0}\int_{0}^{z'}q(\tau)d\tau} - \frac{1}{k_{0}q(z')} \int_{-\infty}^{+\infty} \frac{1}{q(z)} \frac{dq}{dz} \frac{dG(z,z')}{dz} E_{y}(z) dz$$
(5.13.19)

Substituting (5.13.15) into (5.13.19), we have

$$E_{y}(z') = \frac{1}{q(z')} e^{-jk_{0} \int_{0}^{z'} q(\tau)d\tau} + \frac{e^{-jk_{0} \int_{0}^{z'} q(\tau)d\tau}}{2q(z')} \int_{-\infty}^{z'} \frac{dq}{dz} e^{jk_{0} \int_{0}^{z} q(\tau)d\tau} E_{y}(z) dz - \frac{e^{jk_{0} \int_{0}^{z'} q(\tau)d\tau}}{2q(z')} \int_{z'}^{+\infty} \frac{dq}{dz} e^{-jk_{0} \int_{0}^{z} q(\tau)d\tau} E_{y}(z) dz$$
(5.13.20)

This is an integral equation for E_y which is equivalent to the coupled integration (5.10.22) we derived in Section 5.10. The first term on the righthand side is proportional to the WKB approximation for the incident wave. The additional two terms are due to multiple reflections in the medium. As $z' \rightarrow -\infty$, the second integral on the right-hand side vanishes and (5.13.20) has the form given by (5.13.10). Comparing the two equations, we conclude that

$$R = \frac{-1}{2} \int_{-\infty}^{+\infty} \frac{dq}{dz} e^{-jk_0 \int_{0}^{z} q(x)dx} E_y(z) dz \qquad (5.13.21)$$

This is an exact formal expression for the reflection coefficient. For the first-order approximation, we can substitute the WKB solution for E_y in (5.13.21) and obtain

$$R = -\frac{1}{2} \int_{-\infty}^{+\infty} \frac{1}{q(z)} \frac{dq}{dz} e^{-j2k_0 \int_0^z q(\tau)d\tau} dz \qquad (5.13.22)$$

This is the contribution from the waves that have undergone partial reflection once in the medium.

5.14 Reflection Coefficients for Stratified Media—Very Low Frequency Approximation

Up to this point, the discussion has been concentrated on the WKB solutions of the wave equation for which the properties of the medium vary slowly as compared to the wavelength. In some practical situations, however, the properties of the medium may change significantly in one wavelength. A typical example in radio wave propagation problems is the propagation of very low frequency (VLF) radio waves around the earth. The frequency is so low that the wavelength is comparable to the typical dimension of the ionosphere. For these cases the WKB solution is no longer a good approximation. In particular, the phase integral method of calculating the reflection coefficient is no longer applicable. In the following we shall introduce another approximation technique based on the very fact that the properties of the medium vary rapidly as compared to the wavelength (Brekhovskikh, 1960; Wait, 1962).

Let us consider a stratified medium characterized by $\varepsilon(z)$, and $\mu(z)$. As $z \to -\infty$, $\varepsilon(z) \to \varepsilon_0$ and $\mu(z) \to \mu_0$. As $z \to +\infty$, $\varepsilon(z) \to \varepsilon_1$ and $\mu(z) \to \mu_1$, a horizontally polarized wave is incident from $z = -\infty$ with an incident angle θ_i . The electric field can be written as

$$E_{y} = [A(z) + B(z)]e^{-jk_{0}n\sin\theta x}$$
(5.14.1)

where A(z) is the wave going upward and B(z) is the reflected wave. Since the medium varies rapidly, reflection takes place continuously in the medium. The Snell's law $n \sin \theta = \sin \theta_i$ determines the angle θ . $n^2 = \varepsilon \mu c^2$ is the

refractive index. The associated magnetic field components can be written in the following forms:

$$H_x = -(\varepsilon(z)/\mu(z))^{1/2} \cos\theta \left[A(z) - B(z)\right] e^{-jk_0 xn \sin\theta}$$

$$H_z = [k_0 \sin\theta_i/\omega\mu(z)][A(z) + B(z)] e^{-jk_0 xn \sin\theta}$$
(5.14.2)

Equations (5.14.1) and (5.14.2) should satisfy the Maxwell equations. Substituting them into (6.8.3a) and (5.8.3e), we obtain

$$\frac{dA}{dz} + \frac{dB}{dz} = -jk_0 n \cos\theta(A - B)$$

$$\frac{dA}{dz} - \frac{dB}{dz} = -\left(\frac{\mu}{\varepsilon}\right)^{1/2} \frac{1}{\cos\theta} \frac{d}{dz} \left(\left(\frac{\varepsilon}{\mu}\right)^{1/2} \cos\theta\right)(A - B)$$

$$-j\omega\left(\frac{\mu}{\varepsilon}\right)^{1/2} \frac{1}{\cos\theta} \left(\varepsilon - \frac{k_0^2 \sin^2\theta_i}{\omega^2\mu}\right)(A + B) \quad (5.14.4)$$

Defining two functions,

$$\beta(z) = \omega(\varepsilon \mu)^{1/2} \cos \theta \qquad (5.14.5)$$

and

$$\delta(z) = \frac{\mu}{2\beta} \frac{d}{dz} \left(\frac{\beta}{\mu}\right)$$
(5.14.6)

Equations (5.14.3) and (5.14.4) may be put into the form

$$dA/dz + j\beta A + \delta(A - B) = 0 \qquad (5.14.7)$$

$$dB/dz - j\beta B + \delta(B - A) = 0 \qquad (5.14.8)$$

If now we define the reflection coefficient as

$$R(z) = B(z)/A(z)$$
 (5.14.9)

(5.14.7) and (5.14.8) are readily combined to yield an equation for R:

$$dR/dz = 2j\beta R + \delta(1 - R^2)$$
 (5.14.10)

This is a general equation for the reflection coefficient. Up to this point, it is exact. For different physical situations, it may be used to derive the reflection coefficients under different approximations. For example, the WKB approximation (5.13.22) derived in the last section may also be obtained from (5.14.10) if high frequency approximation is made. In the following, we shall consider the case of very low frequency. To obtain a

solution, let us write

$$R = \frac{g(z)v(z) - g_1}{g(z)v(z) + g_1}$$
(5.14.11)

where

$$g(z) = (\varepsilon/\varepsilon_0 n^2) (n^2 - \sin^2 \theta_i)^{1/2}$$
 (5.14.12)

$$g_1 = \lim_{z \to \infty} g(z) = (\varepsilon_1 / \varepsilon_0 n_1^2) (n_1^2 - \sin^2 \theta_i)^{1/2}$$
 (5.14.13)

Since for $z \to \infty$, the reflection coefficient must vanish, it follows that $\lim_{z\to\infty} v(z) = 1$. A differential equation for v(z) can be obtained by substituting (5.14.11) into (5.14.10). We have, after some algebraic manipulation

$$-\frac{dv}{dz} = \frac{jk_0 n^2 g_1 \varepsilon_0}{\varepsilon} \left(1 - \frac{g^2}{g_1^2} v^2\right)$$
(5.14.14)

An iterative solution of (5.14.14) may be obtained in an ascending series in powers of k_0 . The *n*th term has the form

$$v_n = 1 + jk_0 g_1 \int_z^\infty \frac{n^2 \varepsilon_0}{\varepsilon_1} \left(1 - \frac{g^2}{g_1^2} v_{n-1}^2 \right) d\tau \qquad (5.14.15)$$

where the limit of integration corresponds to the boundary condition.

The zeroth-order solution is $v_0 = 1$. Substituting this into (5.14.11), we have

$$R = \frac{g - g_1}{g + g_1} = \frac{\cos \theta_i - (\varepsilon_1 / \varepsilon_0 n_1^2) (n_1^2 - \sin^2 \theta_i)^{1/2}}{\cos \theta_i + (\varepsilon_1 / \varepsilon_0 n_1^2) (n_1^2 - \sin^2 \theta_i)^{1/2}} \quad (5.14.16)$$

This is the Fresnel reflection coefficient for the horizontal polarization at a sharp boundary [see (5.7.10) for comparison]. Therefore we see that in this iterative procedure, the zeroth-order approximation is the reflection coefficient for a sharp boundary. This is to be expected in our formulation, since at the sharp boundary, the medium varies fastest; in fact, the properties have discontinuities there. The higher order terms in the ascending series in powers of k_0 account for the "gradualness" of the boundary. For a given medium, the properties of the medium vary more rapidly for lower frequency waves (waves with smaller values of k_0); thus the iterative procedure works better for these cases. This is exactly the opposite situation to that of the WKB method.

The same technique can be generalized to vertically polarized waves by applying the duality between E and H and ε and μ . The reflection coefficient written in the form of (5.14.11) has been applied to the theory of normal modes propagation of VLF waves in the earth-ionosphere waveguide (Wait, 1962).

5.15 Signal Propagation and Reflection in Stratified Media

In the previous sections, we have only considered the propagation of monochromatic plane waves in a stratified medium. In practice, however, time-dependent signals in the form of pulses are often found in many experimental situations. In this section we shall discuss this problem by following essentially the same technique used in Sections 4.2 and 4.3 in the discussion of transient waves in homogeneous plasma. We shall only consider the high frequency signals so that WKB solutions are applicable.

Let us consider a stratified medium occupying the half-space z > 0. Again, let us assume that a horizontally polarized signal is incident vertically on the medium. The Fourier transform of the incident signal at z = 0 is given by $E_y(0, \omega)$. The WKB approximation for the signal propagating upward into the stratified medium is given by [see (4.3.1) and (5.10.15); Budden (1961); Ginzberg (1964)]

$$E(z, t) = \frac{1}{2\pi} \int_{C} E(0, \omega) [n(z, \omega)]^{-1/2} \exp\left\{j\omega \left[t - \frac{1}{c} \int_{0}^{z} n(s, \omega) \, ds\right]\right\} d\omega$$
(5.15.1)

where the contour C is the Laplace contour discussed in Chapters 2 and 4. Equation (5.15.1) may be considered as the summation of all the WKB solutions; each corresponds to one frequency component of the original signal.

If $z_0(\omega)$ is the turning point (of order one) for $n(z, \omega)$ at frequency ω , then the WKB approximation for the reflection coefficient is given by (5.13.8)

$$R(\omega) = j \exp\left[-j2 \frac{\omega}{c} \int_{0}^{z_{0}(\omega)} n(s, \omega) \, ds\right]$$
(5.15.2)

Therefore, the WKB approximation for the reflected signal may be written as

$$E_{r}(z,t) = \frac{1}{2\pi} \int_{C} E(0,\omega)[n(z,\omega)]^{-1/2}R(\omega)$$

$$\times \exp\left\{j\omega\left[t + \frac{1}{c}\int_{0}^{z}n(s,\omega)\,ds\right]d\omega\right\}$$

$$= \frac{j}{2\pi} \int_{C} E(0,\omega)[n(z,\omega)]^{-1/2} \exp\left\{j\omega\left[t + \frac{1}{c}\int_{0}^{z}n(s,\omega)\,ds\right] - \frac{2}{c}\int_{0}^{z_{0}(\omega)}n(s,\omega)\,ds\right]\right\}d\omega$$
(5.15.3)

We note that in general the turning point z_0 depends on the frequency. Therefore, each harmonic component of the signal is reflected at a different level.

To carry out the Laplace inversions in (5.15.1) and (5.15.3) for a given refractive index function $n(z, \omega)$, the technique described in Section 4.3 may, in principle, be applied. Transient behavior of the transmitted and reflected signals may then be obtained. In the following, instead of studying the detailed transient behavior of the signals for some particular cases, we shall discuss certain general features of the transmitted and reflected signals which are found useful in many experimental situations.

For a pulsed signal, $E(0, \omega)$ is peaked at the carrier frequency ω_0 . For large values of ω_0 , the major contributions to the integrals (5.15.1) and (5.15.3) come from the region where the phase in the integrand is stationary at $\omega = \omega_0$ such that the harmonic components of the signal near $\omega = \omega_0$ all have the same phase. For (5.15.1), this requires

$$\frac{\partial}{\partial \omega} \left[\omega t - \frac{1}{c} \int_0^z n(s, \omega) \omega \, ds \right] = 0 \quad \text{for} \quad \omega = \omega_0 \quad (5.15.4)$$

Hence the position of the pulse at time t is given by

$$t - \frac{1}{c} \int_0^z \left[\frac{\partial}{\partial \omega} (n\omega) \right]_{\omega = \omega_0} ds = 0$$

and the upward velocity is

$$v_{gz} = dz/dt = c/[(\partial/\partial\omega)(n\omega)]_{\omega=\omega_0}$$
(5.15.5)

This is the upward group velocity for the pulse. Comparing (5.15.5) with (2.12.4a) we see that the same expression for the group velocity applies for both homogeneous and stratified media. The quantity c/v_{gz} is defined as the "group refractive Index" (5.7.5)

$$n' = c/v_{gz} = (\partial/\partial\omega)(n\omega) = n + \omega \,\partial n/\partial\omega \qquad (5.15.6)$$

From (5.15.3), the reflected signal that reaches the level z = 0 is given by

$$E_{y}(0, t) = (j/2\pi) \int_{C} E(0, \omega) [n(0, \omega)]^{-1/2}$$
$$\times \exp\left\{j \left[\omega t - \frac{2}{c} \int_{0}^{z_{0}(\omega)} \omega n(s, \omega) ds\right]\right\} d\omega \quad (5.15.7)$$

The major contribution of this integral again comes from the neighbor-

hood where the phase in the integral is stationary at $\omega = \omega_0$. Therefore, we have

$$\frac{\partial}{\partial \omega} \left[\omega t - \frac{2}{c} \int_0^{z_0(\omega)} \omega n(s, \omega) \, ds \right]_{\omega_0} = 0$$

or

$$t = \frac{2}{c} \left[\frac{\partial}{\partial \omega} \int_{0}^{z_{0}(\omega)} \omega n(s, \omega) \, ds \right]_{\omega_{0}}$$
(5.15.8)

If at time t = 0 the incident pulse is at z = 0, (5.15.8) gives the time delay for the signal to travel up to the reflection level $z = z_0(\omega_0)$ and come back to the level z = 0. Carrying out the differentiation of (5.15.8) and applying the relation $n(z_0) = 0$, we obtain

$$t = \frac{2}{c} \int_{0}^{z_{0}(\omega)} n'(z, \omega_{0}) dz \qquad (5.15.9)$$

Half of this quantity is the time it takes the pulse to travel up to $z = z_0(\omega_0)$.

During this length of time, if the signal is propagating in the free space, it will travel a distance ct/2. Thus we have

$$h'(\omega) = ct/2 = \int_0^{z_0(\omega)} n'(z, \omega_0) \, dz \qquad (5.15.10)$$

 $h'(\omega)$ defined by (5.15.10) is called the equivalent distance of reflection, or virtual height in ionospheric terminology. The true distance of reflection is $z_0(\omega_0)$ and is called the true height.

Given a refractive index $n(z, \omega)$, $z_0(\omega)$ may be found and (5.15.10) yields the virtual height by integration. If, instead, the virtual height $h'(\omega)$ is given as a function of frequency, is it possible from (5.15.10) to determine the refractive index uniquely? In general this problem belongs to a class of very difficult "inverse scattering" problems. Analytical solutions are possible only for very few special cases. In this following section, one such case shall be considered.

5.16 The True Height Problem—Ionosonde

Let us consider an isotropic, stratified plasma with electron density varying as a function of z. The refractive index is $n = (1 - \omega_p^2/\omega^2)^{1/2}$ where ω_p is the angular electron plasma frequency. The group refractive index

can be obtained from (5.15.6) and is found to be

$$n' = 1/(1 - \omega_p^2 / \omega^2)^{1/2}$$
(5.16.1)

For this medium, the virtual height given by the formula (5.15.10) becomes

$$h'(\omega) = \int_{0}^{z_{0}(\omega)} \frac{dz}{(1 - \omega_{p}^{2}/\omega^{2})^{1/2}} = \int_{0}^{z_{0}(\omega)} \frac{\omega \, dz}{\omega^{2} - \omega_{p}^{2}} \qquad (5.16.2)$$

We shall assume that $\omega_p(z)$ is a monotonic function of z. Therefore there is a unique inverse $z(\omega_p)$. For a given signal frequency ω_0 , the reflection level z_0 is obtained by setting $\omega_p = \omega_0$ in the expression $z(\omega_p)$ as shown in Fig. 5.16-1.



Fig. 5.16-1. Plasma frequency as a monotonic function of z.

In (5.16.2), we change variables by letting

$$u(z) = [\omega_p(z)]^2, \quad v = \omega^2$$
 (5.16.3)

and we have

$$v^{-1/2}h'(v^{1/2}) = \int_0^{z_0} \frac{dz}{[v - u(z)]^{1/2}}$$
(5.16.4)

This is known as the Abel's integral equation. It can be solved by the following procedure. Multiplying both sides by $(1/\pi)(w - v)^{-1/2}$ where w is a positive constant, and integrating in v from 0 to w, we obtain

$$(1/\pi) \int_{0}^{w} v^{-1/2} (w - v)^{-1/2} h'(v^{1/2}) dv$$

= $(1/\pi) \int_{0}^{w} dv \int_{0}^{z_{0}} dz / (w - v)^{1/2} [v - u(z)]^{1/2}$ (5.16.5)

5.16 The True Height Problem—Ionosonde

Changing the order of integration in (5.16.5), we obtain

$$(1/\pi) \int_{0}^{w} d\nu \int_{0}^{z_{0}} dz/(\nu - u)^{1/2} (w - \nu)^{1/2}$$
$$= (1/\pi) \int_{0}^{z_{0}(\omega)} dz \int_{u}^{w} d\nu/(w - \nu)^{1/2} (\nu - u)^{1/2}$$
(5.16.6)

where $z_0(w)$ is the value of z at which u(z) = w. We note that on the curve in Fig. 5.16-1, v = u.

The ν -integration in (5.16.6) may be carried out by first changing the variable:

$$v = w\cos^2\theta + u\sin^2\theta \tag{5.16.7}$$

Therefore

$$\int_{u}^{w} dv/(w-v)^{1/2}(v-u)^{1/2} = 2 \int_{0}^{\pi/2} d\theta = \pi \qquad (5.16.8)$$

and (5.16.6) becomes $z_0(w)$. Substituting this into (5.16.5), we have

$$z_0(w) = (1/\pi) \int_0^w v^{-1/2} (w - v)^{-1/2} h'(v^{1/2}) \, dv \qquad (5.16.9)$$

Changing the variable $v = w \sin^2 \alpha$, (5.16.9) becomes

$$z_0(w) = (2/\pi) \int_0^{\pi/2} h'(w^{1/2} \sin \alpha) \, d\alpha \qquad (5.16.10)$$

where $h'(w^{-1/2} \sin \alpha)$ means that the argument w in the known function h'(w) is replaced by $w^{1/2} \sin \alpha$.

This is the solution for the Abel's equation. Given a value w, the height at which the plasma frequency is $\omega_p = w^{1/2}$ is given by (5.16.10). This in turn yields the electron density as a function of height.

If the electron density is not a monotonic function of z, then the analytical method just described cannot be applied. In general, for more complicated media, (5.16.2) can only be solved numerically.

We have just described a technique to use electromagnetic waves as a probing tool to measure the electron density in a plasma. A device known as "ionosonde" has been used since the 1920's (Breit and Tuve, 1926) to obtain worldwide electron density in the ionosphere. Simply described, an ionosonde is a device that consists of an automatically sweeping pulse transmitter and receiver. Pulses in the frequency band from about 1 to 20 MHz are sent to the ionosphere (vertically or obliquely) and the time delay of the reflected signal is recorded as a function of the pulse frequency. The graphical display of such a record with frequency as one axis and the equivalent height as the other axis is called the ionogram. This essentially



is the function $h'(\omega)$ in (5.16.2). Electron density as a function of height can then be determined, usually by numerical methods. A typical ionogram is shown in Fig. 5.16-2. The two traces correspond to ordinary mode and extraordinary mode, respectively.

5.17 Wave Propagation in Stratified Magnetoplasma—Försterling's Coupled Equations

In the remaining part of this chapter, we shall extend our discussion on wave propagation in stratified media to anisotropic cases. An immediate example of this type is the earth ionosphere where the medium is a stratified magnetoplasma. The general problem of propagation in such a medium is a fairly complicated one and will be discussed in a later section. In this section we shall first consider a special case, namely, the vertical incidence case.

Let us consider a stratified plasma in a constant static magnetic field. The plasma density varies as a function of z. For vertical incidence, the wave propagate in the z-direction. Without loss of generality, the static magnetic field is assumed to be in the yz-plane. For high frequency waves, at any level z the Appleton-Hartree formula (4.14.9) yields two values, n_0^2 and n_x^2 , corresponding to the two characteristic modes, ordinary and extraordinary, respectively. For a homogeneous medium, the polarization of the wave is defined by

$$R = E_x/E_y = D_x/D_y = -H_y/H_x$$
(4.14.5)

and (4.14.14) gives the two characteristic polarizations R_0 and R_x , respectively:

$$R = \frac{j}{\cos\theta} \left[\frac{Y\sin^2\theta}{2(1-X)} \mp \left(\frac{Y^2\sin^4\theta}{4(1-X)^2} + \cos^2\theta \right)^{1/2} \right] \quad (4.14.14)$$

where θ is the angle the static magnetic field makes with the z-axis.

Equations (4.14.5) and (4.14.14) are used to define the characteristic polarizations for a stratified medium at any level z. From (4.14.12), we know that $R_0R_x = 1$. Therefore, from the definitions, we can write the components of the electric field for the characteristic modes, in the stratified magnetoplasma in the following way:

$$E_x^{(0)} = R_0 E_y^{(0)} = E_y^{(0)} / R_x, \qquad E_x^{(x)} = R_x E_y^{(x)} = E_y^{(x)} / R_0 \qquad (5.17.1)$$

where the superscripts "0" and "x" indicate the two modes, respectively.

Any wave may be expressed as the linear combination of the two characteristic modes. We have

$$E_{x} = E_{x}^{(0)} + E_{x}^{(x)} = E_{x}^{(0)} + R_{x}E_{y}^{(x)}$$

$$E_{y} = E_{y}^{(0)} + E_{y}^{(x)} = R_{x}E_{x}^{(0)} + E_{y}^{(x)}$$
(5.17.2)

For the two characteristic modes at any level z, the displacement vector is related to the electric field by (Chapter 2)

$$D_x^{(0)} = \varepsilon_0 n_0^2 E_x^{(0)}, \qquad D_y^{(0)} = \varepsilon_0 n_0^2 E_y^{(0)}$$

and

$$D_x^{(x)} = \varepsilon_0 n_x^2 E_x^{(x)}, \qquad D_y^{(x)} = \varepsilon_0 n_x^2 E_y^{(x)}$$
(5.17.3)

Therefore, the total displacement vector is written as

$$D_x = \varepsilon_0 (n_0^2 E_x^{(0)} + n_x^2 E_x^{(x)}) = \varepsilon_0 (n_0^2 E_x^{(0)} + n_x^2 R_x E_y^{(x)})$$
(5.17.4)

$$D_{y} = \varepsilon_{0}(n_{0}^{2}E_{y}^{(0)} + n_{x}^{2}E_{y}^{(x)})$$

= $\varepsilon_{0}(n_{0}^{2}R_{x}E_{x}^{(0)} + n_{x}^{2}E_{y}^{(x)})$ (5.17.5)

Starting from Maxwell equations

$$\nabla \times \mathbf{E} = -j\omega\mu_0 \mathbf{H}$$

$$\nabla \times \mathbf{H} = j\omega \mathbf{D}$$
(5.17.6)

and noting that for vertical incidence $\partial/\partial x = \partial/\partial y = 0$, (5.17.6) yields

immediately $H_z = D_z = 0$. Eliminating H_x and H_y from the remaining four component equations, we obtain

$$\frac{\frac{d^2 E_x}{dz^2} + \frac{k_0^2}{\varepsilon_0} D_x = 0}{\frac{d^2 E_y}{dz} + \frac{k_0^2}{\varepsilon_0} D_y = 0}$$
(5.17.7)

The quantities defined in (5.17.3) and (5.17.5) should satisfy the Maxwell equations and hence (5.17.7). Therefore, substituting (5.17.3)–(5.17.5) into (5.17.7), we obtain

$$\frac{d^{2}E_{x}^{(0)}}{dz^{2}} + R_{x}\frac{d^{2}E_{y}^{(x)}}{dz^{2}} + 2\frac{dR_{x}}{dz}\frac{dE_{y}^{(x)}}{dz} + \frac{d^{2}R_{x}}{dz^{2}}E_{y}^{(x)} + k_{0}^{2}n_{0}^{2}E_{x} + k_{0}^{2}R_{x}n_{x}^{2}E_{y}^{(x)} = 0 \quad (5.17.8a)$$

$$R_{x}\frac{d^{2}E_{x}^{(0)}}{dz^{2}} + \frac{d^{2}E_{y}^{(x)}}{dz^{2}} + 2\frac{dR_{x}}{dz}\frac{dE_{x}^{(0)}}{dz} + \frac{d^{2}R_{x}}{dz^{2}}E_{x}^{(0)} + k_{0}^{2}n_{0}^{2}R_{x}E_{x}^{(0)} + k_{0}^{2}n_{x}^{2}E_{y}^{(x)} = 0 \quad (5.17.8b)$$

Thus by defining the characteristic modes in the stratified magnetoplasma as in (5.17.1), we obtain from the Maxwell equations the two coupled equations. They may be simplified if we define new dependent variables in the following manner:

$$F_{+} = (R_{x}^{2} - 1)^{1/2} E_{x}^{(0)}, \qquad F_{-} = (R_{x}^{2} - 1)^{1/2} E_{y}^{(x)} \qquad (5.17.9)$$

where F_+ and F_- may be considered as the two characteristic modes. Further, let

$$\psi = \frac{1}{k_0} \left(R_x^2 - 1 \right)^{-1/2} \frac{dR_x}{dz} = \frac{1}{2k_0} \frac{d}{dz} \ln \left(\frac{R_x - 1}{R_x + 1} \right) \quad (5.17.10)$$

Then (5.17.8) becomes

$$\frac{d^2F_+}{dz^2} + k_0^2(n_0^2 + \psi^2)F_+ = k_0\frac{d\psi}{dz}F_- + 2k_0\psi\frac{dF_-}{dz}$$
 (5.17.11)

$$\frac{d^2F_-}{dz^2} + k_0^2(n_x^2 + \psi^2)F_- = k_0\frac{d\psi}{dz}F_+ + 2k_0\psi\frac{dF_+}{dz} \quad (5.17.12)$$

These are the Försterling's (1942) coupled equations for vertical incidence. The terms on the right-hand side are coupling terms. The variable ψ is

called the coupling parameter. We note that in general $\psi \neq 0$ and consequently the two characteristic modes are coupled. Equations (5.17.11) and (5.17.12) may be used as the basis for an iterative procedure to obtain approximate solutions in cases where coupling is not too strong.

A special case for the Försterling's coupled equations is for $\theta = 0$ in (4.14.14). For this case, the static magnetic field is also vertical. Equation (4.14.14) yields

$$R_0 = -j, \quad R_x = j$$
 (5.17.13)

which are constants throughout the stratified region. From (5.17.10), the coupling parameter ψ for this case vanishes. Therefore (5.17.11) and (5.17.12) reduce to

$$d^2F_+/dz^2 + k_0^2 n_0^2 F_+ = 0 (5.17.14)$$

$$d^2F_{-}/dz^2 + k_0^2 n_x^2 F_{-} = 0 (5.17.15)$$

where, from (5.17.9) and (5.17.2),

$$F_{+} = j\sqrt{2}E_{x}^{(0)} = (j/\sqrt{2})(E_{x} + jE_{y})$$
(5.17.16)

$$F_{-} = j\sqrt{2}E_{y}^{(x)} = (1/\sqrt{2})(E_{x} - jE_{y})$$
(5.17.17)

Also, from (4.14.19), we have

$$n_0^2 = 1 - \frac{X}{1+Y}, \quad n_x^2 = 1 - \frac{X}{1-Y}$$
 (5.17.18)

Thus, we see that for this special case, the two characteristic modes are circularly polarized and propagate independently of each other. In the next section, an example will be considered for this special case.

Finally, we note that the discussion in this section has been for lossless magnetoplasma. Collisional loss in the plasma may be taken into account in exactly the same manner as was done in Chapter 4. Instead of (4.14.14) and (4.14.19) in our discussion, (4.14.21) and (4.14.23) should be used, respectively. The rest of our discussion then follows.

5.18 An Application of Försterling's Coupled Equations

As an example, let us consider a plasma with an exponential electron density profile. The magnetic field is vertical; the collision frequency is assumed to be constant. Therefore

$$X = \exp(az)$$

$$Y = \text{constant}$$

$$U = 1 - j\nu/\omega = \text{constant}$$

(5.18.1)

A wave is incident vertically upon this plasma. For this case, (5.7.14) and (5.7.15) may be used.

$$\frac{d^2F_+}{dz^2} + k_0^2 \left[1 - \frac{e^{az}}{U+Y} \right] F_+ = 0$$
 (5.18.2)

$$\frac{d^2 F_-}{dz^2} + k_0^2 \left[1 - \frac{e^{az}}{U - Y} \right] F_- = 0$$
 (5.18.3)

For later convenience, we redefine $F_+ = E_x + jE_y$, $F_- = E_x - jE_y$. This does not affect (5.18.2) and (5.18.3) since they are homogeneous equations.

Define a new variable

$$\zeta = az/2 + c \tag{5.18.4}$$

where c is a constant to be determined.

In the new variable, (5.18.2) and (5.18.3) become

$$\frac{d^2F_+}{d\zeta^2} + \left[-\frac{a^2k_0^2}{4} \frac{e^{-2c}}{U+Y} e^{2\zeta} + \frac{a^2}{4} k_0^2 \right] F_+ = 0 \qquad (5.18.5)$$

$$\frac{d^2 F_-}{d\zeta^2} + \left[-\frac{a^2 k_0^2}{4} \frac{e^{-2c}}{U-Y} e^{2\zeta} + \frac{a^2}{4} k_0^2 \right] F_- = 0 \qquad (5.18.6)$$

Equations (5.18.5) and (5.18.6) may be put into a standard form of Bessel's equation (Watson, 1944),

$$\frac{d^2y}{d\zeta^2} + (e^{2\zeta} - p^2)y = 0$$
 (5.18.7)

by letting

$$-\frac{a^2k_0^2}{4} \frac{e^{-2c}}{U \pm Y} = 1$$
 (5.18.8)

and

$$(a^2/4)k_0^2 = -p^2 \tag{5.18.9}$$

where the upper sign is for F_+ and lower sign for F_- . From (5.18.8) and (5.18.9) we have

$$c = \ln \frac{2k_0}{a} \left(\frac{-1}{U \pm Y}\right)^{1/2}$$
(5.18.10)
$$p = 2jk_0/a$$

The general solutions of (5.18.5) and (5.18.6) are (Watson, 1944)

$$F_{+} = Z_{p} \left[\frac{2k_{0}}{a} \left(\frac{-1}{U+Y} \right)^{1/2} e^{az/2} \right]$$
(5.18.11)

$$F_{-} = Z_{p} \left[\frac{2k_{0}}{a} \left(\frac{-1}{U - Y} \right)^{1/2} e^{az/2} \right]$$
(5.18.12)

where Z_p is the Bessel's function of any kind of order p.

To choose the right kind of Bessel's function for our solution, let us consider the boundary conditions. The wave should be a decaying solution for $z \rightarrow \infty$. From the theory of Bessel's function, we know

$$\lim_{\substack{r \to \infty \\ r \to \infty}} H_p^{(1)}(re^{j\theta}) \to 0 \qquad \text{for} \quad 0 \le \theta \le \pi \qquad (5.18.13)$$

where $H_p^{(1)}$ and $H_p^{(2)}$ are Hankel functions of the first and second kind, respectively. The phase angles of the arguments of Z_p in (5.18.11) and (5.18.12) are given by

$$\theta = \arg\left(\frac{-1}{U \pm Y}\right)^{1/2} = \arg\left(\frac{-1}{1 \pm Y - jZ}\right)^{1/2}$$
(5.18.14)

where $Z = \nu/\omega$.

For low frequency propagation such that $Y \gg 1$, (5.8.14) may be approximated by

$$\theta \simeq \frac{1}{2} [\pi \pm \tan^{-1} Z/Y]$$
 (5.18.15)

By convention, we can choose

$$0 < \tan^{-1} Z/Y < \pi$$

for either Y positive or Y negative. Therefore

$$0 \le \theta \le \pi \tag{5.18.16}$$

Hence for $z \rightarrow \infty$, the decaying solution is given by the Hankel function of the first kind. Therefore

$$F_{\pm} = H_{2jk/a}^{(1)} \left(\frac{2jk_0}{a} \sqrt{l_{\pm}} e^{az/2} \right)$$
(5.18.17)

where $l_{\pm} = 1/(U \pm Y)$.

We note that at the level $z = z_0$ such that

$$e^{az_0} = U + Y$$

where z_0 is complex, the incident F_+ wave is reflected. To find the reflection coefficient, we compute the field at $z \rightarrow -\infty$. For this case, the argument of the Hankel function approaches zero. We can use the approximate expression for small arguments for Hankel functions. By definition (Watson, 1944)

$$H_p^{(1)}(x) = j \csc(p\pi) [e^{-jp\pi} J_p(x) - J_{-p}(x)]$$
 (5.18.18)

For $x \to 0$,

$$J_p(x) \cong (x/2)^p / \Gamma(p+1)$$
 (5.18.19)

where $\Gamma(z)$ is the gamma function.

Substituting (5.18.19) into (5.18.20) and applying the result to (5.18.17), we obtain after some manipulation

$$F_{\pm} \cong \frac{-j}{\sin\left(\frac{2jk_{0}}{a}\pi\right)} \Gamma\left(1 - \frac{2jk_{0}}{a}\right) \left(\frac{k_{0}j}{a}\sqrt{l_{\pm}}\right)^{(2jk_{0}/a)} \times \left[e^{-jk_{0}z} - e^{(2\pi k_{0}/a)}\right] \left[\frac{k_{0}j}{a}\sqrt{l_{\pm}}\right]^{(4jk_{0}/a)} \frac{\Gamma\left(1 - \frac{2jk_{0}}{a}\right)}{\Gamma\left(1 + \frac{2jk_{0}}{a}\right)} e^{jk_{0}z}$$
(5.18.20)

The first term in the bracket represents the incident wave, the second term represents the reflected wave. We can define the reflection coefficient as

$$R_{\pm} = -e^{(2\pi k_0/a)} \left[\frac{k_0 j}{a} \sqrt{l_{\pm}} \right]^{(4jk_0/a)} \frac{\Gamma\left(1 - \frac{2jk_0}{a}\right)}{\Gamma\left(1 + \frac{2jk_0}{a}\right)}$$
$$= -(l_{\pm})^{(j2k_0/a)} (k_0/a)^{(j4k_0/a)} \frac{\Gamma\left(1 - \frac{2jk_0}{a}\right)}{\Gamma\left(1 + \frac{2jk_0}{a}\right)}$$
(5.18.21)

The magnitude of reflection coefficient is

$$|R_{\pm}| = |l_{\pm}^{(j2k_0/a)}|$$
(5.18.22)

If the medium is lossless, i.e., v = 0, $l_+ = 1/(1 + Y)$, then

$$|R_{+}| = 1 \tag{5.18.23}$$

 F_+ is totally reflected. For F_- , we have $l_- = 1/(1-Y) \cong (1/Y)e^{j\pi}$ for

 $Y \gg 1$; then

$$|R_{-}| = |[(1/Y)e^{j\pi}]^{(j2k_{0}/a)}| = e^{-2\pi k_{0}/a}$$
(5.18.24)

for a slowly varying medium $(2\pi k_0/a) \gg 1$. Hence $|R_-|$ is very small. Most parts of F_- propagate through the plasma. This corresponds to the whistler mode discussed in Section 4.11. This can also be seen from (5.18.3). The quantity $1 - e^{az}/(U - Y)$ does not vanish for any values of z for $Y \gg 1$ —hence no total reflection for F_- .

As $z \rightarrow -\infty$, from (5.18.20), we have

$$E_x = \frac{1}{2}(F_+ + F_-) = A[e^{-jk_0 z} + \frac{1}{2}(R_+ + R_-)e^{ik_0 z}]$$

= $A[e^{-jk_0 z} + {}_{\parallel}R_{\parallel}e^{ik_0 z}]$ (5.18.25)

$$E_{y} = \frac{1}{2j} (F_{+} - F_{-}) = \frac{A}{2j} (R_{+} - R_{-})e^{jk_{0}z}$$
$$= A_{\parallel}R_{\perp}e^{jk_{0}z}$$
(5.18.26)

where

 $_{\parallel}R_{\parallel} = (R_{+} + R_{-})/2, \quad _{\parallel}R_{\perp} = (R_{+} - R_{-})/2j$ (5.18.27)

The presubscript in the reflection coefficient denotes whether the incident electric vector is parallel or perpendicular to the x-axis, and the post-subscript refers in the same way to the reflected electric field (see Section 5.7).

It can be shown from (5.18.21) and (5.18.27) that for $Y \ge 1$

$$|_{\parallel}R_{\parallel}| = \frac{1}{2} |R_{+} + R_{-}| \cong e^{-k_{0}\pi/a} \cosh\left[\frac{k_{0}}{a}\left(\pi - 2\tan^{-1}\frac{Z}{Y}\right)\right] \quad (5.18.28)$$

and

$$|_{\parallel}R_{\perp}| = \frac{1}{2} |R_{+} - R_{-}| \cong e^{-k_{0}\pi/a} \sinh\left[\frac{k_{0}}{a}\left(\pi - 2\tan^{-1}\frac{Z}{Y}\right)\right] \quad (5.18.29)$$

Equation (5.18.28) gives the magnitude of the reflected wave having the same polarization of the incident wave while (5.12.52) gives the magnitude of the reflected wave having a polarization perpendicular to that of the incident wave. Equations (5.18.28) and (5.18.29) have been used to interpret experimental data of absorption measurements at long and very long wavelengths reflected by the ionosphere. Information about the electron density profile and collision frequency in the *D* region of the ionosphere can be obtained (Stanley, 1950). 292 5. Wave Propagation in Inhomogeneous Media

5.19 Wave Propagation in Stratified Anisotropic Media—General Coupled Equations

In the last two sections, wave propagation in stratified magnetoplasma under certain special conditions was studied. In this section, we shall treat the problem of propagation in general stratified anisotropic media. The technique we shall introduce is the so-called coupled equation method. This method may be applied to many other practical problems in addition to propagation in the ionosphere (see Chapter 8).

Let us assume that for electromagnetic waves the medium is characterized by a dielectric tensor

$$\boldsymbol{\varepsilon}(z) = \varepsilon_0 \, \mathbf{K}(z) \tag{5.19.1}$$

The medium is nonmagnetic so that $\mu = \mu_0$. Maxwell equations become (for $e^{j\omega t}$ dependence)

$$\nabla \times \mathbf{E} = -j\omega\mu_0 \mathbf{H}$$

$$\nabla \times \mathbf{H} = j\omega\boldsymbol{\epsilon} \cdot \mathbf{E}$$
(5.19.2)

Since the coefficients are functions for z only, we can assume plane wave solution of the form

$$e^{-j(k_x x + k_y y)} e^{j\omega t}$$
 (5.19.3)

for the unknown components of the fields. Substituting (5.19.3) into (5.19.2) and putting it into component form, we have

$$-jk_yE_z - dE_y/dz = -j\omega\mu_0H_x \tag{5.19.4a}$$

$$dE_x/dz + jk_xE_z = -j\omega\mu_0H_y \tag{5.19.4b}$$

$$-jk_x E_y + jk_y E_x = -j\omega\mu_0 H_z \tag{5.19.4c}$$

$$-jk_yH_z - dH_y/dz = j\omega\varepsilon_0(K_{11}E_x + K_{12}E_y + K_{13}E_z) \quad (5.19.4d)$$

$$dH_x/dz + jk_xH_z = j\omega\varepsilon_0(K_{21}E_x + K_{22}E_y + K_{23}E_z) \qquad (5.19.4e)$$

$$-jk_xH_y + jk_yH_x = j\omega\varepsilon_0(K_{31}E_x + K_{32}E_y + K_{33}E_z) \qquad (5.19.4f)$$

where the K_{ij} 's are the elements of the relative dielectric tensor.

It is now possible to eliminate E_z and H_z from (5.19.4) and obtain a set of four first-order differential equations. In matrix form, this set may be written as

$$d\mathbf{e}/dz = -jk_0\mathbf{T}\cdot\mathbf{e} \tag{5.19.5}$$

where e is a column vector defined by

$$\mathbf{e} = \begin{bmatrix} E_x \\ E_y \\ \eta_0 H_x \\ \eta_{\varsigma} H_y \end{bmatrix}$$
(5.19.6)

and **T** is a 4×4 matrix defined by Eq. (5.19.7). $\eta_0 = (\mu_0/\epsilon_0)^{1/2}$ is the free space intrinsic impedance.

To solve (5.19.5), we introduce a transformation

$$\mathbf{e} = \mathbf{R} \cdot \mathbf{f} \tag{5.19.8}$$

Substituting (5.19.8) into (5.19.5) and assuming that **R** is nonsingular, we obtain

$$d\mathbf{f}/dz + jk_0(\mathbf{R}^{-1} \cdot \mathbf{T} \cdot \mathbf{R}) \cdot \mathbf{f} = -\mathbf{R}^{-1} \cdot d\mathbf{R}/dz \cdot \mathbf{f} \quad (5.19.9)$$

where \mathbf{R}^{-1} is the inverse of \mathbf{R} .

The idea here is to choose the transformation matrix **R** such that the matrix $\mathbf{R}^{-1} \cdot \mathbf{T} \cdot \mathbf{R}$ becomes diagonal. From the matrix theory, this can be done for any particular value of z, if the roots q_i (i = 1, 2, 3, 4) of the quartic eigenvalue equation

$$\det[\mathbf{T} - q\mathbf{I}] = 0 \tag{5.19.10}$$

are distinct. In that case,

$$\mathbf{R}^{-1} \cdot \mathbf{T} \cdot \mathbf{R} = \begin{bmatrix} q_1 & 0 & 0 & 0 \\ 0 & q_2 & 0 & 0 \\ 0 & 0 & q_3 & 0 \\ 0 & 0 & 0 & q_4 \end{bmatrix} = \mathbf{\Delta}$$
(5.19.11)

In a homogeneous medium, the right-hand side of (5.19.9) vanishes; therefore (5.19.9) becomes uncoupled and yields four solutions of the form $e^{-jk_0q_iz}$ each propagating independently. For a stratified medium, the equations are in general coupled. If, under certain conditions, $|\mathbf{R}^{-1} \cdot d\mathbf{R}/dz|$ is small, then (5.19.9) serves as the basis for an interative procedure to obtain approximate solutions.

From the matrix theory, the condition (5.19.11) does not determine the matrix **R** uniquely. **R** is constructed by taking its *i*th column as the *i*th eigenvector \mathbf{v}_i of the eigenvalue problem

$$(\mathbf{T} - q_i \mathbf{I}) \cdot \mathbf{v}_i = 0 \tag{5.19.12}$$

$$\mathbf{T} = \begin{bmatrix} -\frac{k_{x}K_{31}}{k_{0}K_{33}} & -\frac{k_{x}K_{32}}{k_{0}K_{33}} & \frac{k_{x}k_{y}}{k_{0}^{2}K_{33}} & 1 - \frac{k_{x}^{2}}{k_{0}^{2}K_{33}} \\ -\frac{k_{y}K_{31}}{k_{0}K_{33}} & -\frac{k_{y}K_{32}}{k_{0}K_{33}} & -1 + \frac{k_{y}^{2}}{k_{0}^{2}K_{33}} & -\frac{k_{x}k_{y}}{k_{0}^{2}K_{33}} \\ -K_{21} + \frac{K_{23}K_{31}}{K_{33}} - \frac{k_{x}k_{y}}{k_{0}^{2}} & -K_{22} + \frac{K_{23}K_{32}}{K_{33}} + \frac{k_{x}^{2}}{k_{0}^{2}} & -\frac{k_{y}K_{23}}{k_{0}K_{33}} & \frac{k_{x}K_{23}}{k_{0}K_{33}} \\ K_{11} - \frac{K_{13}K_{31}}{K_{33}} - \frac{k_{y}^{2}}{k_{0}^{2}} & K_{12} - \frac{K_{13}K_{32}}{K_{33}} + \frac{k_{x}k_{y}}{k_{0}^{2}} & \frac{k_{y}K_{13}}{k_{0}K_{33}} & -\frac{k_{x}K_{13}}{k_{0}K_{33}} \end{bmatrix}$$
(5.19.7)

The normalization of the \mathbf{v}_i is still arbitrary. One way to define them uniquely is to require that the diagonal terms of the matrix \mathbf{R}^{-1} . $d\mathbf{R}/dz$ on the right-hand side vanish. This additional condition on \mathbf{R} will make the choice of \mathbf{R} unique. Equation (5.19.9) now becomes

$$d\mathbf{f}/dz + jk_0 \mathbf{\Delta} \cdot \mathbf{f} = -\mathbf{R}^{-1} \cdot d\mathbf{R}/dz \cdot \mathbf{f}$$
 (5.19.13)

where Δ is the diagonal matrix given by (5.19.11). The right-hand side involves the coupling terms. The matrix $-\mathbf{R}^{-1} \cdot d\mathbf{R}/dz$ is sometimes referred to as the coupling matrix. Its elements indicate the strength of coupling between any two modes. The vanishing of the diagonal terms indicates that there is no self-coupling term. Without going into details, we shall indicate a way to solve (5.19.13) in successive approximation. Following the same technique as used in Section 5.10, (5.19.13) can be transformed into an integral equation

$$\mathbf{f}(z) = \mathbf{f}_0(z) - \mathbf{F}^{-1}(z) \cdot \int^z \left[\mathbf{F}(\tau)\right] \cdot \left[\mathbf{R}^{-1} \cdot d\mathbf{R}/d\tau \cdot \mathbf{f}(\tau)\right] d\tau \quad (5.19.14)$$

where

$$f_{0i} = e^{-jk_0 \int^z q_i(\tau)d\tau}$$

$$F_{mn} = e^{+jk_0 \int^z q_m(\tau)d\tau} \quad \text{if} \quad m = n \quad (5.19.15)$$

$$= 0 \quad \text{if} \quad m \neq n$$

 f_0 may be considered as the independent characteristic mode in the medium. Equation (5.19.14) may be used as the basis for the iterative procedure to obtain higher order solutions corresponding to coupled wave equations provided \mathbf{R}^{-1} . $d\mathbf{R}/dz$ is small. In the regions where $\mathbf{R}^{-1} \cdot d\mathbf{R}/dz$ is not small, this procedure breaks down. Such regions exist in the neighborhood of those values of z (usually complex) for which the matrix \mathbf{R} is singular. This happens, in particular, when two of the eigenvalues q_i are equal. These points in the complex z-plane are called "reflection" or "coupling" points. The solution of the coupled equation in the neighborhood of these points requires more detailed analysis and will not be considered here (Clemmow and Heading, 1954; Budden and Clemmow, 1957).

If, under certain conditions in the quartic equation for q [(5.19.10)] the coefficients of q and q^3 terms vanish, then (5.19.10) becomes quadratic in q^2 . Therefore at any level there exist two characteristic waves associated with two values of q which differ only in sign, corresponding to a pair of up-going and down-going characteristic waves. For this case, the four first-order equations in (5.19.5) can be combined to yield two second-order coupled equations.

Let us suppose for this case that

$$q_2 = -q_1, \qquad q_4 = -q_3 \tag{5.19.16}$$

From (5.19.14) and (5.19.15), we see that f_1 and f_2 correspond to up-going and down-going waves, respectively and so do f_3 and f_4 . Let us now introduce new variables h_i which are linear combinations of f_1 , f_2 and f_3 , f_4 , respectively. They can be written as, in particular,

$$\mathbf{h} = \mathbf{S} \cdot \mathbf{f} \tag{5.19.17}$$

where

$$\mathbf{S} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$
(5.19.18)

with $S^{-1} = \frac{1}{2}S$.

Substituting $\mathbf{f} = \mathbf{S}^{-1} \cdot \mathbf{h}$ into (5.19.13), we obtain

$$d\mathbf{h}/dz + jk_0(\mathbf{S} \cdot \mathbf{\Delta} \cdot \mathbf{S}^{-1}) \cdot \mathbf{h} = -(\mathbf{S} \cdot \mathbf{R}^{-1} \cdot d\mathbf{R}/dz \cdot \mathbf{S}^{-1}) \cdot \mathbf{h} \quad (5.19.19)$$

where

$$\mathbf{S} \cdot \mathbf{\Delta} \cdot \mathbf{S}^{-1} = \begin{bmatrix} 0 & q_1 & 0 & 0 \\ q_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & q_3 \\ 0 & 0 & q_3 & 0 \end{bmatrix}$$
(5.19.20)

To save some writing, let us define

$$\mathbf{S} \cdot \mathbf{R}^{-1} \cdot d\mathbf{R}/dz \cdot \mathbf{S}^{-1} = \mathbf{W}$$
(5.19.21)

and partition the matrices in (5.19.19) in the following manner

$$\mathbf{U} = \begin{bmatrix} h_1 \\ h_3 \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} h_2 \\ h_3 \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} q_1 & 0 \\ 0 & q_3 \end{bmatrix}$$
$$\mathbf{A} = \begin{bmatrix} w_{11} & w_{13} \\ w_{31} & w_{33} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} w_{12} & w_{14} \\ w_{32} & w_{34} \end{bmatrix}, \quad (5.19.22)$$
$$\mathbf{C} = \begin{bmatrix} w_{21} & w_{23} \\ w_{41} & w_{43} \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} w_{22} & w_{24} \\ w_{42} & w_{44} \end{bmatrix}$$

where w_{ij} 's are the elements of the matrix **W**.

Equation (5.19.19) can then be separated into two pairs:

$$d\mathbf{u}/dz + jk_0\mathbf{Q}\cdot\mathbf{v} = -\mathbf{A}\cdot\mathbf{u} - \mathbf{B}\cdot\mathbf{v} \qquad (5.19.23)$$

$$d\mathbf{u}/dz + jk_0\mathbf{Q}\cdot\mathbf{u} = -\mathbf{C}\cdot\mathbf{u} - \mathbf{D}\cdot\mathbf{v} \qquad (5.19.24)$$

Solving (5.19.23) for v, we have

$$\mathbf{v} = -\mathbf{G}^{-1} \cdot (d\mathbf{u}/dz + \mathbf{A} \cdot \mathbf{u}) \tag{5.19.25}$$

where, for brevity, we have written

$$\mathbf{G} = jk_0\mathbf{Q} + \mathbf{B} \tag{5.19.26}$$

Substituting (5.19.25) into (5.19.24), we obtain finally the second-order coupled equations

$$d^{2}\mathbf{u}/dz^{2} + k_{0}^{2} \mathbf{Q} \cdot \mathbf{Q} \cdot \mathbf{u} = -\left[\mathbf{A} - \frac{d\mathbf{G}}{dz} \cdot \mathbf{G}^{-1} + \mathbf{G} \cdot \mathbf{D} \cdot \mathbf{G}^{-1}\right] \cdot d\mathbf{u}/dz$$
$$-\left[\frac{d\mathbf{A}}{dz} - \left(\frac{d\mathbf{G}}{dz} \cdot \mathbf{G}^{-1} - \mathbf{G} \cdot \mathbf{D} \cdot \mathbf{G}^{-1}\right) \cdot \mathbf{A}$$
$$-\mathbf{G} \cdot \mathbf{C} - jk_{0}\mathbf{B} \cdot \mathbf{Q}\right] \cdot \mathbf{u} \qquad (5.19.27)$$

Similar equations can be derived for v. For homogeneous medium, the right-hand side vanishes. We have

$$d^{2}h_{1}/dz^{2} + k_{0}^{2}q_{1}^{2}h_{1} = 0$$

$$d^{2}h_{3}/dz^{2} + k_{0}^{2}q_{3}^{2}h_{3} = 0$$
(5.19.28)

The equations become uncoupled.

The coupled equation method discussed in this section is quite general. It can be used to solve many physical problems whenever the field equations can be put into the form (5.19.5). It is especially useful when solutions are being computed numerically. It also forms the basis for the analytical discussion on techniques of obtaining approximate solutions. In the next section, the coupled equation technique will be applied to the problem of wave propagation in a magnetoplasma under general conditions.

5.20 Application of the Coupled Equations Method to Wave Propagation in a Stratified Magnetoplasma

In this section, we shall consider wave propagation in a cold magnetoplasma; the electron density is a function of z. The dielectric tensor for this medium can be obtained in the same manner as was done in Section 4.5 for homogeneous cold plasma by considering the motion of mutually noninteracting charged particles. The susceptibility tensor is given by (4.5.11) where now the X_{α} 's are functions of z.

$$\mathbf{X} = \sum_{\alpha} \mathbf{X}_{\alpha}$$

$$= -\sum_{\alpha} \frac{X_{\alpha}}{1 - Y_{\alpha}^{2}} \begin{bmatrix} 1 - Y_{\alpha x}^{2} & -Y_{\alpha x}Y_{\alpha y} + jY_{\alpha z} & -Y_{\alpha x}Y_{\alpha z} - jY_{\alpha y} \\ -Y_{\alpha x}Y_{\alpha y} - jY_{\alpha z} & 1 - Y_{\alpha y}^{2} & -Y_{\alpha y}Y_{\alpha z} + jY_{\alpha x} \\ -Y_{\alpha x}Y_{\alpha z} + jY_{\alpha y} & -Y_{\alpha y}Y_{\alpha z} - jY_{\alpha x} & 1 - Y_{\alpha z}^{2} \end{bmatrix}$$
(4.5.11)

 X_{α} and Y are defined in Section 4.5.

For high frequency waves, only the contribution from electrons is important. Therefore, the relative dielectric tensor may be written as

$$\mathbf{K} = \mathbf{I} + \mathbf{\chi} \tag{5.20.1}$$

where the subscript $\alpha = e$ on **x** is omitted.

Substituting (5.20.1) into (5.19.7) and for waves propagating in the xz-plane, the elements of the matrix **T** are obtained:

$$T_{11} = -(k_x/k_0)AX(Y_xY_z - jY_y)$$

$$T_{12} = -(k_x/k_0)AX(Y_yY_z + jY_x)$$

$$T_{14} = 1 - (k_x^2/k_0^2)A(1 - Y^2)$$

$$T_{23} = -1$$

$$T_{31} = -AX[Y_xY_y + jY_z(1 - X)]$$

$$T_{32} = -A[1 + X^2 - Y^2 - X(2 - Y_y^2 - Y_z^2)] + k_x^2/k_0^2$$

$$T_{34} = (k_x/k_0)AX(Y_yY_z - jY_x)$$

$$T_{41} = A[1 + X^2 - Y^2 - X(2 - Y_x^2 - Y_z^2)]$$

$$T_{42} = AX[Y_xY_y - jY_z(1 - X)]$$

$$T_{44} = -(k_xk_0)AX(Y_xY_z + jY_y)$$

$$T_{13} = T_{21} = T_{22} = T_{24} = T_{33} = T_{43} = 0$$
(5.20.2)

where A is defined as

$$A = [1 - Y^2 - X(1 - Y_z^2)]^{-1}$$

Substituting (5.20.2) into the quartic equation (5.19.10), we obtain

$$q^4 + aq^3 + bq^2 + cq + d = 0 (5.20.3)$$

where

$$a = -(T_{11} + T_{44})$$

$$b = T_{11}T_{44} - T_{14}T_{41} + T_{32}$$

$$c = T_{34}T_{42} + T_{31}T_{12} - T_{32}(T_{11} + T_{44})$$

$$d = T_{12}T_{34}T_{31} + T_{41}T_{42}T_{14} + T_{32}T_{11}T_{44}$$

$$- T_{14}T_{32}T_{41} - T_{11}T_{34}T_{32} - T_{31}T_{12}T_{44}$$
(5.20.4)

Equation (5.20.3) is known as the Booker quartic equation (Booker, 1938). In general (5.20.3) yields four distinct roots for q. At any level, (5.20.3) gives the four characteristic waves in the stratified magnetoplasma of which two are up-going waves and two are down-going waves. When these waves propagate in the stratified magnetoplasma, they are coupled through (5.19.14). The elements $\mathbf{R}^{-1} \cdot d\mathbf{R}/dz$ in (5.19.14) may be obtained by solving the eigenvalue problem (5.19.12). They are given by Budden and Clemmow (1957) and Budden (1961). Various numerical methods have been developed to solve the coupled integral equation (5.19.14) for wave propagation in the ionosphere (Budden, 1969).

As an illustrative example, let us consider the case of vertical incidence again, using the coupled equation method. For vertical incidence, $k_x = 0$, and the magnetic field may be assumed to be in the yz-plane. Therefore in (5.20.2) $T_{11} = T_{12} = T_{34} = T_{44} = 0$ and $T_{14} = 1$.

Equation (5.20.3) becomes

$$q^{4} + (T_{32} - T_{41})q^{2} + (T_{31}T_{42} - T_{32}T_{41}) = 0$$
 (5.20.5)

which yields two roots for q^2 . Using (5.20.2), it is not difficult to show that these two roots coincide with the two refractive indices n_0^2 and n_x^2 obtained from the Appleton-Hartree formula (4.14.19). For this case, (5.19.11) becomes

$$\boldsymbol{\Delta} = \begin{bmatrix} n_0 & 0 & 0 & 0 \\ 0 & -n_0 & 0 & 0 \\ 0 & 0 & n_x & 0 \\ 0 & 0 & 0 & -n_x \end{bmatrix}$$
(5.20.6)

The solution of the eigenvalue problem (5.19.12) yields the eigenvectors \mathbf{v}_i

$$\mathbf{v}_{1} = (A_{1}F_{1})^{-1/2}[-T_{42}, n_{0}^{2} - T_{41}, n_{0}(n_{0}^{2} - T_{41}), -T_{42}n_{0}]$$

$$\mathbf{v}_{2} = (A_{2}F_{2})^{-1/2}[-T_{42}, n_{0}^{2} - T_{41}, -n_{0}(n_{0}^{2} - T_{41}), T_{24}n_{0}]$$

$$\mathbf{v}_{3} = (A_{3}F_{3})^{-1/2}[-T_{42}, n_{x}^{2} - T_{41}, n_{x}(n_{x}^{2} - T_{41}), -T_{42}n_{x}]$$

$$\mathbf{v}_{4} = (A_{4}F_{4})^{-1/2}[-T_{42}, n_{x}^{2} - T_{41}, -n_{x}(n_{x}^{2} - T_{41}), T_{42}n_{x}]$$
(5.20.7)

where

$$A_{1} = A_{2} = n_{0}^{2} - T_{41}, \qquad A_{3} = A_{4} = n_{x}^{2} - T_{41}$$

$$F_{1} = -F_{2} = 2n_{0}(n_{0}^{2} - n_{x}^{2}), \qquad F_{3} = -F_{4} = -2n_{x}(n_{x}^{2} - n_{x}^{2})$$
(5.20.8)

The matrix \mathbf{R} is obtained by taking \mathbf{v}_i 's as its columns. Therefore,

$$\mathbf{R} = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4) \tag{5.20.9}$$

Substituting (5.20.9) into (5.19.8) and (5.19.13), we obtain

$$f_{1} = [2n_{0}(R_{x}^{2} - 1)]^{-1/2}[n_{0}E_{x} - n_{0}R_{x}E_{y} - R_{x}\eta_{0}H_{x} + \eta_{0}H_{y}]$$

$$f_{2} = -j[2n_{0}(R_{x}^{2} - 1)]^{-1/2}[n_{0}E_{x} - n_{0}R_{x}E_{y} + R_{x}\eta_{0}H_{x} - \eta_{0}H_{y}]$$

$$f_{3} = -j[2n_{x}(1 - R_{0}^{2})]^{-1/2}[-n_{x}E_{x} + n_{x}R_{0}E_{y} + R_{x}\eta_{0}H_{x} - \eta_{0}H_{y}]$$

$$f_{4} = [2n_{x}(1 - R_{0}^{2})]^{-1/2}[-n_{x}E_{x} + n_{x}R_{0}E_{y} - R_{x}\eta_{0}H_{x} + \eta_{0}H_{y}]$$
(5.20.10)

and

$$f_{1}' + jk_{0}n_{0}f_{1} = -(n_{0}'/2jn_{0})f_{2} + (j/2)k_{0}\psi(n_{0} + n_{x})(n_{0}n_{x})^{-1/2}f_{3} + \frac{1}{2}k_{0}\psi(n_{0} - n_{x})(n_{0}n_{x})^{-1/2}f_{4}$$

$$f_{2}' - jk_{0}n_{0}f_{2} = (n_{0}'/2jn_{0})f_{1} + \frac{1}{2}k_{0}\psi(n_{0} - n_{x})(n_{0}n_{x})^{-1/2}f_{3} - (j/2)k_{0}\psi(n_{0} + n_{x})(n_{0}n_{x})^{-1/2}f_{4}$$

$$(5.20.11)$$

$$f_{3}' + jk_{0}n_{x}f_{3} = (-j/2)k_{0}\psi(n_{0} + n_{x})(n_{0}n_{x})^{-1/2}f_{1} - \frac{1}{2}k_{0}\psi(n_{0} - n_{x})(n_{0}n_{x})^{-1/2}f_{2} + (n_{x}'/2jn_{x})f_{4}$$

$$f_4' - jk_0n_xf_4 = -\frac{1}{2}k_0\psi(n_0 - n_x)(n_0n_x)^{-1/2}f_1 + (j/2)k_0\psi(n_0 + n_x)(n_0n_x)^{-1/2}f_2 - (n_x'/2jn_x)f_3$$

where prime indicates d/dz.

These are the coupled first-order equations for the new variable f. The function ψ is the coupling parameter defined in (5.17.10). The normalization of the \mathbf{v}_i 's is such that there are no self-coupling terms on the right-hand side of (5.20.11).

When the coupling is weak, the right-hand side of (5.20.11) may be neglected and (5.20.12) becomes uncoupled. The solutions of this set of uncoupled equations gives the four independent characteristic waves and are sometimes referred to as the WKB solutions for the vertically incident waves (Budden, 1961).

We note from (5.20.5) that for the vertical incidence case, the eigenvalue equation becomes quadratic in q^2 . For this case, from the discussion in the last section, it is possible to derive a set of second-order coupled equations. In fact, in Section 5.17 we have already derived the Försterling coupled equations from the Maxwell equation directly. As an example, let us now apply the general technique discussed in the last section to derive the Försterling equations from the set of coupled first-order equations (5.20.11). To do this, we first define a set of new variables

$$g_1 = n_0^{-1/2} f_1, \qquad g_2 = j n_0^{-1/2} f_1 g_3 = -j n_x^{-1/2} f_3, \qquad g_4 = -n_x^{-1/2} f_4$$
(5.20.12)

Substituting (5.20.12) into (5.20.11), we obtain an equation for $g(g_1, g_2, g_3, g_4)$:

$$d\mathbf{g}/dz + jk_0 \mathbf{\Delta} \cdot \mathbf{g} = \mathbf{\Gamma} \cdot \mathbf{g} \tag{5.20.13}$$

where Δ is given by (5.20.6) and

$$\mathbf{\Gamma} = \begin{bmatrix} \frac{-n_{0}'}{2n_{0}} & \frac{n_{0}'}{2n_{0}} & -\left(\frac{k_{0}\psi}{2}\right)\left(1+\frac{n_{x}}{n_{0}}\right) & -\left(\frac{k_{0}\psi}{2}\right)\left(1-\frac{n_{x}}{n_{0}}\right) \\ \frac{n_{0}'}{2n_{0}} & \frac{n_{0}'}{2n_{0}} & -\left(\frac{k_{0}\psi}{2}\right)\left(1-\frac{n_{x}}{n_{0}}\right) & -\left(\frac{k_{0}\psi}{2}\right)\left(1+\frac{n_{x}}{n_{0}}\right) \\ -\left(\frac{k_{0}\psi}{2}\right)\left(1+\frac{n_{0}}{n_{x}}\right) & -\left(\frac{k_{0}\psi}{2}\right)\left(1-\frac{n_{0}}{n_{x}}\right) & \frac{-n_{x}'}{2n_{x}} & \frac{n_{x}'}{2n_{x}} \\ -\left(\frac{k_{0}\psi}{2}\right)\left(1-\frac{n_{0}}{n_{x}}\right) & -\left(\frac{k_{0}\psi}{2}\right)\left(1+\frac{n_{0}}{n_{x}}\right) & \frac{n_{x}'}{2n_{x}} & \frac{-n_{x}'}{2n_{x}} \end{bmatrix}$$

$$(5.20.14)$$

Following the method discussed in Section 5.19, we define

$$\mathbf{h} = \mathbf{S} \cdot \mathbf{g} \tag{5.20.15}$$

where **S** is defined by (5.19.18). The matrix **W** in (5.19.21) is then given by

$$\mathbf{W} = -\mathbf{S} \cdot \mathbf{\Gamma} \cdot \mathbf{S}^{-1} = \begin{bmatrix} 0 & 0 & k_0 \psi & 0 \\ 0 & n_0'/n & 0 & k_0 \psi n_x/n_0 \\ k_0 \psi & 0 & 0 & 0 \\ 0 & k_0 \psi n_0/n_x & 0 & n_x'/n_x \end{bmatrix}$$
(5.20.16)

The matrices defined in (5.19.22) now become

$$\mathbf{A} = \begin{bmatrix} 0 & k_0 \psi \\ k_0 \psi & 0 \end{bmatrix}, \quad \mathbf{B} = \mathbf{C} = 0, \quad \mathbf{D} = \begin{bmatrix} n_0'/n & k_0 \psi n_x/n_0 \\ k_0 \psi n_0/n_x & n_x'/n_x \end{bmatrix}$$

$$\mathbf{Q} = \begin{bmatrix} n_0 & 0 \\ 0 & n_x \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} h_1 \\ h_3 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} h_2 \\ h_4 \end{bmatrix}$$
(5.20.17)

The matrix **G** in (5.19.26) now has a simple form

$$\mathbf{G} = jk_0\mathbf{Q} + \mathbf{B} = jk_0\begin{bmatrix} n_0 & 0\\ 0 & n_x \end{bmatrix}$$
 (5.20.18)

Using (5.20.17) and (5.20.18), we can compute

$$\mathbf{G}' \cdot \mathbf{G}^{-1} = \begin{bmatrix} n_0'/n_0 & 0\\ 0 & n_x'/n_x \end{bmatrix}$$

$$\mathbf{G} \cdot \mathbf{D} \cdot \mathbf{G}^{-1} = \begin{bmatrix} n_0'/n_0 & k_0 \psi\\ k_0 \psi & n_x'/n_x \end{bmatrix}$$

$$\mathbf{A} - \mathbf{G}' \cdot \mathbf{G}^{-1} + \mathbf{G} \cdot \mathbf{D} \cdot \mathbf{G}^{-1} = \begin{bmatrix} 0 & 2k_0 \psi\\ 2k_0 \psi & 0 \end{bmatrix}$$

$$\mathbf{A}' - (\mathbf{G}' \cdot \mathbf{G}^{-1} - \mathbf{G} \cdot \mathbf{D} \cdot \mathbf{G}^{-1}) \cdot \mathbf{A} - \mathbf{G} \cdot \mathbf{C} - jk_0 \mathbf{B} \cdot \mathbf{Q} = \begin{bmatrix} k_0^2 \psi^2 & k_0 \psi'\\ k_0 \psi' & k_0^2 \psi^2 \end{bmatrix}$$

Substituting (5.20.19) into (5.19.27), we obtain the matrix equation

$$\begin{bmatrix} h_{1}'\\ h_{3}'' \end{bmatrix} + k_{0}^{2} \begin{bmatrix} n_{0}^{2} & 0\\ 0 & n_{x}^{2} \end{bmatrix} \begin{bmatrix} h_{1}\\ h_{3} \end{bmatrix} = - \begin{bmatrix} 0 & 2k_{0}\psi & h_{1}'\\ 2k_{0}\psi & 0 & h_{3}' \end{bmatrix} - \begin{bmatrix} k_{0}^{2}\psi^{2} & k_{0}\psi'\\ k_{0}\psi & k_{0}^{2}\psi' \end{bmatrix} \begin{bmatrix} h_{1}\\ h_{3} \end{bmatrix}$$
(5.20.20)

where

$$h_1 = g_1 + g_2 = n_0^{-1/2} (f_1 + jf_2)$$

= $\sqrt{2} (R_x^2 - 1)^{-1/2} (E_x - R_x E_y)$ (5.20.21)

$$h_3 = g_3 + g_4 = n_x^{-1/2} (-jf_3 - f_4)$$

= $\sqrt{2}(1 - R_0^2)^{-1/2} (E_x - R_0 E_y)$ (5.20.22)

 h_1 and h_3 may be related to the two characteristic modes F_+ and F_- defined in (5.17.9) by substituting (5.17.2) into (5.20.21) and (5.20.22). We have

$$h_1 = \sqrt{2}(R_x^2 - 1)^{1/2} E_x^{(0)} = \sqrt{2} F_+ \qquad (5.20.23)$$

$$h_3 = \sqrt{2} (R_x^2 - 1)^{1/2} E_y^{(x)} = \sqrt{2} F_- \qquad (5.20.24)$$

Problems

Therefore (5.20.20) may now be written as

$$d^{2}F_{+}/dz^{2} + k_{0}^{2}(n_{0}^{2} + \psi^{2})F_{+} = k_{0}(d\psi/dz) F_{-} + 2k_{0}\psi \, dF_{-}/dz \quad (5.20.25)$$

$$d^{2}F_{-}/dz^{2} + k_{0}^{2}(n_{x}^{2} + \psi^{2})F_{-} = k_{0}(d\psi/dz) F_{+} + 2k_{0}\psi dF_{+}/dz \quad (5.20.26)$$

which are exactly the Försterling's equations defined in Section 5.17.

Problems

1. Find the equation of rays for an isotropic medium whose refractive index is spherically symmetric, i.e., n = n(r).

2. The curvature vector of a ray is defined by

$$\mathbf{K} = \frac{d\hat{s}}{ds} = \frac{1}{p}\,\hat{v}$$

where p is the radius of curvature and \hat{v} is the unit principal normal at a typical point of the ray. From the ray equation (5.2.15), prove that

$$|\mathbf{K}| = 1/p = \hat{v} \cdot \nabla \ln n$$

What does this relation imply in regard to the ray path?

3. Derive the Fresnel's formulas for reflection and transmission coefficients for vertically polarized waves.

4. In Fig. 5.7-1, if medium 2 is a uniaxial medium characterized by

$$oldsymbol{arepsilon} oldsymbol{arepsilon} = egin{bmatrix} arepsilon_1 & 0 & 0 \ 0 & arepsilon_1 & 0 \ 0 & 0 & arepsilon_2 \end{bmatrix}$$

a vertically polarized wave in free space is incident upon the sharp boundary at z = 0 with an incident angle θ_i (Fig. 5.7-1). Find the reflection and transmission coefficients for this case.

5. In a horizontally stratified ionosphere without the magnetic field, if an obliquely incident wave and a vertically incident wave are reflected at the same level, prove

$$f_{\rm ob} = f_{\rm v} \sec \theta_0$$

where f_{ob} and f_v are respectively the frequency of the oblique wave and vertical wave and θ_0 is the incident angle of the oblique wave. This is known as the secant law (Fig. 5A-1).



Fig. 5A-1. Geometry of an obliquely reflected ray in a plane stratified isotropic ionosphere.

6. In Fig. 5A-1 we assume the ionosphere is horizontally stratified. Neglecting the earth magnetic field, a ray propagates from the transmitter T to the receiver R via the ionosphere. The actual ray path is *TABCR* where the reflection occurs at B. However, if one observes only the angle α of the ray above the horizontal at T and R, the ray appears to follow the path *TAB'CR*.

Prove: the length TAB'CR of the apparent path is equal to the group path TABCR, or $\overline{TABCR} = D/\sin \theta_0$. This is known as the theorem of Breit and Tuve. Therefore, one only needs to know the ground range $D = \overline{TR}$ and the angle of incidence θ_0 to determine the group path.

7. Again referring to Fig. 5A-1, prove that the height h'' of the equivalent triangular vertex is the same as the virtual height measured at the equivalent vertical incidence frequency. The equivalent vertical incidence frequency is defined as the frequency that is related to the frequency of an oblique ray through secant law. This is known as the equivalence theorem.

8. The problem of wave propagation in isotropic stratified media in general can be reduced to the equation of the standard form

$$d^2u/dz^2 + h^2q^2u = 0$$

For a lossless medium q is real. Starting from this equation, prove the following relation:

$$Im[(du/dz) u^*] = constant$$

where Im indicates imaginary part and * indicates complex conjugate.

9. In a lossless isotropic plasma, the electron density varies as a function of z. For a horizontally polarized wave propagating in the direction of z, use the result of Problem 8 to prove

$$\operatorname{Re}(\mathbf{E} \times \mathbf{H}^*) = \operatorname{constant}$$

which indicates that the power flow is constant in the lossless medium.

10. If the WKB solution of the equation

$$\frac{d^2u}{dz^2} + h^2q^2u = 0$$

is given by

$$u = Ay_1 + By_2$$

where

$$y_1 = q^{-1/2} e^{-j\hbar \int_a^z q ds}$$
$$y_2 = q^{-1/2} e^{j\hbar \int_a^z q ds}$$

(a) within the accuracy of the WKB solution, does the law of the conservation of energy hold if q is real and positive (*hint*: compute $Im(u^* du/dz)$?

(b) what will happen if $A = \pm B$?

(c) discuss the results of (a) and (b) for the horizontally polarized wave in Problem 9.

11. Discuss the same questions as in Problem 10 for the case of real and negative q.

12. In the ionosphere, the electron density distribution as a function of height is given by

$$N=ah^2, \qquad h\geq 0$$

where a is a constant. For a pulse with carrier frequency f_0 transmitted vertically from the ground (h = 0) at t = 0, find

- (a) the true height;
- (b) the phase height;
- (c) the virtual height.

13. The study of oblique ionospheric propagation invariably leads us to the study of the Booker quartic (5.20.3). The Booker quartic may also be derived by assuming in each homogeneous slab the wave function dependence of the form $e^{-j(k_xx+k_xz)}$. Here the wave is assumed to be propagating in

xz-plane and its incident angle from free space is θ . Due to Snell's law $k_x = k_0 \sin \theta$. Let $k_z = k_0 q$. Then the Booker quartic can be derived by noting $k^2 = k_x^2 + k_z^2$ or $n^2 = \sin^2 \theta + q^2$ where *n* is the fractive index given by the Appleton-Hartree formula [H. G. Booker, Propagation of wave-packets incident obliquely upon a stratified refracting ionosphere. *Phil. Trans. Roy. Soc. London*, Ser. A 237, 411, (1938)].

14. Show, for the special case of propagation perpendicular to the magnetic meridian, that the Booker quartic results in the solution

$$q^{2} = c^{2} - X \left[U - \frac{Y_{x}^{2} + s^{2}Y_{z}^{2}}{2(U - X)} \pm \left(\frac{(Y_{x}^{2} + s^{2}Y_{z}^{2})^{2}}{4(U - X)^{2}} + \frac{Y_{z}^{2}(c^{2}U - X)}{U - X} \right)^{1/2} \right]^{-1}$$

where $c = \cos \theta$, $s = \sin \theta$, and $\theta =$ incident angle. The wave is propagating in xz-plane. Note that the above relation reduces to the Appleton-Hartree formula for the special case of vertical incidence.

References

- BOOKER, H. G. (1938). Propagation of wave-packets incident obliquely upon a stratified refracting ionosphere, *Phil. Trans. Roy. Soc. London Ser. A* 237, 411-451.
- BREIT, G., and TUVE, M. A. (1926). A test of the existence of the conducting layer, *Phys.* Rev. 28, 554.
- BREKHOVSKIKH, L. M. (1960). "Waves in Layered Media," Academic Press, New York.
- BREMMER, H. (1951). The WKB approximation as the first term of a geometric optical series. In "The Theory of Electromagnetic Waves," Wiley (Interscience), New York. BRILLOUIN, L. (1926). C. R. Acad. Sci. 183, 24-26.
- BUDDEN, K. G., and CLEMMOW, P. C. (1957). Coupled form of the differential equations governing radio propagation in the ionosphere, II. Proc. Cambridge Phil. Soc. 53, 669-682.
- BUDDEN, K. G. (1961). "Radio Waves in the Ionosphere." Cambridge Univ. Press, London and New York.
- BUDDEN, K. G. (1969). Full wave solution for radio waves in a stratified magnetoionic medium. *Alta Freq.* 38, 167–179.
- CLEMMOW, P. C., and HEADING, J. (1954). Coupled forms of the differential equations governing radio propagation in the ionosphere, I. Proc. Cambridge Phil. Soc. 50, 319-333.
- COURENT, R., and HILBERT, D. (1953). "Method of Mathematical Physics," Wiley, New York.
- FORSGREN, S. K. H. (1951). Some calculations of ray paths in the ionosphere. Trans. Chalmers Univ. Technol., Gothenburg. No. 104.
- FörsterLing, K. (1942). Über die Ausbreitung elektromagnetischer Wellen in einem magnetisierten Medium bei senkrechter Incidenz. Hoch. Elek. 59, 110.

References

- FRÖMAN, N., and FRÖMAN, P. O. (1965). "JWKB Approximation," North-Holland, Amsterdam.
- GANS, R. (1915). Fortpflanzung des Lichts durch ein inhomogenes Medium, Ann. Phys. (Leipzig) 47, 709.
- GINZBURG, V. L. (1964). "Propagation of Electromagnetic Waves in Plasmas." Pergamon, Oxford.
- GOURSAT, E., and HEDRICK, E. R. (1959). "A Course in Mathematical Analysis." Dover, New York.
- HASELEGROVE, J. (1954). Ray theory and a new method for ray tracing. Rep. Conf. Phys. Ionosphere, p. 355., London, Phys. Soc. London.
- HASELEGROVE, J. (1957). Oblique ray paths in the ionosphere. Proc. Phys. Soc. London, Sect. B 70, 653.
- HEADING, J. (1962). "An Introduction to Phase Integral Method." Wiley, New York.
- JEFFREVS, H. (1923). On certain approximate solutions of linear differential equations of the second order. *Proc. London Math. Soc.* 23, 428.
- JEFFREYS, H., and JEFFREYS, B. (1956). "Methods of Mathematical Physics," 3rd ed. Cambridge Univ. Press, London and New York.
- KARBOWIAK, A. E. (1967). Wave propagation in inhomogeneous media. In "Electromagnetic Wave Theory" (J. Brown, ed.), p. 1-10. Pergamon, Oxford.
- KLINE, M., and KAY, I. W. (1965). "Electromagnetic Theory and Geometrical Optics." Wiley (Interscience), New York.
- KRAMERS, H. A. (1926). Z. Phys. 39, 828-840.
- POEVERLEIN, H. (1948). Strahlwege von Radiowellen in der Ionosphare, Sitzungsber. Bayer. Akad Wiss. p. 175.
- STANLEY, J. P. (1950). The absorption of long and very long waves in the ionoshpere.2. Atmos. Terr. Phys. 1, 65.
- STOKES, G. G. (1858). On the discontinuity of arbitrary constants which appear in divergent developments. *Trans. Cambridge Phil. Soc.* 10, Part I, 106.
- WAIT, J. R. (1962). "Electromagnetic Waves in Stratified Media." Pergamon, Oxford.
 WATSON, G. N. (1944). "Theory of Bessel Functions," 2nd ed. Cambridge Univ. Press, London and New York.
- WENTZEL, G. (1926). Z. Phys. 38, 518.

6. Wave Propagation in Random Media

6.1 Mathematical Background

In order to study wave propagation phenomena in a random medium, some knowledge of the theory of stochastic process is required. In this section, a brief introduction of the theory will be given. The discussion will be illustrative rather than mathematically rigorous. For more detailed discussion, refer to the books by Papoulis (1965) and Yaglom (1962).

Let us now consider the experiment of tossing a die. There are six different outcomes from this experiment, namely, a "one" showing, a "two" showing, etc. Each experimental outcome is called an event. We denote the set of events by Ω . To every event we assign a number $\xi(\Omega)$. This number $\xi(\Omega)$ is called a random variable. The probability for the occurrence of one event, say Ω_1 , is denoted by $P(\Omega_1)$. This probability can be obtained experimentally by repeating the experiment many times. For a discussion on the various definitions and developments of probability theory, the reader is referred to, for example, the book by Papoulis (1965). In each experiment, the random variable takes a particular value $\xi(\Omega_i)$ which we call a realization of the random variable ξ . The repeated experiments then yield an "ensemble" of realizations. From this ensemble, the statistical characteristics of the random variable can be determined. For example, the simplest statistical characteristic of a random variable is the mean value. This is obtained by

$$\langle \xi \rangle = \sum_{i=1}^{6} \xi(\Omega_i) P(\Omega_i)$$
 (6.1.1)

In the experiment of tossing the die, if we let $\xi(\Omega_i) = i$, then for a fair die such that $P(\Omega_i) = 1/6$, we have

$$\langle \xi \rangle = 3.5 \tag{6.1.2}$$

We next define the distribution function of the random variable ξ by

$$F_{\xi}(x) = P\{\xi \le x\}$$
(6.1.3)

where $P\{\xi \leq x\}$ means the probability such that the random variable ξ is not greater than the number x. x is a real number and $P\{\xi \leq x\}$ is a positive number depending on x. In the above-mentioned example, we see that

$$F_{\xi}(x) = \begin{cases} 0 & x < 1 \\ 1/6 & 1 \le x < 2 \\ 1/3 & 2 \le x < 3 \\ 1/2 & 3 \le x < 4 \\ 2/3 & 4 \le x < 5 \\ 5/6 & 5 \le x < 6 \\ 1 & 6 \le x \end{cases}$$
(6.1.4)

The derivative

$$f_{\xi}(x) = dF_{\xi}(x)/dx$$
 (6.1.5)

of the distribution function $F_{\xi}(x)$ is defined as the density function of the random variable ξ . We note that in general the density function may contain Dirac delta functions.

The following are certain important properties of the distribution and density functions:

(i) $F(-\infty) = 0$, $F(+\infty) = 1$.

(ii) F(x) is monotonically increasing, and f(x) is nonnegative.

(iii)
$$\int_{-\infty}^{+\infty} f(x) dx = F(\infty) - F(-\infty) = 1$$

(iv) $F(x) = \int_{-\infty}^{x} f(t) dt$, $F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(t) dt$

and

$$P\{x_1 \le \xi \le x_2\} = \int_{x_1}^{x_2} f(t) \, dt$$
where in F(x) and f(x) the subscript ξ has been dropped. These properties can be proved from the definitions and the probability theory.

We next introduce the conditional distribution and density. For an event Ω_m such that the probability $P(\Omega_m) \neq 0$, the conditional probability of another event Ω_a assuming Ω_m is given by

$$P(\Omega_a | \Omega_m) = P(\Omega_a \Omega_m) / P(\Omega_m)$$
(6.1.6)

where $P(\Omega_a \Omega_m)$ is the joint probability for the events Ω_a and Ω_m to occur. The conditional distribution function for the random variable ξ assuming Ω_m is defined by

$$F_{\xi}(x \mid \Omega_m) = P\{\xi \le x \mid \Omega_m\} = P\{\xi \le x_1 \Omega_m\} / P(\Omega_m) \qquad (6.1.7)$$

The conditional density function for ξ assuming Ω_m is

$$f_{\xi}(x \mid \Omega_m) = dF_{\xi}(x \mid \Omega_m)/dx \qquad (6.1.8)$$

Given a random variable ξ , then any function of this random variable is also a random variable. For example,

$$\eta = g(\xi) \tag{6.1.9}$$

is a new random variable with distribution

$$F_{\eta}(y) = P\{\eta \le y\} = P\{g(\xi) \le y\}$$
(6.1.10)

and density

$$f_n(y) = dF_n(y)/dy$$
 (6.1.11)

The mean value of $g(\xi)$ can be obtained in terms of the density function of the random variable ξ .

$$\langle \eta \rangle = \langle g(\xi) \rangle = \int_{-\infty}^{+\infty} g(x) f_{\xi}(x) \, dx$$
 (6.1.12)

In particular, if $\eta = \xi$, we have

$$\langle \xi \rangle = \int_{-\infty}^{+\infty} x f_{\xi}(x) \, dx$$
 (6.1.13)

We note that (6.1.13) reduces to (6.1.1) for the case of the die tossing experiment because of the Dirac delta functions in the density function $f_{\xi}(x)$.

Conditional mean value is defined in the similar manner. We have

$$\langle \eta \mid \Omega_m \rangle = \langle g(\xi) \mid \Omega_m \rangle = \int_{-\infty}^{+\infty} g(x) f(x \mid \Omega_m) \, dx \qquad (6.1.14)$$

as the conditional mean value of $g(\xi)$ assuming Ω_m .

If instead of one random variable, we are given two, then any function, of these two random variables is also a random variable. The joint distribution of the two random variables ξ and η is defined by

$$F_{\xi\eta}(x, y) = P\{\xi \le x, \eta \le y\}$$
(6.1.15)

and the joint density function is

$$f_{\xi\eta}(x, y) = \partial^2 F_{\xi\eta}(x, y) / \partial x \, \partial y \tag{6.1.16}$$

Then the mean value of a function

$$\zeta = g(\xi, \eta) \tag{6.1.17}$$

is given by

$$\langle \zeta \rangle = \langle g(\xi,\eta) \rangle = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y) f_{\xi\eta}(x,y) \, dx \, dy \qquad (6.1.18)$$

Similar definitions hold for cases where we have functions of more than two random variables.

Next, let us consider the notion of random functions. In an experiment, to every outcome Ω we assign, according to a certain rule, a function $\xi(t, \Omega)$, real or complex, which depends on the parameter t. We have thus created a family of functions, one for each Ω . These functions are random functions of the parameter t. The family is called a "stochastic process." A simple example is provided by considering the die tossing experiment. If we assign $\xi(t) = \sin t$ for $\Omega = \text{odd}$ and $\xi(t) = t^2$ for $\Omega = \text{even}$, we have a stochastic process.

For a specific t, $\xi(t, \Omega)$ is a random variable. The distribution function is given by

$$F(x, t) = P\{\xi(t) \le x\}$$
(6.1.19)

and the density function is

$$f(x,t) = \frac{\partial F(x,t)}{\partial x}$$
(6.1.20)

which are, obviously, dependent on t. These are the first-order distribution and density of the process $\xi(t)$. For two different values t_1 and t_2 , let us consider the two random variables $\xi(t_1)$ and $\xi(t_2)$. Their joint distribution is

$$F(x_1, x_2; t_1, t_2) = P\{\xi(t_1) \le x_1; \xi(t_2) \le x_2\}$$
(6.1.21)

and density is

$$f(x_1, x_2; t_1, t_2) = \partial^2 F(x_1, x_2; t_1, t_2) / \partial x_1 \partial x_2$$
(6.1.22)

These are the second-order distribution and density of the process, respectively.

In the similar manner, higher order distribution and density functions of the process $\xi(t)$ can be defined. We see that the process is characterized by an infinite sequence of distribution and density functions. The determination of these higher order functions is in general very difficult. In most part of what follows, we shall instead concentrate on the first two characteristics f(x, t) and $f(x_1, x_2; t_1, t_2)$ of the process.

The most important statistical property of the random process is its mean value

$$\langle \xi(t) \rangle = \int_{-\infty}^{+\infty} x f(x;t) \, dx$$
 (6.1.23)

The next simplest quantity is the autocorrelation function

$$\varrho_{\xi}(t_1, t_2) = \langle \xi(t_1)\xi(t_2) \rangle = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_1 x_2 f(x_1, x_2, t_1, t_2) \, dx_1 \, dx_2 \qquad (6.1.24)$$

which is a function of t_1 and t_2 .

The autocovariance of the random variable $\xi(t)$ is defined by

$$C_{\xi}(t_1, t_2) = \langle [\xi(t_1) - \langle \xi(t_1) \rangle] [\xi(t_2) - \langle \xi(t_2) \rangle] \rangle$$
(6.1.25)

From (6.1.24) and (6.1.25), we see that

$$C_{\xi}(t_1, t_2) = \varrho_{\xi}(t_1, t_2) - \langle \xi(t_1) \rangle \langle \xi(t_2) \rangle \tag{6.1.26}$$

In the above definitions, the random process is assumed to be real.

A process $\xi(t)$ is said to be stationary if its mean value is a constant and its autocorrelation function depends on $t_1 - t_2$:

$$\langle \xi(t) \rangle = \text{constant}$$

$$\varrho_{\xi}(t_1, t_2) = \langle \xi(t_1)\xi(t_2) \rangle = \varrho_{\xi}(t_1 - t_2) = \varrho_{\xi}(\tau) \qquad (6.1.27)$$

For a complex stationary process, we have in a slightly different form

$$\varrho_{\xi}(\tau) = \langle \xi(t+\tau)\xi^{*}(t) \rangle = \varrho(\tau) \tag{6.1.28}$$

where the subscript ξ has been dropped.

Clearly

$$\varrho_{\xi}(-\tau) = \varrho_{\xi}^{*}(\tau) \tag{6.1.29}$$

for the complex process, and

$$\varrho_{\xi}(-\tau) = \varrho_{\xi}(\tau) \tag{6.1.30}$$

for the real process.

The autocovariance for a stationary process is

$$C_{\xi}(\tau) = \langle [\xi(t+\tau) - \langle \xi \rangle] [\xi^{*}(t) - \langle \xi^{*} \rangle] \rangle$$

= $\varrho_{\xi}(\tau) - |\langle \xi \rangle|^{2}$ (6.1.31)

The power spectrum (or spectral density) $S_{\xi}(\omega)$ of a stationary process $\xi(t)$ is the Fourier transform of its autocorrelation

$$S_{\xi}(\omega) = \int_{-\infty}^{+\infty} e^{-j\omega\tau} \varrho_{\xi}(\tau) \, d\tau \tag{6.1.32}$$

From the Fourier inversion formula, we have

$$\varrho_{\xi}(\tau) = (1/2\pi) \int_{-\infty}^{+\infty} S_{\xi}(\omega) e^{j\omega\tau} d\omega \qquad (6.1.33)$$

With $\tau = 0$, (6.1.33) yields

$$\varrho_{\xi}(0) = (1/2\pi) \int_{-\infty}^{+\infty} S_{\xi}(\omega) \, d\omega = \langle |\xi(t)|^2 \rangle > 0 \qquad (6.1.34)$$

Thus the integral $\int_{-\infty}^{+\infty} S_{\xi}(\omega) d\omega$ is nonnegative and is proportional to the average "power" of the random process $\xi(t)$; hence the name power spectrum.

The stationary random process $\xi(t)$ itself can be represented in the form of a random Fourier-Stieltjes integral with random amplitude $d\phi(\omega)$:

$$\xi(t) = \int_{-\infty}^{+\infty} e^{j\omega t} \, d\phi(\omega) \tag{6.1.35}$$

Substituting (6.1.35) into (6.1.28), we have

$$\varrho_{\xi}(\tau) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{j\omega_{1}(t+\tau) - j\omega_{2}t} \langle d\phi(\omega_{1}) d\phi^{*}(\omega_{2}) \rangle$$
$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{j(\omega_{1} - \omega_{2})t} e^{j\omega_{1}\tau} \langle d\phi(\omega_{1}) d\phi^{*}(\omega_{2}) \rangle \qquad (6.1.36)$$

Comparing (6.1.36) with (6.1.33), we can write

$$\langle d\phi(\omega_1) \, d\phi^*(\omega_2) \rangle = (1/2\pi) \, \delta(\omega_1 - \omega_2) S_{\xi}(\omega_1) \, d\omega_1 \, d\omega_2 \qquad (6.1.37)$$

which relates the random amplitude $d\phi(\omega)$ to the power spectrum of the process.

One important topic in the theory of random process is the ergodicity of the process. A process is said to be ergodic if time averages are equal to ensemble averages. In our discussion, we shall always assume that the processes are ergodic.

On later occasions, we shall make use of the following theorem of autocorrelation functions for the stationary process. The theorem was formulated by Khinchin (1938). Let us define a random process by

$$\xi_T(t) = \xi(t)$$
 for $-T \le t \le T$
 $\xi_T(t) = 0$ for other values of t (6.1.38)

The Fourier transform of $\xi_T(t)$ is

$$\xi_T(\omega) = \int_{-\infty}^{+\infty} \xi_T(t) e^{-j\omega t} dt = \int_{-T}^{T} \xi(t) e^{-j\omega t} dt$$
 (6.1.39)

Then the theorem states that the power spectrum of the process $\xi(t)$ can be written as

$$S_{\xi}(\omega) = \lim_{T \to \infty} (1/2T) \left| \xi_T(\omega) \right|^2$$
(6.1.40)

To prove (6.1.40), let us introduce the function

$$\varrho_T(\tau) = (1/2T) \int_{-T}^{T} \xi_T(t) \xi_T(t+\tau) dt \qquad (6.1.41)$$

We note that due to ergodicity of the process, $\lim_{T\to\infty} \rho_T(\tau)$ is the auto-

correlation function of the process $\xi(t)$. The Fourier transform of $\rho_T(\tau)$ is

$$\int_{-\infty}^{+\infty} \varrho_T(\tau) e^{-j\omega\tau} d\tau = (1/2T) \int_{-\infty}^{+\infty} e^{-j\omega\tau} d\tau \int_{-\infty}^{+\infty} \xi_T(t) \xi_T(t+\tau) dt$$
$$= (1/2T) \int_{-\infty}^{+\infty} \xi_T(t) e^{j\omega t} dt \int_{-\infty}^{+\infty} \xi_T(t+\tau) e^{-j\omega(t+\tau)} d\tau$$
$$= (1/2T) \xi_T^*(\omega) \xi_T(\omega) = (1/2T) |\xi_T(\omega)|^2 \qquad (6.1.42)$$

Now taking the limit $T \rightarrow \infty$ on both sides of (6.1.42), we have proved (6.1.40).

Up to now, our implication in the discussion of stochastic process is that the parameter t is taken as time. However, the results discussed above are equally applicable to the random functions where the parameter is the spatial coordinate. More generally, if the random function depends not only on one but on several parameters, such as the three spatial coordinates, we define the random function as a random field. The definitions discussed above still hold. But instead of a stationary process, we define a homogeneous field for which the mean value of the field is still constant and the autocorrelation

$$\varrho_{\xi}(\mathbf{r}_{1},\mathbf{r}_{2}) = \langle \xi(\mathbf{r}_{1})^{*}(\mathbf{r}_{2}) \rangle = \varrho_{\xi}(\mathbf{r}_{1}-\mathbf{r}_{2})$$
(6.1.43)

is only a function of $\mathbf{r}_1 - \mathbf{r}_2$ where \mathbf{r}_1 and \mathbf{r}_2 are two arbitrary spatial points. Moreover, if the autocorrelation is a function of the distance between \mathbf{r}_1 and \mathbf{r}_2 , $|\mathbf{r}_1 - \mathbf{r}_2|$, then the field is said to be homogeneous and isotropic.

The power spectrum is defined again by the Fourier transform of the autocorrelation,

$$S_{\xi}(\mathbf{k}) = \iiint_{-\infty}^{+\infty} e^{j\mathbf{k}\cdot\mathbf{R}} \varrho_{\xi}(\mathbf{R}) d\mathbf{R}$$
 (6.1.44)

where $R = \mathbf{r}_1 - \mathbf{r}_2$, and its inverse transform

$$\varrho_{\xi}(\mathbf{R}) = \{1/(2\pi)^3\} \iiint_{-\infty}^{+\infty} S_{\xi}(\mathbf{k}) e^{-j\mathbf{k}\cdot\mathbf{R}} d\mathbf{k}$$
(6.1.45)

Again we have

$$\varrho_{\xi}(0) = \langle | \xi(\mathbf{r}) |^2 \rangle = \{1/(2\pi)^3\} \iint_{-\infty}^{+\infty} S_{\xi}(\mathbf{k}) \, d\mathbf{k} \qquad (6.1.46)$$

We note that in our discussion above we have used Greek letters such as ξ and $\xi(t)$ to represent random variables and random functions. When they take specific values in some realization, they are represented by Roman letters such as x and x(t). However, to simplify our notation, such distinctions are not followed in our later discussions.

This concludes our brief introduction to the theory of random processes. In the following discussion on wave propagation in random media, we shall make use of the definitions and theorems frequently.

6.2 Wave Propagation in Random Media

By random medium, we mean a medium whose properties are random functions of time and position. The randomness may be due to the fluctuations of the thermodynamic quantities of the medium, or due to the presence of the irregular scatterers in the medium. The medium is characterized by its statistical properties. When a wave propagates through this medium, random scatterings take place and the scattered fields interfere with each other in a very complex way. The resultant field also becomes random. The problem of wave propagation in a random medium is the study of the statistical characteristics of the wave. Many physical problems such as scattering of sound waves by turbulent gas, underwater acoustic wave propagation. scattering of radio waves by tropospheric turbulence and by ionospheric irregularities, radio star scintillations, twinkling of stellar images, molecular scattering of light, radiative transfer, and fluctuations of energy levels in a semiconductor all belong to this category. Our discussion will concentrate on the radio waves. However, the same techniques used in our discussion can be applied immediately to other physical problems.

There are two different aspects in the investigation of wave propagation in a random medium. One is to consider the medium as a continuum. The properties of the medium are characterized by their dielectric permittivity. The other is concerned with the scattering of waves by randomly distributed discrete scatterers. Our discussion will be mainly on the first aspect. In most cases, we shall only talk about high frequency electromagnetic waves such that the period of the wave is much shorter than the typical time constant of the random variation of the medium. Therefore, we shall take the permittivity of the medium as a random function of position only.

The major task we shall be concerned with is to solve a partial differential equation with random coefficients. This indeed is a difficult problem. For-

tunately in practice in most situations the randomness of the medium can be considered as weak, i.e., the random part of the dielectric permittivity is small compared with its average value. For those cases, we can apply some perturbation techniques to obtain approximate solutions for the problem. In the following sections, we shall first study the "Born approximation" solution for the various aspects of the problem. This approximation is also called the single scattering approximation because it only takes into account the effect of waves being scattered once. In this context, we shall discuss the power scattered by irregularities, fluctuations, and correlations of phase and amplitudes of the waves in random media. Some applications will also be discussed. Higher order approximations correspond to taking multiple scattering into consideration. We shall discuss certain techniques in dealing with multiple scatterings. In particular, a diagram technique borrowed from field theory will be used to study the propagation of coherent waves in the random medium. The technique is still in its developing stage. Lots of unanswered questions still exist. We shall keep our discussion to the well-developed elementary part of the theory.

6.3 Scattering of Electromagnetic Waves by Irregularities

Suppose in a uniform medium, irregular scatterers are randomly distributed in a localized region. Due to the presence of irregularities, the dielectric permittivity of the medium in the irregular region is expressed as

$$\varepsilon(\mathbf{r}) = \varepsilon_0 \langle K \rangle + \Delta \varepsilon(\mathbf{r}) \tag{6.3.1}$$

where $\varepsilon_0 \langle K \rangle$ is the average dielectric permittivity and $\Delta \varepsilon(\mathbf{r})$ is the fluctuating part, and is a random field of position. Note that the quantities in general will depend on the frequency ω for dispersive medium. $\Delta \varepsilon(\mathbf{r})$ will be assumed to be a homogeneous random field with zero mean value. We also assume throughout our discussion that $|\Delta \varepsilon| \ll \varepsilon_0 \langle K \rangle$, corresponding to weak random irregularities. The susceptibility for this region is given by

$$\chi(\mathbf{r}) = (1/\varepsilon_0)[\varepsilon(\mathbf{r}) - \varepsilon_0] = (\langle K \rangle - 1) + \Delta \varepsilon(\mathbf{r})/\varepsilon_0 \qquad (6.3.2)$$

The dielectric polarization is related to the electric field through χ . We have (Chapter 2)

$$\mathbf{P}(\mathbf{r}) = \varepsilon_0 \chi(\mathbf{r}) \mathbf{E}(\mathbf{r})$$

= [\varepsilon_0(\lap{K}\rangle - 1) + \Delta\varepsilon(\mathbf{r})] \mathbf{E}(\mathbf{r}) \qquad (6.3.3)

The average and fluctuating parts of the polarization are, respectively,

 $\Delta \mathbf{P}(\mathbf{r})$ is due to the presence of the irregularities. In (6.3.4) we have taken the electric field \mathbf{E} as the incident field in the absence of the irregularities since the scattering due to the fluctuating part of the electric field is of second order and is neglected. This corresponds to the so-called Born approximation.



Fig. 6.3-1. Geometry showing transmitter at T, receiver at R and scatterer at S in a volume v'.

Let us now consider the geometry shown in Fig. 6.3-1. A linearly polarized spherical wave radiated from the transmitter T falls on the localized region of irregularities, volume ν' . We like to calculate the scattered power received at a point R. To do this, we first need to know the scattered field generated by the polarization $\Delta P(\mathbf{r})$. The incident field at a scatterer $S(\mathbf{r}')$ is given by

$$\mathbf{E} = (\mathbf{E}_0 / r_1) e^{-jkr_1} \tag{6.3.5}$$

where $k = \sqrt{\langle K \rangle} k_0$ is the propagation constant of the medium in the absence of the irregularities and r_1 is the distance between the transmitter T and the scatterer S. The fluctuating part of the polarization is obtained by substituting (6.3.5) into (6.3.4)

$$\Delta \mathbf{P}(\mathbf{r}) = \Delta \varepsilon (\mathbf{E}_0 / \mathbf{r}_1) e^{-jkr_1}$$
(6.3.6)

The corresponding polarization current is $j\omega \Delta \mathbf{P}$. Hence the vector potential

318

at the receiving point R due to this fluctuating polarization current is given by the integral (Chapter 2);

$$\mathbf{A}(\mathbf{r}_{2}^{0}) = \frac{j\omega\mu_{0}\langle\varepsilon\rangle}{4\pi} \int_{v'} \frac{\varDelta\varepsilon(\mathbf{r}')}{\langle\varepsilon\rangle} \frac{\mathbf{E}_{0}e^{-jk(\mathbf{r}_{1}+\mathbf{r}_{2})}}{\mathbf{r}_{1}\mathbf{r}_{2}} dv' \qquad (6.3.7)$$

where $\langle \varepsilon \rangle = \varepsilon_0 \langle K \rangle$ and r_2 is the distance between the scatterer and the receiver. The integral is over the volume ν' .

We shall assume that r_1^0 and r_2^0 , the distances of T and R from the origin O, respectively, are much greater than the dimension of the region where irregularities exist. From Fig. 6.3-1, we can make the following approximations for the distances.

(i)
$$1/r_1 r_2 \simeq 1/r_1^0 r_2^0$$

(ii) $r_1 = r_1^0 + (r_1 - r_1^0) \simeq r_1^0 - \hat{r}_1^0 \cdot \mathbf{r}'$
 $r_2 = r_2^0 + (r_2 - r_2^0) \simeq r_2^0 - \hat{r}_2^0 \cdot \mathbf{r}'$
(6.3.8)

where P_1^0 and P_2^0 are the unit vectors along the direction OT and OS, respectively.

Using (ii) of (6.3.8), we have

$$r_1 + r_2 = r_1^{0} + r_2^{0} - (f_2^{0} + f_1^{0}) \cdot \mathbf{r}'$$
(6.3.9)

where $P_1^{0} + P_2^{0}$ is the vector that bisects the angle TOR.

Substituting (6.3.9) into (6.3.7), we obtain

$$\mathbf{A}(\mathbf{r}_{2}^{0}) = \frac{j\omega \mathbf{E}_{0}}{4\pi c^{2}} \frac{e^{-j\mathbf{k}(\mathbf{r}_{1}^{0}+\mathbf{r}_{2}^{0})}}{\mathbf{r}_{1}^{0}\mathbf{r}_{2}^{0}} \int \frac{\Delta\varepsilon(\mathbf{r}')}{\langle\varepsilon\rangle} e^{j(\mathbf{k}_{s}-\mathbf{k}_{i})\cdot\mathbf{r}'} d\nu' \qquad (6.3.10)$$

where \mathbf{k}_i and \mathbf{k}_s are the wave vectors of the incident wave and the scattered wave propagating in the direction TO and OR, respectively. This is

$$\mathbf{k}_i = -k\mathbf{\hat{r}}_1^0 \quad \text{and} \quad \mathbf{k}_s = k\mathbf{\hat{r}}_2^0 \quad (6.3.11)$$

The corresponding magnetic field and electric field of the scattered wave can be obtained through the relation $\mathbf{B} = \nabla \times \mathbf{A}$ and the Maxwell's equations. Since the time-averaged scattered power at the receiver is due to the far field components of **B** and **E**, we only need these expressions. We have, for far field, the electric field in the θ -direction (with z-axis taken along \mathbf{E}_0),

$$E_s = -\frac{k^2 E_0}{4\pi} \sin\beta \frac{e^{-jk(r_1^0 + r_2^0)}}{r_1^0 r_2^0} I \qquad (6.3.12)$$

where

$$I = \int_{v'} \left(\Delta \varepsilon(\mathbf{r}') / \langle \varepsilon \rangle \right) e^{j\mathbf{b} \cdot \mathbf{r}'} \, dv' \tag{6.3.13}$$

and β is the angle between \mathbf{E}_0 and the unit vector \hat{k}_s and $\mathbf{b} = \mathbf{k}_s - \mathbf{k}_i$. The scattered power density at R is

$$\frac{|E_s|^2}{2\eta} = \frac{k^4 E_0^2}{(4\pi)^2 2\eta} \sin^2\beta \,\frac{|I|^2}{(r_1^0 r_2^0)^2} \tag{6.3.14}$$

where $\eta = (\mu_0/\langle \varepsilon \rangle)^{1/2}$ is the intrinsic impedance of the medium in the absence of the irregularities. The power scattered per unit solid angle in the direction of R is therefore

$$(r_2^{0})^2 \frac{\mid E_s \mid^2}{2\eta} = \frac{k^4 E_0^2}{(4\pi)^2 2\eta} \sin^2 \beta \frac{\mid I \mid^2}{(r_1^{0})^2}$$

The power density of the incident field is

$$\left(\frac{E_0}{r_1}\right)^2 \frac{1}{2\eta} \simeq \left(\frac{E_0}{r_1^0}\right)^2 \frac{1}{2\eta}$$

Therefore the power scattered per unit solid angle, per unit incident power density, per unit volume is

$$\sigma = \frac{k^4}{(4\pi)^2 \nu'} \sin^2 \beta |I|^2 = \frac{\pi^2 \sin^2 \beta}{\nu' \lambda^4} |I|^2$$
(6.3.15)

where λ is the wavelength in the medium.

The quantity σ is called the scattering cross section and is a measure of the strength of the scattered power in the direction of R.

We note from (6.3.13) that the integral I is the Fourier transform of the homogeneous random field $\Delta \varepsilon(\mathbf{r})/\langle \varepsilon \rangle$ in the limit of $\nu' \to \infty$. Applying Khinchin's theorem, we have

$$S_{\varepsilon}(\mathbf{b}) = \lim_{\nu' \to \infty} \frac{1}{\nu'} |I(\mathbf{b})|^2$$
(6.3.16)

where $S_{\varepsilon}(\mathbf{b})$ is the power spectrum for the random field $\Delta \varepsilon(\mathbf{r})/\langle \varepsilon \rangle$, evaluated at **b**.

Therefore, (6.3.15) becomes approximately

$$\sigma = \{\pi^2 \sin^2 \beta / \lambda^4\} S_s(\mathbf{b}) \tag{6.3.17}$$

The direction **b** is called the mirror vector since it reflects the incident direction to the receiving direction. We see then from (6.3.17) that the

320

power scattered in a particular direction depends on the components of the spectrum of the irregularities in the associated mirror direction.

In the particular case of back-scattering such that $\beta = \pi/2$ and $\mathbf{k}_s = -\mathbf{k}_i$, (6.3.17) reduces to

$$\sigma_B = (\pi^2 / \lambda^4) S_{\varepsilon}(2\mathbf{k}_s) \tag{6.3.18}$$

If we assume an autocorrelation function of the form

$$\varrho_{\varepsilon}(x, y, z) = \frac{\langle (\Delta \varepsilon)^2 \rangle}{\langle \varepsilon \rangle^2} \exp\left\{-\frac{1}{2} \left(\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{d^2}\right)\right\} \quad (6.3.19)$$

where $\langle (\Delta \varepsilon)^2 \rangle$ is a constant. The power spectrum is

$$S_{\varepsilon}(2\mathbf{k}_{s}) = (2\pi)^{3/2} \left\langle \left| \frac{\Delta\varepsilon}{\langle \varepsilon \rangle} \right|^{2} \right\rangle abd \exp\left[-2(k_{x}^{2}a^{2} + k_{y}^{2}b^{2} + k_{z}^{2}d^{2})\right] \quad (6.3.20)$$

where $(k_x, k_y, k_z) = \mathbf{k}_s$. Substituting (6.3.20) into (6.3.18), the back-scattering cross section for this case is given by

$$\sigma_B = (2\pi)^{3/2} \frac{\pi^2}{\lambda^4} \left\langle \left| \frac{\Delta \varepsilon}{\langle \varepsilon \rangle} \right|^2 \right\rangle abd \exp\left[-2(k_x^2 a^2 + k_y^2 b^2 + k_z^2 d^2)\right] \quad (6.3.21)$$

Let us now apply (6.3.21) to a homogeneous, isotropic plasma of density N, the dielectric permittivity is

$$\varepsilon = \varepsilon_0 (1 - \omega_p^2 / \omega^2)$$

If the plasma density has a fluctuating part due to the presence of the irregularities, $N = N_0 + \Delta N(\mathbf{r})$, then the dielectric permittivity is given by

$$\varepsilon = \varepsilon_0 (1 - \omega_{p0}^2 / \omega^2) - \varepsilon_0 (\omega_{p0}^2 / \omega^2) (\Delta N / N_0)$$
(6.3.22)

where ω_{p0} is the plasma frequency corresponding to N_0 . Comparing (6.3.22) with (6.3.1), we have

$$\langle K \rangle = 1 - \omega_{p0}^2 / \omega^2, \qquad \Delta \varepsilon(\mathbf{r}) = -\varepsilon_0 (\omega_{p0}^2 / \omega^2) (\Delta N / N_0) \qquad (6.3.23)$$

Therefore, (6.3.21) can be written for this plasma medium

$$\sigma_{B} = (2\pi)^{3/2} \frac{\pi^{2}}{\lambda^{4}} \frac{(\omega_{p0}/\omega)^{4}}{(1-\omega_{p0}^{2}/\omega^{2})^{2}} \left\langle \left| \frac{\Delta N}{N_{0}} \right|^{2} \right\rangle abd \\ \times \exp[-2(k_{x}^{2}a^{2}+k_{y}^{2}b^{2}+k_{z}^{2}d^{2})] \\ = \frac{\omega_{p0}^{4}}{4\sqrt{2\pi}c^{4}} \left\langle \left| \frac{\Delta N}{N_{0}} \right|^{2} \right\rangle abd \exp[-2(k_{x}^{2}a^{2}+k_{y}^{2}b^{2}+k_{z}^{2}d^{2})] \right.$$
(6.3.24)

We note that (6.3.24) is valid for $\omega \neq \omega_{p0}$.

For the special case in which the irregularities are of spherical shapes, a = b = d = L, (6.3.24) reduces to

$$\sigma_B = \frac{\omega_{p_0}^4}{4\sqrt{2\pi} c^4} \left\langle \left| \frac{\Delta N}{N_0} \right|^2 \right\rangle L^3 \exp(-2k_s^2 L^2)$$
(6.3.25)

This is the back-scattering cross section which is the back-scattered energy flux per unit solid angle, per unit incident power density, per unit volume, from an isotropic plasma having isotropic irregularities whose correlation function is given by (6.3.19) with a = b = d = L.

Hence, in this section, using Born's approximation, we have derived the scattering cross section of electromagnetic waves due to random irregularities. The back-scattering cross-section formula has been used in the interpretation of radar reflection data from the aurora (Booker, 1956) and many other experiments.

Next, as an example of applying (6.3.17), let us calculate the scattering cross section for electromagnetic waves scattered by the thermal fluctuations in an isotropic plasma. For the plasma, (6.3.17) can be written as

$$\sigma = \frac{\pi^2 \sin^2 \beta(\omega_{p0}/\omega)^4}{\lambda^4 (1 - \omega_{p0}^2/\omega^2)^2} S_N(\mathbf{b})$$
(6.3.26)

where $S_N(\mathbf{b})$ is the power spectrum for the plasma density fluctuation $(\Delta N/N_0)$. In Chapter 3, we have computed the plasma density fluctuation due to plasma waves using some simple model. The power spectral for the density fluctuation due to electron plasma wave alone is given by

$$S_N(\mathbf{p},\omega) = \frac{\pi T p^2 N_0}{m_e(\omega_{pe}^2 + \gamma T p^2/m_e)} \left[\delta(\omega - \omega(\mathbf{p})) + \delta(\omega + \omega(\mathbf{p}))\right] \quad (3.6.16)$$

where $\omega(\mathbf{p})$ is the dispersion relation for the plasma wave. For high frequency incident waves, we can neglect the time variation of the density fluctuation. Hence, we take the inverse Fourier transform of (3.6.16), and setting the time interval equal to zero, we have

$$S_N(\mathbf{p}, 0) = \frac{N_0 T p^2 / m_e}{(\omega_{pe}^2 + \gamma T p^2 / m_e)}$$
(6.3.27)

When the motions of ions are included, the total electron density fluctuation is the sum of the fluctuation due to electron plasma wave and that due to ion plasma wave. The total power spectrum can be obtained by adding (3.6.16) and (3.6.17). We have

$$S_N(\mathbf{p}) = \frac{1 + \lambda_D^2 p^2}{2 + \lambda_D^2 p^2} N_0$$
(6.3.28)

where λ_D is the Debye length defined by $\lambda_D = (\epsilon_0 T/N_0 e^2)^{1/2}$. In deriving (6.3.28), we have taken $\gamma = 1$ to make our formula coincide with that derived from kinetic theory.

Substituting (6.3.28) into (6.3.26), we obtain the scattering cross section for electromagnetic waves due to the density fluctuation of the plasma

$$\sigma = \frac{\pi^2 \sin^2 \beta(\omega_{p0}/\omega)^4}{\lambda^4 (1 - \omega_{p0}^2/\omega^2)^2} \frac{1 + \lambda_D^2 |\mathbf{b}|^2}{2 + \lambda_D^2 |\mathbf{b}|^2} N_0$$
(6.3.29)

Since $|\mathbf{b}| = 4\pi\lambda^{-1}\sin(\theta/2)$ where θ is the angle between the incident direction and scattered direction, and $\lambda = 2\pi/[k_0(1-\omega_{p0}^2/\omega^2)^{1/2}]$, (6.3.29) can be put in the form

$$\sigma = \sigma_e N_0 \frac{[4\pi\lambda_D \sin(\theta/2)]^2 + \lambda^2}{[4\pi\lambda_D \sin(\theta/2)]^2 + 2\lambda^2}$$
(6.3.30)

where

$$\sigma_e = (\mu_0 e^2 \sin \alpha / 4\pi m_e)^2 = 8 \times 10^{-30} \sin^2 \alpha \quad (m)^2. \tag{6.3.31}$$

is the classical scattering coefficient for a single free electron.

For $\lambda \ll 4\pi\lambda_D \sin(\theta/2)$, (6.3.30) reduces to $\sigma = N_0\sigma_e$. Since the wavelength is very small compared to the Debye length, the wave sees individual electrons in the plasma as if they are independent scatterers; hence the cross section is the sum of individual free electron cross sections. For sufficiently large values of λ , (6.3.30) reduces to $\sigma = \frac{1}{2}\sigma_e N$. Due to the collective action of the electrons in the plasma, the cross section is thus reduced.

An experiment that takes advantages of the above scattering process is called the "incoherent scattering" experiment. Essentially the experiment involves the following procedure. A strong electromagnetic wave is sent into the plasma; then the scattered signal in certain directions is received. From the power spectrum of the received signal, certain parameters of the plasma such as density, temperature, velocity, etc., can be determined. The technique has been used successfully in probing the ionosphere. The theory behind the experiment is essentially described in the preceding paragraph, although one important modification must be made. In deriving a scattering cross-section formula, the time variation of the fluctuating density must be taken into account. We shall not go further into this subject here and interested readers are referred to various papers (Farley and Dougherty, 1960; Fejer, 1960; Sitenko, 1967).

In previous discussions of this section, the background medium was assumed to be isotropic. It is now of interest to generalize the results to include the case of anisotropic background. The mathematical technique to be used is fairly general and can be used to study scattering from irregularities imbedded in any anisotropic medium. However, in order to be specific, the background is supposed to be a homogeneous magnetoplasma. This is the case of interest in ionospheric studies and laboratory plasma experiments. Due to the presence of irregularities, the dielectric tensor deviates from the mean value and becomes a random function of position; i.e.,

$$\boldsymbol{\varepsilon} = \langle \boldsymbol{\varepsilon} \rangle + \boldsymbol{\varDelta} \boldsymbol{\varepsilon} \tag{6.3.32}$$

where in a weakly random medium $|\Delta \varepsilon|$ is small when compared with $|\langle \varepsilon \rangle|$. The dielectric tensor in a cold, lossless, electronic plasma is given by (4.5.14). The ionic contributions can be ignored if the radio frequency is high when compared with ionic plasma frequencies. Let us also assume that the perturbations in ε are caused entirely by fluctuations in electron density. Then

$$\Delta \boldsymbol{\varepsilon}(\mathbf{r}) = \varepsilon_0 \, \Delta \boldsymbol{\chi}(\mathbf{r}) = \varepsilon_0 \mathbf{M} \, \Delta X(\mathbf{r}) \tag{6.3.33}$$

According to (6.3.32), $\Delta \varepsilon$ and consequently $\Delta \chi$ and ΔX have zero mean. The deterministic tensor **M** is given by

$$\mathbf{M} = -\frac{1}{1 - Y^2} \begin{bmatrix} 1 - Y_x^2 & -Y_x Y_y + j Y_z & -Y_x Y_z - j Y_y \\ -Y_x Y_y - j Y_z & 1 - Y_y^2 & -Y_y Y_z + j Y_x \\ -Y_x Y_z + j Y_y & -Y_y Y_z - j Y_x & 1 - Y_z^2 \end{bmatrix}$$
(6.3.34)

With reference to Fig. 6.3-1, let a spherical wave be transmitted at T with a given characteristic polarization propagating toward the scattering volume ν' . At the scatterer S, the electric field of the incident wave is given by

$$\mathbf{E}_{i}(\mathbf{r}') = (A_{0}/r_{1})\mathbf{a}_{i}e^{-j\mathbf{k}_{i}\cdot\mathbf{r}_{1}}$$
(6.3.35)

where \mathbf{a}_i is the normalized characteristic vector of mode *i*. It should be cautioned that \mathbf{k}_i is the propagation vector of *i*th mode and its corresponding energy propagation is along \mathbf{r}_1 . The angle between \mathbf{k}_i and \mathbf{v}_g ($|| \mathbf{r}_1$) is denoted by α_i which is zero in any isotropic medium but not necessarily zero in an

anisotropic medium. The phase factor of (6.3.35) can be alternately written as

$$\mathbf{k}_i \cdot \mathbf{r}_1 = k_i r_1 \cos \alpha_i \tag{6.3.36}$$

The incident electric field (6.3.35) at S induces a polarization density due to electron density fluctuations.

$$\Delta \mathbf{P}(\mathbf{r}') = \varepsilon_0 \,\Delta \mathbf{\chi}(\mathbf{r}') \cdot \mathbf{E}_i(\mathbf{r}') \tag{6.3.37}$$

The associated polarization current density is $j\omega \Delta \mathbf{P}$, or

$$\Delta \mathbf{J}(\mathbf{r}') = j\omega\varepsilon_0 \,\Delta \mathbf{\chi}(\mathbf{r}') \cdot \mathbf{E}_i(\mathbf{r}')$$

= $j\omega\varepsilon_0 A_0 \,\Delta X(\mathbf{r}')(\mathbf{M} \cdot \mathbf{a}_i/r_1)e^{-j\mathbf{k}_i \cdot \mathbf{r}_1}$ (6.3.38)

The induced current (6.3.38) in the volume v' radiates and gives rise to a scattered field at R. The scattered field, like the incident field, is assumed to propagate in the background homogeneous medium. Such an assumption is identical to the first scatter or Born solution to be discussed later on and is valid when the medium is weakly random. The problem of excitation of fields in an anisotropic medium was discussed in Section 2.13. Making use of these results, the scattered field at R can be expressed as

$$\mathbf{E}_{\mathcal{S}}(\mathbf{r}_{2}^{0}) = -j\omega\mu_{0}\int_{v'}\mathbf{\Gamma}(\mathbf{r}_{2}^{0},\mathbf{r}')\cdot\Delta\mathbf{J}(\mathbf{r}')\,d\mathbf{r}' \qquad (6.3.39)$$

where the dyadic Green's function in a lossless magnetoplasma is given by

$$\mathbf{\Gamma}(\mathbf{r}_{2}^{0},\mathbf{r}') = \frac{1}{(2\pi)^{3}} \sum_{\alpha} \int \frac{\mathbf{a}_{\alpha} \mathbf{a}_{\alpha}^{*}}{k^{2} - k_{\alpha}^{2}} e^{-j\mathbf{k} \cdot (\mathbf{r}_{2}^{0} - \mathbf{r}')} d\mathbf{k} \qquad (6.3.40)$$

The evaluation of the integral in (6.3.40) in general is very difficult. Since in our case the receiver R is very far away from the scattering volume ν' , the inversion in k-space may be carried out asymptotically. The result is, according to Section 2.14,

$$\boldsymbol{\Gamma}(\mathbf{r}_{2}^{0},\mathbf{r}') = \frac{1}{4\pi |\mathbf{r}_{2}^{0} - \mathbf{r}'|} \sum_{\alpha} \frac{C_{\alpha} \mathbf{a}_{\alpha} \mathbf{a}_{\alpha}^{*}}{k_{\alpha} (|K_{\alpha}|)^{1/2} \sec \alpha_{\alpha}} e^{-j\mathbf{k}_{\alpha} \cdot (\mathbf{r}_{2}^{0} - \mathbf{r}')} \quad (6.3.41)$$

The summation is carried out over all characteristic modes and those saddle points of the dispersion surface whose normal is in the direction $\mathbf{r}_2^0 - \mathbf{r}'$. In a cold magnetoplasma, there are only two modes, but there

6. Wave Propagation in Random Media

may be as many as three saddle points on the dispersion surface whose corresponding group velocities are parallel. The exponential phase factor in (6.3.41) must be interpreted in the same manner as that of (6.3.35), i.e., \mathbf{k}_{α} 's of (6.3.41) must be those values of the propagation vector whose corresponding group velocities are all parallel to the vector $\mathbf{r}_{2}^{0} - \mathbf{r}'$. Therefore, \mathbf{k}_{α} as well as \mathbf{a}_{α} , K_{α} , and α_{α} all depend on \mathbf{r}' . If the angle subtended by the scattering volume v' is small at T and R, we may assume \mathbf{k}_{α} , \mathbf{a}_{α} , K_{α} , α_{α} , a_{i} , and \mathbf{k}_{i} to take constant value. Making such an assumption, the scattered electric field (6.3.39) now becomes

$$\mathbf{E}_{s}(\mathbf{r}_{2}^{0}) = \frac{k_{0}^{2}A_{0}}{4\pi} \sum_{\alpha} \frac{C_{\alpha}\mathbf{a}_{\alpha}(\mathbf{a}_{\alpha}^{*} \cdot \mathbf{M} \cdot \mathbf{a}_{i})}{k_{\alpha}(|K_{\alpha}|)^{1/2} \sec \alpha_{\alpha}} \int_{v'} \frac{\Delta X(\mathbf{r}')}{|\mathbf{r}_{1}||\mathbf{r}_{2}^{0} - \mathbf{r}'||} e^{-j\mathbf{k}_{\alpha} \cdot (\mathbf{r}_{2}^{0} - \mathbf{r}') - j\mathbf{k}_{i} \cdot \mathbf{r}_{1}} d\mathbf{r}'$$

The approximations (6.3.8), (6.3.9) can be used again in the above expression to produce

$$\mathbf{E}_{s}(\mathbf{r}_{2}^{0}) = \frac{k_{0}^{2}A_{0}}{4\pi r_{1}^{0}r_{2}^{0}} \sum_{\alpha} \frac{C_{\alpha}\mathbf{a}_{\alpha}(\mathbf{a}_{\alpha}^{*} \cdot \mathbf{M} \cdot \mathbf{a}_{i})}{k_{\alpha}(|K_{\alpha}|)^{1/2} \sec \alpha_{\alpha}} e^{-j(k_{\alpha}r_{2}^{0}\cos\alpha_{\alpha}+k_{i}r_{1}^{0}\cos\alpha_{i})} \\ \times \int_{v'} \Delta X(\mathbf{r}') e^{j(\mathbf{k}_{\alpha}-\mathbf{k}_{i})\cdot\mathbf{r}'} d\mathbf{r}'$$
(6.3.42)

For later convenience, we write (6.3.42) as

$$\mathbf{E}_{s}(\mathbf{r}_{2}^{0}) = \sum_{\alpha} \mathbf{a}_{\alpha} E_{\alpha}(\mathbf{r}_{2}^{0})$$
(6.3.43)

The associated scattered magnetic field is

$$\mathbf{H}_{s}(\mathbf{r}_{2}^{0}) = -\frac{1}{j\omega\mu_{0}} \nabla \times \mathbf{E}_{s}(\mathbf{r}_{2}^{0})$$

where the symbol ∇ operates on \mathbf{r}_2^0 coordinates. Carrying out the operation,

$$\mathbf{H}_{s}(\mathbf{r}_{2}^{0}) = \frac{1}{\omega\mu_{0}} \sum_{\alpha} \mathbf{k}_{\alpha} \times \mathbf{a}_{\alpha} E_{\alpha}(\mathbf{r}_{2}^{0})$$
(6.3.44)

The scattered time-averaged power flow density at R is just the Poynting vector. Let the process ΔX be stationary with the correlation function B, i.e.,

$$B(\mathbf{r}_{1}' - \mathbf{r}_{2}') = \langle \Delta X(\mathbf{r}_{1}') \, \Delta X(\mathbf{r}_{2}') \rangle / \langle (\Delta X)^{2} \rangle \tag{6.3.45}$$

Let the difference of two propagation vectors of two scattered modes be h:

$$\mathbf{h} = \mathbf{k}_{\alpha} - \mathbf{k}_{\beta} \tag{6.3.46}$$

Then, the statistically averaged Poynting vector involves a double volume integral of the form

$$\iint B(\mathbf{r}_1' - \mathbf{r}_2') e^{j(\mathbf{k}_{\alpha} - \mathbf{k}_i) \cdot (\mathbf{r}_1' - \mathbf{r}_2')} \cdot e^{j\mathbf{h} \cdot \mathbf{r}_2'} d\mathbf{r}_1' d\mathbf{r}_2'$$
$$= \iint B(\boldsymbol{\xi}) e^{j(\mathbf{k}_{\alpha} - \mathbf{k}_i) \cdot \boldsymbol{\xi}} \cdot e^{j\mathbf{h} \cdot \boldsymbol{\eta}} d\boldsymbol{\xi} d\boldsymbol{\eta}$$

where coordinate transformations $\boldsymbol{\xi} = \mathbf{r}_1' - \mathbf{r}_2'$, $\boldsymbol{\eta} = \mathbf{r}_2'$ have been introduced. If the dimension of the scattering volume is much larger than 1/h, the integral with respect to $\boldsymbol{\eta}$ vanishes unless $\mathbf{h} = 0$, i.e., $\mathbf{k}_{\alpha} = \mathbf{k}_{\beta}$. Physically then if the differential phase shift of the two scattered modes is large over the scattering volume, the total scattered energy flow is equal to the sum of the partial energy flows in the individual ordinary and extraordinary modes. When this is the case, the statistically averaged Poynting vector reduces to

where S_X is the power spectrum of the process ΔX defined by

$$S_{\mathcal{X}}(\mathbf{k}) = \int B(\boldsymbol{\xi}) \, e^{i\mathbf{k}\cdot\boldsymbol{\xi}} \, d\boldsymbol{\xi} \tag{6.3.48}$$

The power density of the incident field is

$$\operatorname{Re}[A_0^2/2\omega\mu_0(\mathbf{r}_1^0)^2] \mid \mathbf{a}_i \times (\mathbf{k}_i \times \mathbf{a}_i^*) \mid$$

The scattered power density per unit solid angle is just $(r_2^0)^2 S$. The scattering cross section is defined as the scattered power density in α th mode per unit solid angle per unit incident power in the *i*th mode per unit volume, or

$$_{i}\sigma_{\alpha} = \frac{k_{0}^{4}}{(4\pi)^{2}} \frac{\operatorname{Re}|\mathbf{a}_{\alpha} \times (\mathbf{k}_{\alpha} \times \mathbf{a}_{\alpha}^{*})|}{\operatorname{Re}|\mathbf{a}_{i} \times (\mathbf{k}_{i} \times \mathbf{a}_{i}^{*})|} \frac{|\mathbf{a}_{\alpha}^{*} \cdot \mathbf{M} \cdot \mathbf{a}_{i}|^{2}}{k_{\alpha}^{2}|K_{\alpha}|\operatorname{sec}^{2}\alpha_{\alpha}} S_{X}(\mathbf{k}_{\alpha} - \mathbf{k}_{i})$$
(6.3.49)

We note that the vector

$$\operatorname{Re}[\mathbf{a}_{\alpha} \times (\mathbf{k}_{\alpha} \times \mathbf{a}_{\alpha}^{*})] = (\mathbf{a}_{\alpha} \cdot \mathbf{a}_{\alpha}^{*})\mathbf{k}_{\alpha} - \operatorname{Re}(\mathbf{a}_{\alpha} \cdot \mathbf{k}_{\alpha})\mathbf{a}_{\alpha}^{*} \quad (6.3.50)$$

is parallel to the α th mode Poynting vector. We wish now to evaluate this vector. The characteristic vector \mathbf{a}_{α} is normalized so that its magnitude

transverse to \mathbf{k}_{α} is unity. In a coordinate system in which \mathbf{k}_{α} is along the z-axis and \mathbf{B}_0 is the steady magnetic field vector in the yz-plane (see Fig. 4.14-1), it is given by

$$\mathbf{a}_{\alpha} = (1/[1+|R_{\alpha}|^2]^{1/2})[\hat{x}R_{\alpha} + \hat{y} + \hat{z}R_{\alpha}Q_{\alpha}]$$
(6.3.51)

where R and Q are both pure imaginary and are given by (4.14.14) and (4.14.16), respectively. The second term on the right-hand side of (6.3.50) is then computed to be

$$\operatorname{Re}(\mathbf{a}_{\alpha} \cdot \mathbf{k}_{\alpha})\mathbf{a}_{\alpha}^{*} = [k_{\alpha}/(1+\mid R_{\alpha}\mid^{2})][\hat{y}Q_{\alpha}R_{\alpha} + \hat{z}\mid Q_{\alpha}\mid^{2} \mid R_{\alpha}\mid^{2}]$$

Using this expression, (6.3.50) can be reduced to

$$\operatorname{Re} \left| \mathbf{a}_{\alpha} \times (\mathbf{k}_{\alpha} \times \mathbf{a}_{\alpha}^{*}) \right| = \left| -\hat{y} Q_{\alpha} R_{\alpha} k_{\alpha} / (1 + |R_{\alpha}|^{2}) + \hat{z} k_{\alpha} \right|$$
$$= k_{\alpha} \sec \alpha_{\alpha} \qquad (6.3.52)$$

The last equality is obtained because the angle between \mathbf{k}_{α} and the Poynting vector is α_{α} . Similar computations for the incident wave yield

$$\operatorname{Re} \left| \mathbf{a}_{i} \times (\mathbf{k}_{i} \times \mathbf{a}_{i}^{*}) \right| = k_{i} \sec \alpha_{i}$$
(6.3.53)

Substituting (6.3.52) and (6.3.53) in (6.3.49), we obtain finally

$$_{i}\sigma_{\alpha} = \frac{k_{0}^{4}}{(4\pi)^{2}} \frac{|\mathbf{a}_{\alpha}^{*} \cdot \mathbf{M} \cdot \mathbf{a}_{i}|^{2} \cos \alpha_{i} \cos \alpha_{\alpha}}{k_{i}k_{\alpha} | K_{\alpha} |} S_{X}(\mathbf{k}_{\alpha} - \mathbf{k}_{i}) \quad (6.3.54)$$

This is the scattering cross section for an *i*th incident mode scattered into the α th mode. The quantity $\mathbf{a}_{\alpha}^* \cdot \mathbf{M} \cdot \mathbf{a}_1$ can be viewed as the projection of the scattered mode on the induced polarization and is in general complex, but it becomes real if scattering takes place in the magnetic meridian plane. It should be remembered that \mathbf{k}_i is the wave vector of the incident wave whose energy propagates from T to 0 of Fig. 6.3-1 and \mathbf{k}_{α} is the wave vector of the scattered wave whose energy propagates from 0 to R. However, the direction $\mathbf{k}_{\alpha} - \mathbf{k}_{i}$ can still be identified as the mirror direction since a perfect reflector normal to this vector will reflect energy incident from the transmitter at T to the receiver at R. Consequently, the scattering is still dependent on the Fourier content in the spectrum of N in the mirror direction. The scattering process effectively picks out those irregularities in the fluctuation Fourier spectrum that are in the mirror direction; all other irregularities are ineffective as far as the scattering is concerned. The expression (6.3.54) is rather involved and has been studied numerically in detail by Simonich and Yeh (1971).

6.4 Fluctuation of Electromagnetic Waves in Random Media-Geometrical Optics

In the previous section, the power scattered by localized random irregularities has been calculated in the limit of Born's approximation. In many cases, the fluctuations of the other propagation parameters of the wave such as amplitude, phase, direction of arrival, and frequency are of practical importance. On the one hand, these are pertinent parameters in designing a transmitting or a receiving device; on the other hand, they may yield valuable information about the statistical characteristics of the medium. Therefore, a large portion of the literature of wave propagation in random media has been devoted to this area. In the next three sections, we shall discuss the theory as well as some applications of this topic.

Let us consider a random medium occupying the infinite half space z > 0. Again we assume the medium is characterized by the dielectric permittivity given by (6.3.1). Starting from the Maxwell's equations, assuming time dependence $e^{j\omega t}$, we can write the wave equation for the electric field in the form

$$\nabla^{2}\mathbf{E} + k_{0}^{2}(\varepsilon/\varepsilon_{0})\mathbf{E} + \nabla[\mathbf{E}\cdot\nabla\ln(\varepsilon/\varepsilon_{0})] = 0$$
(6.4.1)

Before we go into the discussion of fluctuations of waves, we note that (6.4.1) can be taken as the starting equation in the discussion of scattering from irregularities instead of the procedure described in the last section. From (6.4.1), scattered electric field can be obtained directly.

We now come back to the problem of this section. Let us assume that the wavelength λ is small compared to the typical dimensions of the irregularities, *l*. In this case we can neglect the last term of (6.4.1) and we obtain instead of the vector wave equation three scalar wave equations,

$$\nabla^2 u + k_0^2 (\varepsilon/\varepsilon_0) u = 0 \tag{6.4.2}$$

where u is any one of the components of **E**. This assumption is equivalent to neglecting the effect of depolarization.

Setting

$$u(\mathbf{r}) = A(\mathbf{r})e^{-jS(\mathbf{r})} \tag{6.4.3}$$

and substituting it into (6.4.2) under the assumptions of geometrical optics (Chapter 5) we obtain

$$(\nabla S)^2 = k_0^2 \varepsilon(\mathbf{r}) / \varepsilon_0$$

$$\nabla^2 S + 2\nabla S \cdot \nabla \ln A = 0$$
(6.4.4)

Again we assume in $\varepsilon = \varepsilon_0 \langle K \rangle + \Delta \varepsilon(\mathbf{r}), |\Delta \varepsilon| \ll \varepsilon_0 \langle k \rangle$. We then set

$$S = S_0 + S_1$$

$$\ln A = \ln A_0 + \chi$$
(6.4.5)

where S_1 and χ are of the order of $\Delta \varepsilon$.

Substituting (6.4.5) into (6.4.1) and equating terms of the same order of smallness, we obtain for the zeroth order

$$(\nabla S_0)^2 = k^2 \langle K \rangle = k^2$$

$$\nabla^2 S_0 + 2\nabla \ln A_0 \cdot \nabla S_0 = 0$$
(6.4.6)

and for the first order

$$2\nabla S_0 \cdot \nabla S_1 = k_0^2 [\Delta \varepsilon(r) / \varepsilon_0]$$

$$\nabla^2 S_1 + 2\nabla \ln A_0 \cdot \nabla S_1 + 2\nabla \chi \cdot \nabla S_0 = 0$$
(6.4.7)

where $|\nabla S_1| \ll |\nabla S_0|$ or $|\nabla S_1| \ll k$ has been assumed. This implies that the change of phase in one wavelength is very small.

We now consider a plane incident wave in the z-direction. Thus the solutions of (6.4.6) are $S_0 = kz$ and $A_0 = \text{constant.}$ Equation (6.4.7) becomes

$$\frac{\partial S_1}{\partial z} = \frac{1}{2}k[\Delta \epsilon(\mathbf{r})/\langle \epsilon \rangle]$$

$$\nabla^2 S_1 + 2k \, \partial \chi/\partial z = 0$$
(6.4.8)

where $\langle \varepsilon \rangle = \varepsilon_0 (\langle K \rangle)^{1/2}$ as before. Integrating (6.4.8), we have

$$S_{1}(x, y, L) = (1/2)k \int_{0}^{L} \left[\Delta \varepsilon(x, y, z) / \langle \varepsilon \rangle \right] dz$$

$$\chi(x, y, L) = -(1/2k) \int_{0}^{L} \nabla^{2} S_{1}(x, y, z) dz$$

$$(6.4.9)$$

$$= -\frac{1}{2k} \left\{ \left(\frac{\partial S_1}{\partial z} \right)_{z=0}^{z=L} + \int_0^L \left(\frac{\partial^2 S_1}{\partial x^2} + \frac{\partial^2 S_1}{\partial y^2} \right) dz \right\} \quad (6.4.10)$$

where

$$\left(\frac{\partial S_1}{\partial z}\right)_{z=0}^{z=L} = \frac{k}{2\langle \varepsilon \rangle} \left[\varDelta \varepsilon(x, y, L) - \varDelta \varepsilon(x, y, 0) \right]$$
(6.4.11)

Equations (6.4.9) and (6.4.10) relate the fluctuations of the phase and amplitudes of the wave to the fluctuating part of the dielectric permittivity. Obviously, $\langle S_1 \rangle = \langle \chi \rangle = 0$ since $\langle \Delta \varepsilon \rangle = 0$. In the following we shall compute the mean square fluctuation of the phase. To do this, we first derive the

330

expression for the autocorrelation of S_1 in the plane z = L. We have, from (6.4.9)

$$\varrho_{S}(x_{1} - x_{2}, y_{1} - y_{2}; L) = \langle S_{1}(x_{1}, y_{1}, L) S_{1}(x_{2}, y_{2}, L) \rangle$$

$$= \frac{k^{2}}{4} \int_{0}^{L} \int_{0}^{L} \langle \Delta \varepsilon(x_{1}, y_{1}, z_{1}) \Delta \varepsilon(x_{2}, y_{2}, z_{2}) / (\langle \varepsilon \rangle)^{2} \rangle$$

$$\times dz_{1} dz_{2} \qquad (6.4.12)$$

The notation $\varrho_S(x_1 - x_2, y_1 - y_2; L)$ indicates that the correlation is computed at the plane z = L. For the case $\Delta \varepsilon(\mathbf{r})$ is a homogeneous field, the autocorrelation is a function of $\mathbf{r_1} - \mathbf{r_2}$, (6.4.12) can be written as

$$\varrho_{S}(\xi,\eta;L) = \frac{k^{2}}{4} \left\langle \left| \frac{\Delta\varepsilon}{\langle\varepsilon\rangle} \right|^{2} \right\rangle \int_{0}^{L} \int_{0}^{L} \varrho_{\varepsilon}(\xi,\eta,\zeta) \, dz_{1} \, dz_{2} \qquad (6.4.13)$$

where $\xi = x_1 - x_2$, $\eta = y_1 - y_2$; and $\zeta = z_1 - z_2$, and

$$\langle [\varDelta \varepsilon(\mathbf{r}_1) / \langle \varepsilon \rangle] [\varDelta \varepsilon(\mathbf{r}_2) / \langle \varepsilon \rangle] \rangle = \langle |\varDelta \varepsilon / \langle \varepsilon \rangle |^2 \rangle \varrho_{\varepsilon}(\mathbf{R})$$
(6.4.14)

Changing the variables of integration in (6.4.13)

$$z_1 = \zeta + z_2, \qquad z_2 = z_2$$
 (6.4.15)

we have

$$\varrho_{\mathcal{S}}(\xi,\eta;L) = \frac{k^2}{4} \left\langle \left| \frac{\Delta\varepsilon}{\langle\varepsilon\rangle} \right|^2 \right\rangle \int_0^L dz_2 \int_{-z_2}^{z_1-z_2} \varrho_{\varepsilon}(\xi,\eta,\zeta) \, d\zeta \quad (6.4.16)$$

Changing the order of integration and carrying out the z_2 -integration, we obtain after some algebra

$$\varrho_{S}(\xi,\eta;L) = \frac{k^{2}}{4} \left\langle \left| \frac{\Delta\varepsilon}{\langle\varepsilon\rangle} \right|^{2} \right\rangle \\ \times \left[\int_{-L}^{0} (L+\zeta) \varrho_{\varepsilon}(\xi,\eta,\zeta) \, d\zeta + \int_{0}^{L} (L-\zeta) \varrho_{\varepsilon}(\xi,\eta,\zeta) \, d\zeta \right] \\ = \frac{k^{2}}{2} \left\langle \left| \frac{\Delta\varepsilon}{\langle\varepsilon\rangle} \right|^{2} \right\rangle \int_{0}^{L} (L-\zeta) \varrho_{\varepsilon}(\xi,\eta,\zeta) \, d\zeta$$
(6.4.17)

where $\varrho_{\epsilon}(\xi,\eta,\zeta) = \varrho_{\epsilon}(\xi,\eta,-\zeta)$ has been used.

Since $\rho_{\epsilon}(\xi, \eta, \zeta)$ becomes very small as ζ increases beyond the correlation length, (6.4.17) can be approximately written for large value of L as

$$\varrho_{\mathcal{S}}(\xi,\eta;L) \simeq \frac{k^2}{2} \left\langle \left| \frac{\Delta \varepsilon}{\langle \varepsilon \rangle} \right|^2 \right\rangle L \int_0^\infty \varrho_{\varepsilon}(\xi,\eta,\zeta) \, d\zeta \qquad (6.4.18)$$

If the power spectrum of $\rho_{\varepsilon}(\xi, \eta, \zeta)$ is $S_{\varepsilon}(K_1, K_2, K_3)$, then

$$S_{\varepsilon}(K_1, K_2, K_3) = \iiint_{-\infty}^{+\infty} \varrho_{\varepsilon}(\xi, \eta, \zeta) e^{j(K_1\xi + K_2\eta + K_3\zeta)} d\xi d\eta d\zeta \quad (6.4.19)$$

We can define a two-dimensional spectral density in the plane $z = \zeta$ by

$$F_{\varepsilon}(K_1, K_2, \zeta) = \int_{-\infty}^{+\infty} \varrho_{\varepsilon}(\xi, \eta, \zeta) e^{j(K_1\xi + K_2\eta)} d\xi d\eta \qquad (6.4.20)$$

Therefore, from (6.4.19) and (6.4.20), we have

$$S_{\varepsilon}(K_1, K_2, K_3) = \int_{-\infty}^{+\infty} F_{\varepsilon}(K_1, K_2, \zeta) e^{jK_3\zeta} d\zeta \qquad (6.4.21)$$

and

$$S_{\varepsilon}(K_1, K_2, 0) = \int_{-\infty}^{+\infty} F_{\varepsilon}(K_1, K_2, \zeta) d\zeta \qquad (6.4.22)$$

Applying this definition to (6.4.18), the two-dimensional spectral density of $\rho_s(X, Y, L)$ becomes

$$F_{S}(K_{1}, K_{2}, L) \simeq \frac{k^{2}}{4} \left\langle \left| \frac{\Delta \varepsilon}{\langle \varepsilon \rangle} \right|^{2} \right\rangle L \int_{-\infty}^{+\infty} F_{\varepsilon}(K_{1}, K_{2}, \zeta) d\zeta$$
$$= \frac{k^{2}}{4} \left\langle \left| \frac{\Delta \varepsilon}{\langle \varepsilon \rangle} \right|^{2} \right\rangle LS_{\varepsilon}(K_{1}, K_{2}, 0) \qquad (6.4.23)$$

Equations (6.4.18) and (6.4.23) relate the autocorrelation and twodimensional power spectrum of the phase fluctuation of the wave in the plane z = L to the autocorrelation and power spectrum of the dielectric permittivity, respectively.

Now, a few words about the range of validity for the geometric optics. First of all, in order to apply this method, we should have

$$\lambda \ll l \tag{6.4.24}$$

where *l* is the dimension of the irregularities. This length is closely related to the correlation length of $\varrho_{\epsilon}(\mathbf{r})$. In addition, we should be able to neglect the diffraction effects. For an obstacle of dimension *l*, the angle of divergence of the diffracted wave is of the order of $\theta \sim \lambda/l$. At a distance *L* from the obstacle the size of the diffracted image will be of the order of $\theta L \sim L\lambda/l$.

In order for the geometrical shadow of the irregularity not to be appreciably changed, we require

 $L\lambda/l \ll l$

or

$$(\lambda L)^{1/2} \ll l$$
 (6.4.25)

Equations (6.4.24) and (6.4.25) give the range of validity of the application of geometric optics.

The mean square fluctuation of the phase is obtained by setting $\xi = \eta = 0$ in (6.4.18):

$$\langle S^2 \rangle = \frac{k^2}{2} \left\langle \left| \frac{\Delta \varepsilon}{\langle \varepsilon \rangle} \right|^2 \right\rangle L \int_0^\infty \varrho_\varepsilon(0, 0, \zeta) \, d\zeta \tag{6.4.26}$$

We note that this fluctuation increases with the distance the wave has traveled in the medium.

Similar computation can be made for the amplitude function χ .

6.5 Fluctuation of Electromagnetic Waves in Random Media—Wave Theory

We now turn to the wave theory of propagation in random media. The starting point of this section is again the scalar wave equation

$$\nabla^2 u + k_0^2 (\varepsilon/\varepsilon_0) u = 0 \tag{6.4.2}$$

Let us define a new function $\psi(\mathbf{r})$ by

$$u(\mathbf{r}) = u_0(\mathbf{r})e^{\psi(\mathbf{r})} \tag{6.5.1}$$

where $u_0(\mathbf{r})$ satisfies (6.4.2) with $k_0^2 \varepsilon / \varepsilon_0$ replaced by

$$k^2 = k_0^2 \langle \varepsilon \rangle / \varepsilon_0 \tag{6.5.2}$$

It is the solution of the wave equation in the absence of the random irregularities.

We note by comparing (6.4.3) and (6.5.1) that

Re
$$\psi(\mathbf{r}) = \chi(\mathbf{r})$$

Im $\psi(\mathbf{r}) = S_1(\mathbf{r})$ (6.5.3)

where $\chi(\mathbf{r})$ and $S_1(\mathbf{r})$ are the fluctuations of the logarithmic amplitude and

phase of $u(\mathbf{r})$, respectively. Substituting (6.5.2) into (6.4.2), we obtain for ψ

$$\nabla^2 \psi + (\nabla \psi)^2 + 2 \frac{\nabla u_0}{u_0} \cdot \nabla \psi + k^2 \frac{\Delta \varepsilon}{\langle \varepsilon \rangle} = 0$$
 (6.5.4)

Setting

$$\psi = vw \tag{6.5.5}$$

and substituting it in (6.5.4), we obtain

$$\nu \nabla^2 w + (\nabla^2 \nu) w + (\nabla \nu)^2 w^2 + 2(\nabla u_0/u_0) \cdot \nabla(\nu w) + (\nu^2 \nabla w + 2\nu w \nabla \nu + 2 \nabla \nu) \cdot \nabla w + k^2 (\Delta \varepsilon / \langle \varepsilon \rangle) = 0 \quad (6.5.6)$$

Since now instead of ψ , we have two functions v and w, we have an additional freedom in choosing v and w. In particular if we require

$$v \nabla w + w \nabla v + 2(\nabla v/v) + 2(\nabla u_0/u_0) = 0$$
(6.5.7)

then (6.5.6) becomes

$$\nabla^2 w + [\nabla^2 \ln v - (\nabla \ln v)^2] w + k^2 (\Delta \varepsilon / \langle \varepsilon \rangle) / v = 0 \qquad (6.5.8)$$

which is void of the term ∇w . Equation (6.5.7) yields after integration

$$2\ln v + vw = -2\ln u_0 \tag{6.5.9}$$

Equations (6.5.8) and (6.5.9) are equivalent to (6.5.4). They are two coupled nonlinear partial differential equations. The first-order approximation gives the so-called Rytov's solution. To derive it, we neglect the non-linear term vw in (6.5.9) and obtain

$$\ln v = -\ln u_0$$
 or $v = 1/u_0$ (6.5.10)

Substituting (6.5.10) into (6.5.8), we have

$$\nabla^2 w + k^2 w = -k^2 (\Delta \varepsilon / \langle \varepsilon \rangle) u_0 \tag{6.5.11}$$

where (6.5.2) has been used. The solution of (6.5.11) can be written in terms of the Green's function (2.14.15)

$$w(\mathbf{r}) = k^2 \int \left[\Delta \varepsilon(\mathbf{r}') / \langle \varepsilon \rangle \right] u_0(\mathbf{r}') G(\mathbf{r}, \mathbf{r}') \, d\mathbf{r}' \tag{6.5.12}$$

where

$$G(\mathbf{r},\mathbf{r}') = \frac{1}{4\pi |\mathbf{r}-\mathbf{r}'|} e^{-jk|\mathbf{r}-\mathbf{r}'|}$$
(6.5.13)

Combining (6.5.5), (6.5.10), and (6.5.12), we obtain for the Rytov's approximation

$$\psi(\mathbf{r}) = (k^2/4\pi) \int d\mathbf{r} [u_0(\mathbf{r}')/u_0(\mathbf{r})] [\Delta \varepsilon(\mathbf{r}')/\langle \varepsilon \rangle] e^{-jk|\mathbf{r}-\mathbf{r}'|} |\mathbf{r}-\mathbf{r}'| \quad (6.5.14)$$

Higher order approximation can be obtained by integrating (6.5.8) and (6.5.9). The function $u_0(\mathbf{r})$ is the incident wave. It can be a plane wave, a spherical wave, or even a beam wave, depending on the source and boundary conditions of the problem. In this section, we shall discuss the amplitude and phase fluctuations of a plane wave propagating in the positive z-direction. For this case

$$u_0(\mathbf{r}) = A_0 e^{-jkz} \tag{6.5.15}$$

and (6.5.14) becomes

$$\psi(\mathbf{r}) = (k^2/4\pi) \int d\mathbf{r} [\Delta \varepsilon(\mathbf{r}')/\langle \varepsilon \rangle] e^{-jk[(z'-z)+|\mathbf{r}-\mathbf{r}'|]}/|\mathbf{r}-\mathbf{r}'| \qquad (6.5.16)$$

where the x' and y' integration extend to $\pm \infty$ while the z' integration is from 0 to z.

In the case we are interested in, $\lambda \ll l$. As discussed in the last section, the angle of scattering by the irregularities is of the order of λ/l which is very small. Therefore we can make the so-called "forward scattering" assumption under which the contribution of the scattered field at an observation point **r** comes mainly from the scattering from irregularities in a small cone with vertex at **r**, with axis directed towards the scatterer, and with aperture $\theta_0 = \lambda/l_0 \ll 1$. Therefore, in the integrand of (6.5.16), we can approximate

 $|\mathbf{r}-\mathbf{r}'|\simeq z-z'$

in the denominator and

$$|\mathbf{r} - \mathbf{r}'| = (z - z')(1 + [(x - x')^2 + (y - y')^2]/(z - z')^2)^{1/2}$$

$$\cong (z - z') + [(x - x')^2 + (y - y')^2]/(z - z')$$

in the exponential. With these approximations, (6.5.16) becomes

$$\psi(\mathbf{r}) \simeq (k^2/4\pi) \int d\mathbf{r}' [\Delta \varepsilon(\mathbf{r}')/\langle \varepsilon \rangle] e^{-jk[(x-x')^2 + (y-y')^2]/2(z-z')}/(z-z') \quad (6.5.17)$$

To study the fluctuations of the phase and amplitudes of the plane wave,

6. Wave Propagation in Random Media

we take the real and imaginary parts of (6.5.17). We have

$$\chi(\mathbf{r}) = \operatorname{Re} \psi(\mathbf{r})$$

= $(k^2/4\pi) \int d\mathbf{r}' \cos\{k[(x-x')^2 + (y-y')^2]/2(z-z')\}$
× $[\Delta \varepsilon(\mathbf{r}')/\langle \varepsilon \rangle]/(z-z')$ (6.5.18)

$$S_{1}(\mathbf{r}) = \operatorname{Im} \psi(\mathbf{r})$$

$$= (k^{2}/4\pi) \int d\mathbf{r}' \sin\{k[(x-x')^{2}+(y-y')^{2}]/2(z-z')\}$$

$$\times [\Delta \varepsilon(\mathbf{r}')/\langle \varepsilon \rangle]/(z-z') \qquad (6.5.19)$$

The autocorrelations for χ and S_1 are then easily formed from (6.5.18) and (6.5.19):

$$\varrho_{\chi}(\mathbf{r}_{1},\mathbf{r}_{2}) = \langle |\Delta \varepsilon / \langle \varepsilon \rangle |^{2} \rangle \int_{0}^{z_{1}} \int_{0}^{z_{2}} \int_{-\infty}^{+\infty} d\mathbf{r}_{1}' d\mathbf{r}_{2}' \varrho_{\varepsilon}(\mathbf{r}_{1}'-\mathbf{r}_{2}') \\ \times \Phi_{2}(\varrho_{1},z_{1}-z_{1}') \Phi_{2}(\varrho_{2},z_{2}-z_{2}')$$
(6.5.20)

$$\varrho_{\mathcal{S}}(\mathbf{r}_{1},\mathbf{r}_{2}) = \langle | \Delta \varepsilon / \langle \varepsilon \rangle |^{2} \rangle \int_{0}^{z_{1}} \int_{0}^{z_{2}} \int_{-\infty}^{+\infty} d\mathbf{r}_{1}' d\mathbf{r}_{2}' \, \varrho_{\varepsilon}(\mathbf{r}_{1}'-\mathbf{r}_{2}') \\ \times \Phi_{1}(\varrho_{1},z-z_{1}') \Phi_{1}(\varrho_{2},z_{2}-z_{2}')$$
(6.5.21)

where

$$\Phi_{1}(\varrho, z - z') = \frac{1}{4\pi} \frac{1}{z - z'} \sin[\varrho^{2}/2(z - z')]$$

$$\Phi_{2}(\varrho, z - z') = \frac{1}{4\pi} \frac{1}{z - z'} \cos[\varrho^{2}/2(z - z')]$$

$$\varrho^{2} = (x - x')^{2} + (y - y')^{2}$$
(6.5.22)

and

$$\langle |\Delta \varepsilon / \langle \varepsilon \rangle |^2 \rangle \varrho_{\varepsilon}(\mathbf{r} - \mathbf{r}') = \langle [\Delta \varepsilon (\mathbf{r}) / \langle \varepsilon \rangle] [\Delta \varepsilon (\mathbf{r}') / \langle \varepsilon \rangle] \rangle$$

In (6.5.20) and (6.5.21) we have normalized distance with respect to the wavelength such that $kx \rightarrow x$, $ky \rightarrow y$, $kz \rightarrow z$, and $k^3 dx dy dz \rightarrow dx dy dz$.

To evaluate (6.5.20) and (6.5.21), let us introduce relative coordinates

$$\xi = x_1' - x_2', \quad \eta = y_1' - y_2', \quad \zeta = z_1' - z_2' \quad (6.5.23)$$

and center of mass coordinates

$$X = (x_1' + x_2')/2, \qquad Y = (y_1' + y_2')/2, \qquad Z = (z_1' + z_2')/2 \qquad (6.5.24)$$

Then

$$\varrho_1 = [(\frac{1}{2}\xi + X - x_1)^2 + (\frac{1}{2}\eta + Y - y_1)^2]^{1/2}
\varrho_2 = [(\frac{1}{2}\xi - X + x_1)^2 + (\frac{1}{2}\eta - Y + y_1)^2]^{1/2}$$
(6.5.25)

For the case where the observation point is at (0, 0, L) such that $x_1 = y_1 = 0$ and $z_1 = L$, (6.5.20) and (6.5.21) become the mean square fluctuations of amplitudes and phase of the wave at z = L, respectively. We have

$$\langle \chi^{2} \rangle = \langle | \Delta \varepsilon / \langle \varepsilon \rangle |^{2} \rangle$$

$$\times \int_{0}^{L} \int_{0}^{L} \int_{-\infty}^{+\infty} d\xi \, d\eta \, dX \, dY \, dz_{1}' \, dz_{2}'$$

$$\times \Phi_{1} \{ [(X + \xi/2)^{2} + (Y + \eta/2)^{2}]^{1/2}, L - z_{1}' \}$$

$$\times \Phi_{1} \{ [(X - \xi/2)^{2} + (Y - \eta/2)^{2}]^{1/2}, L - z_{2}' \} \cdot \varrho_{s} (| \mathbf{r}_{1}' - \mathbf{r}_{2}' |)$$

$$(6.5.26)$$

$$\begin{aligned} \langle S_{1}^{2} \rangle &= \langle |\Delta \varepsilon / \langle \varepsilon \rangle |^{2} \rangle \int_{0}^{L} \int_{0}^{L} \int_{-\infty}^{+\infty} d\xi \, d\eta \, dX \, dY \, dz_{1}' \, dz_{2}' \\ &\times \Phi_{2} \{ [(X + \xi/2)^{2} + (\bar{Y} + \eta/2)^{2}]^{1/2}, \, L - z_{1}' \} \\ &\times \Phi_{2} \{ [(X - \xi/2)^{2} + (Y - \eta/2)^{2}]^{1/2}, \, L - z_{2}' \} \varrho_{\varepsilon} (|\mathbf{r}_{1}' - \mathbf{r}_{2}' |) \end{aligned}$$

$$(6.5.27)$$

Since ρ_{ϵ} is not a function of X and Y, the X, Y integration of (6.5.26) and (6.5.27) can be integrated immediately to yield

$$\langle \chi^2 \rangle = \frac{1}{2} \langle | \Delta \varepsilon / \langle \varepsilon \rangle |^2 \rangle (I_1 - I_2)$$
 (6.5.28)

$$\langle S_1^2 \rangle = \frac{1}{2} \langle | \Delta \varepsilon / \langle \varepsilon \rangle |^2 \rangle (I_1 + I_2)$$
(6.5.29)

where

$$I_{1} = \int_{0}^{L} \int_{0}^{L} \int_{-\infty}^{+\infty} dz_{1}' dz_{2}' d\xi d\eta \Phi_{1}[(\xi^{2} + \eta^{2})^{1/2}, z_{1}' - z_{2}']\varrho_{s}(|\mathbf{r}_{1}' - \mathbf{r}_{2}'|)$$

$$I_{2} = \int_{0}^{L} \int_{0}^{L} \int_{0}^{+\infty} dz_{1}' dz_{2}' d\xi d\eta \Phi_{1}[(\xi^{2} + \eta^{2})^{1/2}, 2L - (z_{1}' + z_{2}')]$$
(6.5.30)

$$\times \varrho_{\epsilon}(|\mathbf{r}_{1}'-\mathbf{r}_{2}'|)$$
 (6.5.31)

In deriving (6.5.28) and (6.5.29), the formula

$$\iint_{-\infty}^{+\infty} \Phi_{1}[((X + \xi/2)^{2} + (Y + \eta/2)^{2})^{1/2}, L - z_{1}'] \\ \times \Phi_{1}[((X - \xi/2)^{2} + (Y - \eta/2)^{2})^{1/2}, L - z_{2}'] dX dY \\ = \frac{1}{2} \{ \Phi_{1}[(\xi^{2} + \xi^{2})^{1/2}, z_{1}' - z_{2}'] + \Phi_{1}[(\xi^{2} + \eta^{2})^{1/2}, 2L - (z_{1}' + z_{2}')] \}$$
(6.5.32)

and

$$\iint_{-\infty}^{+\infty} \Phi_2[((X + \xi/2)^2 + (Y + \eta/2)^2)^{1/2}, L - z_1'] \\ \times \Phi_2[((X - \xi/2)^2 + (Y - \eta/2)^2)^{1/2}, L - z_2'] dX dY \\ = \frac{1}{2} \{ \Phi_1[(\xi^2 + \eta^2)^{1/2}, z_1' - z_2'] - \Phi_1[(\xi^2 + \eta^2)^{1/2}, 2L - (z_1' + z_2')] \}$$
(6.5.33)

have been used.

 I_1 and I_2 can be simplified further by changing the variables z_1' and z_2' to ζ and Z. Since ϱ_e is not a function of Z, we have

$$I_{1} = \int_{0}^{L} dZ \iint_{-\infty}^{+\infty} \Phi_{1}[(\xi^{2} + \eta^{2})^{1/2}, \zeta] \varrho_{\varepsilon}(\xi, \eta, \zeta) d\xi d\eta d\zeta$$
$$= \frac{L}{4\pi} \iint_{-\infty}^{+\infty} \frac{1}{\zeta} \sin[(\xi^{2} + \eta^{2})/2\zeta] \varrho_{\varepsilon}(\xi, \eta, \zeta) d\xi d\eta d\zeta \quad (6.5.34)$$

and

$$I_{2} = \int_{0}^{L} dZ \, \Phi_{1}[(\xi^{2} + \eta^{2})^{1/2}, 2(L - Z)] \iint_{-\infty}^{+\infty} \varrho_{\varepsilon}(\xi, \eta, \zeta) \, d\xi \, d\eta \, d\zeta \qquad (6.5.35)$$

But

$$\int_{0}^{L} \Phi_{1}[(\xi^{2} + \eta^{2})^{1/2}, 2(L - Z)] dZ$$

= $\frac{1}{4\pi} \int_{0}^{L} \frac{1}{2(L - Z)} \sin\left[\frac{\xi^{2} + \eta^{2}}{2(2L - 2Z)}\right] dZ$
= $\frac{1}{8\pi} \int_{(\xi^{2} + \eta^{2})/4L}^{\infty} \frac{\sin t}{t} dt = -\frac{1}{8\pi} \operatorname{Si}[(\xi^{2} + \eta^{2})/4L]$ (6.5.36)

where

$$\operatorname{Si}(z) = -\int_{z}^{\infty} \frac{\sin t}{t} dt \qquad (6.5.37)$$

is the sine integral.

Therefore (6.5.35) becomes

$$I_{2} = -\frac{1}{8\pi} \iint_{-\infty}^{+\infty} \operatorname{Si}[(\xi^{2} + \eta^{2})/4L] \varrho_{\epsilon}(\xi, \eta, \zeta) \, d\xi \, d\eta \, d\zeta \quad (6.5.38)$$

For the random medium with an isotropic random field, $\varrho_{\epsilon}(\xi, \eta, \zeta)$ is a function of $r' = (\xi^2 + \eta^2 + \zeta^2)^{1/2}$ only. Therefore (6.5.34) and (6.5.38) can be integrated over the angular dependence in a cylindrical coordinate system $(\varrho, \theta, \varphi)$. We obtain

$$I_1 = \frac{1}{2} L \int_0^\infty d\zeta \int_0^\infty \frac{1}{\zeta} \sin(\varrho^2/2\zeta) \varrho_\varepsilon(r') \varrho \, d\varrho \qquad (6.5.39)$$

$$I_2 = -\frac{1}{4} \int_0^\infty d\zeta \int_0^\infty \operatorname{Si}(\varrho^2/4L) \varrho_{\varepsilon}(r') \varrho \, d\varrho \qquad (6.5.40)$$

The inner integral of (6.5.39) can be approximated further in the following manner. Let $q = \rho^2/2\zeta$; then by integration by parts,

$$\int_{0}^{\infty} (1/\zeta) \sin(\varrho^{2}/2\zeta)\varrho_{\varepsilon}(r')\varrho \,d\varrho$$

$$= \int_{0}^{\infty} (\sin q)\varrho_{\varepsilon}(r') \,dq = -(\cos q)\varrho_{\varepsilon}(r') \Big]_{0}^{\infty} + \int_{0}^{\infty} \cos q \,\frac{\partial \varrho_{\varepsilon}(r')}{\partial q} \,dq$$

$$= \varrho_{\varepsilon}(0, 0, \zeta) + \sin q \,\frac{\partial \varrho_{\varepsilon}}{\partial q} \Big]_{0}^{\infty} - \int_{0}^{\infty} \sin q \,\frac{\partial^{2} \varrho_{\varepsilon}(r')}{\partial q^{2}} \,dq$$

$$= \varrho_{\varepsilon}(0, 0, \zeta) - \int_{0}^{\infty} \sin q \,\frac{\partial^{2} \varrho_{\varepsilon}(r')}{\partial q^{2}} \,dq \qquad (6.5.41)$$

The integrand is proportional to $(1/l^2)\varrho_e$ where *l* is the correlation length. Since the size of the irregularities is assumed to be much greater than the wavelength, the normalized correlation length $l \gg 1$. Therefore the integral in (6.5.41) can be neglected as compared to $\varrho_e(0, 0, \zeta)$. With this approximation, (6.5.39) becomes

$$I_1 \simeq \frac{1}{2} L \int_0^\infty d\zeta \, \varrho_\varepsilon(0, 0, \zeta) \tag{6.5.42}$$

Substituting (6.5.40) and (6.5.42) into (6.5.28) and (6.5.29), we have the

339

6. Wave Propagation in Random Media

fluctuations in amplitudes and phase at the point (0, 0, L):

$$\langle \chi^2 \rangle = \frac{1}{4} \langle | \Delta \varepsilon / \langle \varepsilon \rangle |^2 \rangle L \left\{ \int_0^\infty d\zeta \Big[\varrho_\varepsilon(0, 0, \zeta) + \frac{1}{2L} - \int_0^\infty \operatorname{Si}(\varrho^2 / 4L) \varrho_\varepsilon(r') \varrho \, d\varrho \Big] \right\}$$
(6.5.43)

$$\langle S_1^2 \rangle = \frac{1}{4} \langle |\Delta \varepsilon / \langle \varepsilon \rangle |^2 \rangle L \left\{ \int_0^\infty d\zeta \Big[\varrho_\varepsilon(0, 0, \zeta) - \frac{1}{2L} \int_0^\infty \operatorname{Si}(\varrho^2 / 4L) \varrho_\varepsilon(r') \varrho \, d\varrho \Big] \right\}$$
(6.5.44)

If we change back to the original length, we have

$$\langle \chi^2 \rangle = \frac{1}{4} \langle |\Delta \varepsilon / \langle \varepsilon \rangle |^2 \rangle k^2 L \left\{ \int_0^\infty d\zeta \Big[\varrho_\varepsilon(0, 0, \zeta) + \int_0^\infty \operatorname{Si}(t) \varrho_\varepsilon(r') dt \Big] \right\}$$
(6.5.45)

$$\langle S_1^2 \rangle = \frac{1}{4} \langle |\Delta \varepsilon / \langle \varepsilon \rangle |^2 \rangle k^2 L \left\{ \int_0^\infty d\zeta \Big[\varrho_\varepsilon(0,0,\zeta) - \int_0^\infty \operatorname{Si}(t) \varrho_\varepsilon(r') dt \Big] \right\}$$
(6.5.46)

where

$$t = k\varrho^2/4L \tag{6.5.47}$$

We now introduce a wave parameter defined by

$$D = 4L/kl^2 \tag{6.5.48}$$

where *l* is the correlation length of $\varrho_{\epsilon}(r')$.

First, let us consider the case for which $D \gg 1$; this is called the Fraunhofer diffraction region. Since the major contribution to the integral

$$I = \int_0^\infty \operatorname{Si}(t) \varrho_\varepsilon(r') \, dt$$

comes from $\varrho \leq l$ (since the correlation will decrease rapidly for distance greater than l), the corresponding value of t is $t \leq kl^2/4L = 1/D \ll 1$. Therefore Si(t) in the integrand can be approximated by

$$\operatorname{Si}(t) \sim -\pi/2$$
 for $t \ll 1$

The integral becomes

$$-(\pi/2)\int_0^\infty \varrho_\epsilon(r')\,dt \sim (1/D)\varrho_\epsilon(0,0,\zeta) \ll \varrho_\epsilon(0,0,\zeta)$$

which can be neglected as compared to the first term in (6.5.45) and (6.5.46).

6.5 Fluctuation of Electromagnetic Waves in Random Media

Hence for this case, (6.5.45) and (6.5.46) reduce to

$$\langle \chi^2 \rangle = \langle S_1^2 \rangle \simeq \frac{1}{4} \langle |\Delta \varepsilon / \langle \varepsilon \rangle |^2 \rangle k^2 L \int_0^\infty d\zeta \, \varrho_\varepsilon(0, 0, \zeta), \qquad D \gg 1 \qquad (6.5.49)$$

For a Gaussian correlation $\varrho_{\varepsilon}(0, 0, \zeta) = e^{-\zeta^2/l^2}$,

$$\langle \chi^2 \rangle = \langle S_1^2 \rangle = \sqrt{\pi/8} \langle |\Delta \varepsilon / \langle \varepsilon \rangle |^2 \rangle k^2 Ll, \quad D \gg 1 \quad (6.5.50)$$

The amplitude and phase fluctuations are the same and both increase with the distance the wave has traversed in the medium.

Next, we consider the other extreme case for which $D \ll 1$. This is the case where $(4L/k) \ll l^2$ which corresponds to the range of applicability of geometrical optics. For this case, the diffractional effect is not important. We integrate the integral I by parts

$$I = \left\{ [t \operatorname{Si}(t) + \cos t] \varrho_{\varepsilon}(r') \right\}_{0}^{\infty} - \int_{0}^{\infty} [t \operatorname{Si}(t) + \cos t] \left(\partial \varrho_{\varepsilon}(r') / \partial t \right) dt$$
$$= -\varrho_{\varepsilon}(0, 0, \zeta) - \int_{0}^{\infty} [t \operatorname{Si}(t) + \cos t] \left(\partial \varrho_{\varepsilon}(r') / \partial t \right) dt$$

The derivative $\partial \varrho_{\varepsilon} / \partial t$ is of the order $D \varrho_{\varepsilon}(0, 0, \zeta)$; therefore the integral

$$\int_{0}^{\infty} [t \operatorname{Si}(t) + \cos t] (\partial \varrho_{\epsilon}(r')/\partial t) dt$$

$$\simeq \int_{0}^{1/D} [t \operatorname{Si}(t) + \cos t] (\partial \varrho_{\epsilon}(r')/\partial t) dt$$

$$\leq D \varrho_{\epsilon}(0, 0, \zeta) \int_{0}^{1/D} [t \operatorname{Si}(t) + \cos t] dt$$

$$= D \varrho_{\epsilon}(0, 0, \zeta) \frac{1}{2} [t^{2} \operatorname{Si}(t) + t \cos t + \sin t]_{0}^{1/D}$$

$$= (D/2) \varrho_{\epsilon}(0, 0, \zeta) [(1/D^{2}) \operatorname{Si}(1/D) + (1/D) \cos(1/D) + \sin(1/D)]$$

(6.5.51)

Since $(1/D) \gg 1$, the asymptotic expansion for Si(x) can be used. We have

$$\operatorname{Si}(x) = -\frac{\cos x}{x} - \frac{\sin x}{x^2} + O\left(\frac{1}{x^3}\right)$$

Therefore (6.5.51) becomes

$$\int_0^\infty \left[t \operatorname{Si}(t) + \cos t\right] \left(\partial \varrho_{\varepsilon}(r') / \partial t\right) dt \le D^2 \varrho_{\varepsilon}(0, 0, \zeta) \qquad (6.5.52)$$

Hence we have

$$I \simeq -\varrho_{\epsilon}(0, 0, \zeta) \quad \text{for} \quad D \ll 1 \quad (6.5.53)$$

The fluctuation in phase can now be written as

$$\langle S_1^2 \rangle \simeq \frac{1}{2} \langle |\Delta \varepsilon / \langle \varepsilon \rangle |^2 \rangle k^2 L \int_0^\infty d\zeta \varrho_\varepsilon(0, 0, \zeta), \qquad D \ll 1 \qquad (6.5.54)$$

Comparing (6.5.54) with (6.4.26), we see that they are identical. Thus the result agrees with that derived from geometric optics. This is expected since the diffractional effect is neglected in deriving (6.5.54).

As for the amplitude fluctuation, we see that if (6.5.53) is substituted into (6.5.45), $\langle \chi^2 \rangle = 0$. We must then compute the integral *I* to a higher order to obtain a nontrivial expression. This can be done in a manner similar to that outlined above. We will only write the result

$$\langle \chi^2 \rangle = \frac{1}{12} \langle | \Delta \varepsilon / \langle \varepsilon \rangle |^2 \rangle L^3 \int_0^\infty \left[\nabla_T^2 \nabla_T^2 \varrho_\varepsilon \right]_{\xi=\eta=0} d\zeta \qquad (6.5.55)$$

where $\nabla_T^2 = \partial^2/\partial\xi^2 + \partial^2/\partial\eta^2$ is the transverse Laplacian.

For the intermediate range of the wave parameter, no general expressions can be obtained for $\langle S_1^2 \rangle$ and $\langle \chi^2 \rangle$ other than those shown in (6.5.45) and (6.5.46). However, for a Gaussian correlation function

$$\varrho_{s}(\xi,\eta,\zeta) = e^{-(\varrho^{3}+\zeta^{2})/l^{2}}$$
(6.5.56)

we can evaluate $\langle \chi^2 \rangle$ and $\langle S^2 \rangle$ explicitly. They are

$$\langle \chi^2 \rangle = \sqrt{\pi/8} \langle |\Delta \varepsilon / \langle \varepsilon \rangle |^2 \rangle k^2 l L [1 - (1/D) \tan^{-1} D]$$
 (6.5.57)

$$\langle S^2 \rangle = \sqrt{\pi/8} \langle |\Delta \varepsilon / \langle \varepsilon \rangle |^2 \rangle k^2 l L [1 + (1/D) \tan^{-1} D] \qquad (6.5.58)$$

Thus, we have derived the amplitude and phase fluctuations of a plane wave propagating in a homogeneous, isotropic random medium. The fluctuations are expressed in terms of the autocorrelation of the dielectric permittivity function of the medium. Similar computations can be made for the case where the medium is statistically homogeneous but anisotropic. Furthermore, the derivation can be extended to cover the cases for spherical incident wave or beam wave propagations. The comparison of the amplitude fluctuations for the three cases are shown in Fig. 6.5-1.

From (6.5.50), (6.5.54), (6.5.55), (6.5.57), and (6.5.58), we see that the

342



Fig. 6.5-1. Comparison of the mean square amplitude fluctuations for plane, spherical, and beam waves.

fluctuations of the wave increase without limit as the distance the wave has traveled in the medium, L, increases. From experiment, however, the fluctuations always seem to level off to some saturation value for some finite value of L. This apparent discrepancy between theory and experiment comes about because of the inadequacy of our model. The computations we have made are still based on just a single-scattering model. Since saturation occurs mainly because of the interference between multiple scattered waves, our model can not explain this phenomenon. Some recent work in which multiple scattering is taken into account shows good agreement between theory and experiment (DeWolf, 1968; Tatarskii, 1966).

One way of approaching the problem is to start from (6.5.58) and (6.5.59) and iterate them to the next higher order. The computation becomes very involved and is beyond the scope of this book.

6.6 Correlations of Fluctuations and Application to the Ionosphere

In the last section, we derived the mean square fluctuations of the phase and amplitude of an electromagnetic wave propagating in a random medium. The next step in our plan to understand the statistical characteristics of the wave field is to study the various correlation functions of the amplitudes and the phase. By studying the dependence of these correlation functions on the properties of the medium, more information about the medium itself can be obtained. There are many possible correlations between the phase and amplitudes, and phase or amplitudes themselves. Only some of them will be discussed here. The main purpose is to show the general method of approach, as well as to make use of the results in discussing some practical problems.

Instead of treating the problem for a plane wave, we shall assume a spherical incident wave. This on the one hand is closer to real experimental situations and on the other hand illustrates some new techniques in the computation. The starting point is (6.5.14) of the last section. The geometry of the problem is depicted in Fig. 6.6-1. The origin of the coordinates is at the transmitter T, which emits a spherical wave



$$u_0 = A_0 e^{-jkr}/r \tag{6.6.1}$$

The receiver is at B. From (6.5.14), the scattered field at B is

$$\psi(\mathbf{r}) = -\frac{k^2}{4\pi} \int d\mathbf{r}' \frac{\Delta \varepsilon(\mathbf{r}')}{\langle \varepsilon \rangle} \frac{r}{Rr'} e^{-jk(R+r'-r)}$$
(6.6.2)

where the integration is over the whole region where the random irregularities exist. Note that the factor k(R + r' - r) in the exponential is the phase difference between the direct path *TB* and the scattering path *TSB*. The phase and amplitude of the scattered wave are then given by

$$S_{1}(\mathbf{r}) = \frac{rk^{2}}{4\pi} \int d\mathbf{r}' \frac{\Delta \varepsilon(\mathbf{r}')}{\langle \varepsilon \rangle} \frac{1}{Rr'} \sin[k(R+r'-r)] \qquad (6.6.3)$$

$$\chi(\mathbf{r}) = \frac{rk^2}{4\pi} \int d\mathbf{r}' \frac{\Delta\varepsilon(\mathbf{r}')}{\langle\varepsilon\rangle} \frac{1}{Rr'} \cos[k(R+r'-r)] \qquad (6.6.4)$$

The autocorrelations at two points \mathbf{r}_1 and \mathbf{r}_2 are easily formed from (6.6.3) and (6.6.4)

$$\varrho_{S}(\mathbf{r}_{1}, \mathbf{r}_{2}) = \langle S_{1}(\mathbf{r}_{1})S_{1}(\mathbf{r}_{2})\rangle = \frac{r_{1}r_{2}k^{4}}{(4\pi)^{2}} \left\langle \left| \frac{\Delta\varepsilon}{\langle\varepsilon\rangle} \right|^{2} \right\rangle \iint d\mathbf{r}_{1}' d\mathbf{r}_{2}' \varrho_{\varepsilon}(\mathbf{r}_{1}' - \mathbf{r}_{2}') \right. \\ \left. \times \frac{\sin[k(R_{1} + r_{1}' - r_{1})]}{r_{1}'R_{1}} \frac{\sin[k(R_{2} + r_{2}' - r_{2})]}{r_{2}'R_{2}}$$
(6.6.5)

$$\varrho_{\chi}(\mathbf{r}_{1},\mathbf{r}_{2}) = \langle \chi(\mathbf{r}_{1})\chi(\mathbf{r}_{2}) \rangle = \frac{r_{1}r_{2}k^{4}}{(4\pi)^{2}} \left\langle \left| \frac{\Delta\varepsilon}{\langle\varepsilon\rangle} \right|^{2} \right\rangle \int \int d\mathbf{r}_{1}' d\mathbf{r}_{2}' \varrho_{\varepsilon}(\mathbf{r}_{1}' - \mathbf{r}_{2}') \right\rangle \\ \times \frac{\cos[k(R_{1} + r_{1}' - r_{1})]}{r_{1}'R_{1}} \frac{\cos[k(R_{2} + r_{2}' - r_{2})]}{r_{2}'R_{2}}$$
(6.6.6)

Let us consider the case $\mathbf{r}_1 = (-d/2, 0, L)$, $\mathbf{r}_2 = (d/2, 0, L)$. For $l \gg \lambda$, the forward-scattering approximation can be applied. Under this approximation, in the denominator $r' \simeq z'$, R = z - z' while in the phase

$$R_1 + r_1' - r_1 = [y_1'^2 + (x_1' - dz_1'/2L)^2]/2q_1'$$
(6.6.7)

$$R_2 + r_2' - r_2 = [y_2'^2 + (x_2' + dz_2'/2L)^2]/2q_2'$$
(6.6.8)

where $q_i' = z_i'(L - z_i')/L$, i = 1, 2.

Define two integrals:

$$I_{1} = \frac{4\pi}{\langle |\Delta \varepsilon / \langle \varepsilon \rangle |^{2} \rangle} \left[\varrho_{S}(\mathbf{r}_{1}, \mathbf{r}_{2}) + \varrho_{\chi}(\mathbf{r}_{1}, \mathbf{r}_{2}) \right]$$
(6.6.9)

$$I_{2} = \frac{4\pi}{\langle |\Delta \varepsilon / \langle \varepsilon \rangle |^{2} \rangle} \left[\varrho_{S}(\mathbf{r}_{1}, \mathbf{r}_{2}) - \varrho_{\chi}(\mathbf{r}_{1}, \mathbf{r}_{2}) \right]$$
(6.6.10)

Substituting (6.5.5) and (6.6.6) into (6.6.9) and (6.6.10), and normalizing distance with respect to the wavelength, we obtain

$$I_{1} = \frac{1}{\pi} \iint \frac{\varrho_{\varepsilon}(\mathbf{r}_{1}' - \mathbf{r}_{2}')}{4q_{1}'q_{2}'} \times \cos\left[\frac{y_{1}'^{2} + (x_{1}' - dz_{1}'/2L)^{2}}{2q_{1}'} - \frac{y_{2}'^{2} + (x_{2}' + dz_{2}'/2L)^{2}}{2q_{2}'}\right] d\mathbf{r}_{1}' d\mathbf{r}_{2}'$$

$$I_{2} = \frac{-1}{\pi} \iint \frac{\varrho_{\varepsilon}(\mathbf{r}_{1}' - \mathbf{r}_{2}')}{4q_{1}'q_{2}'} \times \cos\left[\frac{y_{1}'^{2} + (x_{1}' - dz_{1}'/2L)^{2}}{2q_{1}'} + \frac{y_{2}'^{2} + (x_{2}' + dz_{2}'/2L)^{2}}{2q_{2}'}\right] d\mathbf{r}_{1}' d\mathbf{r}_{2}'$$

$$(6.6.12)$$

345
Using the relative coordinates defined in (6.5.23) and center of mass coordinates defined in (6.5.24), the X and Y integration can be carried out first without any difficulty.

$$I_{1} = \int_{Z-a}^{a+b} \int_{\xi=-b}^{b} \int_{\eta=-\infty}^{+\infty} \int_{\xi=-\infty}^{+\infty} \frac{\varrho_{\epsilon}(\mathbf{r}_{1}'-\mathbf{r}_{2}')}{2(q_{1}'-q_{2}')} \sin\left[\frac{\eta^{2}+(\xi+dZ/L)^{2}}{2(q_{1}'-q_{2}')}\right] \\ \times d\xi \, d\eta \, d\zeta \, dZ \tag{6.6.13}$$

$$I_{2} = \int_{Z-a}^{a+b} \int_{\xi=-b}^{b} \int_{\eta=-\infty}^{+\infty} \int_{\xi=-\infty}^{+\infty} \frac{\varrho_{\epsilon}(\mathbf{r}_{1}'-\mathbf{r}_{2}')}{2(q_{1}'+q_{2}')} \sin\left[\frac{\eta^{2}+(\xi+dZ/L)^{2}}{2(q_{1}'+q_{2}')}\right]$$

$$\times d\xi d\eta d\zeta dZ \tag{6.6.14}$$

where the limits of integration correspond to the configuration shown in Fig. 6.6-1.

Further integration of I_1 and I_2 depends on the explicit expression of the autocorrelation $\rho_{\epsilon}(\mathbf{R})$. In the following, we shall assume a Gaussian correlation of the form shown in (6.5.56):

$$\varrho_{\epsilon}(\mathbf{r}_{1}' - \mathbf{r}_{2}') = e^{-(\xi^{2} + \eta^{2} + \zeta^{2})/l^{2}}$$
(6.5.56)

Substituting (6.5.56) into (6.6.13) and (6.6.14) and carrying out the ξ and η integration, we obtain

$$I_{1} = \operatorname{Im} \pi \int_{a}^{a+b} dZ \int_{-b}^{b} d\zeta \frac{\exp\left\{-\frac{(dZ/L)^{2}}{l^{2} + j2\zeta(2Z/L-1)} - \frac{\zeta^{2}}{l^{2}}\right\}}{[-2\zeta(1-2Z/L)/l^{2}] - j} \quad (6.6.15)$$

$$I_{2} = \operatorname{Im} \pi \int_{a}^{a+b} dZ \int_{-b}^{b} d\zeta \frac{\exp\left\{-\frac{(dZ/L)^{2}}{jl^{2}(D-\zeta^{2}/Ll^{2}-j)} - \frac{\zeta^{2}}{l^{2}}\right\}}{D-\zeta^{2}/Ll^{2}-j} \quad (6.6.16)$$

where Im indicates imaginary part and

$$D = 4Z(L-Z)/Ll^2 (6.6.17)$$

is the equivalent wave parameter.

The limits for the ζ integration can be extended to $\pm \infty$ since $b \gg l$. Also since $l \gg 1$, terms like $2\zeta(2Z/L-1)/l^2$ and ζ^2/Ll^2 in the integrand can be neglected compared to unity. With these simplifications, we have

$$I_1 = \frac{\pi^2}{2} - \frac{Ll^2}{d} \left\{ \text{erf}[(d/lL)(a+b)] - \text{erf}[(d/lL)a] \right\}$$
(6.6.18)

$$I_2 = \pi^{3/2} l \operatorname{Im} \int_a^{a+b} \frac{\exp\{-d^2 Z^2 / L^2 l^2 (1+jD)\}}{(D-j)} dZ \qquad (6.6.19)$$

If now we assume that the slab of irregularities is thin such that $a \gg b$, then D in (6.6.19) can be taken as a constant average value \overline{D} . In this approximation, I_2 can be integrated.

$$I_{2} = (\pi^{2}l^{2}L/2d) \times \operatorname{Im}\left\{j(1+j\bar{D})^{-1/2}\left[\operatorname{erf}\left(\frac{(a+b)d}{Ll(1+j\bar{D})^{1/2}}\right) - \operatorname{erf}\left(\frac{ad}{Ll(1+j\bar{D})^{1/2}}\right)\right]\right\}$$
(6.6.20)

The correlations for phase and amplitudes can now be calculated from (6.6.9), (6.6.10), (6.6.18), and (6.6.20). We have

$$\varrho_{S}(\mathbf{r}_{1},\mathbf{r}_{2}) = \frac{\langle |\Delta \varepsilon / \langle \varepsilon \rangle |^{2} \rangle}{8\pi} (I_{1} + I_{2})$$
(6.6.21)

$$\varrho_{\chi}(\mathbf{r}_{1},\mathbf{r}_{2}) = \frac{\langle |\Delta \varepsilon / \langle \varepsilon \rangle |^{2} \rangle}{8\pi} (I_{1} - I_{2})$$
(6.6.22)

Two limiting cases will now be considered.

(i) $\overline{D} \gg 1$. Expending (6.20) and keeping only the leading terms, we have for $d \leq lL/a$,

$$\varrho_{\mathcal{S}}(\mathbf{r}_{1},\mathbf{r}_{2}) = \varrho_{\chi}(\mathbf{r}_{1},\mathbf{r}_{2}) = (\pi^{1/2} \langle |\Delta \varepsilon / \langle \varepsilon \rangle |^{2} \rangle lb/8) e^{-a^{2} d^{2} l l^{2} L^{2}} \quad (6.6.23)$$

Both the phase and amplitude are initially Gaussian; the "scale" of the random waves is a factor (L/a) times the "scale" of the fluctuations in the dielectric permittivity.

(ii) $\bar{D} \ll 1$. Again keeping only leading terms, we have

$$\varrho_{\mathcal{S}}(\mathbf{r}_{1},\mathbf{r}_{2}) = (\pi^{1/2} \langle | \Delta \varepsilon / \langle \varepsilon \rangle |^{2} \rangle lb/4) e^{-a^{2}d^{2}/l^{2}L^{2}}$$
(6.6.24)

and

$$\varrho_{\chi}(\mathbf{r}_{1},\mathbf{r}_{2}) = (\pi^{1/2} \langle |\Delta \varepsilon \rangle \langle \varepsilon \rangle |^{2} \rangle l b \bar{D}^{2} / 8) \cdot [1 - (2a^{2} d^{2} / L^{2} l^{2}) + (a^{4} d^{4} / 2L^{4} l^{4})] \\ \times e^{-a^{2} d^{2} / l^{2} L^{2}}$$
(6.6.25)

The correlation in phase is Gaussian while the correlation in amplitude is not. However, for $(ad/Ll) \ll 1$, ϱ_{χ} also approaches the Gaussian. For intermediate values of \overline{D} , ϱ_{S} and ϱ_{χ} must be computed numerically.

We see immediately that from ρ_s and ρ_z , we can obtain information about the correlation length *l* about the medium which is closely related to the size of the irregularities.

 ϱ_s and ϱ_x calculated above are sometimes called the transverse correlations since they are correlations between two points in the plane z = constant. Longitudinal correlations for two points $(0, 0, z_1)$ and $(0, 0, z_2)$ can be computed in a similar manner. In addition, correlations between phase and amplitudes sometimes are also computed. The basic procedure for all these computations are the same. We shall not discuss them in detail. Also, the derivation we made was based on isotropic irregularities. The results can be generalized to cases where the irregularities are anisotropic (Yeh, 1962).

We now turn to the problem of application of the theory to various experimental situations. Specifically, let us consider the application to the ionosphere. In the ionosphere, at certain periods of time, there exist irregularities in electron density. Hey et al. (1946) first discovered that the intensity of the radiation from radio stars fluctuate on certain occasions in the VHF band. At first it was thought that this was due to the variation of the power output of the source. Subsequent spaced-receiver experiments by Smith (1950) and by Little and Lovell (1950) gave convincing proof that the cause of these fluctuations is in the ionosphere. With artificial satellites, more and more data have been recorded for the scintillation of radio signals passing through the ionosphere. It is now generally believed that there exist in the ionosphere blobs with excesses or deficiencies of electrons which scatter waves randomly. These irregularities in electron density are elongated along the earth's magnetic field lines. The mechanism for generating them is still uncertain. By studying the statistics of the fluctuating radio signals, it is hoped that we may obtain critical information about these irregularities and eventually understand the mechanism behind the phenomenon.

The first important statistical quantity is the mean square value. We shall define $(\langle S_1^2 \rangle)^{1/2}$ and $(\langle \chi^2 \rangle)^{1/2}$ as the scintillation indices of the phase and amplitudes, respectively. Measurements of scintillation index may yield information about the seasonal, diurnal, and regional variations of the irregularities. Furthermore, by measuring the scintillation index when the transmitter (satellite) is at different heights, it is possible to determine the height and the thickness of the irregularity slabs (Yeh, 1962).

Figure 6.6-2 indicates the configuration of an idealized satellite —spacedreceiver experiment. Signals received at station B_1 when the satellite is at position A_1 are correlated with signals received at station B_2 when the satellite is at position A_2 . It has been shown (Liu, 1965) by a similar procedure discussed earlier in this section that the correlations for both the amplitude and phase of the signals at B_1 and B_2 are maximum when the two paths A_1B_1 and A_2B_2 cross each other at the center of the irregular slab.

348

The correlation length of the wave is a factor (L/a) times that of the irregularities as discussed earlier. Therefore by properly choosing the satellite path, it is possible from this experiment to determine the height and the thickness of the slab as well as the size of the irregularities (McClure and Swenson, 1964).

Fig. 6.6-2. Geometry showing receivers at B_1 and B_2 . The idealized transmitting satellite moves along the dotted line. A maximum correlation is obtained when A_1B_1 and A_2B_2 crosses at the center of the slab of irregularities.



The formulas we derived are based on the assumption that the irregularities are imbedded in a homogeneous isotropic medium. In the ionosphere, however, the background electron density is not homogeneous and the Earth's magnetic field makes the medium anisotropic. Therefore, strictly speaking, the formulas derived are not applicable. Nevertheless, for high frequency signals, they are very good approximations. It has been shown that when the Earth magnetic field is taken into account, there will be some modifications to the formulas for mean square fluctuations and correlations. The new feature is that there will be depolarization in the scattered field. When the background electron density is taken as a function of height, it has been shown that the fluctuations of the waves are maxima when the irregular slab is at the height of maximum electron density.

6.7 Higher Order Approximations—Perturbation Techniques

In this section, we shall introduce some perturbation techniques to investigate the problem of electromagnetic waves propagating in random media beyond the scope of the single scattering model of Born's approximation. The main interest will be on the average wave itself, sometimes referred to as the theory of the propagation of coherent waves.

We start with a more general random medium characterized by a dielectric tensor $\boldsymbol{\epsilon}(\mathbf{r})$:

$$\boldsymbol{\varepsilon}(\mathbf{r}) = \varepsilon_0 [\mathbf{K}_0 + \mathbf{K}_1(\mathbf{r})] \tag{6.7.1}$$

where \mathbf{K}_0 is a constant tensor and \mathbf{K}_1 is a tensor with elements which are random functions of position. In particular, for the isotropic medium in the last four sections, $\mathbf{K}_0 = \langle K \rangle \mathbf{I}$ and $\mathbf{K}(r) = [\Delta \varepsilon(\mathbf{r})/\varepsilon_0]\mathbf{I}$ where \mathbf{I} is the identity matrix.

For the general medium, the wave equation can be written as

$$\mathbf{L} \cdot \mathbf{E}(\mathbf{r}) = \{ \nabla^2 \mathbf{I} - \nabla \nabla + k_0^2 [\mathbf{K}_0 + \mathbf{K}_1(\mathbf{r})] \} \cdot \mathbf{E}(\mathbf{r}) = j \omega \mu_0 \mathbf{J}(\mathbf{r}) \quad (6.7.2)$$

where $k_0 = \omega(\varepsilon_0 \mu_0)^{1/2}$ is the free space wave number and $\mathbf{J}(\mathbf{r})$ is the external current source. In an infinite medium, (6.7.2) plus the radiation condition determine the fields uniquely. We note that here we are treating the full vector wave equation. Equation (6.7.2) can be put into component form

$$(L_{0ij} + L_{1ij})E_j = j\omega\mu_0 J_i \tag{6.7.3}$$

where

$$L_{0ij} = \delta_{ij} \nabla^2 - (\nabla \nabla)_{ij} + k_0^2 K_{0ij}$$

$$L_{1ij} = k_0^2 K_{1ij}(\mathbf{r})$$
(6.7.4)

 L_0 is a deterministic operator while L_1 is a random one.

With the help of dyadic Green's function (Chapter 2), (6.7.3) can be written as an integral equation

$$E_{j}(\mathbf{r}) = E_{0j}(\mathbf{r}) + \int \Gamma_{jk}(\mathbf{r}, \mathbf{r}') L_{1kn}(\mathbf{r}') E_{n}(\mathbf{r}') d\mathbf{r}' \qquad (6.7.5)$$

where Γ is the dyadic Green's function for L_0 (2.14.2) satisfying $L_0 \cdot \Gamma = -I \, \delta(\mathbf{r} - \mathbf{r}')$ and

$$\mathbf{E}_{\mathbf{0}} = -j\omega\mu_{\mathbf{0}} \int \mathbf{\Gamma}(\mathbf{r},\mathbf{r}') \cdot \mathbf{J}(\mathbf{r}') \, d\mathbf{r}' \qquad (6.7.6)$$

Equation (6.7.5) is the starting point in most calculations for wave propagation in random media. For example, (6.3.7) and (6.5.14) are of this type.

Let us consider first for a moment the scalar wave equation in a random medium. For this case, (6.7.2) and (6.7.5) become, respectively,

$$(\nabla^2 + k_0^2)u(\mathbf{r}) + k_0^2 K_1(\mathbf{r})u(\mathbf{r}) = q$$
 (6.7.2a)

$$u(\mathbf{r}) = u_0(\mathbf{r}) + \int G_0(\mathbf{r}, \mathbf{r}') L_1(\mathbf{r}') u(\mathbf{r}') d\mathbf{r}' \qquad (6.7.5a)$$

where $u(\mathbf{r})$ is the scalar wave function, q is the source, and $G_0(\mathbf{r}, \mathbf{r}') = \exp[-jk_0 | \mathbf{r} - \mathbf{r}' |]/(4\pi | \mathbf{r} - \mathbf{r}' |)$ is the free space Green's function. In (6.7.5a)

$$u_0(\mathbf{r}) = -\int G_0(\mathbf{r}, \mathbf{r}')q(\mathbf{r}') d\mathbf{r}'$$

$$L_1(\mathbf{r}') = k_0^2 K_1(\mathbf{r}') = k_0^2 [\Delta \varepsilon(\mathbf{r})/\varepsilon_0]$$
(6.7.7)

Equation (6.7.5a) can be solved by iteration and we obtain

$$u(\mathbf{r}) = u_0(\mathbf{r}) + \int G_0(\mathbf{r}, \mathbf{r}') \sum_{n=1}^{\infty} \phi_n(\mathbf{r}') d\mathbf{r}' \qquad (6.7.8)$$

where

$$\phi_1(\mathbf{r}) = L_1(\mathbf{r})u_0(\mathbf{r})$$

$$\phi_{n+1}(\mathbf{r}) = L_1(\mathbf{r})\int G_0(\mathbf{r}, \mathbf{r}')\phi_n(\mathbf{r}') d\mathbf{r}'$$
(6.7.9)

It can be shown that if $|L_1| < M$ where M is some large positive number, then a sufficient condition for the convergence of the series solution (6.7.8) is

$$\frac{1}{2}Md^2 < 1$$
 (6.7.10)

where d is the upper bound of the dimension of the random region (Frisch, 1968).

We see that if $\frac{1}{2}Md^2 \ll 1$, the series (6.7.8) will converge rapidly and only the first term in the series will be needed to describe the field, the wellknown Born's approximation, and is exactly the formula we used in the previous sections. However, in an infinite random medium, even if $M \ll 1$, (6.7.10) will never be satisfied and the series becomes divergent. This divergence difficulty is intrinsic in this perturbation procedure and the series solution (6.7.8) breaks down when the wave is far away from the source. This is the reason why $\langle \chi^2 \rangle$ and $\langle S_1^2 \rangle$ increase without limit with L in the last two sections. Similar difficulty arises for the case of vector waves. To avoid this difficulty, we discuss next an alternate perturbation procedure, the so-called diagram technique.

The diagram method was first developed in quantum field theory and has been used successfully in nonequilibrium statistical mechanics. It was first introduced to the study of wave propagation in random media by Bourret followed by Tatarskii. [See reference list in the paper by Frisch (1968).] We shall develop this method for the vector wave equation for an isotropic background medium. For this case, the dyadic Green's function is given by (2.14.28)

$$\mathbf{\Gamma}(\mathbf{r},\mathbf{r}') = \mathrm{PV}\left[\mathbf{I} + \frac{1}{k_0^2} \nabla \nabla\right] \frac{e^{-jk_0R}}{4\pi R} - \frac{\mathbf{I}}{3k_0^2} \,\delta(\mathbf{r} - \mathbf{r}') \quad (6.7.11)$$

where $R = |\mathbf{r} - \mathbf{r}'|$ and PV means principle values will be taken in integrals involving that term. For simplicity, we have taken $\langle K \rangle = 1$. Also $\mathbf{L}_1 = \mathbf{I} k_0^2 K_1(\mathbf{r})$. We shall assume $K_1(\mathbf{r})$ to be statistically homogeneous and isotropic with zero mean. Therefore

$$\langle K_{\mathbf{i}} \rangle = 0, \qquad \varrho(|\mathbf{r} - \mathbf{r}'|) = \langle K_{\mathbf{i}}(\mathbf{r})K_{\mathbf{i}}(\mathbf{r}') \rangle = \langle [\varDelta \varepsilon(\mathbf{r})/\varepsilon_0] [\varDelta \varepsilon(\mathbf{r}')/\varepsilon_0] \rangle$$
(6.7.12)

Equation (6.7.5) can still be solved formally by iteration. We have

$$E_{j}(\mathbf{r}) = E_{0j}(\mathbf{r}) + \int \Gamma_{jk}(\mathbf{r}, 1) L_{1kn}(1) E_{0n}(1) d1$$

+ $\iint \Gamma_{jk}(\mathbf{r}, 2) L_{1kn}(2) \Gamma_{nm}(1, 2) L_{1mp}(1) E_{0p}(1) d1 d2$
+ $\iint \int \Gamma_{jk}(\mathbf{r}, 3) L_{1kn}(3) \Gamma_{nm}(3, 2) L_{1mp}(2) \Gamma_{pr}(2, 1) L_{1rs}(1) E_{0s}(1)$
× $d1 d2 d3 + \cdots$ (6.7.13)

where the vectors 1, 2, ... represent $\mathbf{r}_1, \mathbf{r}_2, \ldots$, respectively. Substituting the expression $\mathbf{L}_1 = k_0^2 K_1(\mathbf{r}) \mathbf{I}$ into (6.7.13), we obtain

$$E_{j}(\mathbf{r}) = E_{0j}(\mathbf{r}) + k_{0}^{2} \int \Gamma_{jn}(\mathbf{r}, \mathbf{1}) K_{1}(\mathbf{1}) E_{0n}(\mathbf{1}) d\mathbf{1}$$

+ $k_{0}^{4} \int \Gamma_{jn}(\mathbf{r}, \mathbf{2}) K_{1}(\mathbf{2}) \Gamma_{np}(\mathbf{2}, \mathbf{1}) K_{1}(\mathbf{1}) E_{0p}(\mathbf{1}) d\mathbf{1} d\mathbf{2}$
+ $k_{0}^{6} \int \Gamma_{jn}(\mathbf{r}, \mathbf{3}) K_{1}(\mathbf{3}) \Gamma_{np}(\mathbf{3}, \mathbf{2}) K_{1}(\mathbf{2}) \Gamma_{ps}(\mathbf{2}, \mathbf{1}) K_{1}(\mathbf{1}) E_{0s}(\mathbf{1})$
× $d\mathbf{1} d\mathbf{2} d\mathbf{3} + \cdots$ (6.7.14)

We now introduce the *p*-point correlation function for the centered random function $K_1(\mathbf{r})$ by means of the cluster expansion as the following. From (6.7.12), we have the two-point correlation function $\varrho(\mathbf{r}_1, \mathbf{r}_2) = \varrho_2(1, 2)$.

The higher-order correlation functions are defined through the moments

$$\langle K_{1}(1)K_{1}(2)K_{1}(3)\rangle = \varrho_{3}(1, 2, 3)$$

$$\langle K_{1}(1)K_{1}(2)K_{1}(3)K_{1}(4)\rangle = \varrho_{2}(1, 2)\varrho_{2}(3, 4)$$

$$+ \varrho_{2}(1, 3)\varrho_{2}(2, 4)$$

$$+ \varrho_{2}(1, 4)\varrho_{2}(2, 3)$$

$$+ \varrho_{4}(1, 2, 3, 4)$$

$$\langle K_{1}(1)\cdots K_{1}(\mathbf{p})\rangle = \Sigma \varrho_{k}(\mathbf{i}_{1}, \ldots, \mathbf{i}_{k})\varrho_{m}(\mathbf{j}_{1}, \ldots, \mathbf{j}_{m})$$

$$(6.7.15)$$

where the summation in the last equation is extended over all possible partitions of the set 1, 2, ..., p into clusters of at least two points and $\varrho_3, \varrho_4, ..., \varrho_p$ are the higher order correlation functions. For a centered Gaussian random function, only the two-point correlation function ϱ_2 is different from zero.

The general property of the correlation function $\varrho_p(1, 2, \ldots, p)$ is that it vanishes whenever the points 1, 2, ..., p are not inside a common sphere of diameter l, the correlation distance.

Let us now take the average of (6.7.14). With the definitions in (6.7.15), we obtain

$$\langle E_{j}(\mathbf{r}) \rangle = E_{0j}(\mathbf{r}) + k_{0}^{4} \int \Gamma_{jn}(\mathbf{r}, 2) \Gamma_{np}(2, 1) \varrho_{2}(1, 2) E_{0p}(1) d1 d2 + k_{0}^{6} \int \Gamma_{jn}(\mathbf{r}, 3) \Gamma_{np}(3, 2) \Gamma_{ps}(2, 1) \varrho_{3}(1, 2, 3) E_{0s}(1) d1 d2 d3 + \cdots$$
(6.7.16)

Equation (6.7.16) can be represented by diagrams defined by the following conventions:

(i) The dyadic Green's function $\Gamma_{np}(2, 1)$ is represented by a solid line whose end points correspond to points 2, 1 and indices n, p, respectively.

(ii) Points belonging to a given cluster (same correlation function) are connected by dotted lines.

(iii) The factor k_0^2 is represented by a vertex point at which a single dotted line and two solid lines meet.

(iv) $E_{0j}(\mathbf{r})$ is represented by double solid lines.

(v) Integrations are performed over the coordinates of all internal vertices of the diagram. Summations are performed over the indices of all internal vertices. With these conventions, (6.7.16) can be written as



Next we introduce the following definitions.

(i) A diagram is said to be weakly connected if it can be divided into two or more diagrams without cutting through any dotted lines. Diagrams of the type



can be divided into



and are therefore weakly connected.

(ii) The remaining diagrams are strongly connected. Diagrams of the type



are strongly connected.

(iii) The mass operator is the sum of all possible strongly connected

354

diagrams in (6.7.17). It is denoted by ${\bf M}$ or the symbol $\otimes.$ Its first few terms are



(iv) The average field is represented by a broad solid line ______ Let us now consider the Dyson equation defined as follows

_____ = _____ + _____ (6.7.19)

This equation can be expanded into

Substituting (6.7.18) into (6.7.20), it is easy to show that (6.7.19) is equivalent to (6.7.17). Therefore, by resuming the terms in the perturbation series (6.7.16), we derive a new integral equation for the average field where the mass operator **M** is the kernel. Explicitly (6.7.19) is written as

$$\langle E_j(\mathbf{r})\rangle = E_{0j}(\mathbf{r}) + \int \Gamma_{jk}(\mathbf{r}, \mathbf{2}) M_{kn}(\mathbf{2}, \mathbf{1}) \langle E_n(\mathbf{1}) \rangle \, d\mathbf{1} \, d\mathbf{2} \qquad (6.7.21)$$

where

$$M_{kn}(2, 1) = k_0^4 \Gamma_{kn}(2, 1)\varrho_2(2, 1) + k_0^6 \int \Gamma_{km}(2, 3)\Gamma_{mn}(3, 1)\varrho_3(1, 2, 3) d3 + k_0^8 \iint \Gamma_{km}(2, 4)\Gamma_{mp}(4, 3)\Gamma_{pn}(3, 1)\varrho_2(2, 3)\varrho_2(4, 1) d4 d3 + \cdots$$
(6.7.22)

The Dyson's equation is first introduced in the study of quantum field theory. For the isotropic homogeneous random medium we are discussing the mass operator and the dyadic Green's function are both convolution operators. Equation (6.7.21) can then be solved by the Fourier transform technique. Taking the Fourier transform of (6.7.21), we have

$$\langle E_j(\mathbf{p}) \rangle = E_{0j}(\mathbf{p}) + \Gamma_{jk}(\mathbf{p}) M_{kn}(\mathbf{p}) \langle E_n(\mathbf{p}) \rangle \qquad (6.7.23)$$

where

$$\Gamma_{jk}(\mathbf{p}) = \frac{1}{p^2 - k_0^2} \left[\delta_{jk} - p_j p_k / k_0^2 \right]$$
(6.7.24)

To avoid new symbols, the transformed functions are represented by the same original functions with argument replaced by p. Equation (6.7.23) can be put in the form

$$[\delta_{nj} - \Gamma_{jk}(\mathbf{p})M_{kn}(\mathbf{p})]\langle E_n(\mathbf{p})\rangle = E_{0j}(\mathbf{p})$$
(6.7.25)

The set of algebraic equations (6.7.25) can be solved if $M_{kn}(\mathbf{p})$ is known. Then the average field may be obtained from the inverse Fourier transform. If we set the determinant of the matrix $\mathbf{I} - \mathbf{\Gamma} \cdot \mathbf{M}$ to zero, the roots of the equation

$$\det[\mathbf{I} - \mathbf{\Gamma} \cdot \mathbf{M}] = 0 \tag{6.7.26}$$

determine the behavior of the average field $\langle E_j(\mathbf{r}) \rangle$ as $|\mathbf{r}| \rightarrow \infty$. Equation (6.7.26) may be defined as the dispersion relation for the average field and the roots are the effective wave numbers, or effective propagation constants, for the different modes of the average field in the random medium.

From (6.7.22) we see that the exact computation of the mass operator **M** is just as difficult as the computation of the original perturbation series (6.7.16). The convergence of the series is not assured in this case either. However, finite order approximations for **M** correspond to partial summations of the complete perturbation series up to terms of any order in $|\mathbf{L}_1|$, and may therefore be considered as better approximations than the Born's solution. A necessary condition for convergence of the series for the mass operator is given by

 $|L_1| l \ll 1$ (6.7.27)

where $|L_1|$ is the norm of the random operator L_1 and l is the correlation length.

The simplest approximation of the Dyson equation is to take only the first term in the mass operator. Therefore in diagram form (6.7.19) becomes

Or, explicitly

$$\langle E_j(\mathbf{r})\rangle = E_{0j}(\mathbf{r}) + k_0^4 \iint \Gamma_{jk}(\mathbf{r}, \mathbf{2}) \Gamma_{kn}(\mathbf{2}, \mathbf{1}) \varrho_2(\mathbf{2}, \mathbf{1}) \langle E_n(\mathbf{1}) \rangle \, d\mathbf{1} \, d\mathbf{2} \quad (6.7.29)$$

356

Now apply the operator L_{0ij} on both sides of (6.7.29). Since Γ is the inverse of \mathbf{L}_0 , we have

$$L_{0ij}\langle E_j(\mathbf{r})\rangle + k_0^4 \int \Gamma_{in}(\mathbf{r}, 1)\varrho_2(\mathbf{r}, 1)\langle E_n(1)\rangle d1 = j\omega\mu_0 J_i \quad (6.7.30)$$

where (6.7.6) has been used.

Equation (6.7.30) is an equation for the averaged field and was also derived by Keller using a nondiagrammatic approach.

Although we have derived the Dyson equation for the vector wave equation in a homogeneous, isotropic background, the results can be used for scalar wave equations just by changing the dyadic Green's function Γ_{ij} 's to G_0 and using the corresponding L_0 and L_1 for scalar wave equations. The results can also be generalized to describe the more complicated case where the background is anisotropic.

As an example, let us consider (6.7.30) for a scalar wave $u(\mathbf{r})$. From (6.7.2a), $L_0 = \nabla^2 + k_0^2$. Taking the Fourier transform on both sides, we have

$$\left[L_0(\mathbf{p}) + k_0^4 \int G_0(\mathbf{q})\varrho_2(\mathbf{p} - \mathbf{q}) \, d\mathbf{q}\right] \langle u(\mathbf{p}) \rangle = j\omega\mu_0 J(\mathbf{p}) \quad (6.7.31)$$

where $L_0(\mathbf{p}) = k_0^2 - p^2$ and $G_0(\mathbf{p}) = -(k_0^2 - p^2)^{-1}$.

The dispersion relation for (6.7.31) is obtained by setting the quantity in the bracket equal to zero. We have,

$$k_0^2 - p^2 - k_0^4 \int \frac{\varrho_2(\mathbf{p} - \mathbf{q})}{k_0^2 - q^2} d\mathbf{q} = 0$$
 (6.7.32)

This is in general a transcendental equation for the wave number p. Further computation depends on the explicit form of the correlation function ϱ_2 . The value p obtained from (6.7.32) is called the effective propagation constant (or wave number). It is the propagation constant for a plane average wave $\langle u(\mathbf{r}) \rangle$ of the form $e^{-jp}_{\text{eff}} \hat{n}^{*r}$ propagating in the \hat{n} -direction in this random medium. Let us consider a special case for which

$$\varrho_2(\mathbf{R}) = \langle K_1^2 \rangle e^{-R/l} \tag{6.7.33}$$

where $\langle K_1^2 \rangle = \langle |\Delta \varepsilon / \langle \varepsilon \rangle |^2 \rangle$ in our early notations. The dispersion relation (6.7.32) now becomes

$$k_0^2 - p^2 + k_0^4 \langle K_1^2 \rangle / [p^2 - (k_0 - j/l)^2] = 0 \qquad (6.7.34)$$

The simplest way of solving (6.7.34) for p is by iteration. Since $\langle K_1^2 \rangle \ll 1$, we can write

$$p^{2} = k_{0}^{2} + \frac{k_{0}^{4} \langle K_{1}^{2} \rangle}{p_{0}^{2} - (k_{0} - j/l)^{2}}$$

$$\simeq k_{0}^{2} + \frac{k_{0}^{4} \langle K_{1}^{2} \rangle}{k_{0}^{2} - (k_{0} - j/l)^{2}}$$

$$= k_{0}^{2} + \frac{l^{2} k_{0}^{4} \langle K_{1}^{2} \rangle}{1 + j 2 k_{0} l}$$
(6.7.35)

Therefore

$$p_{\rm eff} \simeq k_0 + \frac{1}{2} l^2 k_0^3 \langle K_1^2 \rangle / (1 + j 2 k_0 l) \tag{6.7.36}$$

The real part of the effective propagation constant is greater than k_0 and there is a negative imaginary part. This indicates a decrease of phase velocity for the coherent wave and damping of the coherent wave as it propagates into the random medium. Although these results are derived from a specific correlation function $\varrho_2(R)$, they are also true for general cases. Physically a wave propagating in a random medium is continually scattered by the random inhomogeneities. A fluctuating component of the field is generated by incoherent scattering. Energy is continually transferred from the coherent to the incoherent wave via the scattering process. Hence the coherent wave will be damped. The damping constant is given by the imaginary part of p_{eff} . The reciprocal of Im p_{eff} sometimes is defined as "coherent distance" for the field. The decrease of the phase velocity of the coherent wave is also reasonable since the wave has gone through multiple scattering.

In the Born's approximation, the averaged plane wave is proportional to $e^{-jk_0A\cdot\mathbf{r}}$. Comparing the solution $e^{-jp_{eff}A\cdot\mathbf{r}}$ with the Born's solution, we see the effects of multiple scattering on the wave.

6.8 Effective Dielectric Tensor for Coherent Waves

We shall now make use of the results derived in the last section to discuss some general features of coherent vector waves in random media with isotropic background. For this purpose, it is convenient to introduce two new functions as follows

$$\xi(\mathbf{r}) = 3 \frac{K(\mathbf{r}) - K_0}{K(\mathbf{r}) + 2K_0}, \quad \mathbf{F}(\mathbf{r}) = \frac{K(\mathbf{r}) + 2K_0}{3K_0} \mathbf{E}(\mathbf{r}) \quad (6.8.1)$$

where $K(\mathbf{r}) = \varepsilon(\mathbf{r})/\varepsilon_0$ is the relative dielectric permittivity of the medium

358

and K_0 is a quantity to be defined below. For the isotropic, homogeneous background medium, (6.7.2) can be written as

$$[\nabla^2 - \nabla \nabla + k_0^2 K(\mathbf{r})]\mathbf{E} = j\omega\mu_0 \mathbf{J}$$
(6.7.2a)

Let us now define a field E_0 satisfying the equation

$$[\mathcal{V}^2 - \mathcal{V}\mathcal{V} + k_0^2 K_0]\mathbf{E}_0 = j\omega\mu_0\mathbf{J}$$
(6.7.2b)

Subtracting (6.7.2b) from (6.7.2a) we have

$$[\nabla^2 - \nabla \nabla + k_0^2 K_0](\mathbf{E} - \mathbf{E}_0) = -k_0^2 [K(\mathbf{r}) - K_0] \mathbf{E} \qquad (6.8.2)$$

which can be put into an integral equation

$$E_{j}(\mathbf{r}) = E_{0j}(\mathbf{r}) + k_{0}^{2} \int \Gamma_{jk}(\mathbf{r}, \mathbf{r}')(k - k_{0})E_{k}(\mathbf{r}') d\mathbf{r}' \qquad (6.8.3)$$

where

$$\Gamma_{jk}(\mathbf{R}) = \mathrm{PV}\left[\delta_{jk} + \frac{1}{k_0^2 K_0} \frac{\partial}{\partial x_j} \frac{\partial}{\partial x_k}\right] \frac{e^{-jk_0 \sqrt{K_0 R}}}{4\pi R} - \frac{\delta_{jk}}{3k_0^2 K_0} \,\delta(\mathbf{R})$$
$$= \Gamma'_{jk}(\mathbf{R}) - \frac{\delta_{jk}}{3k_0^2 K_0} \,\delta(\mathbf{R}) \tag{6.7.11a}$$

Note that Γ_{jk} in (6.7.11a) is the same as that in (6.7.11) except for the wave number $k_0\sqrt{K_0}$. Substituting (6.7.11a) into (6.8.3) and using the definition (6.8.1) it is easy to show that

$$F_j(\mathbf{r}) = E_{0j}(\mathbf{r}) + k_0^2 K_0 \int \Gamma'_{jk}(\mathbf{r}, \mathbf{r}') \xi(\mathbf{r}') F_k(\mathbf{r}') \, d\mathbf{r}' \qquad (6.8.4)$$

Taking the average of (6.8.4), we obtain

$$\langle F_j(\mathbf{r})\rangle = E_{0j}(\mathbf{r}) + k_0^2 K_0 \int \Gamma'_{jk}(\mathbf{r}, \mathbf{r}') \langle \xi(\mathbf{r}') F_k(\mathbf{r}') \rangle d\mathbf{r}' \qquad (6.8.5)$$

Let us now define an operator $\boldsymbol{\xi}^{\mathrm{eff}}$ by

$$\langle \xi(\mathbf{r})F_k(\mathbf{r})\rangle = \int \xi_{kn}^{\text{eff}}(\mathbf{r},\mathbf{r}')\langle F_n(\mathbf{r}')\rangle d\mathbf{r}'$$
 (6.8.6)

Equation (6.8.4) becomes

$$\langle F_j(\mathbf{r})\rangle = E_{0j}(\mathbf{r}) + k_0^2 K_0 \iint \Gamma'_{jk}(\mathbf{r},\mathbf{r}') \xi_{kn}^{\text{eff}}(\mathbf{r}',\mathbf{r}'') \langle F_n(\mathbf{r}'')\rangle \, d\mathbf{r}' \, d\mathbf{r}'' \qquad (6.8.7)$$

which is an integral equation for $\langle F_j(\mathbf{r}) \rangle$. To complete our derivation, we need to derive $\xi_{kn}^{\text{eff}}(\mathbf{r}', \mathbf{r}'')$ and define K_0 . This can be done by applying the diagram technique to (6.8.4) in the same manner as in the last section. We obtain an integral equation similar to (6.7.21)

$$\langle F_j(\mathbf{r})\rangle = E_{0j}(\mathbf{r}) + \iint \Gamma'_{jk}(\mathbf{r},\mathbf{r}')\overline{M}_{kn}(\mathbf{r}',\mathbf{r}'')\langle F_n(\mathbf{r}'')\rangle \, d\mathbf{r}'' \, d\mathbf{r}' \quad (6.8.8)$$

where

$$\begin{split} \bar{M}_{kn}(\mathbf{1},\mathbf{2}) &= k_0^4 K_0^2 \Gamma'_{kn}(\mathbf{1},\mathbf{2}) \varrho_{\xi_2}(\mathbf{1},\mathbf{2}) \\ &+ k^6 K^3 \int \Gamma'_{km}(\mathbf{2},\mathbf{3}) \Gamma'_{mn}(\mathbf{3},\mathbf{1}) \varrho_{\xi_3}(\mathbf{1},\mathbf{2},\mathbf{3}) \, d\mathbf{3} \\ &+ k_0^8 K_0^4 \int \Gamma'_{km}(\mathbf{2},\mathbf{4}) \Gamma'_{mp}(\mathbf{4},\mathbf{3}) \Gamma'_{pn}(\mathbf{3},\mathbf{1}) \varrho_{\xi_2}(\mathbf{2},\mathbf{3}) \varrho_{\xi_2}(\mathbf{4},\mathbf{1}) \\ &\times d\mathbf{4} \, d\mathbf{3} + \cdots \end{split}$$
(6.8.9)

and

$$\langle \xi \rangle = \{ K(\mathbf{r}) - K_0 \} / \{ K(\mathbf{r}) + 2K_0 \} = 0$$
 (6.8.10)

The condition (6.8.10) is required in our derivation of (6.8.8) and can be taken as the defining equation for the quantity K_0 . The functions ϱ_{ξ_2} , ϱ_{ξ_3} , etc. are the correlation functions for the random function $\xi(\mathbf{r})$.

Comparing (6.8.8) with (6.8.7), we have

$$\xi_{kn}^{\text{eff}}(\mathbf{r}',\mathbf{r}'') = \bar{M}_{kn}(\mathbf{r}',\mathbf{r}'')/k_0^2 K_0 \qquad (6.8.11)$$

From (6.8.1), it follows easily that

 $\langle F_i \rangle = \langle K(\mathbf{r}) E_i(\mathbf{r}) \rangle / 3K_0 + 2 \langle E_i \rangle / 3$

and

$$\langle \xi F_i \rangle = \langle K(\mathbf{r}) E_i(\mathbf{r}) \rangle / K_0 - \langle E_i \rangle$$
 (6.8.12)

If we now define the relative effective dielectric tensor \mathbf{K}^{eff} by

$$\langle K(\mathbf{r})\mathbf{E}(\mathbf{r})\rangle = \mathbf{K}^{\text{eff}}\langle \mathbf{E}(\mathbf{r})\rangle = \int \mathbf{K}^{\text{eff}}(\mathbf{r},\mathbf{r}') \cdot \langle \mathbf{E}(\mathbf{r}')\rangle d\mathbf{r}' \quad (6.8.13)$$

where \mathbf{K}^{eff} is a tensor integral operator defined by the second equality of (6.8.13), then (6.8.12) can be written as

$$\langle F_i \rangle = [K_{ij}^{\text{eff}}/3K_0 + 2 \ \delta_{ij}/3] \langle E_j \rangle$$

$$\langle \xi F_i \rangle = \xi_{ij}^{\text{eff}} \langle F_i \rangle = [K_{ij}^{\text{eff}}/K_0 - \delta_{ij}] \langle E_j \rangle$$

(6.8.14)

For a statistically homogeneous medium, both $\boldsymbol{\xi}^{\text{eff}}$ and $\boldsymbol{K}^{\text{eff}}$ are convolution operators. Therefore, taking the Fourier transform of (6.8.14), we obtain

$$\langle F_i(\mathbf{p}) \rangle = [\mathbf{K}_{ij}^{\text{eff}}(\mathbf{p})/3K_0 + 2 \,\delta_{ij}/3]\langle E_j(\mathbf{p}) \rangle$$

$$\xi_{ij}^{\text{eff}}(\mathbf{p})\langle F_j(\mathbf{p}) \rangle = [K_{ij}^{\text{eff}}(\mathbf{p})/K_0 - \delta_{ij}]\langle E_j(\mathbf{p}) \rangle$$

$$(6.8.15)$$

From (6.8.15) it follows that

$$\frac{1}{3K_0} \xi_{ij}^{\text{eff}}(\mathbf{p}) K_{jk}^{\text{eff}}(\mathbf{p}) - \frac{1}{K_0} K_{ik}^{\text{eff}}(\mathbf{p}) = -\delta_{ik} - \frac{2}{3} \xi_{ik}^{\text{eff}}(\mathbf{p}) \quad (6.8.16)$$

Since the medium is assumed to be also statistically isotropic, $K_{ij}^{\text{eff}}(\mathbf{p})$ and $\xi_{ij}^{\text{eff}}(\mathbf{p})$ must be rotational symmetric with respect to p. From the discussion in Chapter 2 [(2.4.20a)], we can write $K_{ij}^{\text{eff}}(\mathbf{p})$ and $\xi_{ij}^{\text{eff}}(\mathbf{p})$ in the following forms:

$$K_{ij}^{\text{eff}}(\mathbf{p}) = \left(\delta_{ij} - \frac{p_i p_j}{p^2}\right) K_{\perp}^{\text{eff}}(p) + \frac{p_i p_j}{p^2} K_{\parallel}^{\text{eff}}(p)$$

$$\xi_{ij}^{\text{eff}}(\mathbf{p}) = \left(\delta_{ij} - \frac{p_i p_j}{p^2}\right) \xi_{\perp}^{\text{eff}}(p) + \frac{p_i p_j}{p^2} \xi_{\parallel}^{\text{eff}}(p)$$
(6.8.17)

Substituting (6.8.17) into (6.8.16), we obtain the relations

$$K_{\parallel}^{\text{eff}}(p) = K_0 \frac{1 + 2\xi_{\parallel}^{\text{eff}}(p)/3}{1 - \xi_{\parallel}^{\text{eff}}(p)/3}$$

$$K_{\perp}^{\text{eff}}(p) = K_0 \frac{1 + 2\xi_{\perp}^{\text{eff}}(p)/3}{1 - \xi_{\perp}^{\text{eff}}(p)/3}$$
(6.8.18)

The Fourier transform $\xi_{ij}^{\text{eff}}(\mathbf{p})$ can be obtained in principle from (6.8.11). Hence we note that once again, we return to the problem of computing the mass operator. In the following, we shall discuss the simplest case of retaining only the first term in (6.8.9). This approximation is valid if

$$|\xi| k_0^2 K_0 l^2 \ll 1 \tag{6.8.19}$$

where *l* is the correlation length for ξ and $|\xi|$ is the upper bound of the magnitude of ξ . Under this approximation, we have

$$\xi_{ij}^{\text{eff}}(\mathbf{r}',\mathbf{r}'') = k_0^2 K_0 \langle \xi^2 \rangle \Gamma_{ij}'(\mathbf{r}',\mathbf{r}'') C_{\xi}(\mathbf{r}',\mathbf{r}'')$$
(6.8.20)

where $\rho_{\xi^2}(\mathbf{r}, \mathbf{r}') = \langle \xi^2 \rangle C_{\xi}(\mathbf{r}, \mathbf{r}')$ has been used and $C_{\xi}(\mathbf{r}, \mathbf{r}')$ is the normalized

6. Wave Propagation in Random Media

correlation function. The Fourier transform of (6.8.20) is

$$\xi_{ij}^{\text{eff}}(\mathbf{p}) = k_0^2 K_0 \langle \xi^2 \rangle \int \Gamma_{ij}'(\mathbf{R}) C_{\xi}(R) e^{j\mathbf{p} \cdot \mathbf{R}} d\mathbf{R}$$
(6.8.21)

From (6.8.17), we see that the longitudinal and transverse components of $\boldsymbol{\xi}^{\text{eff}}$ are given by

$$\xi_{\parallel}^{\text{eff}}(p) = (p_i p_j / p^2) \xi_{ij}^{\text{eff}}(\mathbf{p})$$

$$\xi_{\perp}^{\text{eff}}(p) = (\delta_{ij} - p_i p_j / p^2) \xi_{ij}^{\text{eff}}(\mathbf{p})$$
(6.8.22)

respectively.

Substituting (6.8.21) into (6.8.22) with $\Gamma'_{ij}(\mathbf{R})$ given by (6.7.11a), we can carry out the angular integration to give

$$\xi_{\parallel}^{\rm eff}(p) = 2\langle \xi^2 \rangle Q(p, k_0 \sqrt{K_0}) \tag{6.8.23}$$

$$\xi_{\perp}^{\text{eff}}(p) = -\langle \xi^2 \rangle Q(p, k_0 \sqrt{K_0}) - \langle k_0^2 K_0/p \rangle \langle \xi^2 \rangle \int_0^\infty C_{\xi}(x) e^{-jk_0 \sqrt{K_0}x} \sin px \, dx$$
(6.8.24)

where

$$Q(p, k_0 \sqrt{K_0}) = PV \int_0^\infty C_{\xi}(x) e^{-jk_0 \sqrt{K_0}x} \left[\frac{k_0^2 K_0 \cos px}{p^2} - j \frac{k_0 \sqrt{K_0}}{p} \sin px - \frac{k_0^2 K_0}{p^2} \sin px - \frac{1}{px} \sin px - \frac{1}{px} \sin px - \frac{3j \frac{k_0 \sqrt{K_0}}{p}}{p} \frac{\cos px}{px} - 3 \frac{\cos px}{p^2 x^2} + 3j \frac{k_0 \sqrt{K_0}}{p} \frac{\sin px}{p^2 x^2} + 3 \frac{\sin px}{p^3 x^3} \right] x^{-1} dx$$
(6.8.25)

Equations (6.8.23), (6.8.24), and (6.8.25) permit us to compute ξ^{eff} for a given correlation function of the random medium; (6.8.18) then gives us the effective dielectric tensor \mathbf{K}^{eff} . But we must first find K_0 from the definition (6.8.10). For the case $|K_1| \ll \langle K \rangle$ where K_1 is the fluctuating part of the relative permittivity, we have from (6.8.10)

$$K_0 \simeq \langle K \rangle - \frac{1}{3} \frac{\langle K_1^2 \rangle}{\langle K \rangle} - \frac{4}{27} \frac{(\langle K_1^2 \rangle)^2}{\langle K \rangle^3} + \cdots \qquad (6.8.26)$$

Also, under the same approximation,

$$\langle \xi^2 \rangle \simeq \langle K_1^2 \rangle / \langle K \rangle^2$$
 (6.8.27)

Equation (6.8.18) then becomes

$$K_{\parallel}^{\text{eff}}(p) \simeq \langle K \rangle - \frac{1}{3} \frac{\langle K_1^2 \rangle}{\langle K \rangle} - \frac{2 \langle K_1^2 \rangle}{\langle K \rangle} Q(p, k_0 \sqrt{K_0})$$
(6.8.28)

$$K_{\perp}^{\text{eff}}(p) \simeq \langle K \rangle - \frac{1}{3} \frac{\langle K_1^2 \rangle}{\langle K \rangle} + \frac{k_0^2}{p} \langle K_1^2 \rangle \int_0^\infty C_{\xi}(x) e^{-jk_0 \sqrt{K_0}x} \sin px \, dx$$
$$+ \frac{\langle K_1^2 \rangle}{\langle K \rangle} Q(p, k_0 \sqrt{K_0})$$
(6.8.29)

Substituting (6.8.28) and (6.8.29) into (6.8.17), we obtain the effective dielectric tensor \mathbf{K}^{eff} for this medium under the present approximation.

Therefore, as far as the average field is concerned, we can treat the random medium as a dispersive medium with an effective dielectric tensor given by (6.8.17), (6.8.28), and (6.8.29). The discussions in Chapter 2 on dispersive media can be applied directly to the present problem. For example, from the discussion in Section 2.6 we obtain the dispersion relations for transverse mode

$$K_{\perp}^{\rm eff}(p) - p^2/k_0^2 = 0 \tag{6.8.30}$$

and longitudinal mode

$$K_{\parallel}^{\rm eff}(p) = 0 \tag{6.8.31}$$

Also the average dyadic Green's function for the dyadic Green's function for the average field is expressed by

$$\langle \Gamma_{ik}(\mathbf{p}) \rangle = (\delta_{ik} - p_i p_k/p^2) G_{\perp}(p) + (p_i p_k/p^2) G_{\parallel}(p) \quad (6.8.32)$$

where

$$G_{\parallel}(p) = 1/k_0^2 K_{\parallel}^{\rm eff}(p) \tag{6.8.33}$$

is proportional to the Green's function for the longitudinal mode and

$$G_{\perp}(p) = 1/p^2 - k_0^2 K_{\perp}^{\rm eff}(p) \tag{6.8.34}$$

is proportional to the Green's function for the transverse mode as can be seen by comparison with (2.13.6).

The dyadic Green's function is obtained by taking the inverse Fourier transform of (6.8.32).

It has been shown (Ryzhov *et al.*, 1965) that in general if the background medium in the absence of irregularities does not admit a longitudinal mode (for example, a cold plasma), then the longitudinal mode for the average field generated by the random scattering will not be a propagating mode.

The method we discussed above can be used to find the effective static dielectric constant for a mixture (Finkelberg, 1964).

There are many unsolved problems concerning the propagation of wave in random media. For example, the convergence of the perturbation series, especially for large correlation length; how to go beyond the first term in the mass operator; the formulation for a finite random medium or a medium with inhomogeneous background; and so forth—these are all outstanding problems yet to be attacked. As far as the application of the theory is concerned, a variety of problems are of both theoretical and practical interests. To name just a few, we have the problem of finding the radiation pattern and radiation resistance of an antenna imbedded in a random medium, the energy loss and radiation of energetic particles passing through a turbulent plasma, the propagation of waves in interplanetary medium and their effects on outer space probing via electromagnetic waves.

Problems

- **1.** For a real stochastic process $\xi(t)$,
 - (a) derive

$$\langle [\xi(t+\tau)\pm\xi(t)]^2 \rangle = 2[\varrho_{\xi}(0)\pm\varrho_{\xi}(\tau)]$$

(b) prove that

$$|\varrho_{\xi}(\tau)| \leq \varrho_{\xi}(0)$$

i.e., $\rho_{\xi}(\tau)$ has a maximum at the origin.

2. In Section 6.3 we discussed the back-scattering cross section. For axially symmetrical irregularities in which the axis of symmetry is the z-axis, we put $a = b = L_t$ and $d = L_l$ in (6.3.24).

Prove that the back-scattering cross section can be written as

$$\sigma_B = \frac{\omega_{p_0}^4}{4\sqrt{2\pi} c^4} \left\langle \left(\frac{\Delta N}{N_0}\right)^2 \right\rangle L_t^2 L_l \exp\left\{-\frac{8\pi^2 L_t^2}{\lambda^2} - \frac{8\pi^2 L_l^2}{\lambda^2} \psi^2\right\}$$

provided that the direction of incidence (l, m, n) is almost perpendicular to the axis of symmetry. Where $n^2 = \sin^2 \psi$, $l^2 + m^2 = \cos^2 \psi$.

3. Prove the relations (6.5.32) and (6.5.33).

4. Derive (6.6.15) and (6.6.16) from (6.6.11) and (6.6.12).

References

5. Using the definitions (6.8.1), prove that the wave equation can be put into an integral equation of the form shown in (6.8.4).

- **6.** Prove (6.8.22).
- 7. Derive (6.8.23), (6.8.24), and (6.8.25).

References

- BOOKER, H. G. (1956). A theory of scattering by nonisotropic irregularities with application to radar reflections from the aurora. J. Atmos. Terr. Phys. 8, 204–221.
- BOOKER, H. G., and GORDON, W. E. (1950). A theory of radio scattering in the troposphere. Proc. IRE 38, 401.
- CHERNOV, L. A. (1967). "Wave Propagation in a Random Medium," Dover, New York.
- DEWOLF, D. A. (1968). Saturation of irradiance fluctuations due to turbulent atmosphere. J. Opt. Soc. Amer. 58, 461.
- FARLEY, D. T., and DOUGHERTY, J. P. (1960). Theory of incoherent radiowave scattering in a plasma. Proc. Roy. Soc., Ser. A 259, 79.
- FEJER, J. A. (1960). Radiowave scattering by a ionized gas in thermal equilibrium. Can. J. Phys. 38, 1114.

FINKELBERG, V. M. (1964). Dielectric constant of mixtures. Sov. Phys. JETP 9, 396.

- FRISCH, V. (1968). Wave propagation in random media. In "Probabilistic Methods in Applied Mathematics" (A. T. Bharmcha-Reid, ed.). Academic Press, New York. (A good list of references on the topic are included there.)
- HEY, J. S., PARSONS, S. J., and PHILLIPS, J. W. (1946). Fluctuations in cosmic radiation at radio frequencies. *Nature (London)* **158**, 234.
- ISHIMARU, A. (1969). Fluctuations of a beam wave propagatihng trough a locally homogeneous medium. *Radio Sci.* 4, 295–306.
- KHINCHIN, A. I. (1938). The theory of correlation of stationary random processes. Sov. Phys. Usp.
- LITTLE, C. G., and LOVELL, A. C. B. (1950). Origin of the fluctuations in the intensity of radio waves from galactic sources. *Nature (London)* 165, 423.
- LIU, C. H. (1965). Cross-correlation function of spherical waves propagating through a slab containing anisotropic irregularities. J. Atmos. Terr. Phys. 28, 385.
- McClure, J. P., and Swenson, G. W., Jr. (1964). Beacon satellite studies of small scale ionospheric inhomogenities, *Tech. Rep. Univ. of Ill.*
- PAPOULIS, A. (1965). "Probability, Random Variables, and Stochastic Processes." McGraw-Hill, New York.
- RYZHOV, YU. A., TAMOIKIN, V. V., and TATARSKII, V. I. (1965). Spatial dispersion of inhomogeneous media. Sov. Phys. JETP 21, 433.
- SIMONICH, D., and YEH, K. C. (1971). Radio wave scattering from random fluctuations in an anisotropic medium. Tech. Rep. No. 42. Ionosphere Radio Lab., Dept. of Elec. Engin., Univ. of Illinois, March 1971.
- SITENKO, A. G. (1967). "Electromagnetic Fluctuations in Plasma." Academic Press, New York.

- SMITH, F. G. (1950). Origin of the fluctuations in the intensity of radio waves from galactic sources. *Nature (London)* 165, 422.
- TATARSKII, V. I. (1966). On strong fluctuations of light wave parameters in a turbulent medium, Sov. Phys. JETP 22, 1083.
- TATARSKII, V. I. (1967). "Wave Propagation in a Turbulent Medium." Dover, New York.
- YAGLOM, A. M. (1962). "An Introduction to the Theory of Stationary Random Functions." Prentice Hall, Englewood Cliffs, New Jersey.
- YEH, K. C. (1962). Propagation of spherical waves through an ionosphere containing anisotropic irregularities. J. Res. Nat. Bur. Stand. Sect. D. 66, 621.

7. Nonlinear Wave Propagation

7.1 Introduction

In the last two chapters we extended our discussion on wave propagation to inhomogeneous and random media. In the discussion we assumed that the amplitude of the waves are small so that the properties of the media are not affected by the passage of the waves. When the amplitude of the waves increases, we may find in many situations that this is no longer true. Properties of the medium may become dependent on the wave amplitude and the propagation problem becomes nonlinear. In the ionosphere, for example, nonlinear phenomena such as cross-modulation are observed even for waves of moderate strength. Other nonlinear phenomena such as self-interaction, detuning, mixing, harmonic generation, wave-wave interaction, wave breaking, and shock formation have been observed and studied in one form or another in various material media. In this chapter we shall address ourselves to some of these problems. The main purpose is to introduce the different physical ideas behind these phenomena and to demonstrate the techniques to deal with them. Therefore instead of treating specific examples, we shall use simple models in our discussion.

The first step in the study of nonlinear wave propagation is to find in a self-consistent way the effects of the medium due to the waves. This in general involves the investigation of the various microscopic processes, classically or quantum mechanically. Many of the useful results, however, may be obtained from some nonrigorous elementary considerations. This is the route we shall follow in our discussion.

One of the most significant features in linear wave propagation is the fact that arbitrary perturbation can be expressed as a superposition of independent normal modes. When the medium becomes weakly nonlinear one would expect that any arbitrary perturbation may still be expressed as a superposition of the linear normal modes. But now due to the nonlinearity, interactions among these modes occur. Energy exchange takes place among the waves. The amplitudes of these modes will vary slowly in time and eventually assume values quite different from those predicated by the linear theory. This essentially is what happens in the wave-wave interaction process. We shall study this in detail via a simple example.

When the nonlinearity is strong, the superposition picture becomes invalid. The waveform changes rapidly and discontinuities in the wave profile may occur. One example of these types of phenomena is the breaking of the wave, and this is the first topic in our discussion.

7.2 Breaking of Waves

One of the important features of nonlinear wave propagation is the breaking of wave profiles. A ready example of this phenomenon is the "breakers" of water waves on a sloping beach. Another example occurs when the finite amplitude sound wave in a gas propagates in the direction of decreasing density of the gas: discontinuity of the wave profile will occur and the wave breaks. Essentially, wave breaking is a nonlinear phenomenon. The cause can be traced to redistribution of the wave energy spectrum through nonlinear effects. What starts out as a signal consisting of the superposition of a few waves, feeds energy into waves with higher and higher wave numbers. The wave profile will change and eventually result in the breaking of the wave. In this section, we shall discuss this phenomenon by considering the problem of the breaking of finite amplitude plasma waves. The main purpose is to demonstrate the idea of wave breaking rather than a rigorous pursuit of the theory. Therefore, our model will be a very simple one.

Let us consider a homogeneous cold plasma of infinite extent with density N. In Chapter 3 we have studied the one-dimensional longitudinal electron oscillation of small amplitude for this model. As was done in Chapter 3, we consider the electrons in a plane with equilibrium position denoted by x_0 . Suppose we displace this plane by a finite distance $\xi = \xi(x_0)$, say, to

the right of x_0 such that $\xi(x_0) > 0$. $\xi(x_0)$ will in general be a single-valued continuous function of the equilibrium position x_0 . In moving this plane by a distance $\xi(x_0)$, we pass over an amount of positive charge $eN\xi(x_0)$ per unit area. If the ordering of the electrons is maintained, i.e., electrons to the right of x_0 at equilibrium are still to the right of the plane after the displacement, then there will be an excess positive charge of $eN\xi(x_0)$ per unit area to the left of this plane and a negative charge of $-eN\xi(x_0)$ per unit area to the right of this plane. Applying Gauss's law across this plane, we have for the electric field

$$E = eN\xi(x_0)/\varepsilon_0 \tag{7.2.1}$$

The equation of motion for the electrons becomes

$$\ddot{\xi}(x_0) = (-e^2 N/m\varepsilon_0)\xi(x_0)$$
(7.2.2)

which has the solution

$$\xi(x_0) = \xi_1(x_0) \sin \omega_p t + \xi_2(x_0) \cos \omega_p t$$
 (7.2.3)

where $\omega_p = (Ne^2/m\epsilon_0)^{1/2}$ is the plasma frequency defined in Section 3.3 and $\xi_1(x_0)$, $\xi_2(x_0)$ are initial conditions for the displacement.

Let us now consider an initial disturbance in the plasma of the type characterized by $\xi_1(x_0) = 0$ and $\xi_2(x_0) = A \sin kx_0$. That is, at time t = 0, the electrons are displaced from their equilibrium planes sinusoidally with amplitude A and wave number k. The corresponding electric field generated by this disturbance is given by (7.2.1).

$$E = (eN/\varepsilon_0)A \sin kx_0 \cos \omega_p t \qquad (7.2.4)$$

To find the field E as a function of spectral coordinate x, we may use the relation

$$x = x_0 + \xi(x_0)$$

= $x_0 + A \sin kx_0$ (7.2.5)

For small amplitude $(A \ll 1/k)$, x is approximately equal to x_0 and the field E will behave approximately as sin kx with the maximum occurring at $kx \cong \pi/2 \pmod{2\pi}$. As the amplitude A increases, the field E as a function of x will change its shape. The maximum of the field occurs at $kx = \pi/2 + A$ and is displaced towards $kx = \pi$. When the amplitude becomes larger than 1/k, there is a crossover as shown in Fig. 7.2-1. The E field is double valued here and the wave profile begins to break. At this point, the



Fig. 7.2-1. The breaking of waves.

electron planes begin to cross each other, violating the original assumption of no crossover. The analysis beyond this point becomes impossible and further discussion must resort to numerical techniques. We shall not go into the details of the numerical results. Interested readers are referred to the paper by Dawson (1959).

Thus, by considering this simple model of plasma longitudinal oscillation, we have seen how the shape of the electric field changes when the amplitude of oscillation increases. Eventually, when the amplitude is greater than some critical value, the wave profile begins to break. Fourier analysis of the power spectrum of the wave shows that energy is being fed to modes of higher wave numbers from the initial mode in the process of wave breaking.

7.3 Nonlinear Effects in a Plasma in an Electromagnetic Field

Another important aspect of nonlinear wave propagation in a material medium is the phenomenon of wave interaction. This includes self-interaction, mixing, harmonic generation, and cross-modulation of the waves. Basically what happens is that when the amplitude of the wave becomes large, the properties of the medium are changed by the presence of the wave. The propagation parameters of the wave therefore depend on the amplitude of the wave and nonlinear interactions occur. In order to discuss these phenomena, the first step is to calculate the changes of the properties of the medium due to the field. In this section, such a calculation will be made for a plasma in an electromagnetic field. In a plasma, due to slowness of energy transfer from electrons to heavy particles, electrons may acquire considerable amounts of energy from the wave field even for relatively small amplitude waves. Consequently the electrons are heated and the complex dielectric tensor ε becomes dependent on the field strength. To obtain this dependence we use an elementary model for computation in the following.

The equation of motion for electrons in an electromagnetic field is [see (4.6.1)]

$$m\dot{\mathbf{u}} = -e(\mathbf{E} + \mathbf{u} \times \mathbf{B}_0) - m\nu(T_e)\mathbf{u}$$
(7.3.1)

where **u** is the directed velocity of the electrons due to the presence of the field; \mathbf{B}_0 , the external dc magnetic field; \mathbf{E} , the electric field of the wave; and v, the effective collision frequency for the electrons. v depends on the electron temperature T_e , hence, on the electric field if it is strong. In writing (7.3.1), the magnetic field associated with the wave is neglected. This is a good approximation for nonrelativistic electrons. The total velocity of the electrons is the sum of the directed velocity **u** and the random thermal velocity v_0 . The random velocity is related to the temperature through the relation

$$\frac{3}{2}T_e = \frac{1}{2}m\langle v_0^2 \rangle \tag{7.3.2}$$

where the angular brackets indicate an averaging process and the temperature is expressed in energy units. (Note the difference in definition of the velocity v_0 and the thermal velocity v_T defined in Section 3.3.)

The energy balance for the electrons in the presence of the field may be considered as follows. The electric field induces an electron current $\mathbf{J} = -eN\mathbf{u}$ in the plasma. Therefore the field does an amount of work $\mathbf{J} \cdot \mathbf{E} = -eN\mathbf{u} \cdot \mathbf{E}$ on the plasma per unit time, where N is the electron density. On the other hand, an electron loses an average energy $\frac{3}{2} \delta v(T_e - T)$ per unit time in collisions with heavy particles. Here T is the temperature of the electrons in the absence of the field and δ is the mean fraction of energy transferred by the electron in a collision with heavy particles. In elastic collisions, the fractional energy transfer is $\delta \cong 2m/M$, where M is the mass of the heavy particle. Since $M \gg m$, δ is very small and electrons keep most of the energy at each collision with heavy neutral particles. The energy balance equation for electrons can therefore be written as

$$(d/dt)(\frac{3}{2}NT_e) = -eN\mathbf{u} \cdot \mathbf{E} - \frac{3}{2}\delta\nu N(T_e - T)$$

or

$$dT_e/dt = -\frac{2}{3}e\mathbf{u} \cdot \mathbf{E} - \delta v(T_e - T)$$
(7.3.3)

Equations (7.3.1) and (7.3.3) are the basic equations to solve for **u** and T_e as functions of the field **E**. Before we proceed, however, let us consider the solutions of (7.3.1) and (7.3.3) in the absence of the fields. In this case, solutions are simply given by

$$\mathbf{u}(t) = \mathbf{u}(0)e^{-\nu t} \tag{7.3.4}$$

$$T_e - T = (T_e - T)_{t=0} e^{-\delta \nu t}$$
(7.3.5)

In obtaining (7.3.5), $\delta v = \text{constant}$ has been assumed. We see from (7.3.4) and (7.3.5) that the momentum relaxation time 1/v is much shorter than the temperature relaxation time $1/\delta v$ since $\delta \ll 1$ in general.

Let us now solve (7.3.1) and (7.3.3) with the electric field given by $\mathbf{E} = \mathbf{E}_0 \times \cos \omega t$. For simplicity, we first assume that $\mathbf{B}_0 = 0$. Both ν and δ may depend on T_e but are assumed to be independent of time. Under these conditions, (7.3.1) yields after straightforward integration

$$\mathbf{u}(t) = \frac{-e\mathbf{E}_0}{m} \frac{1}{\nu^2 + \omega^2} \left(\nu \cos \omega t + \omega \sin \omega t\right) + c e^{-\nu t} \qquad (7.3.6)$$

where the last term is the initial transient and becomes negligible for t greater than the relaxation time $1/\nu$.

Substituting (7.3.6) into (7.3.3), we have the equation for the temperature

$$\frac{dT_e}{dt} + \delta \nu T_e = \frac{e^2 E_0^2}{3m(\nu^2 + \omega^2)} \left(\nu + \nu \cos 2\omega t + \omega \sin 2\omega t\right) + \delta \nu T \quad (7.3.7)$$

This equation is difficult to solve in general since ν and δ are functions of the unknown T_e . In the following, we shall consider a limiting case of practical interest. We shall seek the first-order solution of (7.3.7) under the assumption that $\omega \gg \delta \nu$ and $\delta \ll 1$.

The formal solution of (7.3.7) can be written as

$$T_{e} = e^{-\delta \nu t} \left[\frac{e^{2} E_{0}^{2}}{3m(\nu^{2} + \omega^{2})} \times \int^{t} e^{\delta \nu \tau} (\nu + \nu \cos 2\omega \tau + \omega \sin 2\omega \tau + \delta \nu T) d\tau + c_{1} \right]$$
(7.3.8)

where c_1 is constant.

For $\omega \gg \delta \nu$, the first order-solution is obtained from (7.3.8) by straightforward integration

$$T_e - T = \frac{e^2 E_0^2}{3m\delta(\nu^2 + \omega^2)}$$
(7.3.9)

where the transient has been neglected.

7.3 Nonlinear Effects in Plasma in Electromagnetic Field

Therefore, to the first order, for $\omega \gg \delta \nu$ and $\delta \gg 1$, the electron temperature is independent of time and is proportional to E_0^2 . Physically, this result is what one would expect, since the temperature relaxation time $1/\delta \nu$ is much greater than the period of the wave $2\pi/\omega$. Under our assumption, $\omega \gg \delta \nu$, the temperature just simply cannot keep up with the fast variation of the field. Instead, it takes some constant mean value given by (7.3.9).

Equation (7.3.9) may be written in another form

$$\frac{T_{e}}{T} = 1 + \frac{e^{2}E_{0}^{2}}{3mT\,\delta(\nu^{2} + \omega^{2})} = 1 + \left(\frac{E_{0}^{2}}{E_{p}}\right)\frac{\omega^{2} + \nu_{0}^{2}}{\omega^{2} + \nu^{2}} \quad (7.3.10)$$

where

$$E_p = (3Tm \ \delta(v_0^2 + \omega^2)/e^2)^{1/2} \tag{7.3.11}$$

is called the plasma field and v_0 is the effective collision frequency in the absence of the field.

We see from (7.3.10) that the plasma field E_p is a measure of the importance of nonlinear effects of the field \mathbf{E}_0 . The quantity E_p depends on the properties of the medium and the frequency of the wave. When $E_0 \gg E_p$, we get $T_e \gg T$; the electrons are intensely heated. Therefore changes in collisional frequency and hence the conductivity and dielectric constant are important. Nonlinear effects are profound. If the converse is true, the nonlinear effects are unimportant.

In the presence of an external dc magnetic field we can proceed in a similar manner and obtain

$$\frac{T_e}{T} = 1 + \left(\frac{E_0^2}{E_p}\right)(\omega^2 + \nu_0^2) \\ \times \left\{ \frac{\cos^2\beta}{\omega^2 + \nu^2} + \frac{\sin^2\beta}{2[(\omega - \omega_B)^2 + \nu^2]} + \frac{\sin^2\beta}{2[(\omega + \omega_B)^2 + \nu^2]} \right\}$$
(7.3.12)

where β is the angle between the field **E** and **B**₀.

The dependence of the effective collision frequency on the temperature may be obtained by considering the collision processes between various particles. For example, for collision between electrons and molecules, the effective collision frequency is given by

$$\nu(T_e) = \nu_0 (T_e/T)^{1/2} \tag{7.3.13}$$

For collisions with ions,

$$\nu(T_e) = \nu_0 (T/T_e)^{3/2} \tag{7.3.14}$$

7. Nonlinear Wave Propagation

Substituting (7.3.10) or (7.3.12) into (7.3.13) and (7.3.14), we obtain the dependence of the effective collision frequency on the field E_0 . Since the temperature does not depend on time, in the absence of the dc magnetic field, the dielectric constant is given by (4.1.24) and (4.1.25).

$$\varepsilon(\omega) = \varepsilon'(\omega) - j\varepsilon''(\omega) \tag{7.3.15}$$

where

$$\varepsilon'(\omega) = \varepsilon_0 \left[1 - \frac{\omega_p^2/\omega^2}{1 + \nu^2/\omega^2} \right]$$
(4.1.24)

$$\varepsilon''(\omega) = \varepsilon_0 \frac{\omega_p^2 \nu / \omega^3}{1 + \nu^2 / \omega^2}$$
(4.1.25)

The dependence of the dielectric constant on the field is obvious through (7.3.13), (7.3.14), and (7.3.10).

If external dc magnetic field is included, the dielectric tensor derived in Section 4.6 should be used.

7.4 Self-Interaction of Waves

From the discussion of the previous section, we have seen that when the condition $\omega \gg \delta v$ holds, the electron temperature in a field of any strength is constant in time to the first order of approximation. Consequently the polarization current J varies with the same frequency as the field E. Therefore, the dielectric constant discussed in Chapter 4 may be used directly to study wave propagation under this approximation. Two related non-linear phenomena may be discussed under the present assumption. One is the self-interaction effect and the other is the cross-modulation of waves. In this and the next section, we shall discuss these two topics with the ionospheric propagation condition in mind.

Let us consider a plane monochromatic electromagnetic wave propagating in a horizontally stratified isotropic plasma medium. For simplicity, we assume the wave to be normally incident. At the boundary of the plasma, say, z = 0, the electric field is given by $E_0(0) \cos \omega t$. From the discussions in Chapter 5 and the previous section, we know that under the assumption $\omega \gg \delta v$, the electric field inside the plasma is governed by the scalar wave equation

$$d^{2}E/dz^{2} + k_{0}^{2}\varepsilon(\omega, z, E_{0})E = 0$$
(7.4.1)

where $\varepsilon(\omega, z, E_0)$ is given by (7.3.15). For a weak field, ε does not depend on E_0 and (7.4.1) reduces to the wave equation that has been studied extensively in Chapter 5. For a strong field, ε depends on the amplitude E_0 of the field and (7.4.1) is nonlinear. If the properties of the medium are slowly varying, we may write the formal WKB solution of (7.4.1) in the form

$$E = C e^{-jk_0 \int_0^x n d\tau} e^{-k_0 \int_0^x x d\tau}$$
(7.4.2)

where to the first order C is a constant and

$$\varepsilon(\omega, z, E_0) = (n - j\varkappa)^2 \tag{7.4.3}$$

n and \varkappa are the real and imaginary parts of the refractive index, respectively, and both depend on ω , *z*, and E_0 . Note that in writing (7.4.2) we have neglected factor $[\varepsilon(0)/\varepsilon(z)]^{1/4}$ in the amplitude. For our present purpose this factor is not essential.

From (7.4.2) we note that the amplitude of the field is

$$E_0 = C e^{-k_0 \int_0^2 \kappa(\omega, \tau, E_0) d\tau}$$
(7.4.4)

Or, in another form

$$dE_0/dz + k_0 \varkappa(\omega, z, E_0) E_0 = 0 \tag{7.4.5}$$

This is a nonlinear differential equation for the amplitude of the field in the plasma. For a given \varkappa it will yield the amplitude as a function of z.

To obtain the explicit expression for \varkappa , we substitute (4.1.24) and (4.1.25) into (7.4.3) and solve for *n* and \varkappa . Under the condition that $|\varepsilon'| \gg |\varepsilon''|$ (this condition is satisfied in most cases for high frequency waves except at the reflection level where $\varepsilon = 0$), we obtain

$$\varkappa(z, E_0) = \varkappa_0(z) \frac{(\nu/\nu_0)(1 + \omega^2/\nu_0^2)}{(\nu/\nu_0)^2 + (\omega^2/\nu_0^2)}$$
(7.4.6)

where v and v_0 are the effective collision frequencies with and without strong field, respectively, and $\kappa_0(z)$ is the absorption coefficient in a weak field and is given by

$$\kappa_0(z) = \omega_p^2 \nu_0 / 2\omega (\omega^2 + \nu_0^2) [1 - \omega_p^2 (\omega^2 + \nu_0^2)]^{1/2}$$
(7.4.7)

In the following we also assume that ν_0 and the equilibrium electron temperature T are independent of z. We first consider the solution of (7.4.5) for the case where the dominant role is played by the electron-molecule collision. For this case $\nu/\nu_0 = [T_e(E_0)/T]^{1/2}$ as given by (7.3.13).

Defining a new variable

$$\tau = [T_e(E_0)/T]^{1/2} \tag{7.4.8}$$

we may write (7.3.10) as

$$\left(\frac{E_0}{E_p}\right)^2 = (\tau^2 - 1) \frac{\omega^2 + \nu_0^2 \tau^2}{\omega^2 + \nu_0^2}$$
(7.4.9)

Using (7.4.9), we may combined Eqs. (7.4.5) and (7.4.6) to yield an equation for τ :

$$\frac{d\tau}{dz}\left(\frac{1}{\tau^2-1}+\frac{2{\nu_0}^2}{\omega^2+{\nu_0}^2}\right)+k_0\varkappa_0(z)=0$$
 (7.4.10)

Its solution is

$$\frac{\tau - 1}{\tau + 1} \exp\left(\frac{4\nu_0^2}{\omega^2 + \nu_0^2} \tau\right) = \frac{\tau_0 - 1}{\tau_0 + 1} \exp\left(\frac{4\nu_0^2}{\omega^2 + \nu_0^2} \tau_0\right) \exp\left[-2K(z)\right]$$
(7.4.11)

where

$$\tau_0 = \tau(0) = \left(-\frac{T_e[E_0(0)]}{T}\right)^{1/2}$$
(7.4.12)

$$\tau = \tau(z) = \left(\frac{T_{\ell}[E_0(z)]}{T}\right)^{1/2}$$
(7.4.13)

and

$$K(z) = k_0 \int_0^z \varkappa_0(\tau) \, d\tau$$
 (7.4.14)

is the total absorption at the level z.

It is seen from (7.4.11) that τ_0 is the maximum value of τ . With increasing z, hence K(z), τ diminishes monotonically. Deep into the plasma where $K(z) \gg 1$, τ tends to unity. From (7.4.9), this means the amplitude of the wave becomes very small.

With the solution for τ given by (7.4.11), the amplitude of the field may be obtained from (7.4.9). It is best represented in the form

$$E_0(z) = E_0(0)e^{-K(z)}P[E_0(0)/E_p, \omega/\nu_0, K(z)]$$
(7.4.15)

where P is called the self-interaction factor and in general is a function of $E_0(0)/E_p$, ω/v_0 , and K(z). Obviously, P is very close to unity for a weak field so that the wave is attenuated in the plasma according to the ordinary

376

absorption law. Deep into the plasma where $K(z) \gg 1$, we obtain a simple expression for P

$$P = 2 \frac{E_p}{E_0(0)} \left(\frac{\tau_0 - 1}{\tau_0 + 1}\right)^{1/2} \exp\left\{\frac{2\nu_0^2}{\omega^2 + \nu_0^2} (\tau_0 - 1)\right\} \quad (7.4.16)$$

For high frequency strong waves such that $\omega^2 \gg \nu_0^2 \tau_0$ and $\tau_0 \gg 1$ the factor P is independent of τ_0 and the amplitude deep in the plasma becomes

$$E_0(z) = 2E_p e^{-K(z)} \tag{7.4.17}$$

which is independent of $E_0(0)$, the amplitude at the boundary.

If the opposite condition $\omega^2 \ll v_0^2 \tau_0$ is satisfied, then for a very strong field such that $\tau_0 \gg 1$, the amplitude deep in the plasma becomes

$$E_0(z) = 2E_p \exp\left[2\left\{\left(\frac{\nu_0^2}{\omega^2 + \nu_0^2}\right)^{3/2} \frac{E_0(0)}{E_p}\right\}^{1/2}\right] \exp\left[-K(z)\right] (7.4.18)$$

which increases exponentially as $E_0(0)$ increases. Thus we see that for the low frequency waves, deep into the plasma, the absorption of the wave is very much reduced for a strong field. This is due to the effects of self-interaction. At low frequencies, the absorption coefficient of the wave in a plasma decreases as the electrons are heated.

The general expression for the self-interaction factor P for arbitrary values of z may be obtained from (7.4.9) and (7.4.11), but is very complicated. A simple expression may be derived for the high frequency case for which $\omega^2 \gg v_0^2 \tau$. We have

$$P = \frac{2E_p}{E_0(0)} \left(\frac{\tau_0 - 1}{\tau_0 + 1}\right)^{1/2} \frac{1}{1 - [(\tau_0 - 1)/(\tau_0 + 1)]e^{-2K(z)}}$$
(7.4.19)

We see that for a very weak field, $\tau_0 \cong 1 + \frac{1}{2} [E_0(0)/E_P]^2$ and $P \cong 1$ as expected.

The above discussion is for electron-molecule collisions. For the case where electron-ion collisions are dominant, similar types of analysis can be made and will not be discussed here.

Up to this point, we have assumed that the incident wave is monochromatic. If instead the wave is amplitude-modulated at a low frequency Ω , then the self-interaction effect in the plasma may change substantially the modulation of the wave. If the modulation frequency Ω is very low (much less than δv_0), the problem of propagation of an amplitude-modulated wave in a plasma is essentially identical with the one considered above, that of propagation of an unmodulated wave. At the boundary z = 0,

7. Nonlinear Wave Propagation

the modulated wave is $E_0(0, t) = E_0(0)(1 + M \cos \Omega t)$, where M is the modulation index. All expressions derived above will be valid with the substitution of $E_0(0, t)$ for $E_0(0)$. In particular, the amplitude of the wave in the plasma may be written as

$$E_0(z, t) = E_0(0, t)e^{-K(z)} P\left[\frac{E_0(0, t)}{E_p}, \frac{\omega}{v_0}, K(z)\right]$$
(7.4.20)

Because of the nonlinear interaction, not only the modulation index is changed but also harmonics with frequencies 2Ω , 3Ω , ... are introduced. The wave shape is hence changed. An example showing the form of strong waves deep in the plasma is shown in Fig. 7.4-1.



Fig. 7.4-1. The effect of selfinteraction on an amplitude-modulated wave. [From Ginzburg and Gurevich (1960).]

If the modulation frequency is not small compared with δv_0 , then the discussion given is not valid. One has to go back to (7.3.3) and (7.4.5) to solve the problem over again.

To summarize the results of this section, we note that for very strong waves, with amplitudes much greater than the plasma field, the absorption of the wave in the plasma differs from that of a weak field even qualitatively. This is due to the self-interaction of the strong wave. For an amplitudemodulated wave, the modulation form can be affected tremendously. Finally, our discussion has been on the amplitude of the wave only. The effects of self-interaction on the phase of the wave may also be discussed in a similar manner.

7.5 Cross-Modulation Phenomenon

In the last section we have seen that when a strong electromagnetic wave propagates through a plasma, the perturbation it causes in the plasma affects the propagation of the wave itself. Obviously, if other waves propagate at the same time through the perturbed region, they will also be affected. If the intense wave is amplitude-modulated with a low frequency Ω , the perturbation in the plasma is also modulated as discussed in the previous section. When a second wave propagates through this perturbed region, it will also be modulated. This is the well-known Luxembourg effect in the ionosphere, also called the cross-modulation phenomenon. In this section, we shall treat this problem in an elementary fashion.

Let us assume that an amplitude-modulated wave is propagating in the z-direction. In the approximation of geometric optics (Chapter 5), the field may be written as

$$E_{1}(z, t) = \left[\frac{\varepsilon_{1}(0)}{\varepsilon_{1}(z)}\right]^{1/4} E_{01}(0)(1 + M_{1} \cos \Omega t)e^{-K_{1}(z)}$$
$$\times \cos\left[\omega_{1}t - k_{1} \int_{0}^{z} n_{1} d\tau\right] \cdot P$$
(7.5.1)

where $k_1 = \omega_1/c$, $K_1(z)$ is the absorption at z for ω_1 , and P is the self-interaction factor. The amplitude of the wave is given by

$$E_{01}(z,t) = \left[\frac{\varepsilon_1(0)}{\varepsilon_1(z)}\right]^{1/4} E_{01}(0) [1 + M_1 \cos \Omega t] e^{-K_1(z)} \cdot P \quad (7.5.2)$$

For simplicity, in the following computation of the perturbation of the plasma we shall neglect the self-interaction effect and set P = 1. At first glance, this seems to be contradictory, since the wave is assumed to be strong and the self-interaction certainly is important. But it so happens that in a plasma cross-modulation can be easily observed even for $(T_e - T)/T \ll 1$. Therefore the case we shall consider is the one in which the wave is strong enough to cause perturbation in the plasma—yet at the same time weak enough so that self-interaction may be neglected. This case includes many practically observable situations, especially in the ionosphere.

For a very small modulation frequency Ω , the amplitude of (7.5.1) may be considered as quasi-steady. Since the momentum relaxation time is very short, in computing the velocity $\mathbf{u}(t)$ (7.3.6) may be used directly with the substitutions of \mathbf{E}_0 by the amplitude in (7.5.1) and ωt by $\omega_1 t - k_1 \int_0^z n_1 d\tau$, respectively. For the temperature T_e , a better approximation must be made since T_e relaxes much more slowly. For $\omega \gg \Omega$, (7.3.7) may be written as

$$\frac{dT_e}{dT} + \delta \nu_0 T_e = \frac{e^2 E_{01}(z, t)}{3m(\nu_0^2 + \omega^2)} \nu_0 + \delta \nu_0 T$$
(7.5.3)

where v_0 is used since self-interaction is neglected and $E_{01}(z, t)$ is given by (7.5.2).

Solving (7.5.3), we see that the part of the temperature perturbation $\Delta T_e = T_e - T$ which varies with frequencies Ω and 2Ω is given by

$$\frac{\Delta T_e}{T} = \frac{2C}{\omega_1^2 + \nu_0^2} \left(\frac{\varepsilon_1(0)}{\varepsilon_1(z)}\right)^{1/2} e^{-2K_1(z)} \times [A_1 \cos(\Omega t - \phi_1) + A_2 \cos(2\Omega t - \phi_2)]$$
(7.5.4)

where

$$C = M_1 e^2 E_{01}^2(0)/3Tm\delta$$

$$A_1 = \delta \nu_0 / (\delta^2 \nu_0^2 + \Omega^2)^{1/2}, \quad A_2 = M_1 \delta \nu_0 / 4 (\delta^2 \nu_0^2 + 4\Omega^2)^{1/2} \quad (7.5.5)$$

$$\phi_1 = \tan^{-1}(\Omega/\delta\nu_0), \quad \phi_2 = \tan^{-1}(2\Omega/\delta\nu_0)$$

The change in temperature in turn causes a change in the collision frequency. Again, let us consider the case where electron-molecule collisions are dominant. For this case $v = v_0 (T_e/T)^{1/2}$. Therefore

$$\Delta \nu \simeq \nu_0 \, \Delta T_e/2T \tag{7.5.6}$$

Now suppose a weak, unmodulated plane wave with frequency ω_2 propagates through the perturbed region. The amplitude of this wave in the geometric approximation is given by

$$E_{02}(z, t) = \left[\frac{\varepsilon_2(0)}{\varepsilon_2(z)}\right]^{1/4} E_{02}(0)e^{-K_1(z)}$$
(7.5.7)

where the absorption is given by

$$K_{2}(z) = (\omega_{2}/c) \int_{0}^{z} \varkappa_{2}(\tau) d\tau \qquad (7.5.8)$$

As is shown in (7.4.7), the absorption coefficient \varkappa_2 depends on the collision frequency. Therefore in the perturbed region \varkappa_2 is modified. For small

 $\Delta \nu$, we may write

$$\varkappa_2 \cong \varkappa_2(\nu_0) + (\partial \varkappa_2 / \partial \nu_0) \, \varDelta \nu \tag{7.5.9}$$

The amplitude of the second wave may now be written as

$$E_{02}(z, t) = \left[\frac{\varepsilon_2(0)}{\varepsilon_2(z)}\right]^{1/4} E_{02}(0) \exp\left[-\frac{\omega_2}{c} \int_0^z \varkappa_2(\nu_0, \tau) d\tau\right] \\ \times \left[1 - \frac{\omega_2}{2c} \int_0^z \nu_0 \frac{\partial \varkappa_2(\nu_0)}{\partial \nu_0} \frac{\Delta T_e}{T} d\tau\right]$$
(7.5.10)

Substituting (7.5.4) into (7.5.10), we obtain

 $E_{02}(z, t) = \text{constant} \times [1 - M_{\Omega} \cos(\Omega t - \phi_1) - M_{2\Omega} \cos(2\Omega t - \phi_2)]$ (7.5.11)

where

$$M_{\Omega} = C \frac{\omega_2}{c} \int_0^z \nu_0 \frac{\partial \varkappa_2(\nu_0)}{\partial \nu_0} \frac{A_1}{\omega_1^2 + \nu_0^2} \left[\frac{\varepsilon_1(0)}{\varepsilon_1(\tau)} \right]^{1/2} e^{-2\varkappa_1(\tau)} d\tau \qquad (7.5.12)$$

$$M_{2\Omega} = M_1 C \frac{\omega_2}{c} \int_0^z \nu_0 \frac{\partial \varkappa_2(\nu_0)}{\partial \nu_0} \frac{A_2}{\omega_1^2 + \nu_0^2} \frac{[\varepsilon_1(0)]^{1/2}}{\varepsilon_1(\tau)} e^{-2\varkappa_1(\tau)} d\tau \quad (7.5.13)$$

We see that due to the modulation of the collision frequency caused by the passing of the first modulated wave, the second wave is also amplitude-modulated at frequencies Ω and 2Ω . The modulation indices are given by M_{Ω} and $M_{2\Omega}$ for the two frequencies, respectively. The depth of modulation (indicated by M_{Ω} and $M_{2\Omega}$) and also the phase depend on the parameters of the plasma as well as the geometry of the problem. Here we just note that both A_1 and A_2 decrease as Ω increase. Since M_{Ω} and $M_{2\Omega}$ are proportional to A_1 and A_2 , respectively, the depth of the cross-modulation decreases as Ω increases. Also, in general the modulation of the second harmonic is smaller than that of the first harmonic. This is evident by examining the following relation obtained from (7.5.12) and (7.5.13).

$$M_{2\Omega} = M_{\Omega} \frac{M_1}{4} \frac{[(\delta \nu_0)^2 + \Omega^2]^{1/2}}{[(\delta \nu_0)^2 + 4\Omega^2]^{1/2}}$$
(7.5.14)

In our computation, we have assumed normal incidence for both waves. The case of oblique incidence can be treated in the similar manner. When the external dc magnetic fields are present, the computation becomes more complicated. However results similar to those obtained above for isotropic plasma may also be obtained.
7.6 Wave-Wave Interaction

In the previous two sections we have discussed the nonlinear effects of wave propagation in a plasma where collisions play a significant role. The dependence of the effective collision frequency on the electromagnetic fields causes the self-interaction as well as cross-modulation of waves. Other phenomena such as harmonic generation, demodulation, etc., may also occur due to a similar mechanism. In this section we shall discuss another aspect of the problem of nonlinear wave propagation. We shall study in general the nonlinear phenomena that arise independent of the collisional effects. The physical processes underlying this type of nonlinear phenomena may be divided into two catagories. The first one is the resonant interaction of waves in which the energy and momentum of the interacting waves are conserved. The second is the resonant interaction between waves and particles in which the total energy and total momentum of waves and particles are conserved. In the following we shall concentrate on the discussion of the first class. Since our purpose is mainly to indicate the ideas and demonstrate the techniques rather than to study a specific physical problem, we choose to investigate the following dimensionless nonlinear partial differential equation

$$\frac{\partial^2 \psi}{\partial t^2} + \frac{\partial^4 \psi}{\partial x^4} + \frac{\partial^2 \psi}{\partial x^2} + \psi = \varepsilon \psi^2$$
(7.6.1)

This equation may represent a dispersive wave system with the nonlinear term given on the right-hand side. We shall assume in the following that ε is a small parameter corresponding to weak interactions. Using (7.6.1), we shall first indicate qualitatively the physical processes underlying the wave-wave interaction phenomenon; then we shall introduce one of the mathematical techniques that are used to study such problems. The main reason for choosing (7.6.1) is just because it represents one of the simplest nonlinear dispersive wave systems for which three-wave resonant interaction is possible. Armed with ideas and results obtained from considering (7.6.1) we can then go on to discuss some wave-wave interaction phenomena that may occur in the real physical world.

Let us first consider (7.6.1) without the nonlinear term on the righthand side. This is a one-dimensional linear dispersive wave system and from the discussion in Chapter 2 we know that for plane waves of the form

$$\psi(x,t) \propto e^{j(\omega t - kx)} \tag{7.6.2}$$

to exist, ω and k must satisfy the dispersion relation

$$\omega^2(k) = 1 - k^2 + k^4 \tag{7.6.3}$$

In the linear system, plane waves satisfying the dispersion relation will propagate in the medium independently with constant amplitudes; there is no energy exchange between the modes. When the nonlinear term on the right-hand side is included, however, interaction among the waves occurs. Energy exchange between the modes does exist. Therefore the amplitudes of the waves will change as time goes on. For weak nonlinearity one would expect to see a slow change in the amplitudes as compared with the fast changing phase. The interaction becomes most important when two waves beat together such that their sum or difference frequency and wave number just match the frequency and wave number of a third wave. We shall call this case resonant wave-wave interaction. Mathematically the resonant condition may be written as

$$\omega_1(k_1) + \omega_2(k_2) + \omega_3(k_3) = 0$$

$$k_1 + k_2 + k_3 = 0$$
(7.6.4)

If one interprets ω and k as the energy and momentum of a quantum associated with the wave, (7.6.4) may be regarded as the principles of conservation of energy and momentum in the process involving the three quanta.

Let us now consider the case of three-wave interaction for the system (7.6.1). Assume a solution of the form

$$\psi(x,t) = \sum_{n=1}^{3} \left[a(k_n,t) e^{-jk_n x} + a(-k_n,t) e^{jk_n x} \right]$$
(7.6.5)

where $a^*(k_n, t) = a(-k_n, t)$ for real $\psi(x, t)$. This is a combination of three interacting waves with wave numbers satisfying the condition

$$k_1 + k_2 + k_3 = 0 \tag{7.6.6}$$

Substituting (7.6.5) into (7.6.1) and equating the terms on both sides with the same exponential factor, we obtain

$$\frac{d^2a(k_n,t)}{dt^2} + \omega_n^2 a(k_n,t) = \varepsilon \sum_{m \neq n, m-1, 2, 3} \frac{a^*(k_m,t)a^*(-k_n-k_m,t)}{n=1, 2, 3}, \quad (7.6.7)$$

where

$$\omega_n^2 = \omega^2(k_n) = 1 - k_n^2 + k_n^4, \quad n = 1, 2, 3$$
 (7.6.8)

A similar set of equations may be obtained for $a(-k_n, t)$.

We note that for $\varepsilon = 0$, there is no nonlinear interaction and the $a(k_i, t)$'s are independent of each other. The solutions are $\exp(\pm j\omega_n t)$ for this case and (7.6.5) represents three independent waves.

For $\varepsilon \neq 0$, (7.6.7) represents three coupled nonlinear equations. In general they are difficult to solve. For weak nonlinearity such that $\varepsilon \ll 1$, approximate solutions can be obtained by several different techniques. The central idea behind all these techniques is that for weak nonlinearity, the function $a(k_i, t)$ can be expressed as the product of a slowly time-varying amplitude function and the exponential $\exp(\pm j\omega_i t)$. Therefore two different time scales exist, a slow one on which the amplitude varies and a fast one on which the phase changes. In the following we shall introduce one of the techniques, the so-called "method of averaging," to solve (7.6.7).

To apply the method of averaging we first transform (7.6.7) into the socalled "standard form" (to be defined later) by defining a vector

$$\boldsymbol{\xi}(k_n, t) = \begin{bmatrix} a(k_n, t) \\ da(k_n, t)/dt \end{bmatrix} = \begin{bmatrix} \xi_1(k_n, t) \\ \xi_2(k_n, t) \end{bmatrix}$$
(7.6.9)

Then (7.6.7) can be put into the form

$$d\boldsymbol{\xi}/dt + \boldsymbol{A} \cdot \boldsymbol{\xi} = \varepsilon \mathbf{X}(k_n, t, \boldsymbol{\xi})$$
(7.6.10)

where

$$\mathbf{A}(k_n) = \begin{bmatrix} 0 & -1\\ \omega_n^2 & 0 \end{bmatrix}$$
(7.6.11)

and

$$\mathbf{X}(k_n, t, \mathbf{\xi}) = \begin{bmatrix} 0 \\ \sum_{m \neq n, m = 1, 2, 3} \xi_1^*(k_m, t) \xi_1^*(-k_n - k_m, t) \end{bmatrix}$$
(7.6.12)

In order to transform (7.6.10) into the standard form, we introduce the transformation

$$\boldsymbol{\xi}(k_n, t) = \boldsymbol{\Phi}(k_n, t) \cdot \boldsymbol{\eta}(k_n, t)$$
(7.6.13)

where $\mathbf{\Phi}(k_n, t)$ is a 2 \times 2 matrix and $\mathbf{\eta}$ is a column vector. If we choose $\mathbf{\Phi}$ to be

$$\mathbf{\Phi}(k_n, t) = \begin{bmatrix} e^{j\omega_n t} & e^{-j\omega_n t} \\ j\omega_n e^{j\omega_n t} & -j\omega_n e^{-j\omega_n t} \end{bmatrix}$$
(7.6.14)

then by substituting (7.6.13) and (7.6.14) into (7.6.10) we have

$$d\mathbf{\eta}(k_n, t)/dt = \varepsilon \mathbf{\Phi}^{-1}(k_n, t) \cdot \mathbf{X}(k_n, t, \mathbf{\Phi} \cdot \mathbf{\eta}), \qquad n = 1, 2, 3 \quad (7.6.15)$$

384

which is the standard form of the equation. Φ^{-1} is the inverse of Φ and is given by

$$\mathbf{\Phi}^{-1}(k_n, t) = \frac{1}{-2j\omega_n} \begin{bmatrix} -j\omega_n e^{-j\omega_n t} & -e^{-j\omega_n t} \\ -j\omega_n e^{j\omega_n t} & e^{j\omega_n t} \end{bmatrix}$$
(7.6.16)

 Φ is called the fundamental matrix for the homogeneous equation

 $d\xi/dt + \mathbf{A} \cdot \mathbf{\xi} = 0$

and we note that the columns of Φ are vector solutions of this equation.

Using (7.6.12), (7.6.14), and (7.6.16), we can write (7.6.15) explicitly in its component form. For example, for n = 1, we have

$$d\eta_{1}(k_{1}, t)/dt = (\varepsilon/j\omega_{1})\{\eta_{1}^{*}(k_{2}, t)\eta_{1}^{*}(k_{3}, t)e^{-j(\omega_{1}+\omega_{2}+\omega_{3})t} + \eta_{1}^{*}(k_{3}, t)\eta_{2}^{*}(k_{2}, t)e^{-j(\omega_{1}+\omega_{3}-\omega_{2})t} + \eta_{1}^{*}(k_{2}, t)\eta_{2}^{*}(k_{3}, t)e^{-j(\omega_{1}+\omega_{2}-\omega_{3})t} + \eta_{2}^{*}(k_{2}, t)\eta_{2}^{*}(k_{3}, t)e^{j(\omega_{2}+\omega_{3}-\omega_{1})t}\}$$
(7.6.17)
$$d\eta_{2}(k_{1}, t)/dt = -(\varepsilon/j\omega_{1})\{\eta_{1}^{*}(k_{2}, t)\eta_{1}^{*}(k_{3}, t)e^{-j(\omega_{2}+\omega_{3}-\omega_{1})t} + \eta_{1}^{*}(k_{3}, t)\eta_{2}^{*}(k_{3}, t)e^{-j(\omega_{3}-\omega_{2}-\omega_{1})t} + \eta_{1}^{*}(k_{2}, t)\eta_{2}^{*}(k_{3}, t)e^{-j(\omega_{3}-\omega_{3}-\omega_{1})t} + \eta_{2}^{*}(k_{2}, t)\eta_{2}^{*}(k_{3}, t)e^{-j(\omega_{1}+\omega_{2}+\omega_{3})t}\}$$
(7.6.18)

Similar equations for $\eta_1(k_2, t)$, $\eta_2(k_2, t)$, $\eta_1(k_3, t)$, and $\eta_2(k_3, t)$ can be derived. They all have the same form. The subscripts on k and ω are cyclicly symmetric. Thus we have six coupled nonlinear equations in the standard form for the vector $\eta(k_n, t)$. Once η is solved, the solution $a(k_n, t)$ is obtained through (7.6.13). We have

$$a(k_n, t) = \xi_1(k_n, t) = \eta_1(k_n, t)e^{j\omega_n t} + \eta_2(k_n, t)e^{-j\omega_n t} \quad (7.6.19)$$

For small ε , from (7.6.17) and (7.6.18) we see that η_1 and η_2 vary slowly with respect to time. The expression of $a(k_n, t)$ is then in the form of the product of a slowly varying amplitude and a fast varying exponential as previously mentioned. To obtain the solution η from the set of nonlinear differential equations, the method of averaging will be used. This method essentially is based on the fact that η varies on a much slower time scale than the time scale of the exponential function. It can be shown that the

7. Nonlinear Wave Propagation

asymptotic solutions for (7.6.17) and (7.6.18) to the first order of the small parameter ε satisfy the equations obtained by taking the average of the original equations over the fast time scale. In taking this average, η_1 and η_2 are kept constant. When this averaging method is applied to (7.6.17) and (7.6.18), we note that the terms on the right-hand side vanish identically except for the various resonant conditions $\omega_1 + \omega_2 + \omega_3 = 0$, $\omega_1 + \omega_3 - \omega_2 = 0$, $\omega_1 + \omega_2 - \omega_3 = 0$, or $\omega_2 + \omega_3 - \omega_1 = 0$. Each condition corresponds to one particular physical situation. In the following we shall study in detail the case where three traveling waves interact with each other such that all η_2 's are absent. For this case, the equations after averaging become

$$d\eta_{1}(k_{1}, t)/dt = (\varepsilon/j\omega_{1})\eta_{1}^{*}(k_{2}, t)\eta_{1}^{*}(k_{3}, t)$$

$$d\eta_{1}(k_{2}, t)/dt = (\varepsilon/j\omega_{2})\eta_{1}^{*}(k_{1}, t)\eta_{1}^{*}(k_{3}, t)$$

$$d\eta_{1}(k_{3}, t)/dt = (\varepsilon/j\omega_{3})\eta_{1}^{*}(k_{1}, t)\eta_{1}^{*}(k_{2}, t)$$

(7.6.20)

with the resonant conditions

$$k_1 + k_2 + k_3 = 0, \qquad \omega_1 + \omega_2 + \omega_3 = 0$$

Two possible resonant trios for the dispersion equation $\omega_n = 1 - k_n^2 + k_n^4$ are shown in Fig. 7.6-1. Before we solve (7.6.20) in detail, let us first



Fig. 7.6-1. Two possible resonant trios for $\omega^2(k) = 1 - k^2 + k^4$. Waves traveling in positive x-direction are represented by dots; waves traveling in negative x-direction are represented by crosses. [From F. P. Bretherton (1964). Resonant interaction between waves. J. Fluid Mech. 20, 457-479. By permission of Cambridge Univ. Press]. combine them with their respective complex conjugate equations. We can write

$$\omega_{1}(d/dt) | \eta_{1}(k_{1}, t) |^{2} = \omega_{2}(d/dt) | \eta_{1}(k_{2}, t) |^{2}$$

= $\omega_{3}(d/dt) | \eta_{1}(k_{3}, t) |^{2}$
= $-2 \operatorname{Im} \{ \varepsilon \eta_{1}(k_{1}, t) \eta_{1}(k_{2}, t) \eta_{1}(k_{3}, t) \}$ (7.6.21)

where Im indicates the imaginary part.

For ω_1 and ω_2 both positive, ω_3 must be negative according to the resonant condition. (7.6.21) then indicates that when the amplitudes of waves 1 and 2 increase with time, that of wave 3 must decrease with time. In general, if one amplitude is increasing with time, then at least one of the other amplitudes must be decreasing.

To solve (7.6.20), we write

$$\eta_1(k_n, t) = \alpha_n e^{j\phi_n}, \quad n = 1, 2, 3$$
 (7.6.22)

where α_n and ϕ_n are real functions of k_n and t.

Substitute (7.6.22) into (7.6.20) and equating the real and imaginary parts, we obtain

$$d\alpha_1/dt = -(\varepsilon/\omega_1)\alpha_2\alpha_3\sin\theta$$

$$d\alpha_2/dt = (\varepsilon/\omega_2)\alpha_1\alpha_3\sin\theta$$
 (7.6.23)

$$d\alpha_3/dt = (\varepsilon/\omega_3)\alpha_1\alpha_2\sin\theta$$

and

$$d\phi_1/dt = -(\epsilon \alpha_2 \alpha_3/\omega_1 \alpha_1) \cos \theta$$

$$d\phi_2/dt = -(\epsilon \alpha_1 \alpha_3/\omega_2 \alpha_2) \cos \theta$$

$$d\phi_3/dt = -(\epsilon \alpha_1 \alpha_2/\omega_3 \alpha_3) \cos \theta$$

(7.6.24)

where

$$\theta = \phi_1 + \phi_2 + \phi_3 \tag{7.6.25}$$

From (7.6.24), we obtain immediately

$$d\theta/dt = \cos\theta(d/dt)\ln(\alpha_1\alpha_2\alpha_3) \tag{7.6.26}$$

which yields a relation after integration

$$\alpha_1 \alpha_2 \alpha_3 \cos \theta = \Gamma = \text{constant}$$
 (7.6.27)

From (7.6.23), we have

$$\omega_1 \, d\alpha_1^2/dt - \omega_2 \, d\alpha_2^2/dt = 0$$

$$\omega_1 \, d\alpha_1^2/dt - \omega_3 \, d\alpha_3^2/dt = 0$$

$$\omega_2 \, d\alpha_2^2/dt - \omega_3 \, d\alpha_3^2/dt = 0$$

(7.6.28)

Integrating (7.6.28) with respect to t, the following can be obtained

$$\omega_{1}\alpha_{1}^{2} - \omega_{2}\alpha_{2}^{2} = m_{1}$$

$$\omega_{1}\alpha_{1}^{2} - \omega_{3}\alpha_{3}^{2} = m_{2}$$

$$\omega_{2}\alpha_{2}^{2} - \omega_{3}\alpha_{3}^{2} = m_{3}$$

(7.6.29)

where m_1 , m_2 , and m_3 are constants.

Let us now consider the first equation of (7.6.23). Using (7.6.27) and (7.6.29), we can write

$$d\alpha_{1}^{2}/dt = -(2\varepsilon/\omega_{1})\alpha_{1}\alpha_{2}\alpha_{3}\sin\theta$$

= $-(2\varepsilon/\omega_{1})[\alpha_{1}^{2}\alpha_{2}^{2}\alpha_{3}^{2}(1-\cos^{2}\theta)]^{1/2}$
= $-(2\varepsilon/\omega_{1})[\alpha_{1}^{2}(\omega_{1}\alpha_{1}^{2}/\omega_{2}-m_{1}/\omega_{2})(\omega_{1}\alpha_{1}^{2}/\omega_{3}-m_{2}/\omega_{3})-\Gamma^{2}]^{1/2}$
(7.6.30)

Defining

$$\alpha_1^2 = n_1, \qquad \alpha_2^2 = n_2, \qquad \alpha_3^2 = n_3$$
 (7.6.31)

(7.6.30) may be put in the form

$$dn_1/dt = -(2\varepsilon/\omega_1)[(n_1 - n_{1a})(n_1 - n_{1b})(n_1 - n_{1c})]^{1/2} \quad (7.6.32)$$

where n_{1a} , n_{1b} , and n_{1c} are the roots for the equation

$$n_1(\omega_1 n_1/\omega_2 - m_1/\omega_2)(\omega_1 n_1/\omega_3 - m_2/\omega_3) - \Gamma^2 = 0 \qquad (7.6.33)$$

and are arranged in the following manner

$$n_{1c} \ge n_{1b} \ge n_{1a} \ge 0$$

Equation (7.6.32) may be integrated formally to give

$$\frac{2\varepsilon}{\omega_1}(t_0-t) = \int_{n_1(t_0)}^{n_1(t)} \frac{dn_1}{[(n_1-n_{1c})(n_1-n_{1b})(n_1-n_{1a})]^{1/2}} \quad (7.6.34)$$

388

This equation can be transformed into the standard elliptic integral by the following change of variable. Let

$$y(t) = \left[\frac{n_1(t) - n_{1a}}{n_{1b} - n_{1a}}\right]^{1/2}$$
(7.6.35)

$$\beta = \left[\frac{n_{1b} - n_{1a}}{n_{1c} - n_{1a}}\right]^{1/2}$$
(7.6.36)

Substituting the new variable y into (7.6.34), we may reduce it to an elliptic integral,

$$\frac{\varepsilon}{\omega_1} (t_0 - t)(n_{1c} - n_{1a})^{1/2} = \int_0^{y(t)} \frac{dy}{((1 - y^2)(1 - \beta^2 y^2))^{1/2}} (7.6.37)$$

where $y(t_0) = 0$ has been assumed. This relation may be taken as the definition for t_0 .

The solution of (7.6.37) is given by the so-called Jacobian elliptic function $sn(u, \beta)$ (Gradshteyn and Ryzhik, 1965).

$$y(t) = \operatorname{sn}\left[\frac{\varepsilon}{\omega_1} (n_{1c} - n_{1a})^{1/2} (t_0 - t), \beta\right]$$
(7.6.38)

and, by the definition of y, the solution for $n_1(t)$, hence $\alpha_1^2(t)$, is given by

$$\alpha_1^{2}(t) = n_1(t) = n_{1a} + (n_{1b} - n_{1a}) \operatorname{sn}^2 \left[\frac{\varepsilon}{\omega_1} (n_{1c} - n_{1a})^{1/2} (t_0 - t), \beta \right] (7.6.39)$$

The amplitude functions α_2 and α_3 may be obtained from (7.6.29). Once α_1 , α_2 , and α_3 are known, (7.6.27) determines θ and (7.6.24) may be used to to calculate ϕ_1 , ϕ_2 , and ϕ_3 . The general solutions $\eta_1(k_n, t)$ are then obtained through (7.6.22). The elliptic function $\operatorname{sn}(u, \beta)$ is a tabulated special function. It reduces to the common trigonometric function sin u when $\beta = 0$.

To visualize the physical significance of our derivation we consider the following example. At t = 0, let us assume that the amplitude of the first wave $\alpha_1(0) = 0$. Also, we assume that $\alpha_2^2(0) \gg \alpha_3^2(0)$; the second wave has dominating amplitude at t = 0. Without loss of generality, we can put $\Gamma = 0$ in (7.6.27). From (7.6.29), we have

$$m_1 = -\omega_2 n_2(0), \qquad m_2 = -\omega_3 n_3(0)$$
 (7.6.40)

7. Nonlinear Wave Propagation

Substituting (7.6.40) into (7.6.33), for $\Gamma = 0$, we have

$$n_{1c} = \left(\frac{\omega_2}{\omega_1}\right) n_2(0) = \frac{\omega_2}{\omega_2 + \omega_3} n_2(0)$$

$$n_{1b} = \left(\frac{\omega_3}{\omega_1}\right) n_3(0) = \frac{\omega_3}{\omega_2 + \omega_3} n_3(0)$$
(7.6.41)
$$n_{1a} = 0$$

For this case, since $n_2(0) \gg n_3(0)$, β is very small. Therefore the elliptic function $\operatorname{sn}(u, \beta)$ may be approximated by $\sin u$. The solutions are therefore given by

$$\alpha_{1}^{2}(t) = n_{1}(t) = \frac{\omega_{3}}{\omega_{2} + \omega_{3}} n_{3}(0) \sin^{2} \left[\frac{\varepsilon}{\omega_{1}} t \left(\frac{\omega_{2}}{\omega_{2} + \omega_{3}} n_{2}(0) \right)^{1/2} \right]$$

$$\alpha_{2}^{2}(t) = n_{2}(t) = n_{2}(0) - \frac{\omega_{3}}{\omega_{2}} n_{3}(0) \sin^{2} \left[\frac{\varepsilon}{\omega_{1}} t \left(\frac{\omega_{2}}{\omega_{2} + \omega_{3}} n_{2}(0) \right)^{1/2} \right]$$

$$\alpha_{3}^{2}(t) = n_{3}(t) = n_{3}(0) \cos^{2} \left[\frac{\varepsilon}{\omega_{1}} t \left(\frac{\omega_{2}}{\omega_{2} + \omega_{3}} n_{2}(0) \right)^{1/2} \right]$$
(7.6.42)

In Fig. 7.6-2, the functions α_1^2 , α_2^2 , and α_3^2 are plotted against time. We see that although initially the wave with frequency ω_1 has zero amplitude, through the interaction of waves 2 and 3, the wave with frequency ω_1 is generated.



Fig. 7.6-2. Amplitudes of the three interacting waves.

Thus, using the relatively simple equation (7.6.1), we have demonstrated one of the techniques in treating the problem of wave-wave interaction. The example is for one-dimensional waves but it can be extended immediately for the case of three-dimensional waves. The resonant conditions then become

$$\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = \mathbf{0}$$

$$\omega_1(\mathbf{k}_1) + \omega_2(\mathbf{k}_2) + \omega_3(\mathbf{k}_3) = \mathbf{0}$$
(7.6.43)

and the rest of the above discussion applies. Furthermore, our discussion has been on the simplest case of a three-wave system. Similar consideration may be made for multiwave systems. In some media where more than one type of wave can exist, then ω_1 , ω_2 , and ω_3 in (7.6.43) may satisfy the different dispersion relations corresponding to the different modes, respectively.

The technique we discussed above can be used to study various wavewave interaction phenomena that occur in real physical systems. For instance, in a plasma there may exist many types of waves as discussed in Chapters 3 and 4. The technique may be used to study the resonant interactions among these waves. It is possible under certain conditions that two transverse waves will interact to generate a longitudinal plasma wave or ion acoustic wave; or two transverse waves may generate another transverse wave propagating in different direction with the beat frequency. It has been suggested (Harker and Crawford, 1969) that resonant interaction among the whistler modes may be the source of low frequency noise in the magnetosphere. The technique has also been used successfully in the study of weak turbulence in a plasma as well as surface waves in the ocean. In both cases, large numbers of waves interact with each other resonantly in a random fashion.

A special case in the resonant interaction phenomenon is when one wave has very low frequency as compared with the other two waves. For this case, in some physical systems, the resonant interaction may serve as a mechanism in trapping the waves. The low frequency wave does not participate in the energy exchanging process; rather it acts as some sort of catalyst to promote the interaction of the two high frequency waves. They interact in such a way that the energy is exchanged from one wave to another continuously and the propagation is confined in a spatial duct. This type of interaction may occur in the ocean for internal gravity waves (Phillips, 1968), or in the atmosphere for acoustic-gravity waves (Yeh and Liu, 1970).

We mentioned in the beginning of this section that another type of nonlinear phenomenon involves the resonant interaction among both waves and particles. This type of interaction occurs quite often in plasma. The well-known Landau damping is a linear example of the phenomenon. Techniques such as the kinetic equations for waves have been developed to study these problems. Interested readers are referred to the book by Sagdeev and Galeev (1969).

7.7 An Averaged Variational Principle

In closing this chapter, we wish to discuss briefly one of the recently developed techniques in treating nonlinear wave propagation problems. The technique was developed by Whitham (1965) for studying changes of a wave train governed by nonlinear partial differential equations. It corresponds to the Krylov-Bogliubov method of averaging for the ordinary differential equations as discussed in the previous section. The key to this approach is to realize that for many nonlinear problems the governing nonlinear equations admit elementary "wave train" solutions given by

$$\psi(\mathbf{r}, t) = a\cos(\omega t - \mathbf{k} \cdot \mathbf{r}) \tag{7.7.1}$$

where a, k, and $\omega = \omega(\mathbf{k})$ are slowly varying functions of r and t. The averaged variational method indicates a way to obtain equations that describes the variations of these quantities.

Let us consider a system for which the dynamic equations can be derived from a variational principle such as the Fermat's principle discussed in Chapter 5. We shall assume that the system admits wave train solutions as given by (7.7.1). The simplest case occurs when there is only one single dependent variable $\psi(\mathbf{r}, t)$ and the system is described by the Lagrangian $L(\psi_t, \mathbf{\psi_r}, \psi)$ where

$$\psi_t = \partial \psi / \partial t, \qquad \mathbf{\psi}_r = \nabla \psi$$
 (7.7.2)

The variational principle takes the form

$$\delta \iint L(\psi_t, \mathbf{\psi}_r, \psi) \, dt \, d\mathbf{r} = 0 \tag{7.7.3}$$

Following the discussion in Section 5.4, we derive the Euler equation for the system as

$$\partial L_1 / \partial t + \nabla \cdot \mathbf{L}_2 - L_3 = 0 \tag{7.7.4}$$

where

$$L_1 = \partial L / \partial \psi_t, \qquad \mathbf{L}_2 = \nabla_{\psi_r} L, \qquad L_3 = \partial L / \partial \psi$$
(7.7.5)

Equation (7.7.4) in general is a nonlinear partial differential equation for the unknown function ψ . Since we have assumed that a wave train solution for the system exists, we may write the solution in the form

$$\psi = \psi_0(\theta, a) \tag{7.7.6}$$

where

$$\theta = \omega t - \mathbf{k} \cdot \mathbf{r} \tag{7.7.7}$$

and a is the amplitude and ψ_0 is periodic in θ .

The three slowly varying parameters ω , **k**, *a* are not independent. In order that (7.7.6) satisfy (7.7.4), a dispersion relation must exist such that

$$D(\mathbf{k},\omega,a) = 0 \tag{7.7.8}$$

We note that for linear problems the dispersion relation does not involve the amplitude a. Such linear examples were amply discussed in earlier chapters. When the problem is nonlinear, the dispersion relation becomes dependent on the amplitude as shown in (7.7.8).

The wave number **k** and the frequency ω may be written, for the general case, as

$$\omega = \partial \theta / \partial t, \quad \mathbf{k} = -\nabla \theta \tag{7.7.9}$$

as defined in Section 2.12. The aim now is to derive equations for the slowly varying parameters \mathbf{k} , ω , and a. To achieve this, let us assume that the period of the wave train solution $\psi = \psi_0(\theta, a)$ is normalized to 2π . We define the averaged Lagrangian of the system by

$$\mathscr{L}(\mathbf{k},\omega,a) = (1/2\pi) \int_0^{2\pi} L \, d\theta \qquad (7.7.10)$$

The calculation of (7.7.10) is carried out in the following manner. We first substitute $\psi = \psi_0(\theta, a)$ in the expression for L; then the integration in (7.7.10) with respect to θ is carried out holding **k**, ω , and *a* constant. The dependence of \mathscr{L} on **k** and ω arises from the substitution of $\psi_t = \omega \psi_0'$, $\nabla \psi = -\mathbf{k} \psi_0'$ into L; the prime indicates differentiation with respect to θ . The average variation principle states that the equations for (**k**, ω , *a*) follow from

$$\delta \iint \mathscr{L}(\mathbf{k}, \omega, a) \, dt \, d\mathbf{r} = 0 \tag{7.7.11}$$

The quantities **k** and ω are related to the phase function θ . Therefore the variation of \mathscr{L} comes from variation of θ and variation of a. Hence, we must derive the Euler equations for (7.7.11) for independent variations $\delta\theta$ and δa . Again, following Section 5.4, we have, from variation δa

$$\mathscr{L}_{a}(\mathbf{k},\omega,a)=0 \tag{7.7.12}$$

from variation $\delta\theta$

$$(\partial/\partial t)\mathcal{L}_{\omega} - \nabla \cdot \nabla_{\mathbf{k}}\mathcal{L} = 0 \tag{7.7.13}$$

where the subscripts indicate differentiations as defined earlier.

Equation (7.7.12) is a functional relation between (\mathbf{k}, ω, a) which is just a nonlinear dispersion relation (7.7.8). Equation (7.7.13) together with the consistency equations

$$\partial \mathbf{k}/\partial t + \nabla \omega = 0, \qquad \partial k_i/\partial x_i - \partial k_i/\partial x_i = 0$$
 (7.7.14)

describe the variation of \mathbf{k} and ω . The consistency equations are derived from (7.7.9). Thus, we see that by applying the averaged variational principle, we obtain the equations for \mathbf{k}, ω , and a. Whitham (1970) has shown that this averaged variational principle may be derived as the first term in a formal perturbation expansion. In the following we shall show this for the onedimensional case; extension to higher dimensional case is straightforward.

The derivation makes use of the so-called "two-timing" technique which recognizes explicitly the fact that the solution of the problem varies on two different scales: the fast oscillation of the wave train and the slow variations of the parameters (\mathbf{k}, ω, a) . The main idea is to express the solution in the form

$$\psi(x,t) = \Psi(\theta, X, T, \varepsilon) \tag{7.7.15}$$

where

$$\theta = \varepsilon^{-1} \Theta(X, T), \quad X = \varepsilon x, \quad T = \varepsilon t \quad (7.7.16)$$

and ε is a small parameter that indicates the ratio between the fast and slow scales. The wave number and the frequency are now expressed by

$$\omega(X,T) = \theta_t = \Theta_T, \qquad k(X,T) = -\theta_x = -\Theta_X \qquad (7.7.17)$$

The independent variables x, t represent the fast scale variation and X, T represent the slow scale variation of the solution. The t and x derivatives now can be written as

$$\frac{\partial}{\partial t} = \Theta_T \frac{\partial}{\partial \theta} + \varepsilon \frac{\partial}{\partial T} = \omega \frac{\partial}{\partial \theta} + \varepsilon \frac{\partial}{\partial T}$$

$$\frac{\partial}{\partial x} = \Theta_X \frac{\partial}{\partial \theta} + \varepsilon \frac{\partial}{\partial X} = \varkappa \frac{\partial}{\partial \theta} + \varepsilon \frac{\partial}{\partial X}$$
(7.7.18)

where the notation $\varkappa = -k$ has been used.

394

7.7 An Averaged Variational Principle

When we apply (7.7.18), Eq. (7.7.4) becomes

$$\frac{\partial}{\partial \theta} \left(\omega L_1 + \varkappa L_2 \right) - L_3 + \varepsilon \frac{\partial L_1}{\partial T} + \varepsilon \frac{\partial L_2}{\partial X} = 0 \qquad (7.7.19)$$

This equation combining with the relation

$$\frac{\partial L}{\partial \theta} = L_1 \frac{\partial \Psi_{\theta}}{\partial t} + L_2 \frac{\partial \Psi_{\theta}}{\partial x} + L_3 \Psi_{\theta}$$
$$= L_1 \omega \, \partial \Psi_{\theta} / \partial \theta + \varepsilon L_1 \, \partial \Psi_{\theta} / \partial T + L_2 \varkappa \, \partial \Psi_{\theta} / \partial \theta + \varepsilon L_2 \, \partial \Psi_{\theta} / \partial X + L_3 \Psi_{\theta}$$
(7.7.20)

yields a conservation equation

$$(\partial/\partial\theta)[(\omega L_1 + \varkappa L_2)\Psi_{\theta} - L] + \varepsilon(\partial/\partial T)(\Psi_{\theta}L_1) + \varepsilon(\partial/\partial X)(\Psi_{\theta}L_2) = 0$$
(7.7.21)

For later convenience, the conservation equation is re written as

$$\partial R/\partial \theta + \varepsilon \, \partial P/\partial T + \varepsilon \, \partial Q/\partial X = 0$$
 (7.7.22)

. ___

where

$$R = (\omega L_1 + \varkappa L_2) \Psi_{\theta} - L$$

$$P = \Psi_{\theta} L_1$$

$$Q = \Psi_{\theta} L_2$$
(7.2.23)

We now formally expand the solution Ψ in a power series in ε ,

$$\Psi = \sum_{0}^{\infty} \varepsilon^{n} \Psi_{n} \tag{7.7.24}$$

and substitute it into the expressions for R, P, and Q. Each of them will also have a power series expansion of the form

$$R = \sum_{0}^{\infty} \varepsilon^{n} R^{(n)}, \quad \text{etc.} \qquad (7.7.25)$$

Equating the terms of equal powers in ε in (7.7.22), the first two terms are

$$\partial R^{(0)}/\partial \theta = 0 \tag{7.7.26}$$

$$\partial R^{(1)}/\partial \theta = - \partial P^{(0)}/\partial T - \partial Q^{(0)}/\partial X \qquad (7.7.27)$$

Equation (7.7.26) may be integrated once to yield

$$R^{(0)} = (\omega L_1^{(0)} + \varkappa L_2^{(0)}) \Psi_{00} - L^{(0)} = A(X, T)$$
(7.7.28)

where A is the integration constant (with respect to θ). This equation can be solved for Ψ_0 . The next order equation, (7.7.27), has to be solved for Ψ_1 . A solution uniformly valid in θ requires Ψ , and hence each Ψ_n , to be periodic in θ with period 2π . This is only possible in (7.7.27) if the integral of the right-hand side with respect to θ over one period is zero. That is, to avoid secular terms in Ψ_1 , we must demand

$$\frac{\partial}{\partial T} \left(\frac{1}{2\pi} \int_0^{2\pi} P^{(0)} d\theta \right) + \frac{\partial}{\partial X} \left(\frac{1}{2\pi} \int_0^{2\pi} Q^{(0)} d\theta \right) = 0 \quad (7.7.29)$$

From (7.7.23), Eq. (7.7.29) becomes

$$\frac{\partial}{\partial T} \left(\frac{1}{2\pi} \int_0^{2\pi} \Psi_{0\theta} L_1^{(0)} d\theta \right) + \frac{\partial}{\partial X} \left(\frac{1}{2\pi} \int_0^{2\pi} \Psi_{0\theta} L_2^{(0)} d\theta \right) = 0 \quad (7.7.30)$$

Using the definition of the averaged Lagrangian (7.7.10), the secular condition (7.7.30) can be written as

$$\partial \mathscr{L}_{\omega} / \partial T + \partial \mathscr{L}_{\varkappa} / \partial X = 0 \tag{7.7.31}$$

Or, going back to the original wave number k,

$$\partial \mathscr{L}_{\omega} / \partial T - \partial \mathscr{L}_{k} / \partial X = 0 \tag{7.7.32}$$

which is just the one-dimensional equivalent of (7.7.13) derived from the averaged Lagrangian principle. Thus, we have shown that the averaged Lagrangian principle may be derived as the first term in a formal perturbation expansion. Higher order terms may be formally obtained from the expansion of (7.7.22).

From (7.7.28), we can solve for $L^{(0)}$ in terms of other variables and obtain

$$L^{(0)} = (\omega L_1^{(0)} + \varkappa L_2^{(0)}) \Psi_{0\theta} - A$$

= $(\partial L^{(0)} / \partial \Psi_{0\theta}) \Psi_{0\theta} - A$ (7.7.33)

The averaged Lagrangian can now be written as

$$\mathscr{L} = \frac{1}{2\pi} \int_{0}^{2\pi} \left(\frac{\partial L^{(0)}}{\partial \Psi_{0\theta}} \Psi_{0\theta} - A \right) d\theta$$
$$= \frac{1}{2\pi} \oint \frac{\partial L^{(0)}}{\partial \Psi_{0\theta}} d\Psi_{0} - A \tag{7.7.34}$$

where the integration is over one complete period of Ψ_0 .

Using this form of $\mathscr L$ in the variational principle, the Euler equations are

$$\mathscr{L}_{A} = 0, \qquad \partial \mathscr{L}_{\omega} / \partial T + \partial \mathscr{L}_{\star} / \partial X = 0$$
 (7.7.35)

Whitham (1970) has shown that for many nonlinear problems, (7.7.34) and (7.7.35) are most effective to use to obtain the dispersion relation $D(k, \omega, A)$ as well as the equation for ω and k. We note that A in (7.7.35) is related to the amplitude as defined in (7.7.1).

We conclude this section by applying the averaged variational principle to a simple nonlinear problem. The equation is the general nonlinear Klein– Gordon equation

$$\psi_{tt} - \psi_{xx} + V'(\psi) = 0 \tag{7.7.36}$$

The Lagrangian is

$$L = \frac{1}{2}\psi_t^2 - \frac{1}{2}\psi_x^2 - V(\psi)$$
 (7.7.37)

where $V(\psi)$ is a function of ψ .

The lowest order approximation yields

$$L^{(0)} = \frac{1}{2}(\omega^2 - k^2)\Psi_{0\theta}^2 - V(\Psi_0)$$
(7.7.38)

Therefore

$$\partial L^{(0)} / \partial \Psi_{0\theta} = (\omega^2 - k^2) \Psi_{0\theta}$$
(7.7.39)

Substituting (7.7.39) into (7.7.33), we obtain

$$L^{(0)} = (\omega^2 - k^2)^{-1} (\partial L^{(0)} / \partial \Psi_{0\theta})^2 - A$$
(7.7.40)

Comparing (7.7.38) with (7.7.40), we have

$$A = \frac{1}{2}(\omega^2 - k^2)^{-1}(\partial L^{(0)}/\partial \Psi_{0\theta})^2 + V(\Psi_0)$$
(7.7.41)

We note that the purpose of this manipulation is to eliminate $\Psi_{0\theta}$ in favor of $\partial L^{(0)}/\partial \Psi_{0\theta}$. Solving (7.7.41) for $\partial L^{(0)}/\partial \Psi_{0\theta}$, we obtain

$$\partial L^{(0)} / \partial \Psi_{0\theta} = \sqrt{2} (\omega^2 - k^2)^{1/2} [A - V(\Psi_0)]^{1/2}$$
 (7.7.42)

Substituting (7.7.42) into (7.7.34), we obtain the averaged Lagrangian

$$\mathscr{L} = \frac{1}{2\pi} \left[2(\omega^2 - k^2) \right]^{1/2} \oint \left[A - V(\Psi_0) \right]^{1/2} d\Psi_0 - A \quad (7.7.43)$$

The Euler equations (7.7.35) become

$$2\pi = \left[\frac{1}{2}(\omega^2 - k^2)\right]^{1/2} \oint d\Psi_0 / [A - V(\Psi_0)]^{1/2}$$
(7.7.44)

and

$$\frac{\partial}{\partial T} \left\{ \frac{1}{2\pi} \frac{\omega}{(\omega^2 - k^2)^{1/2}} \oint [A - V(\Psi_0)]^{1/2} d\Psi_0 \right\} \\ + \frac{\partial}{\partial X} \left\{ \frac{1}{2\pi} \frac{k}{(\omega^2 - k^2)^{1/2}} \oint [A - V(\Psi_0)]^{1/2} d\Psi_0 \right\} = 0 \quad (7.7.45)$$

Equation (7.7.44) gives the dispersion relation between ω , k, and A while (7.7.45) together with the consistency equations (7.7.14) describe their slow variations.

For the linear case such that $V(\Psi) = \psi^2/2$, the integral in (7.7.43) may be evaluated explicitly. We have

$$\oint (A - \frac{1}{2}\Psi_0^2)^{1/2} \, d\Psi_0 = 2 \int_{-\sqrt{2A}}^{\sqrt{2A}} (A - \frac{1}{2}\Psi_0^2)^{1/2} \, d\Psi_0 = (2A)^{1/2} \, \pi$$

Equation (7.7.43) then becomes

$$\mathscr{L} = [(\omega^2 - k^2)^{1/2} - 1]A \tag{7.7.46}$$

Similarly, (7.7.44) and (7.7.45) may be evaluated to yield, respectively,

$$(\omega^2 - k^2)^{1/2} = 1 \tag{7.7.47}$$

$$\partial (A\omega)/\partial T + \partial (Ak)/\partial X = 0$$
 (7.7.48)

The consistency equation (7.7.14) becomes

$$\frac{\partial k}{\partial T} + \frac{\partial \omega}{\partial X} = 0 \tag{7.7.49}$$

Combining (7.7.47), (7.7.48), and (7.7.49), we obtain

$$\partial k/\partial T + v_g(k) \,\partial k/\partial X = 0$$
 (7.7.50)

$$\partial A/\partial T + \nu_g(k) \,\partial A/\partial X + A\nu_g'(k) \,\partial k/\partial X = 0 \qquad (7.7.51)$$

where $v_q(k)$ is the group velocity defined by

$$\nu_g(k) = d\omega/dk \tag{7.7.52}$$

398

Problems

For a periodic solution of the form

$$\Psi_0(\theta) = a\cos\theta \tag{7.7.53}$$

we see from (7.7.28) that the amplitude *a* is related to *A* through

$$A = a^2/2 \tag{7.7.54}$$

Therefore A is proportional to the energy density of the wave.

We note that the curve

$$dX/dT = dx/dt = v_a(k) \tag{7.7.55}$$

is a double characteristic for the two equations (7.7.50) and (7.7.51). Along this characteristic,

$$dk/dT = 0$$
, $dA/dT = -v_g'(\partial k/\partial X)A$ on $dx/dt = v_g(k)$ (7.7.56)

That is, the wave number remains constant for an observer moving with group velocity $v_q(k)$. This is the same result as in Chapter 2 when we discussed the kinematics of the waves. Equations (7.7.50) and (7.7.51) give the slow variations for the amplitude and the wave number of the wave train. The solutions depend on the initial and boundary conditions corresponding to the physical situations.

Thus, in this section we have introduced the averaged variational principle for treating general problems of dispersive waves. The main results are included in the three equations (7.7.12)-(7.7.14). These equations have been used by many authors to study a wide variety of problems, including the extension of the concept of group velocity to nonlinear problems, nonlinear instability, resonant wave interaction, dispersive waves in inhomogeneous media, in moving media, etc. Interested readers may consult the paper by Whitham (1970) for the references on the various topics.

Problems

1. In Section 7.3 we have derived the expression for the electron temperature T_e in the limit of $\omega \gg \delta \nu$, and found it to be independent of time. In the opposite extreme of low frequencies such that $\omega \ll \delta \nu$, find the electron temperature T_e (assuming δ and ν are constants). 2. In the high frequency approximation $(\omega \gg \nu_0)$, assuming the incident wave to be amplitude-modulated with the amplitude given by $E_0(1 + M \times \cos \Omega t)$, $\Omega \ll \delta \nu_0$, find the absorption coefficient for the wave. The dominant collisional process is between electrons and molecules.

3. In the resonant wave system (7.6.20) if one of the waves, say wave 1 (k_1, ω_1) , is excited initially with an amplitude much greater than that of waves 2 and 3, (7.6.20) may be solved by the usual linearization procedure. We start by linearizing (7.6.20) with respect to $\eta_1(k_2, t)$ and $\eta_1(k_3, t)$ but not with respect to $\eta_1(k_1, t)$.

(a) Prove that to this order η_1 remains constant in time.

(b) Prove that the conditions for effective transfer of energy from wave 1 to waves 2 and 3 are

$$|\omega_1| > |\omega_2|$$
 and $|\omega_1| > |\omega_3|$.

4. The internal gravity waves of the form $e^{j(\omega t - \mathbf{k} \cdot \mathbf{r})}$ in a uniformly stratified fluid satisfy the dispersion relation

$$\omega = \omega_b \cos \theta$$

where ω_b is the Brunt-Vaisala frequency and is a constant and θ is the angle between the wave vector **k** and the horizontal plane. The resonant conditions for wave-wave interaction are written in the form equivalent to but slightly different from the ones in the text:

$$\mathbf{k}_1 - \mathbf{k}_2 = \mathbf{k}_3$$
$$\omega_1 - \omega_2 = \omega_3$$

For the case in which $\omega_3 = 0$, find \mathbf{k}_3 , \mathbf{k}_1 , \mathbf{k}_2 , ω_1 , and ω_2 that satisfy the resonant conditions.

References

BOGOLIUBOV, N. N., and MITROPOLSKY, Y. A. (1961). "Asymptotic Methods in the Theory of Nonlinear Oscillations." Gordon Breach, New York.

- BRETHERTON, F. P. (1964). Resonant interaction between waves. J. Fluid Mech. 20, 457-479.
- DAWSON, J. M. (1959). Nonlinear electron oscillations in a cold plasma. *Phys. Rev.* **113**, 383.

References

- GINZBURG, V. L., and GUREVICH, A. V. (1960). Nonlinear phenomena in a plasma located in an alternating electromagnetic field. Sov. Phys. Usp. 3, 175-194.
- GRADSHTEYN, I. S., and RYZHIK, I. W. (1965). "Table of Integrals Series and Products." Academic Press, New York.
- HARKER, K. J., and CRAWFORD, F. W. (1969). Nonlinear interaction between whistlers. J. Geophys. Res. 74, 5029-5040.
- PHILLIPS, O. M. (1968). The interaction trapping of internal gravity waves. J. Fluid Mech. 34, 407-416.
- SAGDEEV, R. Z., and GALEEV, A. A. (1969). "Nonlinear Plasma Theory." Benjamin, New York.
- TSYTOVICH, V. N., (1970). "Nonlinear Effects in Plasma." Plenum, New York.
- WHITHAM, G. B. (1965). Nonlinear dispersive waves. Proc. Roy. Soc. Ser. A 283, 238-261.
- WHITHAM, G. B. (1970). Two-timing, variational principles and waves. J. Fluid Mech. 44, Part 2, 373-395.
- YEH, K. C., and LIU, C. H. (1970). On resonant interactions of acoustic-gravity waves. *Radio Sci.* 5, 39-48.

8. Interaction of Atmospheric Waves with the Ionosphere

The previous chapters were concerned exclusively with the propagation of electromagnetic waves in plasmas although the mathematical techniques developed in studying such waves can be equally applied to studies of other waves. We have seen that in response to the incident electromagnetic energy the plasma particles irradiate in such a manner that propagation of the total electromagnetic energy can be described in terms of characteristic waves.

The terrestrial ionosphere is known to be a weakly ionized plasma. The plasma-neutral-density ratio is at most 1 : 100. If the neutral atmosphere is capable of wave motions of its own, the plasma must respond to it through collisions. One outstanding example of wave motions in the neutral gas is the propagation of sound. In the atmosphere, the gravitational force causes the density to be vertically nonuniformly distributed. For waves of sufficiently low frequency the buoyancy force plays an important role in making otherwise ordinary sound waves anisotropic. In fact, the modification is so drastic that the resulting wave is called the acoustic-gravity wave. This chapter is concerned with the propagation of acoustic-gravity waves and their interaction with the ionosphere. Because of space limitation, we will discuss propagation in the unbounded atmosphere only. Most experimental observations at ionospheric heights tend to support free waves and not guided waves. It should be pointed out that many observed acousticgravity waves in the troposphere are guided.

The earth rotational effects are ignored as tides form a special topic of their own.

8.1 Structure of the Atmosphere

The primary properties of the atmosphere are its density, pressure, temperature, composition, and motion. Conventional meteorology deals mainly with these properties in the lower ten to twenty kilometers. These regions are accessible by kites (as first used in about 1900), balloons, and aircrafts. As experimental techniques improved, the probing of the atmosphere also extended in height. Recently, rockets and satellites have been making direct *in situ* measurements in addition to observing, at a distance, naturally occurring events such as meteors, noctilucent clouds, etc. One of the most striking features of the earth's atmosphere is its vertical diminution of density with height. Therefore, to the first order, the atmosphere may be assumed to be horizontally stratified. The cause of this stratification is the strong gravitational force. The balance of the gravitational force by the pressure gradient force is given by the hydrostatic equation

$$dp = -g\varrho \, dz \tag{8.1.1}$$

Here p is the pressure, g the gravity, ϱ the mass density, and z the height. For an ideal gas, the pressure is related to the number density N and temperature T through

$$p = NT \tag{8.1.2}$$

For convenience, T is expressed in energy units. Divide (8.1.1) by (8.1.2); the following equation is obtained.

$$dp/p = -dz/H \tag{8.1.3}$$

where H is called the scale height and is given by

$$H = T/mg \tag{8.1.4}$$

In (8.1.4), *m* is the mean molecular mass at a given height. Integrating (8.1.3) from a reference height z_0 at which $p = p_0$ to an arbitrary height *z*, we obtain

$$p = p_0 e^{-\int_{z_0}^{z} dz/H}$$
(8.1.5)

The distribution of number density with height can be obtained from (8.1.5) and the ideal gas law (8.1.2) yielding

$$N = (N_0 T_0/T) e^{-\int_{z_0}^{z} dz/H}$$
(8.1.6)

where N_0 and T_0 are, respectively, the number density and temperature at z_0 . It is clear by examining (8.1.5) and (8.1.6) that a knowledge of the scale height variation with height is very crucial in understanding the pressure and density distributions. In the following, we will consider two highly idealized cases, i.e., the isothermal case and the adiabatic case.

We first discuss the isothermal case. If the atmosphere is allowed to reach a thermal equilibrium, the atmospheric temperature eventually will reach a constant value for all heights. Such an atmosphere is called an isothermal atmosphere. Of course, isothermality nearly never happens in the real atmosphere. But in the region of the upper thermosphere above about 300 km the temperature profile shows very little height variation. This is because of the large thermal conductivity which tends to smooth out any temperature gradient. At these heights, there is very little mixing. The atmosphere, being a mixture of gases, is then distributed with partial pressure given by its constituent gas. Let the scale height of the *i*th gas be

$$H_i = T/m_i g \tag{8.1.7}$$

Over a limited region of the atmosphere, the gravity can be assumed to be constant. The height distribution of partial pressures in an isothermal atmosphere is therefore given by

$$p_i = p_{0i} e^{-(z - z_0)/H_i} \tag{8.1.8}$$

Similarly, the density distribution is

$$N_i = N_{0i} e^{-(z-z_0)/H_i} \tag{8.1.9}$$

In an isothermal atmosphere the meaning of scale height is very clear: It is the e-folding distance for pressure and density. A light gas has a large scale height. It is therefore expected that the light gas will predominate at sufficiently great heights.

Next, we wish to discuss the adiabatic case. In the lower atmosphere the thermal conduction is extremely slow. Convective motions are set up when the atmosphere is heated from below. These motions also mix the gas sufficiently so that the average mass m is nearly unchanged with height. Because of the slow conduction, the gas, when transported from one place to

another, expands and contracts nearly adiabatically. The pressure and mass density in an adiabatic atmosphere must then satisfy

$$p = A\varrho^{\gamma} \tag{8.1.10}$$

or

$$dp = A\gamma \varrho^{\gamma-1} \, d\varrho \tag{8.1.11}$$

where A is a constant and γ the ratio of the specific heats. Eliminate p from the hydrostatic equation (8.1.1) and the adiabatic equation (8.1.11) and integrate the resulting equation to find

$$\varrho = [\varrho_0^{\gamma-1} - (g(\gamma-1)/A\gamma)(z-z_0)]^{1/(\gamma-1)}$$
(8.1.12)

The mass density (8.1.12) decreases continuously as height increases and it is reduced to zero when

$$z - z_0 = p_0 \gamma / \varrho_0 g(\gamma - 1) \tag{8.1.13}$$

where we have made the substitution $A = p_0/\rho_0^{\gamma}$. This means that an adiabatic atmosphere must have a limited height. As an example, let us take $p_0 = 1.01 \times 15^5 \text{ N/m}^2$, $\rho_0 = 1.23 \text{ kg/m}^3$, $\gamma = 1.4$ and $g = 9.80 \text{ m/sec}^2$. We get 29.4 km as the limit of the adiabatic atmosphere. The corresponding temperature profile can be obtained by using the ideal gas law (8.1.2) and the adiabatic law (8.1.10).

$$T = Am\varrho_0^{\gamma-1} = Am\varrho_0^{\gamma-1} - (mg(\gamma-1)\gamma^{-1}/\gamma)(z-z_0) \qquad (8.1.14)$$

where the density profile (8.1.12) has been used. The temperature gradient is easily obtained from (8.1.14),

$$dT/dz = -mg(\gamma - 1)/\gamma \tag{8.1.15}$$

The temperature given by (8.1.15) decreases with height at a constant rate. The magnitude of the gradient is often referred to as the lapse rate. If we take the mean molecular weight to be 29, $g = 9.80 \text{ m/sec}^2$, $\gamma = 1.4$, the lapse rate comes out to be $136 \times 10^{-24} \text{ J/km}$ or 9.84°K/km . Except for the first two kilometers above the earth's surface, the troposphere lapses at an average rate of 6.5°K/km . Consequently, the troposphere is not in perfect adiabatic equilibrium.

The behavior of the average temperature profile in the atmosphere is shown in Fig. 8.1-1. Also shown are main names for different atmospheric regions classified according to the temperature behavior. It is obvious that

406 8. Interaction of Atmospheric Waves with the Ionosphere

the real atmosphere is much more complicated than the highly idealized cases discussed here. The reader should consult specialized books for a detailed discussion [see, for example, Ratcliffe (1960)]. Our interest here is to study the wave propagation in the atmosphere. The background material discussed so far will serve to remind us that the real atmosphere is quite complicated.



Fig. 8.1-1. Average temperature profiles in the atmosphere.

It is well known that any sufficiently rapid perturbation in the homogeneous atmosphere will propagate away as a sound or acoustic wave. Being an adiabatic process, the speed of sound c is given by the formula

$$c^{2} = (dp/d\varrho)_{\text{adiabatic}}$$
(8.1.16)

The pressure and density variations for an adiabatic process are related through (8.1.11). The use of (8.1.11) and (8.1.10) reduces (8.1.16) to

$$c^2 = \gamma p/\varrho = \gamma T/m \tag{8.1.17}$$

The speed of sound is seen proportional to the square root of temperature. Corresponding to the average temperature profile of Fig. 8.1-1, the speed of sound varies in a manner shown in Fig. 8.1-2. Because of its dependence on the temperature, the speed of sound will vary diurnally over a wide range in the thermosphere. It is interesting to note that there are two sound



ducts where the speed of sound is a minimum. The possibilities of such ducts to serve as wave guides have been studied by several authors [e.g., Pfeffer and Zarichny (1962)]. In most cases, these studies make use of numerical techniques. We will limit ourselves to analytic studies of simple and idealized problems in order to gain an understanding of the physical processes involved.

8.2 Buoyancy Oscillations

We have seen in Section 8.1 that the gravitational field is responsible, to a large extent, for making the atmosphere inhomogeneous. We will see now that the gravitational field is also responsible for making propagation of atmospheric waves anisotropic. In an ideal fluid, the motion of a small parcel of fluid along the equipotential surface does not require any energy. There is no restoring force involved as long as the parcel is displaced along the equipotential surface. This is no longer true when the displacement is away from the equipotential surface. A vertically displaced fluid parcel experiences a buoyancy force which tends to restore the original equilibrium. The rapidity with which the equilibrium is restored characterizes the atmospheric oscillation frequency. If the frequency of the wave motion is large compared with the characteristic frequency, the propagation is expected to be nearly isotropic and the gravity effect is minimized. Conversely, the wave motion is expected to be anisotropic and the gravity plays an important role if the frequency is low.

Consider a horizontally stratified fluid in hydrostatic equilibrium. An external force is applied to displace an element of fluid vertically upward by a small distance ξ . Now remove the external force. In this position, this element of fluid experiences a force and starts to move according to the equation of motion

$$\varrho \ddot{\xi} = -g \, \varDelta \varrho \tag{8.2.1}$$

The right hand of (8.2.1) is just the buoyancy force. The quantity $\Delta \varrho$ is the difference in density of the displaced fluid element from that of the new environment and is composed of two terms. The first term comes about because of the inhomogeneous nature of the atmosphere and its contribution to $\Delta \varrho$ is

$$-\xi \, d\varrho/dz \tag{8.2.2}$$

The second term comes from the change in pressure. If we assume that the change is taking place adiabatically, then the contribution to $\Delta \varrho$ is, according to (8.1.16),

$$\frac{\Delta p}{c^2} = \frac{1}{c^2} \frac{dp}{dz} \,\xi = -\frac{\varrho g}{c^2} \,\xi \tag{8.2.3}$$

where the hydrostatic equation (8.1.1) has been used in obtaining the last expression. Substitution of (8.2.2) and (8.2.3) into (8.2.1) results in the equation of motion

$$\ddot{\xi} = -\omega_b^2 \xi \tag{8.2.4}$$

with the angular buoyancy frequency (alternately called Vaisala frequency, Brunt frequency, or Brunt-Vaisala frequency) given by

$$\omega_b^2 = -g[d \ln \varrho/dz + g/c^2] = (\gamma - 1)g^2/c^2 + (g/c^2)(dc^2/dz) \qquad (8.2.5)$$

The fluid element therefore executes a simple harmonic motion about the equilibrium position with an angular frequency ω_b provided that ω_b^2 given by (8.2.5) is positive. If ω_b^2 is negative, the initial perturbation will grow exponentially with time and the fluid is unstable. The condition of marginal stability occurs when $\omega_b = 0$ or when

$$dT/dz = -mg(\gamma - 1)/\gamma \qquad (8.2.6)$$

This is just the lapse rate of an adiabatic atmosphere given by (8.1.15).

Therefore, the atmosphere is unstable if its temperature lapses at a rate faster than the lapse rate of an adiatabic atmosphere.

In an isothermal atmosphere with a constant mean molecular mass, the buoyancy frequency reduces to

$$\omega_b^2 = (\gamma - 1)g^2/c^2 \tag{8.2.7}$$

which is a constant, at least over the region within which g may be assumed constant.

In the atmosphere, the buoyancy frequency is less than 1 Hz and it is therefore more convenient to speak of the period of the wave. The buoyancy period in the stratosphere is roughly 6 min; it increases to 12 min in the thermosphere.

8.3 Acoustic Gravity Waves in an Isothermal Atmosphere

The appearance of the buoyancy frequency brings in an entirely new mode of wave motion which does not have its counterpart in the study of ordinary sound in the homogeneous medium. This characteristic frequency is very small as the period is several minutes to ten minutes in the real atmosphere. Only comparably low frequency waves would be affected by the buoyancy force. It is for this reason that such waves are sometimes referred to as infrasonic waves. We will call them acoustic gravity waves.

In discussing the propagation of acoustic gravity waves, we will assume that the process is adiabatic. Its justification can be obtained by the following order of magnitude estimation. For our present purpose, the onedimensional equation of heat conduction may be taken as

$$\varrho c_v \, \partial T / \partial t = \varkappa \, \partial^2 T / \partial z^2 \tag{8.3.1}$$

where c_v is the specific heat at constant volume and has a numerical value 7.2×10^6 erg/gm-deg in cgs units. The coefficient of heat conductivity \varkappa is proportional to the square root of temperature and is found to be given by

$$\kappa = (\sum_{i} A_{i} N_{i} / N) T^{1/2} \quad \text{erg/cm-sec-deg}$$
(8.3.2)

where $A_i = 360$ for O, and $A_i = 180$ for O₂ and N₂. Let the characteristic time of heat conduction be τ and the characteristic length be the scale height H. The velocity of heat conduction has the order of magnitude, estimated

from (8.3.1),

$$v_{\text{heat}} \approx H/\tau = \varkappa/\varrho c_v H$$
 (8.3.3)

On the surface of the earth, this velocity of heat conduction has a numerical value 4×10^{-9} m/sec which is negligibly small when compared with the corresponding speed of sound, roughly 300 m/sec as seen from Fig. 8.1-2. Therefore, the adiabatic assumption is very good. Even at a height of 200 km, the speed of heat conduction is estimated to be 8 m/sec which is still small when compared with the corresponding speed of sound at 800 m/sec. For purposes of studying the propagation of acoustic gravity waves, the adiabatic assumption can be justifiably used up to a height of about 300 km. The effect of heat conduction in this height range is of the second order and it serves as a damping mechanism.

The equations of concern are the fluid equations. For simplicity, the atmosphere is assumed to be nonrotating and stationary. The equations are the equation of continuity, the equation of motion, and the adiabatic equation given, respectively, by

$$\frac{\partial \varrho}{\partial t} + \nabla \cdot \varrho \mathbf{v} = 0 \tag{8.3.4a}$$

$$\rho \, D\mathbf{v}/Dt = -\nabla p + \rho \mathbf{g} \tag{8.3.4b}$$

$$Dp/Dt = c^2 D\varrho/Dt \tag{8.3.4c}$$

where the convective derivative is $D/Dt = \partial/\partial t + v \cdot \nabla$. The gravity is directed downward, i.e., $\mathbf{g} = (0, 0, -g)$. The last equation of (8.3.4) is applicable only to an adiabatic process and its justification has just been examined. For a static, motionless, isothermal atmosphere the mass density must be distributed exponentially as a function of height,

$$\varrho_0 = \varrho_{00} e^{-z/H} \tag{8.3.5}$$

where ρ_{00} is the density at z = 0. The corresponding pressure distribution is

$$p_0 = p_{00} e^{-z/H} \tag{8.3.6}$$

with $p_{00} = \rho_{00}T/m$. Let the isothermal atmosphere be perturbed according to the following scheme:

$$\varrho(\mathbf{r}, t) = \varrho_0(z) + \varrho'(\mathbf{r}, t)$$

$$p(\mathbf{r}, t) = p_0(z) + p'(\mathbf{r}, t)$$

$$\mathbf{v}(\mathbf{r}, t) = 0 + \mathbf{v}(\mathbf{r}, t)$$

(8.3.7)

410

The first terms on the right-hand side of (8.3.7) are the equilibrium quantities and the second terms are the perturbations. Inserting (8.3.7) into (8.3.4)and linearizing the resulting equations, we obtain the following equations:

$$\partial \varrho' / \partial t + \varrho_0 \nabla \cdot \mathbf{v} + \mathbf{v} \cdot \nabla \varrho_0 = 0$$

$$\varrho_0 \partial \mathbf{v} / \partial t = -\nabla p' + \varrho'.$$

$$\partial p' / \partial t + \mathbf{v} \cdot \nabla p_0 = c^2 (\partial \varrho' / \partial t + \mathbf{v} \cdot \nabla \varrho_0)$$
(8.3.8)

Now assume the perturbed quantities to vary like

(function of
$$z$$
) $e^{j(\omega t - k_x x)}$ (8.3.9)

By the judicious choice of coordinate axes, we have restricted ourselves to studying those waves that are not functions of y. In such a case, the y-component of the equation of motion shows that $v_y = 0$. The remaining equations of (8.3.8) reduce to

$$j\omega(\varrho'/\varrho_{0}) - jk_{x}v_{x} + \partial v_{z}/\partial z - H^{-1}v_{z} = 0$$

$$j\omega v_{x} = jk_{x}(p'/\varrho_{0})$$

$$j\omega v_{z} = -(\partial/\partial z)(p'/\varrho_{0}) + H^{-1}(p'/\varrho_{0}) - g(\varrho'/\varrho_{0})$$

$$j\omega mT^{-1}(p'/\varrho_{0}) - H^{-1}v_{z} = \gamma [j\omega(\varrho'/\varrho_{0}) - H^{-1}v_{z}]$$
(8.3.10)

The advantage of the isothermal condition is now clear. As seen in (8.3.10), the four differential equations for ϱ'/ϱ_0 , ν_x , and ν_z have constant coefficients in an isothermal atmosphere. This permits us to seek a solution with z dependence proportional to e^{-jk_xz} . In matrix form, the set of equations can be written as

$$\mathbf{D} \cdot \mathbf{F} = \mathbf{0} \tag{8.3.11}$$

where **F** is the vector $(\varrho'/\varrho_0, p'/\varrho_0, \nu_x, \nu_z)$. The symbol **D** stands for a dyadic differential operator which reduces to an algebraic expression for plane waves in an isothermal atmosphere. In matrix form, it is given by

$$\mathbf{D} = \begin{bmatrix} j\omega & 0 & -jk_x & -jk_z - H^{-1} \\ 0 & -jk_x & j\omega & 0 \\ g & -jk_z - H^{-1} & 0 & j\omega \\ -j\omega c^2 & j\omega & 0 & (\gamma - 1)g \end{bmatrix}$$
(8.3.12)

The set of homogeneous algebraic equations (8.3.11) has a unique solution (outside of a multiplying constant) when the determinant of the coefficient

matrix vanishes identically, i.e.,

$$\det \mid \mathbf{D} \mid = 0 \tag{8.3.13}$$

Expand the determinant (8.3.13), and the following algebraic equation is obtained.

$$\omega^4 - \omega^2 c^2 (k_x^2 + k_z^2) + g^2 (\gamma - 1) k_x^2 + j \omega^2 \gamma g k_z = 0 \qquad (8.3.14)$$

Equation (8.3.14) is known as the dispersion relation. It is complex even for the lossless medium. The choice of real or complex ω , k_x , and k_z depends on the problem at hand. For example, if the problem of interest is concerned with imperfect horizontal ducting, k_x may become complex to show leakage of energy from the duct. Let us consider the forced oscillation case for which ω is real. The wave is supposed to be incident on a horizontally stratified medium. Due to the kinematic boundary conditions at interfaces, k_x must be invariant from stratification to stratification and is therefore real. Since ω and k_x are real, k_z must then be complex. Let

$$k_z = k_z' - jk_z'' \tag{8.3.15}$$

The real part and the imaginary part of (8.3.14) can be easily separated to give the following two equations.

$$\omega^{4} - \omega^{2}c^{2}(k_{x}^{2} + k_{z}^{\prime 2} - k_{z}^{\prime \prime 2}) + g^{2}(\gamma - 1)k_{x}^{2} + \omega^{2}\gamma gk_{z}^{\prime \prime} = 0 \quad (8.3.16a)$$
$$\omega^{2}k_{z}^{\prime}(2c^{2}k_{z}^{\prime \prime} + \gamma g) = 0 \quad (8.3.16b)$$

There are three possibilities resulting from (8.3.16b). We discuss them in turn.

(i) Vertical Wind Shear Case. The first case we wish to discuss is when $\omega = 0$ for which $k_x = 0$ and k_z is arbitrary. For this time independent case, the perturbations in pressure and density, if any, must satisfy the hydrostatic equilibrium, the vertical component of the velocity must be zero (i.e., $v_z = 0$), while the horizontal component v_x is arbitrary. These properties can be easily obtained by examining (8.3.10). This is the case of steady horizontal winds sheared in the vertical direction.

(ii) Surface Wave Case. The second case of interest is when $k_z' = 0$. The wave number k_z is purely imaginary, indicating that the wave has no phase variation along z. The amplitude varies exponentially in the vertical direction, and it may become very large for some values of z. Such waves obviously have difficulty to exist in an unbounded region and they must

412

be confined to sharp boundaries on which properties of the atmosphere change drastically. There is some experimental evidence for the existence of such waves in the real atmosphere. The sharp boundary is thought to be the large temperature gradient in the lower thermosphere and the ground. These waves are occasionally referred to as the Lamb waves.

(iii) Internal Wave Case. If $k'_z \neq 0$ so that there is a phase variation along z, then, from (8.3.16b), k''_z must be a constant given by

$$k_z'' = -\gamma g/2c^2 = -1/2H \tag{8.3.17}$$

The dependence on z of field quantities has the form

$$e^{z/2H}e^{-jk_z'z}$$
 (8.3.18)

which grows exponentially with height. We note that the kinetic energy in a wave is proportional to $\rho_0 \nu \nu^*$. Since ρ_0 varies like $e^{-z/H}$ while $|\nu|$ varies like $e^{z/2H}$, the energy is kept constant as required in a lossless medium. The growth of the amplitude of internal waves shown by (8.3.18) accounts for the importance of such waves in the ionospheric height. Insertion of (8.3.17) in (8.3.16a) yields

$$\frac{k_{2}^{2}}{\{1-\omega_{a}^{2}/\omega^{2}\}/\{1-\omega_{b}^{2}/\omega^{2}\}}+\frac{k_{z}^{2}}{(1-\omega_{a}^{2}/\omega^{2})}=k_{0}^{2} \quad (8.3.19)$$

where $k_0 = \omega/c$, $\omega_a = \gamma g/2c = c/2H$ (acoustic cutoff frequency), and $\omega_b = (\gamma - 1)^{1/2}g/c$ (buoyancy frequency). Equation (8.3.19) is the desired dispersion relation. Its properties as well as other relevant information of internal waves are discussed in the next section. For simplicity the prime on k_z in (8.3.19) has been ignored so that k_z in (8.3.19) is real. The exponentially growing factor shown in (8.3.18) must be remembered to apply to all components of **F**.

8.4 Properties of Internal Waves

We found in the last section that the wave components of the internal waves grow exponentially with height like $e^{z/2H}$. The dispersion relation is given by (8.3.19) which is reproduced in the following:

$$\frac{n_x^2}{\{1-\omega_a^2/\omega^2\}/\{1-\omega_b^2/\omega^2\}} + \frac{n_z^2}{(1-\omega_a^2/\omega^2)} = 1$$
(8.4.1)

where

and

$$n_x = k_x/k_0$$
$$n_z = k_z/k_0$$

In an isothermal atmosphere, we have $\omega_a > \omega_b \operatorname{since} \gamma$ is less than 2. In the atmosphere, we may take $\gamma = 1.4$ for which $\omega_b = 0.904\omega_a$. As mentioned earlier, these frequencies are in the infrasonic range and we frequently use periods rather than frequencies. Typical values in the *E* region heights (about 100 km) are 4.5 min for the acoustic cutoff period and 5 min for the buoyancy period. The corresponding values in the *F* region (about 250 km) are 13 min and 14.4 min, respectively.

We now wish to discuss the properties of the dispersion relation (8.4.1) in different frequency ranges.



Fig. 8.4-1. The refractive index curve for an internal gravity wave. $T_a = 13$ min, $T_b = 14.5$ min, T = 30 min. The resonance angle θ_r is found to be 28.9°. The index surface is obtained by revolving the curve about the n_z -axis.

(i) Gravity Wave Branch. This is the low frequency branch in which $0 < \omega < \omega_b$ or $T > T_b$. The curve given by (8.4.1) in the $n_x n_z$ -plane is a hyperbolic curve. An example is shown in Fig. 8.4-1. Since the vertical gravity provides the only axis of symmetry, the index surface can be obtained by revolving the hyperbola about the n_z axis. The surface intersects the horizontal axis at $\{(\omega_a^2/\omega^2 - 1)/(\omega_b^2/\omega^2 - 1)\}^{1/2}$. The resonance at which n_x , $n_z \to \infty$ occurs when the propagation has a polar angle θ_r given by

$$\sin \theta_r = \omega/\omega_b \tag{8.4.2}$$

In the gravity branch, the refractive index is always greater than 1, showing that the phase velocity of the wave is always less than the velocity of sound.

414

(ii) Cutoff Region. When $\omega_b < \omega < \omega_a$, either n_x or n_z has to be imaginary, contrary to initial assumption. Therefore, the internal waves cannot exist in this frequency range. It does not mean that other wave types cannot exist either. For example, the surface wave can certainly propagate in this frequency region since it has imaginary k_z . In a nonisothermal atmosphere the buoyancy frequency defined by (8.2.5) may become larger than the acoustic cutoff frequency and the cutoff region of the form discussed for the isothermal atmosphere no longer exists.

(iii) Acoustic Branch. In the high frequency branch for which $\omega > \omega_a$, the internal wave again can propagate. The curve given by (8.4.1) in the $n_x n_z$ -plane is an ellipse. The index surface obtained by revolving the ellipse about the n_z -axis is then the surface of an ellipsoid whose major axis is along the n_x -axis with a radius $\{(1 - \omega_a^2/\omega^2)/(1 - \omega_b^2/\omega^2)\}^{1/2}$ and whose minor axis is along the n_z -axis with a radius $\{(1 - \omega_a^2/\omega^2)/(1 - \omega_b^2/\omega^2)\}^{1/2}$. Since the surface is closed, there is no resonance. In the high frequency limit when $\omega \gg \omega_a$, the dispersion relation (8.4.1) reduces to

$$n^2 = 1$$
 (8.4.3)

which is isotropic. The wave becomes just the ordinary sound except that the exponential growth with height of the form (8.3.18) is still present to assure continuity of energy flow.

The regions of propagation of the gravity branch and acoustic branch are distinct in the ωk space as shown by the shaded regions in Fig. 8.4-2.



Fig. 8.4-2. Regions of propagation of the gravity branch and the acoustic branch in an isothermal atmosphere.

These two propagation branches are bounded by curves along which $k_z = 0$. As $k_x \to \infty$, these bounding curves approach asymptotically to $\omega = \omega_b$ for the gravity wave branch and $k_x = \omega/c$ for the acoustic branch. Curves corresponding to $k_z \neq 0$ appear in the interior of the shaded regions.

The remaining parts of this section are concerned with the group velocity and the polarization relations. We will take up the group velocity first.

As discussed in Section 2.12, the group velocity has a direction normal to the index surface. For the gravity branch, the index surface has a crosssection shown in Fig. 8.4-1. An example is drawn for the refractive index vector and its corresponding group velocity in Fig. 8.4-1. It is interesting to note that even though the phase progression of this wave has a downward component, the energy propagation has an upward component. In general, the vertical components of the phase progression and energy progression have opposite signs for the gravity branch and the same signs for the acoustic branch. This can be seen by examining the character of the dispersion surface or by examining the analytic expression for the group velocity. The dispersion relation for the internal wave is (8.3.19) which can also be written as

$$\omega^4 - \omega^2 c^2 (k_x^2 + k_z^2 + \omega_a^2/c^2) + \omega_b^2 c^2 k_x^2 = 0$$
(8.4.4)

Differentiating (8.4.4), we can obtain the horizontal and vertical components of the group velocity as

$$\nu_{gx} = \frac{\partial \omega}{\partial k_x} = \frac{\omega c^2 k_x (\omega^2 - \omega_b^2)}{(\omega^4 - \omega_b^2 k_x^2 c^2)} \qquad (8.4.5a)$$

$$v_{gz} = \partial \omega / \partial k_z = \omega^3 c^2 k_z / (\omega^4 - \omega_b^2 k_x^2 c^2)$$
(8.4.5b)

For the gravity branch, the factors $(\omega^2 - \omega_b^2)$, $(\omega^4 - \omega_b^2 k_x^2 c^2)$ are both negative so that v_{gx} has the same sign as k_x while v_{gz} has the opposite sign as k_z . But these two factors are both positive in the acoustic branch so that v_{gx} and v_{gz} both have the same sign as k_x and k_z , respectively. Of course, the waves are still anisotropic even in the acoustic branch except when $\omega \gg \omega_b$.

The system of linear equations (8.3.11) has a unique solution except for a constant multiplier. Therefore, we may solve three components of F in terms of the remaining component. The four equations are not linearly independent because of the dispersion relation. Let us eliminate the third equation of (8.3.11) and solve the field components in terms of v_z . The resulting equations of interest can be expressed in matrix form as

$$\begin{bmatrix} j\omega & 0 & -jk_x \\ 0 & -jk_x & j\omega \\ -j\omega c^2 & j\omega & 0 \end{bmatrix} \begin{bmatrix} \varrho'/\varrho_0 \\ \varrho'/\varrho_0 \\ v_x \end{bmatrix} = \begin{bmatrix} (ik_z + 1/2H)v_z \\ 0 \\ -(\gamma - 1)gv_2 \end{bmatrix}$$
(8.4.6)

The above set of equations can be solved by a matrix inversion. The determinant of the coefficient matrix is found to be $j\omega(\omega^2 - k_x^2 c^2)$. The inverted set of equations shows that the four components of the vector **F** must have the ratios

$$\begin{aligned} (\varrho'/\varrho_0) : (p'/\varrho_0) : v_x : v_z &= [\omega^2 (k_z - j/2H) + j(\gamma - 1)gk_x^2] \\ &: \omega^2 [k_z c^2 + j(\gamma/2 - 1)g] \\ &: \omega k_x [k_x c^2 + j(\gamma/2 - 1)g] : \omega [\omega^2 - k_x^2 c^2] \end{aligned} (8.4.7)$$

The relations (8.4.7) are sometimes called the polarization relations for the internal waves. We note that the ratio v_x/v_z gives the orbits of the air parcels under the influence of internal waves. The air parcels move in the plane of the propagation vector. In general, the ratio v_x/v_z is complex, indicating that the orbit is elliptical. There are two limiting cases for which the air parcels oscillate linearly. In the high frequency limit, $\omega \gg \omega_a$, we find that $v_x/v_z = k_x/k_z$. Therefore the air parcels oscillate along the propagation vector **k**, i.e., the wave is longitudinal. Near the resonance condition of the gravity branch, both k_x , k_z approach to infinity and we find $v_x/v_z = -k_z/k_x$, i.e., air parcels oscillate in a direction perpendicular to **k**. For other conditions, the orbit of an air parcel describes an ellipse. Define the right-handed and left-handed rotating vectors by

$$\hat{r} = (\hat{x} - j\hat{z}), \quad \hat{l} = (\hat{x} + j\hat{z})$$
 (8.4.8)

The velocity of the air parcel can be expressed in terms of these unit rotating vectors,

$$\mathbf{v} = \{\hat{x}k_{x}[k_{z}c^{2} + j(\gamma/2 - 1)g] + \hat{z}(\omega^{2} - k_{x}^{2}c^{2})\}Ae^{z/2H}e^{j(\omega t - k_{x}x - k_{z}z)}$$

$$= \{\hat{r}[k_{x}k_{z}c^{2} + jk_{x}(\gamma/2 - 1)g + j(\omega^{2} - k_{x}^{2}c^{2})] + \hat{l}[k_{x}k_{z}c^{2} + jk_{x}(\gamma/2 - 1)g - j(\omega^{2} - k_{x}^{2}c^{2})]\}(A/2)e^{z/2H}e^{j(\omega t - k_{x}x - k_{z}z)}$$

$$= (\hat{r}v_{r} + \hat{l}v_{l})(A/2)e^{z/2H}e^{j(\omega t - k_{x}x - k_{z}z)}$$
(8.4.9)

where A denotes the arbitrary amplitude and v_r , v_l are the relative complex amplitudes of the right-handed and left-handed rotating components, respectively. Their phase angles are denoted by θ_r and θ_l . The elliptical orbit of the air parcel is generally characterized by (1) the tilt angle of the ellipse, (2) the sense of rotation, and (3) the axial ratio. They are all determined by
v_r and v_l and are given by

tilt angle =
$$(\theta_r - \theta_l)/2$$

axial ratio = $\left| \frac{|v_r| + |v_l|}{|v_r| - |v_l|} \right|$
sense of rotation: $\begin{cases} \text{right-handed if } |v_r| > |v_l| \\ \text{left-handed if } |v_r| < |v_l| \end{cases}$ (8.4.10)

As an example, let us inquire when is the air-parcel orbit circular. This requires that either $v_r = 0$ or $v_l = 0$. We can find that air parcels describe a right-handed circle when $k_z = 0$ and $\omega = (\gamma/2)^{1/2}\omega_b$ (see problem at the end of this chapter). A few example orbits are shown in Fig. 8.4-3 in the $k_x k_z$ -plane.



Fig. 8.4-3. Sample air-parcel orbits associated with the propagation of internal waves (gravity branch) in an ideal gas. The plot is superimposed on the wave-number plane to show the relation between the orbits and the propagation vector. R = axial ratio, $\psi = tilt$ angle, $Y = \omega_b/\omega$, and $\phi = tan^{-1}k_s/k_x$. [After Chang (1969).]

8.5 Propagation in a Wind-Stratified Isothermal Atmosphere

Winds of high magnitudes have been observed in the upper atmosphere. Measurements indicate that the E region winds can be reduced to a steady component, a tidal component, and a component that varies in a few hours (Spizzichino, 1968). In the F region the solar heating sets up a diurnal bulge, a few hours trailing the subsolar point. As a result, pressure gradients are induced and they together with other forces drive the neutral air to produce winds. These winds are mainly in the horizontal directions. Their magnitudes may be as high as 200 m/sec which is a large fraction of the velocity of sound. Consequently, the winds are expected to modify the propagation of acoustic gravity waves in a profound way.

The effect of a constant horizontal background wind can be taken into account relatively simply by introducing a co-moving coordinate in which the background atmosphere is stationary. The wave properties are assumed known in the stationary atmosphere. The effect of wind can then be taken into account by an appropriate transformation.

Let S be a stationary frame in which the observer is situated. Let S' be a frame moving with the medium with a constant horizontal velocity v_0 . All quantities dependent on the motion and expressed in S are unprimed and those expressed in S' are primed. For a nonrelativistic velocity v_0 , the space-time transformation is given by the Galilean transformation

$$\mathbf{r}' = \mathbf{r} - \mathbf{v}_0 t, \qquad \mathbf{k}' = \mathbf{k}$$

$$t' = t, \qquad \omega' = \omega - \mathbf{k} \cdot \mathbf{v}_0 \qquad (8.5.1)$$

Such a transformation is applicable to acoustic gravity waves when v_0 is strictly horizontal since the isothermal atmosphere is inhomogeneous in the vertical direction. Note the invariance

$$\omega' t' - \mathbf{k}' \cdot \mathbf{r}' = \omega t - \mathbf{k} \cdot \mathbf{r}$$
(8.5.2)

which shows that if a wave is a plane wave in S' then it must be so in S and vice versa. Let us suppose that the dispersion relation in S' is given by

$$\omega' = f(\mathbf{k}') \tag{8.5.3}$$

with a corresponding group velocity

$$v_g' = \nabla_{k'} \omega' \tag{8.5.4}$$

Through Galilean transformation (8.5.1), we find that the dispersion relation in S becomes

$$\omega' = \omega - \mathbf{k} \cdot \mathbf{v}_0 = f(\mathbf{k}) \tag{8.5.5}$$

and the corresponding group velocity becomes

$$\mathbf{v}_g = \boldsymbol{\nabla}_k \boldsymbol{\omega} = \mathbf{v}_0 + \mathbf{v}_g' \tag{8.5.6}$$

The quantity ω' is sometimes called the intrinsic frequency. It is the frequency Doppler shifted due to the relative motion and is observed in S'. Comparing (8.5.3) with (8.5.5) shows that the effect of a background wind on the dispersion relation can be easily taken into account by allowing for a Doppler shift in frequency. The wind also carries the wavepacket bodily with it as indicated by (8.5.6). Actually (8.5.6) is a direct consequence of the Galilean transformation (8.5.1). If we let **r** be the position of the wavepacket in S, its position in S' is then **r**'. According to the first equation of (8.5.1), the wavepacket must move such that $d\mathbf{r}/dt = \mathbf{v}_0 + d\mathbf{r}'/dt$, which is identical to (8.5.6). The formulas (8.5.5) and (8.5.6) can be directly applied to acoustic gravity waves whose dispersion relation has been derived in Section 8.3.

When the background horizontal wind has a vertical shear, a multilayer model is often assumed. The whole region of interest is divided into horizontal layers and within each layer the wind is assumed constant. Such a procedure is valid if the wind shear is not too large. The foregoing results can then be applied to each layer. Appropriate boundary conditions must be applied at the interface of each layer. In the following, let us consider the effect of horizontal winds on the propagation of internal waves in an isothermal atmosphere by using a multilayer model.

Let us suppose that a plane internal wave of gravity branch is incident on a wind-stratified isothermal atmosphere. The effect of wind is to introduce a Doppler shift as indicated by (8.5.5). The dispersion curves plotted in k_xk_z -plane for gravity waves are shown in Fig. 8.5-1. Different curves correspond to different frequencies of the wave. Suppose the incident wave has a propagation vector indicated by the point A. The corresponding energy ray is normal to the curve at A and is shown by an arrow. When the wave penetrates into the first layer, the Doppler shift given by (8.5.5) indicates that the dispersion curve corresponding to a Doppler shifted frequency must be used. Since the kinematic boundary condition requires that the horizontal propagation vector be the same for all layers, the propagation vector k can change from layer to layer only along a vertical path through A as shown by the dotted line in Fig. 8.5-1. If the wind has a positive com-



Fig. 8.5-1. Dispersion curves in an isothermal atmosphere with c = 850 m/sec, $\gamma = 1.4, g = 9.7$ m/sec. The path of the k vector in a wind stratified isothermal atmosphere is along the vertical line through A which corresponds to the incident wave vector. If the horizontal wind is blowing with the wave, the path of k moves downward along the vertical line; while if the wind is blowing against the wave, the path of the k vector moves upward. [After Cowling, Webb, and Yeh (1971).]

ponent along the propagation vector, the Doppler shifted frequency decreases and the tip of the k vector moves downward along the dotted line. This downward motion continues if the wind is assumed to increase continuously in intensity with height. Eventually, the frequency is Doppler shifted to zero for which $\mathbf{k} \to \infty$. The layer at which $\omega' = 0$ is known as the critical layer. At the critical layer the horizontal phase shows that the wave is severely damped if there is present ever so small a loss mechanism. As the operating point is moved downward along the dotted line through A, the slope of the group velocity and hence also the slope of the ray decreases until the ray becomes entirely horizontal at the critical layer. A sample ray path is shown in Fig. 8.5-2.



Fig. 8.5-2. Group path of an internal gravity wave showing the effect of a critical layer. [After Cowling, Webb, and Yeh (1971).]

Let us now consider the opposite case for which the wind has a negative component along the propagation vector and the magnitude of the wind increases steadily with height. The incident wave has k corresponding to point A in Fig. 8.5-1. As the wave penetrates into the atmosphere, the tip of the k vector moves upward along the vertical dotted line through A. The direction of the group velocity in S' steepens until the operating point reaches near the k_x -axis at which the group velocity in S' is entirely horizontal and the wave is reflected. To an observer in S, the ray direction is given by the vector sum of the group velocity in S' and the horizontal wind vector as given by (8.5.6). For the present case, v_g' and v_0 have opposite horizontal components. Near the reflection point, the horizontal component of $v_{g'}$ is very small and the ray bends backward and makes a characteristic loop as shown in Fig. 8.5-3. After reflection, the direction of the ray is still given



Fig. 8.5-3. Group path of an internal gravity wave showing reflection in a wind stratified atmosphere. [After Cowling, Webb, and Yeh (1971).]

by $\mathbf{v}_0 + \mathbf{v}_{g'}$ where $\nu_{g'}$ continues to be determined by the vertical operating line extending to the first quadrant of the $k_x k_z$ -plane. The situation is similar to the incident case and is not shown in Fig. 8.5-1.

If the background wind is not sufficiently strong to give rise to either the critical layer condition or the reflection, the ray must then penetrate the atmosphere. An example is shown in Fig. 8.5-4.

It is therefore clear that background winds in an isothermal atmosphere have a very pronounced effect on the propagation of internal gravity waves. As a result we may classify rays into the following three types: critical trapping, reflection, and penetration. Examples are shown in Fig. 8.5-2 through Fig. 8.5-4 on the vertical plane. If there are present horizontal cross winds,



Fig. 8.5-4. Group path of an internal gravity wave showing penetration in a wind stratified atmosphere. [After Cowling, Webb, and Yeh (1971).]

these rays may deviate out of the vertical plane. Horizontal deviations are not shown in these figures.

The theory of ray tracing of internal waves in a general, slowly varying medium can be developed along the line discussed in Chapter 5. Such a theory is available in the literature (Jones, 1969).

8.6 Effect of Ion Drag

So far in this chapter, our consideration of the propagation of acoustic gravity waves is restricted to the isothermal and ideal atmosphere in which all damping mechanisms have been ignored. We wish now to examine the effect produced by the presence of ions. This case is of interest in the real upper atmosphere because of the presence of the ionosphere.

The propagation of acoustic gravity waves in a coupled neutral atmosphere and ionosphere is rather complex because of the large number of equations. The mathematical problem can be simplified if we view the ionosphere as responding to the propagation of such waves. As a result, the ionosphere is set into motion and therefore the wave is damped. The damping of the wave due to the presence of the ionosphere is not the only effect on the wave. Since the ionosphere is inhomogeneous, there is also the possibility for partial reflection. The partial reflection problem requires solving a differential equation with variable coefficients. We will treat one such case in this section.

In the F region heights, we have the following inequalities:

$$\omega_{Bi} \gg v_{\rm in} \gg \omega \tag{8.6.1}$$

The angular frequency of acoustic gravity waves is very small when compared with the ion-neutral collision frequency v_{in} (approximately 1/sec). The process is quasi-equilibrium and therefore ions and neutrals must move with the same velocity. However, since the angular gyrofrequency of atomic oxygen ions ω_{Bi} is approximately 300 rad/sec which is much larger than v_{in} , the ions essentially spiral about magnetic lines of force. The combined effect of inequalities (1) is that the wave-induced ionic velocity is along the magnetic field and its magnitude is equal to the component of the neutral velocity in the same direction. Mathematically, the ionic velocity is related to the neutral velocity through

$$\mathbf{v}_i = (\mathbf{v} \cdot \hat{B}_0) \hat{B}_0 \tag{8.6.2}$$

where \hat{B}_0 is a unit vector in the direction of the steady magnetic field. In a coordinate system in which \mathbf{B}_0 is in the *xz*-plane with *z*-axis vertically upward and with a dip angle *I*, the ionic velocity given by (8.6.2) can be written in component form as

$$\mathbf{v}_i = (v_x \cos I - v_z \sin I)(\hat{x} \cos I - \hat{z} \sin I)$$
 (8.6.3)

The basic equations are still those given by (8.3.4) except for the addition of a frictionlike ion drag term on the left-hand side of the equation of motion. The equation of motion now reads

$$\varrho \ D\mathbf{v}/Dt = -\nabla p + \varrho \mathbf{g} - \mathbf{v}_{\rm in} \varrho_p (\mathbf{v} - \mathbf{v}_i) \tag{8.6.4}$$

where ρ_p is the plasma mass density. The continuity equation and the adiabatic equation of (8.3.4) are still unchanged. It should be pointed out that strictly speaking, the wave motions are no longer adiabatic in the presence of loss. If the loss is small, the motions are approximately adiabatic and this we assume here. Going through the analysis similar to that carried out in Section 8.3 by linearizing perturbations about a windless, nonrotating isothermal atmosphere, we find that the following matrix equation results.

$$\mathbf{D} \cdot \mathbf{F} = \mathbf{0} \tag{8.6.5}$$

for wave solutions of the form $f(z)e^{j(\omega t - k_x x - k_y y)}$ The field vector F and the coefficient matrix operator are given, respectively, by

$$\mathbf{F} = \begin{bmatrix} p'/\varrho_{0} \\ p'/\varrho_{0} \\ v_{x} \\ v_{y} \\ v_{z} \end{bmatrix}$$
$$\mathbf{D} = \begin{bmatrix} j\omega & 0 & -jk_{x} & -jk_{y} & \partial/\partial z - H^{-1} \\ 0 & -jk_{x} & j\omega + v \sin^{2} I & 0 & v \sin I \cos I \\ 0 & -jk_{y} & 0 & j\omega + v & 0 \\ g & \partial/\partial z - H^{-1} & v \sin I \cos I & 0 & j\omega + v \cos^{2} I \\ -j\omega c^{2} & j\omega & 0 & 0 & (\gamma - 1)g \end{bmatrix}$$
(8.6.6)

where $v = v_{in} \varrho_{p0}/\varrho_0$ is the effective neutral-ion collision frequency. Note the obvious generalization of (8.6.6) when compared with (8.3.12). In general, in the ionosphere, v is height dependent. The operator **D** then has variable coefficients. We will consider one such case later on in this section. Let us first consider the simple case in which v is constant. For this case, the operator **D** has constant coefficient and therefore we may seek solutions with z dependence given by $e^{-jK_x z}$. For such solutions we may replace $\partial/\partial z$ by $-jK_z$ and the matrix **D** becomes algebraic. The necessary and sufficient condition to have nontrivial solution of (8.6.5) is the vanishing of the determinant of **D**. Setting

$$\det \mathbf{D}(k_x, k_y, K_z, \omega) = 0 \tag{8.6.7}$$

we obtain the desired dispersion relation. Since **D** is a 5×5 matrix, the algebra involved is rather lengthy and is left as a problem at the end of this chapter. For the special case I = 0, $k_y = 0$, corresponding to propagation

in the magnetic meridian plane at the magnetic equator, the dispersion relation reduces to

$$K_z^2 = k_0^2 (1 - j\nu/\omega - \omega_a^2/\omega^2) - k_x^2 (1 - j\nu/\omega - \omega_b^2/\omega^2) \quad (8.6.8)$$

where $K_z = k_z + j/2H$. In the absence of collisions, (8.6.8) reduces to (8.3.19). The presence of collisions is to make k_z complex since for a given incident wave, k_x is real.

When v is a function of height, we must go back to (8.6.5). For the special case of propagation in the magnetic meridian plane at the magnetic equator, a single equation for the z component of the velocity can be obtained, viz.,

$$\frac{d^2 v_z}{dz^2} - (1/H) \frac{dv_z}{dz} + [k_0^2 - k_x^2(1 - \omega_b^2/\omega^2) - j(v/\omega)(k_0^2 - k_x^2)]v_z = 0$$
(8.6.9)

Let

$$v_z = u e^{z/2II}$$
 (8.6.10)

The differential equation (8.6.9) can be transformed into

$$\frac{d^2u}{dz^2} + k_z^2(z)u = 0 \tag{8.6.11}$$

Here, the vertical wave number $k_z(z)$ is given by (8.6.8) and is a function of z through the dependence of ν on height. To proceed further analytically, we must assume an appropriate model for ν that can lead to a solution in terms of known functions. One convenient form is the exponential function

$$v = v_0 e^{\alpha(z-z_0)} \tag{8.6.12}$$

Let us define a new height variable ζ by

$$\zeta = (2k_0/\alpha)[j(1-n_x^2)/\omega]^{1/2}e^{i\pi/2+\alpha(z-z_0)/2}$$
(8.6.13)

The differential equation (8.6.11) can be transformed to the standard Bessel's equation

$$\frac{d^2 u}{d\zeta^2} + \frac{1}{\zeta} \frac{d u}{d\zeta} + \frac{\nu_0 - m^2}{\zeta^2} u = 0$$
 (8.6.14)

where the order of the Bessel's equation is given by

$$m = (j/\alpha)[k_0(1 - \omega_a^2/\omega^2) - k_x^2(1 - \omega_b^2/\omega^2)]^{1/2} = (j/\alpha) |k_2(-\infty)| (8.6.15)$$

Compare (8.6.15) with (8.3.19) and note that the expression in the square

bracket is just the vertical wave number in the absence of collisional loss; it occurs for our model collision frequency (8.6.12) when $z \to -\infty$. The two solutions of (8.6.14) are Hankel functions of the *m*th order of the first kind $H_m^{(1)}(v_0^{1/2}\zeta)$ and of the second kind $H_m^{(2)}(v_0^{1/2}\zeta)$. The complete determination of the solution depends on the boundary conditions. As mentioned earlier, we are concerned with the incidence of an internal wave from below and the wave propagates upward. At ground level, ν is negligibly small and the internal wave propagates with little loss. When the wave penetrates into the ionosphere, the collisions become more frequent, and because of the inhomogeneous nature of the collision frequency, there is partial reflection along the way. Therefore, at the ground, there is a reflected wave in addition to an incident wave. When z becomes very large, we require that u be finite on physical grounds. These boundary conditions are now applied to the following two cases:

(i) Acoustic Branch. We found in Section 8.4 that when $1 - n_x^2 > 0$, the wave belongs to the acoustic branch. The boundary conditions are

$$z \to -\infty, \quad u \to e^{-jk_z(-\infty)z} + Re^{jk_z(-\infty)z}$$

 $z \to \infty, \quad u \text{ must be finite}$ (8.6.16)

where R is the reflection coefficient and $k_z(-\infty)$ is the vertical wave number of the incident wave. In this case, from (8.6.13), arg $\zeta = 3\pi/4$. As $z \to \infty$, $|\zeta| \to \infty$.

The asymptotic expressions for large arguments are

$$\begin{split} H_m^{(1)}(x) &= (2/\pi x)^{1/2} e^{j(x-m\pi/2-\pi/4)}, \qquad -\pi < \arg \chi < 2\pi \\ H_m^{(2)}(x) &= (2/\pi x)^{1/2} e^{-j(x-m\pi/2-\pi/4)}, \qquad -2\pi < \arg \chi < \pi \end{split}$$

For finite u as $z \to \infty$, only $H_m^{(1)}$ can be taken as our solution. When $z \to -\infty$ $|\zeta| \to 0$ and we may use the approximate expression of $H_m^{(1)}$ for small argument,

$$H_m^{(1)}(\nu_0^{1/2}\zeta) = \frac{1}{j\sin m\pi} \left[\frac{(\nu_0^{1/2}\zeta/2)^{-m}}{(-m)!} - e^{im\pi} \frac{(\nu_0^{1/2}\zeta/2)^m}{m!} \right]$$

Transforming the expression above back to variable z, we note that the first term is proportional to

$$e^{-m_{\alpha}(z-z_0)/2} = e^{-jk_z(-\infty)z}$$

and the second term is proportional to

$$e^{m_{\alpha}(z-z_0)/2} = e^{jk_z(-\infty)z}$$

The reflection coefficient can be easily obtained as

$$R = \frac{(-m)!}{m!} \left(\frac{k_0}{\alpha}\right)^{2m} (j\nu_0/\omega)(1 - n_x^2)^m e^{-jk_z(-\infty)z_0}$$
(8.6.17)

Since $k_2(-\infty)$ is real if the incident wave is a propagating mode, *m* is therefore purely imaginary. The magnitude of the reflection coefficient is then

$$|R| = (j)^m = e^{-\pi k_2(-\infty)/\alpha}$$
 (8.6.18)

The collision frequency is constant when $\alpha = 0$ for which the reflection coefficient vanishes. The upgoing and downgoing waves propagate independently. This is the case studied earlier in this section. When ν is not a constant, the reflection takes place because of the inhomogeneous nature of the medium. The magnitude of the reflection coefficient depends on the ratio $k_z(-\infty)/\alpha$. If $k_z(-\infty) < \alpha$, the collision frequency varies considerably in one vertical wavelength of the gravity wave, and the reflection is appreciable as expected.

(ii) Gravity Branch. When $1 - n_x^2 < 0$, the wave belongs to the gravity branch. The boundary conditions are now

$$z \to -\infty, \quad u \to e^{jk_z(-\infty)z} + Re^{-jk_z(-\infty)z}$$

 $z \to \infty, \quad u \text{ must be finite}$ (8.6.19)

Here R is the reflection coefficient. As discussed in Section 8.4, a gravity wave with energy propagating upward has a negative vertical wave number. This accounts for the sign difference in the exponent between (8.6.16) for the acoustic branch and (8.6.19) for the gravity branch. In this case, arg $\zeta = 5\pi/4$ and the upper boundary condition requires that the solution be given by $H_m^{(2)}(v_0^{1/2}\zeta)$. The magnitude of the reflection coefficient can be similarly found to be

$$|R| = e^{-\pi |k_z(-\infty)|/\alpha}$$
(8.6.20)

Details are assigned as a homework problem. Again, the reflection is appreciable if the effect of neutral-ion collision frequency varies appreciably over one vertical wavelength.

8.7 Attenuation due to Thermal Conduction and Viscosity

In the last section we saw that the effect of ion drag is to attenuate the wave. Additional loss processes such as thermal conduction and viscosity further attenuate the wave. To discuss these effects fully requires considerable exposition. This is because (1) the wave motions are no longer adiabatic and (2) there exist additional modes. As shown in thermodynamics, the effect of a loss process is to increase the total entropy with time. This means that the adiabatic equation (8.3.4c) is no longer valid and that it must be replaced by the conservation of energy equation. The presence of loss processes also increases the order of equation, and as a result, a total of four modes may appear. They are acoustic gravity mode, thermal condition mode, ordinary viscosity mode, and extraordinary viscosity mode. In general, these modes are coupled and can be studied by the coupled mode technique discussed in Chapter 5. Such a formulation can be found in the literature, and interested readers should consult, for example, Volland



Fig. 8.7-1. The attenuation of internal gravity wave amplitude in the upper atmosphere k = horizontal wavelength, T = period, ϕ = azimuth angle of the wave measured from magnetic north. The ionization profile is as shown. [After Clark *et al.* (1970).]

(1969). The results are invariably very complicated, involving a large number of parameters. Interpretation of the results are difficult except through extensive numerical computations. In general, it has been found that the loss processes begin to have an effect at a height of about 250 km. Figure 8.7-1 is an example which shows the effect of loss to an otherwise exponentially growing amplitude.

8.8 Effect of Internal Waves in the Ionospheric F Region

We have seen that the presence of internal waves introduces perturbations in the neutral density, pressure, and velocity. These perturbations in the neutral atmosphere at ionospheric heights will further excite changes in the ionization density due to modifications in atmospheric processes. To discuss fully all effects requires a careful examination of these processes and is beyond the scope of this book. Fortunately, the inequality (8.6.1)is valid in the F region and thus the approximation (8.6.2) can be made.

Let the electron density be given by

$$N = N_0(z) + N'(\mathbf{r}, t)$$
(8.8.1)

where the unperturbed density N_0 is horizontally stratified. The perturbed density N' satisfies the linearized equation

$$\frac{\partial N'}{\partial t} + \nabla \cdot (N_0 \mathbf{v}_i) = 0 \tag{8.8.2}$$

Equation (8.8.2) is drastically simplified since the only process that has been taken into account is the effect of induced ionization velocity which is assumed to be given by (8.6.2). In a lossless isothermal atmosphere, the perturbed neutral velocity must grow exponentially. In the actual atmosphere, this growth is approximately balanced by the loss processes as seen in Section 8.7. We will assume, for simplicity, that this is the case, i.e.,

$$\mathbf{v} \propto e^{j(\omega t - \mathbf{k} \cdot \mathbf{r})} \tag{8.8.3}$$

Under these conditions, (8.8.2) can be solved for N' to produce

$$N' = (1/\omega)(\mathbf{v} \cdot B_0)(\mathbf{k} \cdot B_0 + j\hat{B}_0 \cdot \hat{z} \,\partial/\partial z)N_0 \tag{8.8.4}$$

We see from (8.8.4) that the ionization perturbation is proportional to the component of neutral velocity along the magnetic field. At the peak of

ionization where $\partial N_0/\partial z = 0$ or above the magnetic equator where $\hat{B}_0 \cdot \hat{z} = 0$, the perturbed ionization density and the neutral velocity are in phase. In general, however, N' and v are out of phase and the phase lag depends on the scale of vertical gradient in background ionization relative to the vertical wavelength of the wave.

More careful analysis shows that a term $N'v_d$ may be important. Here, v_d is the diffusion velocity of the ionization. The inclusion of such a term introduces additional phase shift. It has also been discovered that on occasion, a wave of perturbation magnitude may induce large ionization changes not of perturbation magnitude. Figure 8.8-1 shows iso-ionic contours computed by using (8.8.4) and more careful numerical technique. The phase difference is very apparent.

Wave motions in the ionosphere have been studied fairly extensively by Munro (1950). These waves have been named as traveling disturbances. Evidence is strong that these traveling disturbances are actually driven by the passing internal waves in the manner discussed in this section.



Fig. 8.8-1. Iso-ionic density contour in the presence of an internal gravity wave. The dotted curves are computed by using (8.8.4) and the full curves are computed by numerically solving the equation of continuity. [After Clark *et al.* (1970).]

8.9 Impulse Response of an Isothermal Atmosphere

The discussion of acoustic-gravity waves has been so far concerned with free waves. No attempt has been made to connect these waves with the source or to see the modification of the waves by the presence of boundaries. In the real atmosphere, sources of acoustic-gravity waves are many. Interested readers are referred to papers in the literature [e.g., the *Symposium Proceedings on Acoustic-Gravity Waves in the Atmosphere*, edited by Georges (1968)]. In this book, we will be concerned only with the impulse response. The reasons for discussing the impulse response are: (1) The impulse response is the Green's function which is the starting point in treating more complicated sources. (2) Violent events that occur in the atmosphere may be approximated by impulses. One outstanding example is nuclear detonations in the atmosphere. (3) Some approximate expressions for the Green's function in fairly simple analytic forms can be obtained, and remarkable confirmations with experimental data have been found.

Processes involved in a nuclear detonation in the atmosphere are very involved and certainly nonlinear in the vicinity of the explosion. Associated with these processes, equivalent sources of mass, momentum, and energy may be present. Effects of these processes are felt at a distance as waves diverging from the region of explosion. If the distance is large, we may approximate the small source as a point source.

The relevant linearized equations have been given by (8.3.8) except that now we must add source terms. Addition of source terms modifies (8.3.11)to the following set.

~

212.

$$\mathbf{D}(\mathbf{\nabla}, \partial/\partial t) \cdot \mathbf{F} = \mathbf{S} \tag{8.9.1}$$

212

.

010

where

$$\mathbf{D}(\nabla, \partial/\partial t) = \begin{bmatrix} \partial/\partial t & 0 & \partial/\partial r & \partial/\partial z - 1/H \\ 0 & \partial/\partial r & \partial/\partial t & 0 \\ g & \partial/\partial z - 1/H & 0 & \partial/\partial t \\ -c^2 \partial/\partial t & \partial/\partial t & 0 & (\gamma - 1)g \end{bmatrix}$$
(8.9.2)
$$\mathbf{F} = \begin{bmatrix} \varrho'/\varrho_0 \\ p'/\varrho_0 \\ v_r \\ v_z \end{bmatrix}$$
(8.9.3)

The coordinate r is the horizontal distance in cylindrical coordinates. For plane solutions, the matrix operator (8.9.2) reduces to the algebraic matrix

given by (8.3.12). Let $\hat{\mathbf{D}}$ be the adjoint matrix operator of **D**. The elements of $\hat{\mathbf{D}}$ are formed by the transposed cofactor of **D**. Let det **D** be the determinants of **D** and it is a differential operator. Then from (8.9.1) we have

det
$$\mathbf{D}(\nabla, \partial/\partial t)\mathbf{F} = \hat{\mathbf{D}} \cdot \mathbf{S}$$
 (8.9.4)

The response to any source S is obtained by solving (8.9.4). The differential operator det **D** is found from (8.9.2) to be

det
$$\mathbf{D}\left(\overline{V}, \frac{\partial}{\partial t}\right)$$

= $\frac{\partial^4}{\partial t^4} - c^2 \frac{\partial^2}{\partial t^2} \left(\frac{\partial^2}{\partial r^2} + \frac{\partial^2}{\partial z^2} - \frac{1}{H} \frac{\partial}{\partial z}\right) - \omega_b^2 c^2 \frac{\partial^2}{\partial r^2}$ (8.9.5)

It reduces to the left-hand side of the dispersion relation (8.3.14) for plane wave solutions for which we replace ∇ by $-j\mathbf{k}$ and $\partial/\partial t$ by $j\omega$.

For concreteness, let us assume that the source is caused by the production of mass at a rate of kg/m³/sec. For this case, the source function is $\mathbf{S} = (q/\varrho_0, 0, 0, 0)$. The equation for $\nu_r(\mathbf{r}, t)$ is then, from (8.9.4),

det
$$\mathbf{D}v_r = -c^2 \frac{\partial}{\partial r} \left(\frac{\partial^2}{\partial t^2} + \omega_b^2 \right) \frac{q}{\varrho_0}$$
 (8.9.6)

Equation (8.9.6) is the equation of concern; it can be solved by first taking the Fourier transform with respect to it. In the transformed domain, (8.9.6) reduces to

$$\begin{bmatrix} \omega^4 + \omega^2 c^2 \left(\frac{\partial^2}{\partial z^2} + \frac{\partial^2}{\partial z^2} - \frac{1}{H} \frac{\partial}{\partial z} \right) - \omega_b^2 c^2 \frac{\partial^2}{\partial r^2} \Big] v_r(r, \omega) \\ = \frac{c^2}{\varrho_0} \left(\omega^2 - \omega_b^2 \right) \frac{\partial q(\mathbf{r}, \omega)}{\partial r}$$
(8.9.7)

The term involving the first derivative with respect to z can be eliminated if we multiply both sides of (8.9.7) by $\rho^{1/2}$ and write the equation for $\nu_r \rho_0^{1/2}$. In an isothermal atmosphere, ρ_0 is exponentially distributed and is given by (8.3.5). The relations such as

$$\varrho_0^{1/2} \partial v_r / \partial z = \partial (v_r \varrho_0^{1/2}) / \partial z + (v_r \varrho_0^{1/2}) / 2H$$

can be derived easily. The use of such relations reduces the resulting equation to

$$\left[\left(\frac{\partial^2}{\partial r^2} + \frac{\omega^2}{\omega^2 - \omega_b^2} \frac{\partial^2}{\partial z^2}\right) + \frac{\omega^2}{c^2} \frac{\omega^2 - \omega_a^2}{\omega^2 - \omega_b^2}\right] \varrho_0^{1/2} \nu_r(\mathbf{r}, \omega) = \varrho_0^{-1/2} \frac{\partial q(\mathbf{r}, \omega)}{\partial r}$$
(8.9.8)

434 8. Interaction of Atmospheric Waves with the Ionosphere

The operator on the left-hand side of (8.9.8) looks somewhat like the Helmholtz wave operator except for its cylindrical symmetry rather than the spherical symmetry. The operator can be made spherically symmetric by a scale change in z. Let us introduce the coordinate transformation

$$r' = r, \qquad z' = z(1 - \omega_b^2/\omega^2)^{1/2}$$
 (8.9.9)

In the primed coordinates, the differential equation (8.9.8) transforms to

$$\left[\nabla^{\prime 2} + \frac{\omega^2}{c^2} \frac{\omega^2 - \omega_a^2}{\omega^2 - \omega_b^2} \right] \varrho_0^{1/2} \nu_r(\mathbf{r}^\prime, \omega) = \varrho_0^{-1/2} \frac{\partial q(\mathbf{r}^\prime, \omega)}{\partial r^\prime} \quad (8.9.10)$$

The operator is now reduced to the Helmholtz wave operator whose inversion in terms of Green's function is known. Define a Green's function to satisfy

$$\left(\nabla^{\prime 2} + \frac{\omega^2}{c^2} \frac{\omega^2 - \omega_a^2}{\omega^2 - \omega_b^2} \right) G(\mathbf{r}, \omega) = \frac{\delta(z) \,\delta(r)}{2\pi r}$$
$$= -\left(1 - \omega_b^2 / \omega^2\right)^{1/2} \,\delta(z') \,\delta(r') / 2\pi r'$$
(8.9.11)

The source on the right of (8.9.11) is an impulse point source at the origin. Equation (8.9.11) has the well-known retarded solution

$$G(\mathbf{r}',\omega) = \frac{(1-\omega_b^2/\omega^2)^{1/2}}{4\pi R'} \exp\left[-j\frac{\omega(\omega^2-\omega_a^2)^{1/2}}{c(\omega^2-\omega_b^2)^{1/2}}R'\right] \quad (8.9.12)$$

The spherical radial distance R' in the transformed coordinates is given by

$$R' = [r^{2} + (1 - \omega_{b}^{2}/\omega^{2})z^{2}]^{1/2} = (R/\omega)(\omega^{2} - \omega_{c}^{2})^{1/2}$$

where

$$R = (r^2 + z^2)^{1/2}, \qquad \omega_c = \omega_b z/R \le \omega_b$$
 (8.9.13)

The Green's function in the original coordinates is then

$$G(\mathbf{r},\omega) = \frac{1}{4\pi R} \left(\frac{\omega^2 - \omega_b^2}{\omega^2 - \omega_c^2} \right)^{1/2} \exp\left[-j \frac{R}{c} \left(\frac{(\omega^2 - \omega_a^2)(\omega^2 - \omega_c^2)}{\omega^2 - \omega_b^2} \right)^{1/2} \right]$$
(8.9.14)

As discussed earlier, $\omega_b \leq \omega_a$ for $\gamma \leq 2$. An examination of the exponential factor in (8.9.14) shows that the free propagation is permitted if either $\omega > \omega_a$ or $\omega_c < \omega < \omega_b$. For the special case of a homogeneous atmosphere in a zero gravity, $\omega_a = \omega_b = \omega_c = 0$, and (8.9.14) reduces to

$$G(\mathbf{r},\omega) = (1/4\pi R)e^{-j(R/c)\omega}$$
(8.9.15)

which has the Fourier image in time domain

$$G(\mathbf{r}, t) = (1/4\pi R) \ \delta(t - R/c)$$
(8.9.16)

There is no distortion of the pulse as the medium is not dispersive. The causal nature of (8.9.16) is clear since the response is delayed by R/c which is the time required by the ordinary sound to reach the observer. Causality is still expected to be valid when there is finite gravity. The initial arrival of the signal is called the precursor and its contribution comes mainly from the high frequency components of (8.9.14). Since (8.9.14) reduces to (8.9.15) in the high frequency limit, the arrival of the precursor is still expected to be delayed by a time R/c. The long time behavior of the impulse response depends on the lowest frequency that can propagate in the medium. For (8.9.14), the lowest propagation frequency is ω_e and so the response tends to oscillate with a frequency ω_c as $t \to \infty$. This is significant as the behavior can be checked experimentally. In some experiments, the probing of the atmosphere is confined to a fairly narrow height region. This means that the farther the observer is from the source, the lower is the oscillation frequency in the decay portion of the response. [See (8.9.13) for the definition of ω_c .] This has been found to be true experimentally.

The inversion of the Green's function (8.9.14) is difficult in general. One special case that can be evaluated exactly is when the observer is directly over the source, i.e., r = 0 and thus $\omega_c = \omega_b$. The Green's function is now

$$G(0, z, \omega) = \left(\frac{1}{4\pi z}\right) \exp\left[-j\frac{R}{c} (\omega^2 - \omega_a^2)^{1/2}\right]$$
(8.9.17)

This expression has an exact inverse; it is

$$G(0, z, t) = \frac{1}{4\pi z} \,\delta\!\left(t - \frac{z}{c}\right) - \frac{z\omega_a J_1[\omega_a(t^2 - z^2/c^2)^{1/2}]}{c(t^2 - z^2/c^2)^{1/2}} \,u\!\left(t - \frac{z}{c}\right) \tag{8.9.18}$$

In addition to an impulse which propagates with the velocity of sound, there is an additional term in (8.9.18). For $t \gg z/c$, this additional term oscillates with a frequency ω_a . The plane wave analysis of Section 8.4 has shown that for vertical propagation, only the acoustic branch can propagate. The acoustic branch has a lower cutoff frequency ω_a which dictates the long time behavior.

The exact inversion of (8.9.14) is difficult. Therefore, we will consider the following approximate case. Let us consider the long-time behavior for which the contribution comes mainly from integration for small ω . Further, for observers far away from ground zero, the height of the atmosphere

(roughly 300 km) may be small compared to the ground distance (several thousand kilometers), i.e., $z/R \ll 1$. Hence $\omega_c \ll \omega_b$. This means that as far as the integration around the branch cut from $-\omega_c$ to ω_c is concerned, we may approximately write $\omega_a \approx \omega_b$. Making such approximations reduces (8.9.14) to

$$G(\mathbf{r},\omega) \approx \frac{1}{4\pi R} \frac{j\omega_b}{(\omega^2 - \omega_c^2)^{1/2}} \exp\left[-j\frac{R}{c} (\omega^2 - \omega_c^2)^{1/2}\right] \quad (8.9.19)$$

This approximate expression (8.9.19) has known Fourier inverse

$$G(\mathbf{r}, t) \approx \frac{\omega_b}{4\pi R} J_0 \left(\omega_c \left(t^2 - \frac{R^2}{c^2} \right)^{1/2} \right) u \left(t - \frac{R}{c} \right)$$
(8.9.20)

where u is the unit step function. The zeroth-order Bessel function J_0 oscillates about the origin and zero crossings become periodic for large argument. Equation (8.9.20) shows that at a fixed height z, the farther is the observer, the slower is the oscillation. The expression (8.9.20) is not valid for t near R/c. An improvement of (8.9.20) can be made so that the acoustic branch contribution is taken into account [see Row (1967) and the problem at the end of this chapter].

A very general consideration of the source problem in an isothermal atmosphere can be carried out by using asymptotic methods. Interested readers may wish to consult Pierce and Posey (1970) or Liu and Yeh (1971).

8.10 Effect of a Wind Shear in the Ionospheric E Region

Experimental observations have indicated that on occasions there exist in the *E* region very thin (several kilometers) and intense ionization layers. These layers are called sporadic *E* layers or E_s layers. A certain type of E_s layers in the temperate latitudes seems to correlate very well with the existence of wind shears. A critical review of experimental observations and theories can be found in Whitehead (1970).

It was originally proposed that a shear in horizontal wind may push ionization along the field line which results in accumulation or diminution of ionization at the node. If this is so, the north-south wind would be most effective in producting E_s . A full analysis shows that shears of east-west winds also produce polarization electric field which, through Lorentz force, is more effective in redistributing ionization than the north-south winds. The origin of the wind shear is presumed to come from propagating internal waves.

Let us orient our coordinate axes so that x-, y-, z-directions are given respectively by northward, westward, and upward directions. The xz-plane thus contains the geomagnetic field vector \mathbf{B}_0 . For simplicity, we assume steady state condition and horizontal stratification (i.e., $\partial/\partial t = \partial/\partial x$ $= \partial/\partial y = 0$). Then by Faraday's law, $\nabla \times \mathbf{E} = -\mathbf{B} = 0$, we conclude that both E_x and E_y are constant. The starting point is the equation of motion which for α th species of charged particles is given by

$$0 = Z_{\alpha} e(\mathbf{E} + \mathbf{v}_{\alpha} \times \mathbf{B}_{0}) + m_{\alpha} v_{\alpha} (\mathbf{v}_{n} - \mathbf{v}_{\alpha})$$
(8.10.1)

Here the pressure gradient term has been ignored; its inclusion leads to diffusion effect later and has control on the distribution of ionization density. The neutral wind is assumed to blow in the east-west direction but sheared vertically, i.e., $\mathbf{v}_n = (0, V(z), 0)$. Such an assumption is not entirely consistent with properties of internal waves and is made to simplify analysis. The three component equations of (8.10.1) can be used to solve for \mathbf{v}_{α} . The resulting equation for $v_{\alpha z}$ is

$$\nu_{\alpha z}(\nu_{\alpha}^{2} + \omega_{B\alpha}^{2}) = -Z_{\alpha}\omega_{B\alpha}\nu_{\alpha}V\sin\theta + (Z_{\alpha}\omega_{B\alpha}^{2}e/m_{\alpha}\nu_{\alpha})E_{x}\sin\theta\cos\theta$$
$$-(e\omega_{B\alpha}/m_{\alpha})E_{y}\sin\theta + (Z_{\alpha}e/m_{\alpha}\nu_{\alpha})(\nu_{\alpha}^{2} + \omega_{B\alpha}^{2}\cos^{2}\theta)E_{z}$$
(8.10.2)

where $\omega_{B\alpha}$ is the angular gyrofrequency and θ , the polar angle of the steady magnetic field. Equation (8.10.2) actually represents two equations, the electron equation when $\alpha = e$ and the ion equation when $\alpha = i$. For convenience, ions are assumed to be singly charged. The parameters applicable to *E* region heights have orders of magnitudes given by (4.20.8). For electrons, we have $\omega_{Be} \cos^2 \theta \gg v_e^2$ and thus (8.10.2) can be used to solve for E_z , yielding

$$E_z = -(m_e \nu_e/e \cos^2 \theta) \nu_{ez} + (m_e \nu_e^2 \sin \theta/\omega_{Be} e \cos^2 \theta) V$$

- tan $\theta E_x - (\nu_e \sin \theta/\omega_{Be} \cos^2 \theta) E_y$ (8.10.3)

For ions, the inequality $\omega_i^2 \ll v_i^2$ is valid and (8.10.2) reduces to

$$v_{iz}v_i^2 = -\omega_{Bi}v_iV\sin\nu + (ev_i/m_i)E_z \qquad (8.10.4)$$

Now we assume that the process does not introduce appreciable charge separation and vertical current. Charge neutrality requires $N_i = N_e \equiv N$.

Absence of vertical current and charge neutrality results in $v_{iz} = v_{ez} \equiv v_z$. Substitute (8.10.3) into (8.10.4) and make use of these approximate relations; we obtain

$$v_z = -\frac{\omega_i}{v_i} V \sin \theta - \frac{e \sin \theta}{m_i v_i \cos \theta} E_x - \frac{e v_e \sin \theta}{\omega_{Be} m_i v_i \cos^2 \theta} E_y \quad (8.10.5)$$

As a result of winds, the ionization must move vertically according to (8.10.5).

The ionization in the E region in the absence of wind shear satisfies the equation

$$q = \alpha N_0^2 \tag{8.10.6}$$

where q is the production function and α , the recombination coefficient. In the presence of wind shear, the continuity equation takes the form

$$\partial (v_z N)/\partial z = q - \alpha N^2$$
 (8.10.7)

where the production function is assumed to be unchanged. The difference of (8.10.6) and (8.10.7) is therefore

$$\frac{\partial(\nu_z N)}{\partial z} = \alpha (N_0^2 - N^2) \tag{8.10.8}$$

The ionization profile can be obtained only by solving (8.10.8) and the solution depends on the nature of $v_z(z)$. In the following, we will be concerned only with the stationary value of N.

Let z_0 be used to denote the height at which the ionization is either a maximum or a minimum. Then at z_0 , dN/dz vanishes and (8.10.8) reduces to an algebraic equation

$$[N(z_0)/N_0]^2 - \eta [N(z_0)/N_0] - 1 = 0$$
(8.10.9)

which has solutions

$$N(z_0)/N_0 = \frac{1}{2} [\eta \pm (\eta^2 + 4)^{1/2}]$$
(8.10.10)

The dimensionless parameter

$$\eta = -[dv_z/dz]_{z_0}/\alpha N_0 \tag{8.10.11}$$

represents the ratio of the rate of accumulation of charge by transport process through wind shear to the rate of loss due to recombination. When Problems

 dv_z/dz is large and negative at z_0 , the transport process dominates over the recombination process. When this happens, (8.10.10) simplifies to

$$N(z_0)/N_0 = \begin{cases} \eta & \text{at maximum} \\ 1/\eta & \text{at minimum} \end{cases}$$
(8.10.12)

and the ionization is very large at the maximum, and very small at the minimum.

The observed range of $N(z_0)/N_0$ varies from 1 to 10. With the use of *E* region data and observed horizontal wind shears, $|dv_z/dz|$ has an order of magnitude 0.01 sec⁻¹. If N_0 is taken to be 10⁵/cm³, it will require α to be 10⁻⁸ cm³/sec. This value is lower by one order of magnitude than that inferred by the *E* region eclipse measurement. It is speculated that the discrepancy may be due to the presence of multiple ions with different reaction rates.

Problems

1. Find the height distribution in pressure in an atmosphere whose scale height increases linearly with height.

2. Consider a thick isothermal atmosphere in which the gravity is given by $g = g_0 a^2/(a + z)^2$. Find the height distribution of pressure and density in this atmosphere. Note that the density and pressure are both finite even at infinite height. How can this difficulty be overcome? [See E. A. Milne, *Trans. Cambridge Phil. Soc.* 22, 483 (1922-1923).]

3. Show that the Galilean transformation (8.5.1) is valid by deriving it from fluid equations (8.3.4) in an isothermal atmosphere. Can such a transformation be applied if there is a constant vertical wind in the isothermal atmosphere?

4. Let the atmosphere be horizontally stratified and have a constant lapse rate γ , i.e., $T = T_0 - \gamma z$ where T_0 is the temperature at ground level and z is the height. Consider now the propagation of sound in such an atmosphere.

(a) Show that the ray that makes a grazing incidence at ground is given by $dx = (T/\gamma z)^{1/2} dz$.

(b) Apply the result to ray paths of sound generated by a thunder at a height h. Show that the "critical" ray that strikes the earth at grazing inci-

440 8. Interaction of Atmospheric Waves with the Ionosphere

dence is approximately given by a parabolic curve. Note that because of refraction, observers beyond the critical ray cannot hear the thunder.

(c) For a lapse rate of 7.5° C/km, temperature of 300° K, and height of thunder of 4 km, show that thunder should not be audible beyond about 25 km.

5. The orbits of air parcels perturbed by the propagation of internal waves in an ideal gas have been discussed in Section 8.4. The orbits can be circular only when $k_z = 0$ and when ω has some special values. Find the value of ω when the orbit is right-circular and left-circular, respectively. Express answers in ω_b and ω_a .

6. Derive the dispersion relation for an acoustic gravity wave in the presence of ion drag. This can be obtained by expending (8.6.7). Show that for the special case of propagation in the magnetic meridian plane at the magnetic equator, the dispersion relation reduces to (8.6.8).

7. The effect of ion drag on the propagation of internal waves has been discussed in Section 8.6. Let the effective neutral-ion collision frequency be exponential given by (8.6.12). Find the reflection coefficient for an internal gravity wave incident at $z = -\infty$. Show that the magnitude of the reflection coefficient is given by (8.6.20).

8. Equation (8.8.4) is an approximate formula that gives the gravity wave induced ionization density perturbation. Suppose in a Faraday rotation experiment in which integrated ionization density along the path from, for example, a transmitting satellite to the ground station is observed, the satellite can be assumed to be outside the ionosphere. Compute the perturbed integrated density along the ray path. Show that if the background ionization density is given by an α -Chapman model, i.e.,

$$N_0(z) = N_m \exp\left[\frac{1}{2}\left[1 - \frac{z - z_m}{H} - \exp\left(-\frac{z - z_m}{H}\right)\right]\right]$$

then the fractional perturbation in the integrated ionization density is given by

$$\left| I'/I \right| = (k_h/\omega \cos^2 \chi) \left| \mathbf{v} \cdot \hat{B}_0 \right| \left[(\hat{\mathbf{r}} \times \hat{B}) \times \hat{z} \cdot \hat{k} \right] [\cosh(\pi k H \cos \eta/\cos \chi)]^{-1/2}$$

where k_{\hbar} is the horizontal wave number, χ the zenith angle of the ray, \hat{r} a unit vector along the ray from the transmitter to the receiver, and η the angle between \hat{r} and \hat{k} .

References

9. This problem is concerned with the impulse response of an isothermal atmosphere. A better approximation of (8.5.14) than that given by (8.5.19) is

$$G(\mathbf{r},\omega) = \frac{1}{4\pi R} \left(\frac{\omega^2 - \omega_b^2}{\omega^2 - \omega_c^2} \right)^{1/2} \exp\left[-j \frac{R}{c} \left(\omega^2 - \omega_c^2 \right)^{1/2} \right]$$

This Green's function is composed of two factors f_1 and f_2 defined by

$$f_1(\omega) = (\omega^2 - \omega_b^2)^{1/2}$$

$$f_2(\omega) = \frac{1}{4\pi R(\omega^2 - \omega_c^2)^{1/2}} \exp\left[-j\frac{R}{c}(\omega^2 - \omega_c^2)^{1/2}\right]$$

Both have known Fourier inversion. Therefore, the Fourier inversion can be achieved by applying the convolution theorem to $f_1(t)$ and $f_2(t)$. Find $G(\mathbf{r}, t)$ [R. V. Row, Acoustic-gravity waves in the upper atmosphere due to a nuclear detonation and an earthquake. J. Geophys. Res. 72, 1599-1610 (1967)].

10. In Section 8.10 we discussed the formation of E_s due to vertical shear of EW wind. If, in addition, there exists NS wind, how should (8.10.5) be modified? Show that at temperature latitudes, the NS wind is a factor ω_{Bi}/v_i smaller when compared with the EW wind.

References

- AXFORD, W. I., and CUNNOLD, D. M. (1966). The wind-shear theory of temperate zone sporadic E, *Radio Sci.* 1, 191–198.
- BOOKER, J. R., and BRETHERTON, F. P. (1967). The critical layer for internal gravity waves in a shear flow. J. Fluid Mech. 27, 513.
- CHANG, N. J. F. (1969). Acoustic-gravity waves in the ionosphere and their effects on high-frequency radio propagation. Ph. D. thesis, Univ. of Colorado, Boulder, Colorado.
- CLARK, R. M., YEH, K. C., and LIU, C. H. (1970). Gravity waves in the ionosphere. Tech. Rep. No. 39. Ionosphere Radio Lab., Univ. of Illinois, Urbana.
- COWLING, D. H., WEBB, H. D., and YEH, K. C. (1971). Group rays of internal gravity waves in a wind-stratified atmosphere. J. Geophys. Res. 76, 213-220.
- GEORGES, T. M. (1967). Ionospheric effects of atmospheric waves. Tech. Rep. IER 57-ITSA 54. ESSA, Boulder, Colorado.
- GEORGES, T. M. (ed.) (1968). Acoustic-Gravity Waves in the Atmosphere Symp. Proc. 1968. U. S. Govt. Printing Office, Washington, D. C.
- HINES, C. O. (1960). Internal atmpspheric gravity waves at ionospheric heights. Can. J. Phys. 38, 1441-1481.
- HOOKE, W. H. (1968). Ionospheric irregularities produced by internal atmospheric gravity waves. J. Atmos. Terr. Phys. 30, 795-823.

- JONES, W. L. (1969). Ray tracing for internal gravity waves. J. Geophys. Res. 74, 2028-2033.
- LIU, C. H., and YEH, K. C. (1969). The effect of ion drag on propagation of acousticgravity waves in the atmospheric F region. J. Geophys. Res. 74, 2248-2255.
- LIU, C. H., and YEH, K.C. (1971). Excitation of acoustic-gravity waves in an isothermal atmosphere, *Tellus* 23, Part 2, 150-163.
- MIDGLEY, J. E., and LIEMOHN, H. B. (1966). Gravity waves in a realistic atmosphere. J. Geophys. Res. 71, 3 29-3748.
- MUNRO, G. H. (1950). Traveling disturbances in the inosphere, Proc. Roy. Soc. Ser. A 202, 208-223.
- PFEFFER, R. L., and ZARICHNY, J. (1962). Acoustic-gravity wave propagation from nuclear explosions in the Earth's atmosphere. J. Atmos. Sci. 19, 256–263.
- PIERCE, A. D. (1965). Propagation of acoustic-gravity waves in a temperature- and wind-stratified atmosphere. J. Acoust. Soc. 37, 218-227.
- PIERCE, A. D., and POSEY, J. W. (1970). Theoretical prediction of acoustic-gravity pressure waveforms generated by large explosions in the atmosphere. Final Rep. Dept. of Mech. Eng., MIT, Cambridge, Massachusetts.
- RATCLIFFE, J. A. (ed.) (1960). "Physics of the Upper Atmosphere." Academic Press, New York.
- Row, R. V. (1967). Acoustic-gravity waves in the upper atmosphere due to a nuclear detonation and an earthquake. J. Geophys. Res. 12, 1599-1610.
- SPIZZICHINO, A. (1968). Mesures de Vents par Meteores entre 80 et 110 km. In "Winds and Turbulence in Stratosphere, Mesosphere and Ionosphere" (K. Rawer, ed.), pp. 201–236. Wiley, New York.
- VOLLAND, H. (1969). The upper atmosphere as a multiply refractive medium for neutral air motions, J. Atmos. Terr. Phys. 31, 491-514.
- WHITEHEAD, J. (1970). Production and prediction of sporadic E. Rev. Geophys. Space Phys. 8, 1, 65-144.

Appendix A. The Method of Steepest Descents

A.1 The Method of Steepest Descents

In this appendix we shall discuss briefly the essential practical aspects of what is known as the method of steepest descents in the asymptotic evaluation of integrals.

Consider the integral

$$I = \int_C e^{\lambda f(\xi)} F(\xi) \, d\xi \tag{A.1.1}$$

in the complex ξ -plane, where $f(\xi)$ and $F(\xi)$ are arbitrary analytic functions of ξ and λ is a large positive real parameter. There is no loss of generality in assuming that λ is positive, since the sign may always be included in the function $f(\xi)$. We shall assume that the path of integration C goes from $-\infty$ to $+\infty$. The basic idea of this technique of asymptotic evaluation is to change the path C to a so-called "steepest descent" path such that along this new path the value of the integral is determined mainly by the contribution from a short portion of the path. We should keep in mind that from the theory of complex variables the value of the original integral is not changed after the deformation of the path if there are no singularities between the two paths. Let us write

$$f(\xi) = f_1(\xi) + jf_2(\xi)$$
 (A.1.2)

Hence, the exponential factor in the integrand of (A.1.1) becomes

$$e^{j\lambda f_2 + \lambda f_1}$$
 (A.1.3)

The steepest descent path is a path along which the function f_1 has a maximum at some point on the path and decreases as fast as possible away from this point.

Since f_1 and f_2 are the real and imaginary parts of the analytic function f they have the property that in the complex ξ -plane, the lines of most rapid decrease or increase of f_1 are the lines of constant values of f_2 and vice versa. This can be seen by considering the rate of change of f along any path S which makes an angle α with the ξ_1 axis ($\xi = \xi_1 + j\xi_2$).

$$\frac{\partial f_1}{\partial S} = \cos \alpha \,\frac{\partial f_1}{\partial \xi_1} + \sin \alpha \,\frac{\partial f_1}{\partial \xi_2} \tag{A.1.4}$$

This rate will be stationary with respect to α if

$$\frac{\partial}{\partial \alpha} \left(\frac{\partial f_1}{\partial S} \right) = -\sin \alpha \frac{\partial f_1}{\partial \xi_1} + \cos \alpha \frac{\partial f_1}{\partial \xi_2} = 0 \qquad (A.1.5)$$

But since f_1 and f_2 are the real and imaginary parts of the analytic function f, hence, according to the Riemann-Cauchy relations,

$$\partial f_1 / \partial \xi_1 = \partial f_2 / \partial \xi_2, \qquad \partial f_1 / \partial \xi_2 = -\partial f_2 / \partial \xi_1$$
 (A.1.6)

Therefore in (A.1.5) we have

$$0 = -\sin\alpha \frac{\partial f_2}{\partial \xi_2} - \cos\alpha \frac{\partial f_2}{\partial \xi_1} = -\frac{\partial f_2}{\partial S}$$
(A.1.7)

which is satisfied on a path of constant f_2 . Along this path, the stationary point of f_1 is at $\partial f_1/\partial S = 0$. Since $\partial f_2/\partial S$ also is zero on the path, from the Cauchy-Riemann relations, $\partial f_1/\partial n = \partial f_2/\partial n = 0$ at this stationary point for f_1 , where dn is the element normal to the path. Therefore this point can be found by setting

$$df/d\xi = 0 \tag{A.1.8}$$

and is called the saddle point of f. For a first-order saddle point ξ_0 , we have $f''(\xi_0) \neq 0$; then in general the equation $f_2(\xi) = f_2(\xi_0)$ corresponds to two curves. Along one of them f_1 has a maximum at ξ_0 and decreases most rapidly away from it while along the other f_1 has a minimum at ξ_0 and increases most rapidly away from the saddle point. Therefore for this case

there are two steepest descent directions and two steepest ascent directions from ξ_0 . In Fig. A-1 we have shown the two constant phase lines $(f_2 = \text{constant})$ and the two constant amplitude lines $(f_1 = \text{constant})$ passing through ξ_0 . The height of this saddle point is $f_1(\xi_0)$. The hatched regions are called the valleys since the $f_1(\xi)$ in these regions are smaller than $f_1(\xi_0)$. The two adjacent unhatched regions are called hills since $f_1(\xi)$ there is greater than $f_1(\xi_0)$. ξ_0A and ξ_0B are the two steepest descent directions; $\xi_0 C$ and $\xi_0 D$, two steepest ascent directions.





In general, for a saddle point ξ_0 of order m-1 such that $f^{(m)}(\xi_0)$ is the lowest nonzero derivative of $f(\xi)$ at ξ_0 , there are 2m steepest directions from ξ_0 , m directions of steepest ascent, and m directions of steepest descent. In the following, we shall concentrate first on the case where ξ_0 is a first-order saddle point.

Let us now return to the integral (A.1.1). The most desirable path in our integration will be the one passing through the saddle point along which, for large λ , the magnitude of the integrand decreases rapidly with distance from the saddle point; hence the value of the integral is determined approximately by the portion of the path near the saddle point. The larger the value of λ , the faster the decrease of magnitude away from the saddle point. For Fig. A-1, this corresponds to the path $A\xi_0B$. This is the steepest descent path (SDP). Let us suppose that a saddle point is obtained from (A.1.8) as ξ_0 and the path of integration C can be changed to the steepest descent path through ξ_0 without crossing any singularities. Define the new variable of integration along the SDP by

$$f(\xi) = f(\xi_0) - \zeta^2$$
 (A.1.9)

where ζ is the new variable and varies from $-\infty$ to $+\infty$. The saddle point

Appendix A

corresponds to $\zeta = 0$. The integral (A.1.1) is now written in the new variable

$$I = e^{\lambda f(\xi_0)} \int_{-\infty}^{+\infty} e^{-\lambda \zeta^2} \Phi(\zeta) d\zeta \qquad (A.1.10)$$

where

$$\boldsymbol{\Phi}(\zeta) = F(\xi) \, d\xi / d\zeta \tag{A.1.11}$$

Since λ is large, it is convenient to expand $\Phi(\zeta)$ in the integrand of (A.1.10) in a power series of ζ

$$\Phi(\zeta) = \Phi(0) + \Phi'(0)\zeta + \frac{1}{2}\Phi''(0)\zeta^2 + \cdots$$
 (A.1.12)

Equation (A.1.10) now can be evaluated as a series with terms involving inverse powers of $\hat{\lambda}$

$$I = e^{\lambda f(\xi_0)} (\pi/\lambda)^{1/2} \Big[\Phi(0) + \frac{1}{4\lambda} \Phi''(0) + \cdots \Big]$$
 (A.1.13)

The series (A.1.13) is generally not convergent but is asymptotic. For sufficiently large λ the first term alone usually serves as a good approximation of the integral (A.1.1), and is called the asymptotic expression of I in λ .

The expression (A.1.11) in general does not have a closed form. The difficulty involved is to find ξ as a function of ζ from (A.1.9). In the following we shall indicate a method to obtain $\Phi(\zeta)$ in a power series of ζ . Differentiating (A.1.9) with respect to ξ , we have

$$f'(\xi) = -2\zeta \, d\zeta/d\xi$$

or $f'(\xi) d\xi/d\zeta = -2\zeta$. Therefore

$$\Phi(\zeta) = -2\zeta F(\xi)/f'(\xi) \qquad (A.1.14)$$

Expanding $f(\xi)$ and $F(\xi)$ in powers of $\eta = \xi - \xi_0$

$$f(\xi) = f(\xi_0) + \frac{1}{2} f''(\xi_0) \eta^2 + \frac{1}{6} f'''(\xi_0) \eta^3 + \cdots$$

$$F(\xi) = F(\xi_0) [1 + \eta F'(\xi_0) / F(\xi_0) + \frac{1}{2} \eta^2 F''(\xi_0) / F(\xi_0) + \cdots]$$
(A.1.15)

Then in (A.1.9)

$$\frac{1}{2}f''(\xi_0)\eta^2 + \frac{1}{6}f'''(\xi_0)\eta^3 + \cdots = -\zeta^2$$
 (A.1.16)

Inverting this series, we can represent η in terms of a power series in ζ . Let us set

$$\eta = [\zeta/(-\frac{1}{2}f''(\xi_0))^{1/2}](1 + a_1\zeta + a_2\zeta^2 + \cdots)$$
 (A.1.17)

and substitute it into (A.1.16). Equating the coefficients of like powers of ζ , we obtain

$$a_1 = f^{\prime\prime\prime}(\zeta_0)/12[-\frac{1}{2}f^{\prime\prime}(\xi_0)]^{3/2}$$
 (A.1.18)

etc.

Substituting (A.1.15) into (A.1.14), we have

$$\Phi(\zeta) = -2\zeta F(\xi_0) \frac{1 + \eta F'(\xi_0)/F(\xi_0) + \frac{1}{2}\eta^2 F''(\xi_0)/F(\xi_0) + \cdots}{f''(\xi_0)\eta + \frac{1}{2}f'''(\xi_0)\eta^2 + \cdots}$$

When (A.1.17) is substituted in this expression, we can write $\Phi(\zeta)$ in a power series of ζ after some manipulation. The result is

$$\Phi(\zeta) = \frac{F(\xi_0)}{(-\frac{1}{2}f''(\xi_0))^{1/2}} \times \left\{ 1 + \zeta \left[\frac{F'(\xi_0)}{F(\xi_0)(-\frac{1}{2}f''(\xi_0))^{1/2}} + \frac{f'''(\xi_0)}{6(-\frac{1}{2}f'''(\xi_0)^{3/2}} \right] + \cdots \right\}$$
(A.1.19)

The expressions we need in (A.1.13) are

$$\Phi(0) = \left(-\frac{2}{f''(\xi_0)}\right)^{1/2} F(\xi_0)
\Phi''(0) = 2\Phi(0) \left[\frac{f'''}{(f'')^2} \frac{F'}{F} + \frac{1}{4} \frac{f^{IV}}{(f'')} - \frac{5}{12} \frac{(f''')^2}{(f'')^3} - \frac{F''}{Ff''}\right]$$
(A.1.20)

Higher order terms can be obtained in the similar manner. The asymptotic expression for the integral I now can be written as

$$I \sim e^{\lambda f(\xi_0)} \left(-\frac{2\pi}{f''(\xi_0)} \right)^{1/2} F(\xi_0) \frac{1}{\sqrt{\lambda}}$$
(A.1.21)

If there are more than one saddle point at the same height, then the contribution from each of them should be taken provided that the contour can be deformed to pass through each one of them without crossing any singularity. When poles are crossed in deforming the path, their contributions must be taken into account by the usual residue method.

We see that the method of steepest descents is often cumbersome because of the difficulty in expressing ξ as a function of ζ . A simpler procedure known as the saddle point method will be discussed briefly in the following.

Instead of (A.1.9), let us expand the function $f(\xi)$ by a Taylor series about the saddle point

$$f(\xi) = f(\xi_0) + f''(\xi_0)(\xi - \xi_0)^2/2 + \cdots$$
 (A.1.22)

We now choose the path of integration near ξ_0 to be a straight line on which the second term of (A.1.22) is real and negative. The direction of this line is tangent at ξ_0 to the two directions of steepest descent from ξ_0 . For large λ , we can expand $F(\xi)$ in (A.1.1) about ξ_0 and the first term of the integral becomes

$$F(\xi_0)e^{\lambda f(\xi_0)} \int_P e^{\lambda [f''(\xi_0)(\xi-\xi_0)^2/2+\cdots]} d\xi$$
 (A.1.23)

where P is any path passing through the saddle point (assuming of course the original path C may be changed to P). Let us assume that near ξ_0 the desired path is given by

$$\xi = \xi_0 + \varrho e^{j\theta} \tag{A.1.24}$$

Therefore

$$f''(\xi_0)(\xi - \xi_0)^2 = |f''(\xi_0)| \varrho^2 e^{j(2\theta + \alpha)}$$

where α is the phase angle of $f''(\xi_0)$. The requirement that (A.1.24) be real and negative results in $\theta = -\frac{1}{2}\alpha \pm \frac{1}{2}\pi$. Therefore in the neighborhood of ξ_0 , along this desired path, (A.1.23) may be written as

$$F(\xi_0)e^{\lambda f(\xi_0)} \int_{-\varepsilon}^{+\varepsilon} e^{[-\lambda f''(\xi_0)\varrho^2 + j(\pi-\alpha)]/2} d\varrho \qquad (A.1.25)$$

But for large λ , along the steepest descent path, the major contribution to (A.1.23) comes from the neighborhood of ξ_0 . Therefore, we can approximate (A.1.23) by (A.1.25), which, after change of variable, becomes

$$F(\xi_0)e^{\lambda f(\xi_0)+\frac{1}{2}j(\pi-\alpha)} \left(\frac{2}{\lambda \mid f''(\xi_0) \mid}\right)^{1/2} \int_{-\eta}^{+\eta} e^{-u^2} du \qquad (A.1.26)$$

where

$$\eta = \varepsilon(\lambda | f''(\xi_0) | /2)^{1/2}$$
 (A.1.27)

For large λ , η is large and the limits of integration in (A.1.16) may be extended to $\pm\infty$, and we have

$$F(\xi_0)e^{\lambda f(\xi_0)} \left(\frac{-2}{\lambda |f''(\xi_0)|e^{j\alpha}}\right)^{1/2} \int_{-\infty}^{+\infty} e^{-u^2} du = F(\xi_0)e^{\lambda f(\xi_0)} \left(\frac{-2\pi}{\lambda f''(\xi_0)}\right)^{1/2}$$
(A.1.28)

This is just the asymptotic expression for the integral I as given by (A.1.21) Higher-order terms may be obtained by the same procedure.

For the case where the saddle point ξ_0 is of the order *m*, the asymptotic expression for the integral *I* may be obtained easily by the saddle point method

$$I \sim e^{\lambda f(\xi_0) - j\pi/(m+1)} \left[\frac{\Gamma(M+2)}{f^{(m+1)}(\xi_0)} \right]^{1/(m+1)} \frac{\omega_{m+1}\Gamma\left(\frac{1}{m+1}\right)}{(m+1)\lambda^{1/(m+1)}}$$
(A.1.29)

where $\Gamma(x)$ is the Gamma function and ω_n is the *n*th roots of unity. The proper choice of ω_m in (A.1.29) depends on the direction of the steepest descent path through the saddle point.

A.2 Modified Method of Saddle Point

In some cases, the saddle point of $f(\xi)$ is close to a pole of the function $F(\xi)$. For these cases the Taylor series expansion of $F(\xi)$ about ξ_0 will no longer be valid in a sufficiently large region, and we cannot transform the integral as we have done in (A.1.28). Instead we must use the Laurent-series expansion for $F(\xi)$. Assuming that ξ_0 is a first-order saddle point, then along the steepest descent path defined by (A.1.24) the integral (A.1.1) may be transformed into the following form, except for a multiplying factor,

$$I_1 = \int_0^\infty t^\alpha g(t) e^{\lambda t} dt \qquad (A.2.1)$$

where t = 0 corresponds to the saddle point ξ_0 and $\operatorname{Re} \alpha > -1$. g(t) is analytic at t = 0 but has a pole of order p at $t = t_0$ in the complex *t*-plane. Therefore g(t) is analytic in some sector $\phi_1 < \arg t < \phi_2$. We can expand g(t) by a Laurent series

$$g(t) = \sum_{1}^{p} b_{-s}(t-t_0)^{-s} + g_1(t) \qquad (A.2.2)$$

where $g_1(t)$ is analytic for $|t| < |t_1|$ and t_1 is the next singular point of g(t). The function $g_1(t)$ may then be expanded about the origin in a Taylor series

$$g_1(t) = \sum_{0}^{\infty} C_n t^n \tag{A.2.3}$$

where

$$C_n = \frac{1}{n!} \frac{d^n}{dt^n} \left[g(t) - \sum_{s=1}^p b_{-s}(t-t_0)^{-s} \right]_{t=0}$$
(A.2.4)

Substituting (A.2.2) and (A.2.3) into (A.2.1) and integrating term by term, we have

$$I = \sum_{n=0}^{\infty} C_n \Gamma(\alpha + n + 1) (-\lambda)^{-(\alpha + n + 1)} + \sum_{s=1}^{p} b_{-s} \Gamma(\alpha + 1) (-t_0)^{(\alpha - s)/2} (-\lambda)^{(s - \alpha - 2)/2} W_{\mu,\nu}(\lambda t_0) \quad (A.2.5)$$

$$\mu = -(\alpha + s)/2, \quad \nu = (\alpha - s + 1)/2 \quad (A.2.6)$$

here $W_{-}(\zeta)$ is Whittaker's confluent hypergeometric function (Erdelyi

where $W_{\mu,\nu}(\zeta)$ is Whittaker's confluent hypergeometric function (Erdelyi, 1953).

In particular,

$$W_{-1/4,-1/4}(\zeta) = \sqrt{\pi} \,\zeta^{1/4} e^{\zeta/2} \,\operatorname{erfc} \sqrt{\zeta} \tag{A.2.7}$$

where

$$\operatorname{erfc} x = (2/\sqrt{\pi}) \int_{x}^{\infty} e^{-t^{2}} dt \qquad (A.2.8)$$

is the complementary error function.

Equations (A.2.5) to (A.2.8) are used in Chapter 4 in the computation of transient waves in a plasma.

In certain situations the saddle point is close to a branch point or a zero of the integrand, or two or more saddle points are very close to each other. For all these cases the original saddle point method does not apply. Modified methods have been derived for the various cases. Interested readers are referred to the references.

References

- BLEISTEIN, N. (1967). Uniform asymptotic expansion of integrals with many nearby stationary points and algebraic singularities. J. Math. Mech. 17, 533-560.
- COPSON, E. T. (1965). "Asymptotic Expansions." Cambridge Univ. Press, London and New York.
- ERDELYI, A. (1953). "Higher Transcendental Functions," Vol. 1. McGraw-Hill, New York.
- OLVER, F. W. J. (1970). Why steepest descents? SIAM (Soc. Ind. Appl. Math.) Rev. 12, 228-247.
- VAN DER WAERDER, B. L. (1950). On the method of saddle points. Appl. Sci. Res. B 2, 33-43.

Appendix B. The Distant Radiation Field from a Localized Source

In this section we shall discuss a fairly general technique for asymptotically evaluating the radiation field from a localized source. The discussion essentially follows that by Lighthill (1965).

Let us consider a system in which small disturbances to the undisturbed state are governed by a linear partial differential equation with constant coefficients, such as

$$D(\nabla, \partial/\partial t)\psi(\mathbf{r}, t) = 0 \tag{B.1.1}$$

where D is a polynomial in the partial differential operators ∇ and $\partial/\partial t$; $\psi(\mathbf{r}, t)$ is some function specifying the disturbance. For this system, a plane wave

$$\psi(\mathbf{r},t) = \psi_0 e^{j(\omega t - \mathbf{k} \cdot \mathbf{r})} \tag{B.1.2}$$

can exist if the dispersion relation

$$D(\mathbf{k},\omega) = 0 \tag{B.1.3}$$

is satisfied. We shall assume that the undisturbed state is stable such that no solution of (B.1.3) exists with k real and the imaginary part of ω negative.

Equation (B.1.3) may be solved to yield

$$\omega = \omega_n(\mathbf{k}) \tag{B.1.4}$$

where the subscript indicates the different characteristic modes in the system. For a given ω , (B.14) represents a surface in the three-dimensional **k**-space. This is the dispersion surface. The group velocity is $\nabla_k \omega(\mathbf{k})$ and is normal to the surface.

Our problem is to find the expression for the disturbance, $\psi(\mathbf{r}, t)$, when the system is under the influence of a localized source. In particular, we first consider the case where the source is harmonic in time for which the right-hand side of (B.1.1) is replaced by

$$e^{j\omega_0 t} f(\mathbf{r}) \tag{B.1.5}$$

where $f(\mathbf{r})$ vanishes outside a limited spatial region around, say, the origin. We shall present a technique to determine the form of ψ at distances from the source region large compared with its size.

Since $f(\mathbf{r})$ is localized, it can be written as a Fourier integral

$$f(\mathbf{r}) = (1/(2\pi)^3) \iint_{-\infty}^{+\infty} F(\mathbf{k}) e^{-j\mathbf{k}\cdot\mathbf{r}} d\mathbf{k}$$
(B.1.6)

where

$$F(\mathbf{k}) = \iiint_{-\infty}^{\infty} f(\mathbf{r}) e^{j\mathbf{k}\cdot\mathbf{r}} \, d\mathbf{r}$$
(B.1.7)

is the Fourier transform of $f(\mathbf{r})$ and is a regular function of k_x , k_y , and k_z . The equation

$$D(\nabla, \partial/\partial t)\psi = e^{j\omega_0 t} f(\mathbf{r}) \tag{B.1.8}$$

with the substitution of (B.1.6) on the right-hand side, then has the formal solution

$$\psi(\mathbf{r}, t) = e^{j\omega_0 t} \phi(\mathbf{r}, \omega_0) \tag{B.1.9}$$

where

$$\phi(\mathbf{r},\omega_0) = \frac{1}{(2\pi)^3} \iiint_{-\infty} \frac{F(\mathbf{k})e^{-j\mathbf{k}\cdot\mathbf{r}}}{D(\mathbf{k},\omega_0)} d\mathbf{k}$$
(B.1.10)

The problem is to find the asymptotic expression for ϕ for large values of $|\mathbf{r}|$. Physically, what we want is to determine the wave numbers and amplitudes of the plane waves at large distances along a given direction away from the source region. However (B.1.10) as it stands, does not yield a unique solution that tends to zero at large distances, because in our

system plane wave solutions satisfying (B.1.3) exist for the homogeneous equation (B.1.1) and the denominator in (B.1.10) can vanish for real values of k_x , k_y , and k_z . For this case, many determinations of the integral (B.1.10) are possible. Out of all these, only one is of physical interest, the one commonly described as satisfying the "radiation condition." Physically, this corresponds the the requirement that at any observation point the steadystate wave solution must be "arrivable at by switching on (the source) and waiting" [see Lighthill (1960)]. One way to derive this is to replace the right-hand side of (B.1.8) by $e^{j(\omega_0 - j\varepsilon)t} f(\mathbf{r})$ where ε is very small so that the source has been built up from zero to its present strength gradually during all the time from $t = -\infty$. We then try to find a solution ψ also proportional to $e^{j(\omega_0 - j\varepsilon)t}$. Therefore in (B.1.10), $D(\mathbf{k}, \omega_0)$ is replaced by $D(\mathbf{k}, \omega_0 - j\varepsilon)$. To facilitate the discussion let us first rotate the coordinates such that the observation point is in the direction of one of the axis, say k_x . Then, (B.1.10) becomes

$$\phi(\mathbf{r},\omega-j\varepsilon) = \frac{1}{(2\pi)^3} \int_{\infty}^{\infty} dk_y \int_{\infty}^{\infty} dk_z \int_{\infty}^{\infty} \frac{F(k_x,k_y,k_z)e^{-jk_xx}}{D(k_x,k_y,k_z,\omega-j\varepsilon)} dk_x$$
(B.1.11)

where we have dropped the subscript on ω .

The integration in k_x may be carried out by contour integration in the complex k_x -plane. As shown in Fig. B-1, let us displace the original path of integration to a new one on which the imaginary part of k_x has a negative value -h.

For large values of x, the integrand in (B.1.11) is vanishingly small on this new path (of the order e^{-hx}). Therefore the original integral may be



Fig. B-1. Displacement of path of integration.
estimated by the residues at those poles of the integrand that lie between the two paths. Since $F(\mathbf{k})$ is regular for a localized source, the poles of the integrand come from the zeros of D alone. Hence, for given values of ω , k_y and k_z , we need to find those solutions k_x of $D(k_x, k_y, k_z, \omega - j\varepsilon) = 0$ that have imaginary part with values between 0 and -h. We know from the discussion below (B.1.3) that for our system there are solutions of $D(\mathbf{k}, \omega) =$ 0 with real k_x for given values of k_y , k_z , and ω . By giving ω a small negative imaginary part k_x for which D = 0 will acquire a negative imaginary part if

$$\partial \omega / \partial k_x > 0$$
 (B.1.12)

where $\partial \omega / \partial k_x$ may be computed from (B.1.3) or (B.1.4) and is just the component of the group velocity along the observation direction.

Although D = 0 may have zeros with negative imaginary parts of k_x even when ε is zero, we can take h to be so small that the contribution to the integral (B.1.11) comes only from those zeros that have been displaced to the lower half-plane from the real axis. Therefore we may conclude that the integral (B.1.11) is nonzero only if the condition (B.1.12) is satisfied. This means that wave energy will be found along a particular observation direction only if the component of the group velocity along that direction is positive. This is the physical significance of the radiation condition.

In the limit of $\varepsilon \to 0$, the contribution to the k_x -integration of (B.1.11) from each pole may be written as

$$-2\pi j \frac{F(k_x, k_y, k_z)e^{-jk_x x}}{\partial D(k_x, k_y, k_z, \omega)/\partial k_x}$$
(B.1.13)

where k_x satisfies $D(k_x, k_y, k_z, \omega) = 0$ and $\partial \omega / \partial k_x > 0$. The asymptotic expression for (B.1.11) is then given by

$$\phi(\mathbf{r},\omega) \sim \frac{-j}{(2\pi)^2} \iint_{S} \frac{F(k_x,k_y,k_z)e^{-jk_xx}}{\partial D(k_x,k_y,k_z,\omega)/\partial k_x} dk_y dk_z \quad (B.1.14)$$

for each normal model, where the integration surface S is the portion of the dispersion surface on which $\partial \omega / \partial k_x > 0$. The double integral in (B.1.14) may be estimated by the method of stationary phase for large values of x. The stationary points may be found by solving $2k_x(k_y, k_z, \omega)/\partial k_y = 0$ and $\partial k_x(k_y, k_z, \omega)/\partial k_z = 0$, or equivalently, $\partial D / \partial k_y = 0$ and $\partial D / \partial k_z = 0$. This means that the phase $k_x x$ of (B.1.14) is stationary on the surface S at those points $\mathbf{k}^{(i)}$ where the normal to S is parallel to the k_x -axis. The major

contributions to the integral then come from the neighborhood of these stationary points. At each of these points, if we temporarily choose the k_y and k_z axes along the principal directions of curvature of the surface, then locally, the surface may be represented approximately by

$$k_x = k_x^{(i)} + \frac{1}{2}K_y^{(i)}(k_y - k_y^{(i)})^2 + \frac{1}{2}K_z^{(i)}(k_z - k_z^{(i)})^2$$
(B.1.15)

where $K_y^{(i)} = \partial^2 k_x / \partial k_y^2$ and $K_z^{(i)} = \partial^2 k_x / \partial k_z^2$ are the two associated curvatures at the point $\mathbf{k}^{(i)}$. The curvatures are taken as positive where the surface is concave to the positive k_x -direction and negative when it is convex.

Substituting (B.1.15) into (B.1.14), the integral may be approximated by

$$\phi(\mathbf{r},\omega) \sim \frac{-1}{2\pi x} \sum_{i} \frac{F(\mathbf{k}^{(i)}) \exp[-jk_x^{(i)}x + j(\operatorname{sgn} K_y^{(i)} + \operatorname{sgn} K_z^{(i)})\pi/4]}{(|K_y^{(i)} K_z^{(i)}|)^{1/2} [\partial D(\mathbf{k},\omega)/\partial k_x]_{\mathbf{k}=\mathbf{k}^{(i)}}}$$
(B.1.16)

where sgn x means the sign of x.

Equation (B.1.16) may be put into the form that is invariant under the rotation of axes, viz,

$$\phi(\mathbf{r},\omega) \sim \frac{1}{2\pi} \sum_{i} \frac{C^{(i)} F(\mathbf{k}^{(i)}) \exp[-j\mathbf{k}^{(i)} \cdot \mathbf{r}]}{|K^{(i)}|^{1/2} |\nabla_k D|_{\mathbf{k}=\mathbf{k}^{(i)}} |\mathbf{r}|}$$
(B.1.17)

where

$$|\nabla_k D| = \left[\left(\frac{\partial D}{\partial k_x} \right)^2 + \left(\frac{D}{\partial k_y} \right)^2 + \left(\frac{\partial D}{\partial k_z} \right)^2 \right]^{1/2} = \left[\sum D_x^2 \right]^{1/2}$$
(B.1.18)

$$K = \frac{\sum D_{\alpha}^{2} (D_{\beta\beta} D_{\gamma\gamma} - D_{\beta\gamma}^{2}) + 2 \sum D_{\beta} D_{\gamma} (D_{\alpha\beta} D_{\alpha\gamma} - D_{\alpha\alpha} D_{\beta\gamma})}{(\sum D_{\alpha}^{2})^{2}} \quad (B.1.19)$$

is the Gaussian curvature. The summation is cyclic over the three coordinate axes denoted by α , β , and γ . The factor $C^{(i)}$ comes from the additional phase factor. For $K^{(i)} > 0$, it takes the values ± 1 where the dispersion surface is convex to the direction $\pm \nabla_k D$. For $K^{(i)} < 0$, it takes the values $\pm j$ according as the direction $\pm \nabla_k D$ is parallel or antiparallel to **r**.

To use (B.1.17), let us refer to Fig. B-2. We first draw the dispersion surface $D(k_x, k_y, k_z, \omega) = 0$. At each point on this surface, we draw (or imagine) an arrow normal to the surface, choosing for its direction from the two normal directions the one pointing towards the surface

$$D(k_x, k_y, k_z, \omega + \delta) = 0 \tag{B.1.20}$$

Then for a given observation point **r**, we find the wave vectors $\mathbf{k}^{(i)}$'s in (B.1.17) by taking those points on the dispersion surface $D(\mathbf{k}, \omega) = 0$ where



Fig. B-2. Dispersion surface in $k_x k_z$ -plane. The procedure to determine the wave vectors that contribute to the far field is demonstrated.

the arrows are in the direction of **r**. This way the condition that $(\nabla_k \omega) \cdot \mathbf{r} > 0$ and the stationary phase requirement are satisfied. Using these values of $\mathbf{k}^{(i)}$, Eq. (B.1.17) gives the asymptotic expression for the field ϕ at a distance far away from the source region. Physically, we may interpret this result in the following way. Since the observation point is very far away from the localized source, we may consider the source to be concentrated at the origin. The contribution to the far field only comes from those groups of waves in the source spectrum that have group velocity in the direction of the observation point. On the dispersion surface the normals from a small area dS around a point $\mathbf{k}^{(i)}$ fill a cone whose cross-sectional area increases with distance like $|K^{(i)}| r^2 dS$ (Fig. B-3). In the physical space, this is a cone of rays. The number of rays is conserved in the cone; hence the intensity is diminished by a factor $|K^{(i)}|^{-1}r^{-2}$ as the rays propagate to the observation point. This accounts for the factor $|K^{(i)}|^{-1/2}r^{-1}$ in the amplitude.

The expression (B.1.17) is valid if the Gaussian curvature $K^{(i)} \neq 0$. For the case $K^{(i)} = 0$, the technique may be modified to yield results that fall



Fig. B-3. Physical interpretation of the amplitude factor in the asymptotic expression (B.1.17).

off less rapidly than r^{-1} . Details of the modification may be found in Lighthill (1960).

Substituting (B.1.17) into (B.1.9), we obtain the steady-state far field expression due to a localized source. Using this result, we can also compute the asymptotic transient field for arbitrary time variation of the source. The transient field is given by

$$\psi(\mathbf{r},t) = \frac{1}{2\pi} \int_{\infty}^{\infty} e^{j\omega t} \,\phi(\mathbf{r},\omega) \,d\omega \qquad (B.1.21)$$

Substituting (B.1.17) into (B.1.21), $\psi(\mathbf{r}, t)$ may be put in the form

$$\psi(\mathbf{r},t) \sim \sum_{i} \int_{\infty}^{\infty} A^{(i)}(\mathbf{r},\omega) e^{jq^{(i)}(\mathbf{r},\omega)} d\omega = \sum_{i} \psi_{i} \qquad (B.1.22)$$

where

$$A^{(i)}(\mathbf{r},\omega) = \frac{1}{(2\pi)^2} \frac{C^{(i)}F(\mathbf{k}^{(i)},\omega)}{|K^{(i)}|^{1/2}|\nabla_k D|_{\mathbf{k}^{(i)}}|\mathbf{r}|}$$
(B.1.23)

$$q^{(i)}(\mathbf{r},\omega) = \omega t - r\xi^{(i)}(\omega) \tag{B.1.24}$$

$$r\xi^{(i)}(\omega) = \mathbf{k}^{(i)}(\omega) \cdot \mathbf{r}$$
 (B.1.25)

In general, $\xi^{(i)}(\omega) \to \omega/\nu$ as $\omega \to \infty$, where ν is a characteristic speed for the system (e.g., light speed in vacuum, sound speed in the gas, etc). We shall assume that the system is quiescent for time t < 0. This requires that the path of integration in (B.1.22) be below all singularities in the integrand. If the source is turned on at t = 0, then it is easy to show that at a distance rfrom the source, the field ψ is zero for $t < r/\nu$, in accordance with the causality principle. To show this, we only need to change the path of integration in (B.1.22) to one on which Im ω goes to $-\infty$.

For t > r/v, the field at r for large r may be computed using the saddle point technique discussed in Appendix A.1. The saddle point ω_s is determined by

$$\partial q^{(i)}/\partial \omega = 0$$
 or $\partial \xi^{(i)}/\partial \omega = t/r$ (B.1.26)

Applying (A.1.21) to (B.1.22), we obtain the contribution from one saddle point

$$\psi_{i}(r,t) \sim \left(\frac{2\pi}{r \mid \partial^{2} \xi^{(i)}(\omega_{s})/\partial \omega^{2} \mid}\right)^{1/2} A^{(i)}(\mathbf{r},\omega_{s}) \exp[j[\omega t - \xi^{(i)}(\omega_{s})r \pm \pi/4]]$$
(B.1.27)
$$\partial^{2} \xi^{(i)}(\omega_{s})/\partial \omega^{2} \ge 0$$

If there are more than one saddle points at the same height, we can compute the contribution in the same manner. Equation (B.1.27) then gives the asymptotic expression for the transient field for t > r/v. It is possible, using modified saddle point techniques, to find the asymptotic expression for the transient signal at most observation times starting from the first arrival of the signal when $t \rightarrow r/v$. Details of these computations may be found in Felsen (1969).

References

- FELSON, L. B. (1969). Transients in dispersive media, Part I: Theory. IEEE Trans. Antennas Propagat. AP-17, 191-200.
- LIGHTHILL, M. J. (1960). Studies on magneto-hydrodynamic waves and other anisotropic wave motions. *Phil. Trans. Roy. Soc. London Ser. A* 252, 397–430.
- LIGHTHILL, M. J. (1965). Group velocity. J. Inst. Math. Its Appl. 1, 1-28.
- WHITHAM, G. B. (1961). Group velocity and energy propagation for three-dimensional waves. Commun. Pure Appl. Math. 14, 675-691.

Numbers in italics refer to problems.

A

Abel's equation, 283 Acoustic cutoff frequency, 413 Adiabatic atmosphere, 404, 405 Adjoint matrix operator, 433 Advanced solution, 16 Air-parcel orbits associated with internal waves, 417, 418 Airy equation, 261, 268 Airy functions, 261 Alfvén velocity, 167, 177, 180 Anti-Stokes lines, 262 Appleton-Hartree formula, 183-193, 220 (18), 220 (19) Autocorrelation function, 312 Autocovariance, 313 Averaged Lagrangian, 393 Averaged variational principle, 392-399

B

Bessel's equation, 288, 426
Booker quartic equation, 299, 305 (13), 306 (14)
Born approximation, 317
Breaking of waves, 368-370
Breit and Tuve's theorem, 304 (6)
Brunt-Vaisala frequency, 408
Buoyancy frequency, 408
for isothermal atmosphere, 409

С

Cauchy-Riemann relation, 444 Causality principle, 16 Caustic, 246 Čerenkov cone, 120 Čerenkov radiation, 42 in fluid plasma, 113-123 Characteristic polarization, 42-49 for cold magnetoplasma, 161-165 Characteristic vector, 37 conjugate vector, 38 isotropic medium, 43 orthonormality relation, 38 Characteristic waves, dispersion relation, 37, 38, 41 excitation of, 67-71 in lossless anisotropic medium, 39, 40 in stratified magnetoplasma, 295 in stratified medium, 257 Charge neutrality, 86-89 Circularly polarized waves, 47, 156-158, 165, 166, 182, 183, 188, 189 Classical scattering coefficient for a single electron, 323 Cluster expansion, 352, 353 CMA diagrams, 197 Coherent distance, 358 Coherent kinetic energy flux, 55 for fluid plasma, 98, 99 relation to group velocity, 58 Coherent waves, 349 Collisional frequency, 133 Complementary error function, 144, 145, 146, 450 Compressional mode, 181, 182 Conductivity tensor, 153 Confluent hypergeometric function, 450 Correlation function of plasma density fluctuations, 100

Coupling matrix, 295 Coupling parameter, 286, 287, 300 Coupling points, 295 Critical layer, 421 Cross-modulation phenomenon, 379 Cross-over frequency, 166, 175, 176 Cross-over transition, 204 Cutoff, 41 for circularly polarized waves, 170 for electronic plasma, 191 for extraordinary wave, 177 for ordinary wave, 177 in parameter space, 194–196

D

D region, 3 dc conductivity, 31 for cold magnetoplasma, 154 for plasma, 134 Debye length, 87-89 Debye sphere, 93 Debye wave number, 92 Destructive transition, 201 Diagram method, 351-358 strongly connected, 354 weakly connected, 354 Dielectric tensor, 28, see also Relative dielectric tensor for cold magnetoplasma, 151, 152 for lossy magnetoplasma, 154 principle axes, 81, 82 (6) for warm magnetoplasma, 207 Dispersion, 25-28 frequency, 25-27 spatial, 27, 28 Dispersion relation, 37, 38 for acoustic-gravity waves, 412 for anisotropic media, 37 for cold beam plasma, 104 for cold magnetoplasma, 160, 161 for fluid plasma, 98, 99 for internal waves, 413 for longitudinal waves, 40 for nonlinear waves, 393 for transverse waves, 40 for warm beam plasma, 107 for warm magnetoplasma, 207 Dispersion surface, 60, 193

Dissipation of heat, 50 Distant radiation field, 451-458 for anisotropic medium, 80 Distribution function, 112 Maxwellian distribution, 127 (8) Dominant solution, 262 Doppler shifted frequency, 420 for internal waves, 420 for streaming plasmas, 104 for streaming warm plasmas, 214 Dyadic Green's function, 71-80 for anisotropic medium, 75, 80 for isotropic medium, 74 for longitudinal waves, 78 for transverse waves, 78 Dyadic notation, 7-9 Dyson equation, 355

Е

E region, 3 effect of wind shear in, 436 Effective propagation constant, 357, 358 Eikonal, 226 equation, 226 Electric polarization density, 17-23 for cold magnetoplasma, 151 for fluid plasma, 96 Elliptic integral, 389 Energy conservation of electromagnetic fields, 55 Energy density, 55 for electron plasma waves, 99 Energy velocity, 56-58 Equatorial electrojet, 213 Equivalence theorem, 304 (7) Euler equations of calculus of variation, 232, 233, 392 Extraordinary wave, 176, 177, 189

F

F region, 4 effect of internal waves in, 430 Faraday rotation, 170-173, 220 (22) in ionospheric applications, 218 (12) when propagation is perpendicular, 218 (11) when propagation is quasi-parallel, 192 Fast process, 23-28

460

Fermat's principle, 231–240 for anisotropic medium, 234, 237–240 for isotropic medium, 231, 232 for sharp boundary, 240, 241
Fluctuations, 317 correlations of, 343–349 of electromagnetic waves in random media, 329–333, 333–343 in plasma density, 99
Försterling's coupled equations, 286
Forward scattering approximation, 335
Fraunhofer diffraction region, 340
Fresnel formulas, 135, 248
Frozen-in magnetic field, 4, 167
Fundamental matrix, 385

G

Galilean trasformation, 419, 420 Gauge transformation, 15 Gaussian curvature, 79, 80, 455 for moving plasma, 119 Geometric optics, 224, 225 Group index of refraction, 132, 245 Group path, 245, 246 Group velocity, 58, 59 for electron whistlers, 173, 174 geometric interpretation, 63 for internal waves, 416 kinematic properties of, 65, 66 for plasma waves, 98, 99 for transverse waves in a plasma, 132 Gyrofrequency, 148, 149 Gyroresonance, 167, 168

Η

Hall conductivity, 154, 155 Hamilton's equations, 66, 239 Hybrid resonances, 177, 178 Hydromagnetic waves, 179–183 Hydrostatic equation, 403

I

Incoherent scatter experiment, 323 Index circle, 197, 198 Instability marginal, 106 two-stream, 103–109, 212

Intact transition, 202 Interaction of solid-state plasma with lattice vibrations, 127 (11) Internal waves, 413 dispersion relation, 413 group velocity, 416 interaction with the ionosphere, 430, 431 ion drag effect, 423-428 polarization relation, 417, 418 thermal conduction and viscosity effect, 429, 430 wind effect, 419–423 Intrinsic frequency, 420 Ion drag, 423 effects on acoustic-gravity waves, 423-428 Ionogram, 284 Ionosonde, 178, 283, 284 Ionosphere D region of, 3 E region of, 3, 436 F region of, 4, 430 gyrofrequency in, 148 IEEE definition, 1 interaction of atmospheric waves with, 402 irregularities in, 213 nature of, 2-5 ray tracing, 243, 244 Ion sound waves, 99 Ion whistler waves, 182, 219 (16) Irregularities, 213 caused by equitorial electrojet, 213 in ionosphere, 213 scattering of electromagnetic waves by, 317 Isothermal atmosphere, 404, 439 (2) impulse response, 432-436, 441 (9)

K

Khinchin theorem, 314 Kinematic approach of waves, 63–66 Klein–Gordon equation, 128 (12), 397 Kramers–Kronig relations, 29–33, 215 (1)

L

Lagrangian, 392 Landau damping, 113 for a Maxwellian plasma, 127 (8)

Lapse rate, 405 Longitudinal dielectric constant, 28 for cold beam plasma, 104 for fluid plasma, 97 for warm beam plasma, 107 Longitudinal waves, 42, 43 approximate condition, 48, 49 in cold magnetoplasma, 155–158 dispersion relation, 40, 43 excitation of, 67 Lorentz condition, 18 Lorentz polarization term, 135, 216 (2) Luxembourg effect, 379

Μ

Magnetic dip. 149 Magnetic moment, 17 of earth's dipole, 148 of plasma, 217 (9) Magnetic polarization density, 17-23 Magnetic pressure, 167 Main signal, 143 Mass operator, 354, 355 Material equations, 13 in anisotropic and dispersive medium, 28 in dispersive medium, 27 in free space, 13, 23 in nondispersive medium, 24 in rotationally symmetric medium, 28 Maxwell's equations, 14 Method of averaging, 384 Mirror vector, 320 Mobility tensor, 152-153 Modified saddle point method, 143, 449 Momentum relaxation time, 372 Multiple scattering, 317, 349-358

N

Nonlinear effects in plasma, 370-374

0

Onsager relation, 33-35 Optical path length, 231-232 Ordinary wave, 48, 176, 181, 189

P

Parallel conductivity, 154, 155 Parameter space, 194 Pederson conductivity, 154, 155 Phase velocity, 64, 65 for plasma waves, 98 for transverse waves in plasma, 132 Plasma, 85, 86 cold, 85 fluid model, 94 warm, 86 waves, 98, 99, 213 Plasma distribution in a gravitational field, 124(3)Plasma field, 373 Plasma frequency, 90 of ath species, 96 of cylindrical plasma column, 124 (2) effect produced by collisional loss, 126 (5) Polarization relations, 44-46 for electronic plasma, 184-188 for internal waves, 417 for magnetoplasma, 162-165 for parallel propagation, 47 for perpendicular propagation, 47, 48 Polarization variation along the ray, 230 Power spectrum, 313 for plasma density fluctuation due to electron plasma waves, 102, 322 for plasma density fluctuation due to ion plasma waves, 102 Poynting theorem, 49 Poynting vector, 37, 55 for an electronic plasma, 192, 193 Precursors, 141, 435 Principal wave, 254, 255, 257 Probability, 309 conditional density, 310 conditional distribution, 310 density, 309 distribution, 309 joint distribution, 312 Propagation parallel to magnetic field, 47, 165-170, 188, 189, 208 Propagation perpendicular to magnetic field, 47, 176-179, 189, 208, 209

462

Q

Quasi-parallel propagation, 191, 192 Quasi-perpendicular propagation, 191, 192 Quasi-resonance condition, 210.

R

Radiation condition, 453-454 Random amplitude, 313 Random medium, 316 Random variable, 310 Ray equation, 228 for anisotropic medium, 239 expressed in spherical coordinates, 240 for isotropic medium, 228, 233, 234 for spherically symmetric medium, 303 (1)Rays, 65, 66 curvature vector, 303 (2) effect of boundary, 240-246 group, 65, 66 for internal waves, 423 for isotropic medium, 227, 228 phase, 65 principal normal, 303 (2) reflection and refraction at boundary, 242 Ray surface, 235 Reciprocal vectors, 235 Reflection coefficients, 249 for magnetoplasma, 291 Reflection level, 244 Reflection point, 260 Refractive index, 37, 39-42 for internal waves, 413-415 for lossy plasma, 134, 135 for plasma waves, 98 for propagation parallel to magnetic field, 47 for propagation perpendicular to magnetic field, 47 for transverse waves in plasma, 131 Refractive index surface, 60-63, 193, 194, 235 Relative dielectric tensor, 28 Hermitian and anti-Hermitian parts, 30 real and imaginary parts, 30

relation to gyration vector, 82 (8) symmetry properties, 32, 35, 82 (8) Relative effective dielectric tensor, 360-364 Reshaping transition, 202 Resonance, 41, 221 (23) cones, 42 for electronic plasma, 191, 210 for internal gravity waves, 414 lower hybrid, 178, 218 (14) in parameter space, 196 upper hybrid, 177, 218 (13), 218 (14) Resonant condition for three-wave interaction, 383, 390, 391 Retarded solution, 16 Rytov's solution, 334

S

Saddle point, 444, 445 Saddle point method, 443-449 modified method, 449, 450 Saturation phenomenon, 343 Scale height, 403 Scattering cross section, 320 for magnetoplasma, 328 for single electron, 323 for thermal plasma, 322, 323 Scintillation index, 348 Screened Coulomb potential, 93 produced by moving test particle, 117 Secant law, 303 (5) Self-interaction factor, 376, 377 Slow process, 23-28 Snell's law, 243, 247, 248 generalized, 244, 245, 268 Sommerfeld solution, 139 Spectral density of plasma density fluctuations, 100 electronic branch, 102 ionic branch, 102 Speed of sound, 406 Sporadic E layers, 436 Stability in beam plasma, 105-109 Steepest ascent path, 445 Steepest descent path, 443 Stochastic process, 311 Stokes constant, 263, 264 Stokes lines, 262

Stokes phenomenon, 263 Stopping power, 123 Stratified media, 246 Subdominant solution, 262 Susceptibility tensor, 28 for cold magnetoplasma, 151 for lossy magnetoplasma, 154 for warm magnetoplasma, 206

Т

Temperature relaxation time, 372 Theory of homogeneous functions, 236 Thermal conduction, 429 attenuation of acoustic-gravity waves due to, 429 Thermal velocity, 91 Torsional mode, 181 Transmission coefficients, 249 Transport of amplitude along the ray, 229 Transverse dielectric constant, 28 for plasma, 131 Transverse waves, 39, 40, 43 dispersion relation, 40 excitation of, 67 Trapping of waves, 391 Turning point, 260, 271 Two-stream instability, 103, 104

application to equatorial electrojet, 213-215 for isotropic plasma, 103-109 for magnetoplasma, 212 Two-timing technique, 394

U

Uniaxial medium, 81 (5), 303 (4)

V

Velocity of heat conduction, 409, 410 Velocity of light, 14 Virtual height, 281, 282 Viscosity, 429 attenuation of acoustic-gravity waves due to, 429

W

Wave normal, 227, 234 Wave packet, 58, 59 Wave parameter, 340, 346 Wave train solution, 392 Wave vector surface, 60 Whistlers, 173–176 quasi-parallel propagation of, 192, 220 (20) Wind-stratified atmosphere, 420 WKB solution, 231, 252–260

464

International Geophysics Series

Editor

J. VAN MIEGHEM

Royal Belgian Meteorological Institute Uccle, Belgium

- Volume 1 BENO GUTENBERG. Physics of the Earth's Interior. 1959
- Volume 2 JOSEPH W. CHAMBERLAIN. Physics of the Aurora and Airglow. 1961
- Volume 3 S. K. RUNCORN (ed.). Continental Drift. 1962
- Volume 4 C. E. JUNGE. Air Chemistry and Radioactivity. 1963
- Volume 5 ROBERT C. FLEAGLE AND JOOST A. BUSINGER. An Introduction to Atmospheric Physics. 1963
- Volume 6 L. DUFOUR AND R. DEFAY. Thermodynamics of Clouds. 1963
- Volume 7 H. U. ROLL. Physics of the Marine Atmosphere. 1965
- Volume 8 RICHARD A. CRAIG. The Upper Atmosphere: Meteorology and Physics. 1965
- Volume 9 WILLIS L. WEBB. Structure of the Stratosphere and Mesosphere. 1966
- Volume 10 MICHELE CAPUTO. The Gravity Field of the Earth from Classical and Modern Methods. 1967
- Volume 11 S. MATSUSHITA AND WALLACE H. CAMPBELL (eds.). Physics of Geomagnetic Phenomena. (In two volumes.) 1967
- Volume 12 K. YA. KONDRATYEV. Radiation in the Atmosphere. 1969
- Volume 13 E. PALMEN AND C. W. NEWTON. Atmospheric Circulation Systems: Their Structure and Physical Interpretation. 1969
- Volume 14 HENRY RISHBETH AND OWEN K. GARRIOTT. Introduction to Ionospheric Physics. 1969
- Volume 15 C. S. RAMAGE. Monsoon Meteorology. 1971
- Volume 16 JAMES R. HOLTON. An Introduction to Dynamic Meteorology. 1972
- Volume 17 K. C. YEH AND C. H. LIU. Theory of Ionospheric Waves. 1972