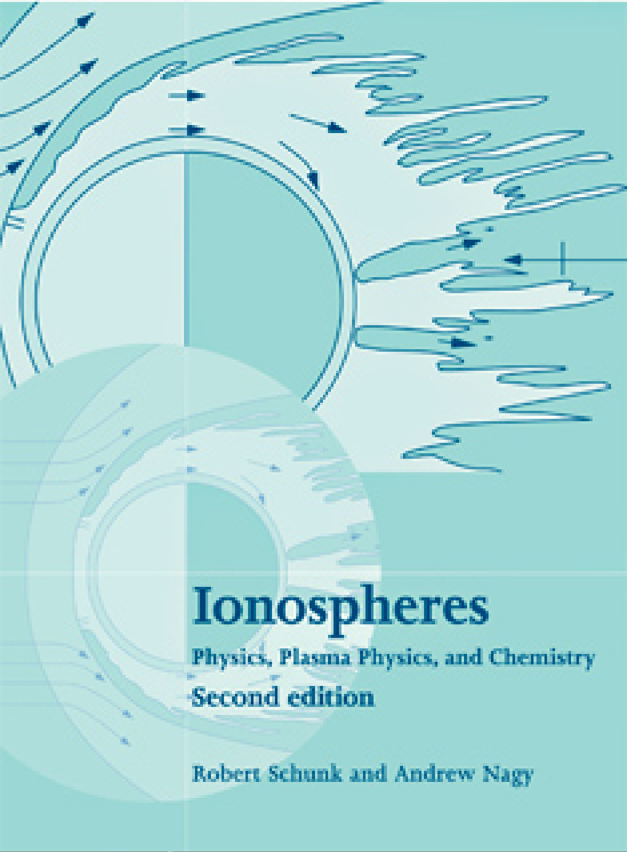
Cambridge Atmospheric and Space Science Series



Ionospheres

Physics, Plasma Physics, and Chemistry

Second edition

Robert Schunk and Andrew Nagy

CAMBRIDGE www.cambridge.org/9780521877060

This page intentionally left blank

Ionospheres

Physics, Plasma Physics, and Chemistry

Second Edition

This combination of text and reference book provides a comprehensive description of the physical, plasma, and chemical processes controlling the behavior of ionospheres, upper atmospheres and exospheres. It describes in detail the relevant processes, mechanisms, and transport equations that are required to solve fundamental research problems. Our current understanding of the structure, chemistry, dynamics, and energetics of the terrestrial ionosphere is summarized in two chapters, while that of other solar system bodies is outlined in a separate chapter. The final chapter of the book is devoted to relevant in-situ and remote measurement techniques.

This second edition incorporates the results, model developments, and interpretations from the last ten years, both in the text and through the addition of new figures and references. In particular, it includes new material on neutral atmospheres, on the terrestrial ionosphere at low, middle, and high latitudes, and on planetary atmospheres and ionospheres, where results from recent space missions have yielded a wealth of new data. Extensive appendices (including an additional four for this edition) provide information about physical constants, mathematical formulas, transport coefficients, and other important parameters needed for ionospheric calculations.

The book forms an extensive and lasting volume for researchers studying ionospheres, upper atmospheres, aeronomy, and plasma physics, and is an ideal textbook for graduate-level courses. Problem sets are provided, for which solutions are available to instructors online at www.cambridge.org/9780521877060.

Cambridge Atmospheric and Space Science Series
Editors: J. T. Houghton, M. J. Rycroft, and A. J. Dessler

This series of upper-level texts and research monographs covers the physics and chemistry of different regions of the Earth's atmosphere, from the troposphere and stratosphere, up through the ionosphere and magnetosphere, and out to the interplanetary medium.

ROBERT W. SCHUNK is Professor of Physics and the Director of the Center for Atmospheric and Space Sciences at Utah State University. He is also a co-founder and the President of Space Environment Corporation, a small high-tech company in Logan, Utah. He has over 35 years of experience in theory, numerical modeling, and data analysis in the general areas of plasma physics, fluid mechanics, kinetics, space physics, and planetary ionospheres and atmospheres. He has been a Principal Investigator on numerous NASA, NSF, Air Force, and Navy grants and has chaired many national committees, international organizations and review panels. Professor Schunk received the D. Wynne Thorne Research Award from USU in 1983, the Governor's Medal for Science & Technology from the State of Utah in 1988, gave the AGU Nicolet Lecture in 2002, is a Fellow of the AGU, and was inducted into the International Academy of Astronautics in 2006.

ANDREW F. NAGY has been on the faculty of the University of Michigan since 1963, serving as a professor of Space Science and Electrical Engineering, Associate Vice President for Research (1987–1990), and Director of the Space Physics Research Laboratory (1990–1992). He has over 40 years of experience in both theoretical and experimental studies of the upper atmospheres, ionospheres, and magnetospheres of the Earth and planets, and has been principal and co-investigator and interdisciplinary scientist on a variety of space missions. Professor Nagy has chaired or been a member of over 40 national and international committees and boards. He was President of the Space Physics and Aeronomy Section of the AGU. He is a Fellow of the AGU, a member of the International Academy of Astronautics, has given the AGU Nicolet Lecture (1998), and received the NASA Public Service Medal (1983) and the Attwood (1998) and the Distinguished Faculty Achievement Awards (2003) from the University of Michigan.

Cambridge Atmospheric and Space Science Series

EDITORS

John T. Houghton

Michael J. Rycroft

Alexander J. Dessler

Titles in print in this series

M. H. Rees
Physics and chemistry of the upper atmosphere

R. Daley
Atmosphere data analysis

J. K. Hargreaves
The solar–terrestrial environment

J. R. Garratt
The atmosphere boundary layer

S. Sazhin
Whistler-mode waves in a hot plasma

S. P. Gary
Theory of space plasma microinstabilities

I. N. James
Introduction to circulating atmospheres

T. I. Gombosi
Gaskinetic theory

M. Walt
Introduction to geomagnetically trapped radiation

B. A. Kagan
Ocean–atmosphere interaction and climate modelling

D. Hastings and H. Garrett
Spacecraft–environment interactions

J. C. King and J. Turner
Antarctic meteorology and climatology

T. E. Cravens
Physics of solar system plasmas

J. F. Lemaire and K. I. Gringauz
The Earth's plasmasphere

T. I. Gombosi
Physics of space environment

J. Green
Atmospheric dynamics

G. E. Thomas and K. Stamnes
Radiative transfer in the atmosphere and ocean

R. W. Schunk and A. F. Nagy
Ionospheres: Physics, plasma physics, and chemistry

I. G. Enting
Inverse problems in atmospheric constituent transport

R. D. Hunsucker and J. K. Hargreaves
The high-latitude ionosphere and its effects on radio propagation

M. C. Serreze and R. G. Barry
The Arctic climate system

N. Meyer-Vernet
Basics of the solar wind

V. Y. Trakhtengerts and M. J. Rycroft
Whistler and Alfvén mode cyclotron masers in space

R. W. Schunk and A. F. Nagy
Ionospheres: Physics, plasma physics, and chemistry, second edition

IONOSPHERES

Physics, Plasma Physics, and Chemistry

Second Edition

ROBERT W. SCHUNK

Utah State University

and

ANDREW F. NAGY

University of Michigan

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521877060

© R. W. Schunk and A. F. Nagy 2000, 2009

This publication is in copyright. Subject to statutory exception and to the provision of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published in print format 2009

ISBN-13 978-0-511-63489-5 eBook (NetLibrary)

ISBN-13 978-0-521-87706-0 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of urls for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

To our parents for their past guidance, encouragement and support, and to our children and AFN's wife (Allison, Lisa, Michael, Robert, and Susan) for their love and understanding.

Contents

Chapter 1

Introduction

1.1 Background and purpose

The ionosphere is considered to be that region of an atmosphere where significant numbers of free thermal ($<$1 eV) electrons and ions are present. All bodies in our solar system that have a surrounding neutral-gas envelope, due either to gravitational attraction (e.g., planets) or some other process such as sublimation (e.g., comets), have an ionosphere. Currently, ionospheres have been observed around all but two of the planets, some moons, and comets. The free electrons and ions are produced via ionization of the neutral particles both by extreme ultraviolet radiation from the Sun and by collisions with energetic particles that penetrate the atmosphere. Once formed, the charged particles are affected by a myriad of processes, including chemical reactions, diffusion, wave disturbances, plasma instabilities, and transport due to electric and magnetic fields. Hence, an understanding of ionospheric phenomena requires a knowledge of several disciplines, including plasma physics, chemical kinetics, atomic theory, and fluid mechanics. In this book, we have attempted to bridge the gaps among these disciplines and provide a comprehensive description of the physical and chemical processes that affect the behavior of ionospheres.

A brief history of ionospheric research is given later in this introductory chapter. An overview of the space environment, including the Sun, planets, moons, and comets, is presented in Chapter 2. This not only gives the reader a quick look at the overall picture, but also provides the motivation for the presentation of the material that follows. Next, in Chapter 3, the general transport equations for mass, momentum, and energy conservation are derived from first principles so that the reader can clearly see where these equations come from. This is followed by a derivation of the collision terms that appear in the transport equations, including those relevant to resonant charge exchange, nonresonant ion-neutral and electron-neutral interactions, and Coulomb collisions (Chapter 4). These general collision terms and transport equations are complicated and in many situations it is possible to use simpler sets of transport equations. Therefore, in Chapter 5, several simplified systems of transport equations are derived, including the Euler, Navier–Stokes, diffusion, and thermal conduction equations. This is followed by a discussion of the wave modes, plasma instabilities, and shocks that can occur in the ionospheres (Chapter 6). In Chapter 7, the magnetohydrodynamic (MHD) equations are derived and then used to describe MHD waves, shocks, and pressure balance.

In Chapter 8, chemical kinetics and a variety of reactions relevant to the ionospheres are discussed and presented, including those involving metastable species and negative ions. Optical emissions are also briefly discussed in this chapter. The relevant ionization and energy exchange processes are detailed in Chapter 9, including those pertaining to both photons and particles. The chapter concludes with a summary of the heating and cooling expressions that are needed for practical applications. Chapter 10 is devoted to a discussion of neutral atmospheres. The Euler and Navier–Stokes equations for neutral gases are presented at the beginning of the chapter, and this is followed by a discussion of atmospheric waves and tides. The rest of Chapter 10 deals with atmospheric structure, escape fluxes, the exosphere, and hot atoms. In Chapters 11 and 12, the general material given in the previous chapters is applied to elucidate the unique characteristics associated with the terrestrial ionosphere at low, middle, and high latitudes. Although much of this material is still of a fundamental nature, an overview of what has been accomplished to date is also provided. Chapter 13 summarizes what is currently known about all of the other ionospheres in the solar system. The most commonly used experimental techniques for measuring ionospheric densities, temperatures, and drifts are briefly described in Chapter 14. Finally, several Appendices are included that contain physical constants, mathematical formulas, some important derivations, and useful tables.

This book is the outgrowth of two decades of numerous joint research endeavors and publications by the authors. Some of the material was used in courses taught by the authors at Utah State University and at the University of Michigan. This book should be useful to graduate students, postdoctoral fellows, and established scientists who want to fill gaps in their knowledge. It also serves as a reference book for obtaining important equations and formulas. A subset of the material can be used for a graduate level course about the upper atmosphere and ionosphere, and plasma physics. At the University of Michigan a one-semester graduate course on the ionosphere and upper atmosphere has been based on Chapters 2, 3, parts of 5, 8, 9, 10, most of 11 and 12, and 13 and 14. At Utah State University, a one-semester course on plasma physics has been based on Chapters 3–7, and a course on aeronomy has been based on Chapters 2, 3, 5, 8–12. To facilitate the use of this book as a text, problems are provided at the end of most of the chapters.

Several people were helpful in the preparation of this book, and we wish to acknowledge them here. The help came in a variety of forms (e.g., providing some unpublished material, reading, or proofing part of the manuscript, etc.), and it certainly improved the book. AFN would especially like to thank (in alphabetical order)

J.R. Barker, T.E. Cravens, J.L. Fox, B.E. Gilchrist, T.I. Gombosi, J.W. Holt, A.J. Kliore, M.W. Liemohn, H. Rishbeth, and C.T. Russell. RWS would like to thank Melanie Oldroyd for typing a preliminary form of some of the chapters. We would both like to thank Shawna Johnson for drawing some of the figures, for digitizing figures, and for overseeing the production of the book. We would both also like to thank Elizabeth Wood for preparing the manuscript in LaTeX. Some of the material in the book comes from lecture notes collected over many years and thus may contain material without appropriate references to their sources, which we have forgotten. This is inadvertent and we apologize to such authors. Also, to keep the bibliographies from becoming unrealistically long, we limited our referencing to only those papers from which figures were taken, to either the latest or original reference for the material discussed, and to review papers. Hence, we omitted many deserving, appropriate, and relevant references. We hope that the readers and scientists working in the field will understand and appreciate our dilemma.

The units used in the book are a mixture of MKSA and Gaussian-cgs because of the corresponding usage by practitioners in the field. Most of the equations and formulas throughout the book are in MKSA units, and some tables and numbers are given in Gaussian-cgs units when this is the common practice. The conversion from one system to the other is briefly discussed in Appendix E.

1.2 History of ionospheric research

The earliest exposure of humankind to a phenomenon originating in the upper atmosphere is the visual aurora. The visual displays of colored light appear in the form of arcs, bands, patches, blankets, and rays, and often the features move rapidly across the night sky. It has been suggested that the earliest records of the aurora can be traced to the Stone Age.[1] References to the aurora appear in the Old Testament, in writings of Greek philosophers, including Aristotle's *Meteorologica*, and possibly in ancient Chinese works from before 2000 BC. In most of these early writings, the auroral displays were interpreted to be manifestations of God.The name *aurora borealis* (northern dawn) appears to have been coined by Galileo at some time prior to 1621.[1] The first recorded observation of the southern hemispheric aurora (*aurora australis*) was by Cook in 1773.

A serious scientific study of auroras began at about 1500 AD.[1] However, the early theories put forth by noted scientists were completely wrong. Edmund Halley, who predicted the reappearance of what is now known as Halley's comet, suggested that the auroras were "watery vapors, which are rarefied and sublimed by subterraneous fire, [and] might carry along with them sulphureous vapors sufficient to produce this luminous appearance in the atmosphere." In 1746, the Swiss mathematician Leonard Euler suggested that "the aurora was particles from the Earth's own atmosphere driven beyond its limits by the impulse of the sun's light and ascending to a height of several thousand miles. Near the poles, these particles would not be dispersed by the Earth's rotation."[2] Benjamin Franklin, who was a respected scientist in his time, thought that the aurora was related to atmospheric circulation patterns.[3] Basically, Franklin argued that the atmosphere in the polar regions must be heavier and lower than in the equatorial region because of the smaller centrifugal force, and therefore, the vacuum–atmosphere interface must be lower in the polar regions. He then further argued that the electricity brought into the polar region by clouds would not be able to penetrate the ice, and hence, would break through the low atmosphere and run along the vacuum toward the equator. The electricity would be most visible at high latitudes, where it is dense, and much less visible at lower latitudes, where it diverges. Franklin claimed such an effect would "give all the appearances of an Aurora Borealis."[1,3]

Numerous other theories of the aurora have been proposed over the last 150 years, including reflected sunlight from ice particles, reflected sunlight from clouds, sulfurous vapors, combustion of inflammable air, luminous magnetic particles, meteoric dust ignited by friction with the atmosphere, cosmic dust, currents generated by compressed cosmic ether, thunderstorms, electric discharges between the Earth's magnetic poles, and electric discharges between fine ice needles. A comprehensive and fascinating account of the aurora in science, history, and the arts is given in Reference 1, and additional theories are presented there.

Although early auroral theories did not fare very well, observations made during the latter half of the 1700s and throughout the 1800s elucidated many important auroral characteristics. In 1790, the English scientist Cavendish used triangulation and estimated the height of auroras at between 52 and 71 miles.[4] In 1852, the relationship among geomagnetic disturbances, auroral displays, and sunspots was clearly established; the frequency and amplitude of these features varied with the same 11-year periodicity.[5,6] In 1860, Elias Loomis drew the first diagram of the region where auroras are most frequently observed and noted that the narrow ring is not centered on the geographic pole, but that its oval form resembles lines of equal magnetic dip, thereby establishing the relationship between the aurora and the geomagnetic field. In 1867, the Swedish physicist Angström made the first measurements of the auroral spectrum.[7] However, a significant breakthrough in auroral physics was not achieved until the end of the nineteenth century, when cathode rays were discovered and identified as electrons by the British physicist J. J. Thomson. Subsequently, the Norwegian physicist Kristian Birkeland proposed that the aurora was caused by a beam of electrons emitted by the Sun. Those electrons reaching the Earth would be affected by the Earth's magnetic field and guided to the high-latitude regions to create the aurora.

Until the discovery of sunspots by Galileo in 1610, the Sun was generally thought to be a quiet, featureless object. Galileo not only discovered the dark spots but also noted their westward movement, which was the first indication that the Sun rotates. In subsequent observations, it was quickly established that the number of sunspots varies with time. It was not until more than two centuries later, however, that an amateur astronomer in Germany, Heinrich Schwabe, noted an apparent 10-year periodicity in his 17 years of sunspot observations.[8] Shortly after Schwabe's discovery, professional astronomers set out to determine whether or not the cycle was real. The leader of this effort was Rudolf Wolf of the Zürich observatory. Wolf conducted an extensive search of past data and was able to establish that the number of sunspots varied with an 11-year cycle that had been present since at least 1700.[9] In 1890, Maunder called attention to the 70-year period from 1645 to 1715, when almost no sunspots were observed.[10] This period, which is known as the Maunder Minimum Period, raises the question whether the sunspot cycle is a universal feature or just a recent phenomenon.

As defined at the beginning of this chapter, the terrestrial ionosphere begins at an altitude of about 60 km and extends beyond 3000 km, with the peak electron concentration occurring at approximately 300 km. The first suggestion of the existence of what is now called the ionosphere can be traced to the 1800s. Carl Gauss and Balfour Stewart hypothesized the existence of electric currents in the atmosphere to explain the observed variations of the magnetic field at the surface of the Earth. Gauss argued:[11]

> It may indeed be doubted whether the seat of the proximate causes of the regular and irregular changes which are hourly taking place in this [terrestrial magnetic] force, may not be regarded as external in reference to the Earth . . . But the atmosphere is no conductor of such [galvanic] currents, neither is vacant space. But our ignorance gives us no right absolutely to deny the possibility of such currents; we are forbidden to do so by the enigmatic phenomena of the Aurora Borealis, in which there is every appearance that electricity in motion performs a principal part.

It had been well established that there was a direct correlation between the solar cycle and magnetic disturbances on the Earth. To account for this strong correlation, Stewart speculated that electrical currents must flow in the Earth's upper atmosphere, and that the Sun's action is responsible for turning air into a conducting medium.[12] It was also concluded that the conductivity of the upper atmosphere is higher at sunspot maximum than at sunspot minimum. This view, however, was not widely accepted and strong counterarguments were presented in 1892 by Lord Kelvin.

The existence of the ionosphere was clearly established in 1901 when G. Marconi successfully transmitted radio signals across the Atlantic. This experiment indicated that radio waves were deflected around the Earth's surface to a much greater extent than could be attributed to diffraction. The following year, A. E. Kennelly and O. Heaviside suggested that free electrical charges in the upper atmosphere could reflect radio waves.[13] That same year, the first *physical* theory of the ionosphere was proposed.[14]

> The observed effect, which if confirmed is very interesting, seems to me to be due to the conductivity . . . of air, under the influence of ultra-violet solar radiation. No doubt electrons must be given off from matter . . . in the solar beams; and the presence of these will convert the atmosphere into a feeble conductor.

In 1903, J. E. Taylor independently suggested that solar ultraviolet radiation was the source of electrical charges, which implied solar control of radio propagation.[15] The first rough measurements of the height of the reflecting layer were made by Lee de Forest and L. F. Fuller at the Federal Telegraph Company in San Francisco from

1912 to 1914. The reflecting layer's height was deduced using a transmitter–receiver spacing of approximately 500 km, which was determined by the circuits of the Federal Telegraph Company.[16] However, the de Forest–Fuller results were not well known, and generally accepted measurements of the height of the reflecting layer were made in 1924 by Breit and Tuve[17] and by Appleton and Barnett.[18] The Breit–Tuve experiments involved a "pulse sounding" technique, which is still in use today, while Appleton and Barnett used "frequency change" experiments, which demonstrated the existence of downcoming waves by an interference technique. These experiments led to a considerable amount of theoretical work, and in 1926 the name "ionosphere" was proposed by R. A. Watson-Watt in a letter to the United Kingdom Radio Research Board, but it did not appear in the literature until three years later.[19] Radio soundings of the ionosphere initially seemed to indicate that the ionosphere consisted of distinct layers; we now know that this is generally not the case and we refer to different regions. These regions are called the *D*, *E*, and *F* regions. The names of these regions originated with Appleton, who stated that in his early work he wrote *E* for the reflected electric field from the first layer that he recognized. Later, when he recognized a second layer at higher altitudes, he wrote *F* for the reflected field. Subsequently, he conjectured that there may be another layer at lower altitudes so he decided to name the first two layers *E* and *F* and the possible lower one *D*, thus allowing the alphabetical designation of other undiscovered layers.[20]

The rocket technology available at the end of World War II was used by scientists to study the upper atmosphere and ionosphere, paving the way for space exploration via satellites. The first rocket-borne scientific payload, which carried instrumentation to make measurements directly in the upper atmosphere and ionosphere, was launched in 1946 on a *V-2* from White Sands, New Mexico. The University of Michigan payload consisted of a Langmuir probe and a thermionic pressure gage; although the *V-2* failed during this flight it marked the beginning of direct exploration of the ionosphere. The first book devoted to the ionosphere was published in 1952 by Rawer.[21]

The rocket technology, coupled with a major advance in ground-based instrumentation, led scientists to realize that a dramatic increase in our knowledge of the terrestrial environment was possible. To take advantage of these new capabilities, the International Geophysical Year (IGY), 1957–1958, was organized.[22,23] This cooperative effort was to begin with the next maximum of the solar cycle. As part of the IGY, scientists proposed to launch artificial satellites, and eventually *Sputnik 1* was launched on October 4, 1957.

Many consider the launch of *Sputnik 1* the beginning of the Space Age, but to some degree it started much earlier. Rockets have been with us ever since the ancient Chinese used them for fireworks. Later variations of "rockets" were used, basically for military purposes, to send payloads from one location to impact at another. Newton developed the scientific basis to describe how an object could be placed in orbit around the Earth, and visionaries like Jules Verne and H. G. Wells dreamt such thoughts.

The modern era of rocket propulsion began in Russia in the 1880s, where Konstantin Tsiolkovsky worked out the fundamental laws of rocket propulsion and