

Space Physics

Springer-Verlag Berlin Heidelberg GmbH

May-Britt Kallenrode

Space Physics

An Introduction
to Plasmas and Particles
in the Heliosphere
and Magnetospheres

With 170 Figures, 9 Tables,
Numerous Exercises and Problems



Springer

Professor Dr. May-Britt Kallenrode

Universität Lüneburg
FB IV – Umweltwissenschaften
Scharnhorststrasse 1
D-21332 Lüneburg, Germany
E-Mail: mbk@uni-lueneburg.de

ISBN 978-3-662-03655-6

Library of Congress Cataloging-in-Publication Data

Kallenrode, May-Britt, 1962–

Space physics: an introduction to plasmas and particles in the heliosphere and magnetospheres /
May-Britt Kallenrode.

p. cm.

Includes bibliographical references and index.

ISBN 978-3-662-03655-6 ISBN 978-3-662-03653-2 (eBook)

DOI 10.1007/978-3-662-03653-2

1. Plasma (Ionized gases) 2. Space plasmas. 3. Heliosphere. 4. Magnetosphere. 5. Space physics. I. Title.
QC718.K28 1998 523-01-DC21 98-16075 CIP

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag Berlin Heidelberg GmbH.

Violations are liable for prosecution under the German Copyright Law.

© Springer-Verlag Berlin Heidelberg 1998

Originally published by Springer-Verlag Berlin Heidelberg New York in 1998

Softcover reprint of the hardcover 1st edition 1998

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting: Camera ready by the author using a Springer TeX macro package

Cover design: *design & production* GmbH, Heidelberg

SPIN 10634356

57/3144 – 5 4 3 2 1 0 – Printed on acid-free paper

Preface

A beginner, student as well as young scientist, in the field of space physics often has to struggle with two problems: (a) space physics basically is plasma physics, but the relevant topics can be quite different from those covered in traditional lectures or textbooks, and (b) space physics is different from laboratory physics in so far as problems are complex and observations of phenomena often are incomplete: there is no chance to run the ‘experiment’ aurora or flare with a different set of well-defined parameters or even to know all relevant parameters of the one just observed.

This book will try to lend the beginner a hand in this struggle and help to bridge the gap between the formal plasma physics background and the manifold phenomena in space. Good books on both topics exist: a very accessible, and in its style and methodic conscientiousness unattainable book on plasma physics is *Plasma Physics and Controlled Fusion* by F.F. Chen [2.2]; a well-written and up-to-date account of the phenomena in space plasmas is given in *Introduction to Space Physics* edited by M.G. Kivelson and C.T. Russell [3.13]. Both books cannot be matched by this one and can serve as valuable supplements. More formal introductions to plasma physics as required by a space scientist are given in *Physics of Space Plasmas* by G.K. Parks [3.16], or *Basic Space Plasma Physics* by W. Baumjohann and R.A. Treumann [3.2] and its sequel *Advanced Space Plasma Physics* [3.23].

The attempt chosen in this book is different because it contains both a more formal background and an introduction to the phenomena: the first chapters introduce the basic principles of plasma physics required in space science. Where possible, simple applications are added. The later chapters are concerned with the phenomena and an introduction to space physics and some of its currently most interesting questions. This division is disadvantageous in so far as the phenomena are not used to illustrate the concepts. But the phenomena often are complex and invoke more than one of the concepts or can be understood properly only in relation to other phenomena. Continuous cross referencing to earlier chapters points to the fundamental concepts and draws our attention to conceptual difficulties and fundamental open questions. In particular, we will see that ‘real space’ essentially is non-linear, while our concepts most often are based on a well-behaved plasma described by linear or at least linearized equations. In addition to plasmas,

we will discuss energetic particles in space and inside the magnetospheres and their interaction with the plasma. This topic often is neglected, either because the books deal with plasmas only or because of its formal difficulties. Nonetheless, this book at least allows a glimpse of these phenomena.

As the interests of the readers might be different, a small  in the margin will provide orientation. ‘Whatnow?’ always is accompanied by a label and a reference to another section. The labels are as follows:

 Whatnow? →goto	Exam _○	This section/paragraph contains an example from space plasmas to illustrate the concept discussed earlier. Such a section might be skipped by the reader who is interested primarily in the physical concepts and less in space science.
	Formal	This section is more formal, but is not vital for an understanding of basic observations. It might be skipped by the reader who is mainly interested in an introduction to space physics.
	Space	The concept introduced in this section is important in space physics but usually is not discussed in standard plasma physics textbooks.
	Suppl	This section is supplementary: although the ideas presented here are important in space physics, the theoretical background is complicated and only briefly sketched. In particular, the beginner in space physics should feel free to skip these sections on first reading and return to them later after becoming more acquainted with the topic.
	Hot	This text points to hotly debated topics and fundamental open problems.

This book is based on lecture notes on cosmic electrodynamics and magnetospheric physics, both courses introduced at the University of Kiel by Prof. Gerd Wibberenz and later continued – and, of course, modified – by myself. I am indebted to him for introducing me to the field of space science and overseeing part of my scientific education. Many other people, too many to name, have contributed directly or indirectly to this book: students participating in the lectures and offering suggestions and criticism, colleagues of the Extraterrestrial Physics Group at the University of Kiel serving as guinea-pigs to test my own understanding or a certain way to introduce a topic, and colleagues from throughout the space science community who discussed the seemingly silliest questions with me and repeatedly challenged my scientific understanding and beliefs. But in writing a book, professional support is one aspect; encouragement and social support are of equal importance. Here the important people can be named more easily: Bernd Heber, who shared my enjoyment in understanding physics and the ways of the world; Günter Virkus, who shared my love of books and gave the necessary support to get this one started; and, last but not least, Ulrich Fischer, who always smiling and shaking his head helped to stay through the final process of writing it. Thank you, all of you. And thank you also to the helpful team at Springer, in particular Claus Ascheron, Andrea Gross, and Gertrud Dimler.

Contents

1. Introduction	1
1.1 Neutral Gases and Plasmas	1
1.2 Plasmas in Space	2
1.3 A Brief History of Space Research	4
Exercises and Problems	8
2. Charged Particles in Electromagnetic Fields	9
2.1 Electromagnetic Fields	9
2.1.1 Maxwell's Equations	10
2.1.2 Transformation of Field Equations	11
2.1.3 Generalized Ohm's Law	12
2.1.4 Energy Equation of the Electromagnetic Field	12
2.2 Particle Motion in Electromagnetic Fields	13
2.2.1 Lorentz Force and Gyration	15
2.3 Drifts of Particles in Electromagnetic Fields	17
2.3.1 The Concept of the Guiding Center	17
2.3.2 Crossed Magnetic and Electric Fields: $\mathbf{E} \times \mathbf{B}$ Drift	17
2.3.3 Magnetic and Gravitational Fields	19
2.3.4 Inhomogeneous Magnetic Fields	19
2.3.5 Curvature Drift	20
2.3.6 Drifts Combined with Changes in Particle Energy	21
2.3.7 Drift Currents in Plasmas	21
2.4 Adiabatic Invariants	22
2.4.1 First Adiabatic Invariant: The Magnetic Moment	23
2.4.2 Magnetic Mirrors and Bottles	25
2.4.3 Second Adiabatic Invariant: Longitudinal Invariant	27
2.4.4 Third Adiabatic Invariant: Flux Invariant	29
2.5 Summary	29
Exercises and Problems	30
3. Magnetohydrodynamics	31
3.1 Basic MHD Equations	32
3.1.1 Momentum Balance	32
3.1.2 Equation of Continuity	37

VIII Contents

3.1.3	Equation of State	38
3.1.4	Basic Equations of MHD	39
3.1.5	Two-Fluid Description	41
3.2	Magnetohydrostatics	44
3.2.1	Magnetic Pressure	45
3.2.2	Magnetic Tension	46
3.3	Magnetohydrokinematics	49
3.3.1	Frozen-in Magnetic Fields	49
3.3.2	Dissipation of Fields	51
3.3.3	Reconnection	53
3.3.4	Deformation of the Field in a Plasma Flow	55
3.4	The Magnetohydrodynamic Dynamo	58
3.5	Debye Shielding	62
3.6	Summary	65
	Exercises and Problems	66
4.	Plasma Waves	67
4.1	What is a Wave?	68
4.2	Magnetohydrodynamic Waves	69
4.2.1	Linearization of the Equations: Perturbation Theory .	69
4.2.2	Alfvén Waves	71
4.2.3	Magneto-Sonic Waves	73
4.3	Electrostatic Waves in Non-Magnetic Plasmas	75
4.3.1	Plasma Oscillations	76
4.3.2	Electron Plasma Waves (Langmuir Waves)	77
4.3.3	Ion-Acoustic Waves (Ion Waves)	78
4.4	Electrostatic Waves in Magnetized Plasmas	79
4.4.1	Electron Oscillations Perpendicular to \mathbf{B} (Upper Hybrid Frequency)	79
4.4.2	Electrostatic Ion Waves Perpendicular to \mathbf{B} (Ion Cyclotron Waves)	80
4.4.3	Lower Hybrid Frequency	81
4.5	Electromagnetic Waves in Non-Magnetized Plasmas	81
4.6	Electromagnetic Waves in Magnetized Plasmas	83
4.6.1	Electromagnetic Waves Perpendicular to \mathbf{B}_0	83
4.6.2	Waves Parallel to the Magnetic Field: Whistler (R-Waves) and L-Waves	85
4.7	Summary	87
	Exercises and Problems	88
5.	Kinetic Theory	89
5.1	The Distribution Function	89
5.1.1	Phase Space and Distribution Function	89
5.1.2	Maxwell's Velocity Distribution	90
5.1.3	Other Distributions	92

5.1.4	Distribution Function and Measured Quantities	93
5.2	Basic Equations of Kinetic Theory	94
5.2.1	The Boltzmann Equation	94
5.2.2	The Vlasov Equation	95
5.2.3	The Fokker–Planck Equation	96
5.3	Collisions	98
5.3.1	Collisions Between Neutrals	99
5.3.2	Collisions Between Charged Particles	100
5.4	Summary	102
	Exercises and Problems	102
6.	Diffusive Transport	103
6.1	Diffusion	103
6.1.1	Spatial Diffusion	104
6.1.2	Pitch-Angle Diffusion	110
6.1.3	Diffusion in Momentum Space	111
6.2	Wave–Particle Interactions	112
6.2.1	Quasi-Linear Theory	112
6.2.2	Resonance Scattering	113
6.2.3	Alfvén Waves and Interplanetary Propagation	114
6.2.4	Electromagnetic Waves	115
6.2.5	Diffusion in Momentum Space	116
6.3	Summary	117
	Exercises and Problems	118
7.	Shock Waves	119
7.1	What is a Shock Wave?	119
7.1.1	Information, Dissipation, and Non-Linearity	120
7.1.2	The Shock’s Rest Frame	122
7.1.3	Collisionless Shock Waves	122
7.2	Shock Conservation Laws	123
7.2.1	Rankine–Hugoniot Equations in Ordinary Shocks	123
7.2.2	Rankine–Hugoniot Equations in MHD Shocks	125
7.2.3	Jump Conditions and Discontinuities	126
7.3	Shock Parameters	127
7.3.1	Shock Geometry	127
7.3.2	Fast and Slow Shocks	127
7.3.3	The Coplanarity Theorem	130
7.3.4	The Shock Normal Direction	130
7.4	Particle Acceleration at Shocks	131
7.4.1	Shock Drift Acceleration (SDA)	131
7.4.2	Diffusive Shock Acceleration	134
7.4.3	Diffusive Shock Acceleration and Self-Generated Turbulence	137
7.4.4	Stochastic Acceleration	139

7.5	The Shock as a Non-Linear System	139
7.6	Summary	141
	Exercises and Problems	142
8.	Sun and Solar Wind:	
	Plasmas in the Heliosphere	143
8.1	The Sun	143
8.1.1	Nuclear Fusion	145
8.1.2	Structure of the Sun	146
8.1.3	The Solar Atmosphere	147
8.1.4	The Coronal Magnetic Field	148
8.2	The Solar Wind	150
8.2.1	Properties	150
8.2.2	Solar Wind Models	151
8.2.3	The Problem: Coronal Heating and Solar Wind Acceleration	155
8.3	The Interplanetary Magnetic Field	156
8.3.1	Spiral Structure	156
8.3.2	Sector Structure	158
8.3.3	The Ballerina Model	159
8.3.4	Corotating Interaction Regions	160
8.4	Plasma Waves in Interplanetary Space	162
8.4.1	Power-Density Spectrum	162
8.4.2	Waves or Turbulence?	163
8.5	The Three-Dimensional Heliosphere	165
8.6	The Active Sun	167
8.6.1	The Solar Cycle	167
8.6.2	A Simple Model of the Solar Cycle	168
8.6.3	The Heliosphere During the Solar Cycle	170
8.7	Flares and Coronal Mass Ejections	171
8.7.1	Electromagnetic Radiation	171
8.7.2	Classes of Flares	175
8.7.3	Coronal Mass Ejections	176
8.7.4	Coronal Mass Ejections, Flares, and Coronal Shocks .	178
8.7.5	Models of Coronal Mass Ejections (CMEs)	179
8.7.6	Models of Flares	182
8.7.7	Magnetic Clouds: CMEs in Interplanetary Space	183
8.7.8	Interplanetary Shocks	184
8.8	Summary	186
	Exercises and Problems	186

9.	Energetic Particles in the Heliosphere	187
9.1	Particle Populations in the Heliosphere	187
9.2	Solar Energetic Particles	190
9.2.1	Classes of Particle Events	191
9.2.2	Interplanetary Propagation	194
9.3	Particles and Traveling Interplanetary Shocks	197
9.3.1	Low Energies (Tens of Kiloelectronvolts)	198
9.3.2	High Energies (Megaelectronvolts)	200
9.3.3	Forbush Decreases	203
9.4	Particles at Planetary Bow Shocks	204
9.5	Galactic Cosmic Rays	205
9.5.1	Modulation	206
9.5.2	Over the Poles	208
9.6	Summary	210
	Exercises and Problems	210
10.	The Terrestrial Magnetosphere	211
10.1	The Terrestrial Magnetic Field	211
10.1.1	The Dipole Field	212
10.1.2	Variability	216
10.1.3	The Terrestrial Dynamo	218
10.2	Topology of the Magnetosphere	220
10.2.1	The Magnetopause	221
10.2.2	Polar Cusps	223
10.2.3	The Tail	223
10.2.4	Magnetosheath and Bow Shock	225
10.2.5	Open or Closed?	227
10.2.6	Magnetospheric Modeling	228
10.3	Plasmas and Currents in the Magnetosphere	229
10.3.1	The Ionosphere	229
10.3.2	The Plasmasphere	231
10.3.3	The Geosphere	233
10.3.4	The Outer Magnetosphere	234
10.3.5	Magnetosphere–Ionosphere Coupling	238
10.4	Particles in the Magnetosphere	241
10.4.1	The Radiation Belts	241
10.4.2	Galactic Cosmic Rays – Størmer Orbits	250
10.4.3	Solar Energetic Particles – Polar Cap Absorption	252
10.5	Summary	253
	Exercises and Problems	254

11. The Dynamic Magnetosphere	255
11.1 Geomagnetic Variability	255
11.1.1 Daily Variations	255
11.1.2 Geomagnetic Disturbances	256
11.1.3 Geomagnetic Indices	257
11.1.4 Geomagnetic Activity Through the Solar Cycle	258
11.1.5 Annual Variations and Recurrent Geomagnetic Activity	259
11.1.6 Magnetic Storms in Response to Transient Phenomena	262
11.1.7 A Simple Model of Geomagnetic Activity	262
11.2 Substorms	263
11.3 Aurorae	264
11.3.1 Historical Excursion	265
11.3.2 Beginning of the Scientific Analysis	267
11.3.3 Modern Interpretation	270
11.4 Summary	274
12. Planetary Magnetospheres	275
12.1 Overview	275
12.1.1 The Planets	275
12.1.2 Structures of Planetary Magnetospheres	276
12.1.3 Sizes	278
12.1.4 Plasma Sources	279
12.1.5 Upstream of the Bow Shock: The Foreshocks	280
12.1.6 Radiation Belts	281
12.2 Planets with a Magnetic Field	283
12.2.1 Mercury	283
12.2.2 Jupiter	285
12.2.3 Saturn	286
12.2.4 Uranus	288
12.2.5 Neptune	289
12.3 Planets Without a Magnetic Field	291
12.4 Summary	292
Exercises and Problems	292
13. Solar–Terrestrial Relationships	293
13.1 Solar–Terrestrial Relationships: Overview	293
13.2 The Atmosphere	295
13.3 Responses of the Upper Atmosphere to Solar Variability	297
13.3.1 Polar Cap Absorptions and Ozone	298
13.3.2 Thermospheric Circulation	300
13.4 The Technical Environment and Solar Activity	303

13.5 The Solar Cycle, Sector Boundaries, Droughts, and Thunderstorms	305
13.5.1 Solar Activity, Climate, and Culture	305
13.5.2 Sun and Weather	307
14. Instrumentation	313
14.1 Field Instruments	313
14.1.1 The Magnetic Field	314
14.1.2 Electric Field Measurements	316
14.1.3 Wave Measurements	317
14.2 Plasma Instruments	318
14.2.1 Instruments for Dense Plasmas	318
14.2.2 Instruments for Rarefied Plasmas	320
14.2.3 Energetic Particle Instruments	321
14.3 Supplementary Ground-Based Observations	323
Appendix	325
A.1 List of Symbols	325
A.2 Useful Relations	327
A.2.1 Vector Identities	327
A.2.2 Vector Calculus	328
A.2.3 Cylindrical Coordinates	329
A.2.4 Spherical Coordinates	329
A.3 Useful Numbers	330
A.3.1 Fundamental Constants	330
A.3.2 Numbers in Plasmas	330
References	331
Index	351

1. Introduction

We shall not cease from exploration.

And the end of all our exploring
will be to arrive where we started
and know the place for the first time.

T. S. Eliot, *Little Gidding*

1.1 Neutral Gases and Plasmas

Our daily experience with matter is based on our dealings with solid bodies, liquids and neutral gases. These interactions have shaped our scientific world view as well as our intuitive understanding of states of matter and their behavior. But moving upwards from the surface of the Earth, our environment changes dramatically: starting at a height of about 80 km, the atmosphere contains an ionized particle component, the ionosphere. With increasing height, the relative importance of the neutral component decreases. Farther out in the magnetosphere and in interplanetary space only ionized gases, called plasmas, exist: the hard electromagnetic radiation from the Sun immediately ionizes all matter.

A plasma differs from a neutral gas in so far as it (also) contains charged particles. The number of charged particles has to be large enough to allow for electromagnetic interactions. In addition, the number of positive and negative charges has to be nearly equal. This is called quasi-neutrality, because viewed from the outside the plasma appears to be electrically neutral.

Owing to the existence of free charges, a plasma is a conductor. But moving electric charges are also currents which induce magnetic fields. These fields, in turn, influence the motion of particles. A plasma, therefore, is not only controlled by external electric and magnetic fields, but also creates fields which modify the external fields and the motion of the charged particles: a plasma can interact with itself. From this special ability, differences compared with neutral gases arise. They become most obvious in the large number of different types of waves in plasmas (cf. Chap. 4). In apparently simple situations, a plasma can behave entirely differently from what our intuition would expect. Pouring milk into our coffee, we expect the milk to heat up and mix with the coffee. A sunspot is a sharply bordered volume of cool gas embedded

in the hot solar photosphere and stays stable for several months prevented by strong magnetic fields from warming or mixing with its environment. A cold and dense volume of gas or liquid in a hot environment sinks. A solar filament is cold and dense compared with the surrounding corona but it is held in position against gravity by strong magnetic fields.

The discrepancy between our daily experience and the behavior of ionized gases clearly shows that plasmas do not belong to our direct environment. Why then do we study such exotic phenomena? Are there applications for plasmas?

First, plasmas are not exotic, they are natural and common. The interplanetary and interstellar medium and the stars are made of ionized gases. Thus about 99% of matter in the universe is plasma. Spatially closest areas dominated by plasmas are the magnetosphere with its radiation belts, the ionosphere, lightning bolts in the atmosphere, and the Earth's core. Thus even in the system Earth a significant part of matter is in the plasma state. Plasma physics, therefore, contributes to the understanding of our environment. In turn, the natural plasma laboratories, i.e. the ionosphere, the magnetosphere, and interplanetary space, help to test the concepts of plasma physics on spatial scales and with densities which cannot be realized in a laboratory.

Second, plasmas can be used for quite worldly applications. One of the most ambitious projects is nuclear fusion: to form helium by merging hydrogen atoms, imitating the processes inside the Sun and the stars, in order to create a clean and long-lasting power source. The main aspects of this project are the production of a plasma with suitable properties (density, temperature, losses) and its confinement inside a magnetic field.

There are also less spectacular applications of plasma physics. Chemistry takes advantage of the fact that reactions in plasmas are different from reactions in neutral gases. For instance, cyan gas can be synthesized by burning coal dust in a nitrogen electric arc plasma. Plasma beams are used for ion implantation in microchip production. Plasma burners and pistols are used to cut, weld, or clean metals. Other technical applications of plasmas are as diverse as lasers, capacitors, oscillators, and particle accelerators.

1.2 Plasmas in Space

The focus of this book is natural plasmas in space. Depending on their location, these plasmas often exhibit different properties as characterized by the plasma parameters (cf. Chap. 3, in particular Fig. 3.13). Stellar interiors consist of hot and extremely dense plasmas with the highest densities inside white dwarfs. The plasma density and temperature decrease in the stellar coronae which are hot enough to be blown away as stellar winds. The combined action of the stellar wind and the star's magnetic field slows down its rotation. Zooming into the solar system, we find the heliosphere filled with

such a plasma, the solar wind. As this interacts with a planetary magnetic field, it creates a magnetosphere: a cavity in the solar wind, dominated by the planet's magnetic field. Disturbances in the solar wind, caused either by temporal and spatial variations or by transient phenomena, shake the magnetosphere, leading to geomagnetic disturbances and aurora. When the supersonic solar wind hits an obstacle, i.e. the magnetosphere, it is abruptly slowed down, forming a bow shock. The basic ingredient of the magnetosphere, the magnetic field, originates in a magnetohydrodynamic dynamo in the plasma inside the Earth's core. A similar process can be found inside other planets; solar and stellar magnetic fields originate in dynamo processes, too.

Space physics not only is concerned with plasmas and fields but also with energetic particles. These have energies much higher than the thermal energy of the plasma. An understanding of the propagation and acceleration of these particles also is an important topic, in particular because it requires the concept of wave-particle interactions.

Basic physical questions to be addressed in the study of space plasmas include the acceleration of particles in space, the origin and acceleration of galactic cosmic radiation, the radiative emission of planets, pulsars, quasars, and radio galaxies, the physics of interplanetary, interstellar and intergalactic plasmas, the formation of organic molecules in plasma-chemical processes in space, and the origin of planetary magnetic fields.

This book does not attempt to address all these topics. Instead, it will limit its scope to the introduction of the basic principles of plasma physics and applications of these concepts to the understanding of the heliosphere and the magnetospheres of the planets.

The heliosphere is structured by the solar wind and the frozen-in solar magnetic field, extending at least about 100 AU away from the Sun.¹ Fluctuations on the Sun, interactions of different plasma streams, and plasma clouds ejected from the Sun excite different kinds of waves. Some of these waves steepen during their propagation, forming shock waves. Energetic particles of solar and galactic origin interact with these waves: this results in spatial scattering as well as scattering in momentum space. These wave-particle interactions are of foremost interest in understanding space plasmas; however, they are also formally difficult because they delve deeply into non-linear processes.

The topology of the magnetosphere is even more complex than that of the heliosphere; however, the physical processes, although on smaller spatial scales, are the same. A magnetosphere is defined as a spatial region where the motion of particles is governed by the planet's magnetic field. This brief definition contains a lot of information. We learn the obvious: to exist, the magnetosphere requires a magnetic field. Particles in the magnetosphere are charged particles: the motion of neutrals would not be influenced by the

¹ AU is short for astronomical unit which is the average distance between the Sun and Earth. 1 AU = 149.6×10^6 km.

magnetic field. In addition, their density is low: in a dense medium, collisions between the particles would determine their motion, and the influence of the magnetic field would be negligible. And we learn that the energy density of the charged particles is small compared with the energy density of the field: otherwise the particle motion would distort the field instead of the field guiding the particles.

The inner boundary of the magnetosphere is determined by the density: getting closer to the planet, the density increases. Brownian motion becomes dominant and the magnetic field no longer guides the particles. In the Earth's magnetosphere this transition happens at a height of a few hundred kilometers. The upper ionosphere, extending to a height of about 1000 km, lies well inside the magnetosphere. The outer boundary of the magnetosphere is the magnetopause. At this boundary layer the energy density of the Earth's magnetic field equals the energy density of the solar wind. Moving towards the Sun, we would detect this boundary at a distance of about 10 Earth's radii. In the opposite direction, the magnetosphere extends far beyond the orbit of the moon. The magnetopause separates the planetary and the interplanetary magnetic fields and prevents most of the solar wind plasma from entering the Earth's magnetosphere and atmosphere.

Earth is not the only planet with such a luxurious shielding. In the solar system all planets except for our two neighbours, Mars and Venus, have a magnetosphere (cf. Chap. 12). But even planets without magnetospheres and comets are shielded against the solar wind: when the supersonic solar wind hits these bodies, it is slowed down and deflected around the obstacle, forming a bow shock in front of it.

1.3 A Brief History of Space Research

Only in the late 1950s and early 1960s, with the advance in satellite technology, did *in situ* observations of plasmas and particles in the magnetosphere and in interplanetary space become possible. Many of these observations are discussed in the broad topic of solar-terrestrial relationships, which are the influence of the Sun and solar activity on the Earth and its environment. Amazingly, solar-terrestrial relationships is also an old field of science: it dates back to the first correlations between sunspots and aurora, in ancient China a few thousand years ago. Aurorae are the prime example of solar-terrestrial relationships: they can be detected easily, even with the naked eye, they are closely correlated with solar activity, and they also have an aesthetic and even mythological quality. But the big steps in the understanding of solar-terrestrial relationships required observations made on board rockets and satellites: measurements of plasmas and particles in the ionosphere, the magnetosphere, and interplanetary space, and the observation of the Sun's electromagnetic radiation in frequency ranges not observable from the ground.

The solar wind and the magnetosphere act as coupling devices in solar–terrestrial relationships. Magnetism was detected more than 2000 years ago when the ancient Greeks found stones that attracted iron. The first reliable description of a compass dates back to the eleventh century when Shon-Kau (1030–1093 AD) wrote in a Chinese encyclopedia: ‘fortune-tellers rub the point of a needle with the stone of the magnet in order to make it properly indicate the south’. The first written account of a compass in Europe dates back to Alexander Neekam (1157–1217 AD), a monk at St. Albans. He describes the compass and its application in navigation as common. Neekam’s compass is a second generation instrument and quite similar to the ones used today: while in the first compass a small piece of magnetic stone floated in water on a piece of wood or cork, in Neekam’s compass a needle is placed on a pivot, allowing it to rotate freely and align itself in the north–south direction. In the fourteenth century the compass was common on ships. The declination, the time- and space-dependent difference between magnetic and geographic north, was well known by the middle of the sixteenth century.

But only in 1600s was the basic ingredient of the magnetosphere, the Earth’s magnetic field, “detected”: in his treatise *De Magnete*, William Gilbert suggested that the north–south alignment of the compass results from the magnetic field of the Earth. In the middle of the nineteenth century, scientists began to understand the terrestrial magnetic field. A global network of observatories started continuous registrations of the magnetic field and its fluctuations. After a detailed analysis of these data, C.-F. Gauss proposed that the Earth’s magnetic field consists of two components, one stemming from its interior and a second one originating in the atmosphere. To first order, the geomagnetic field stemming from the interior can be described as the field of a homogeneously magnetized sphere. Ørsted and Ampere suggested ring currents inside the Earth to explain the field. This dipole-field approach survived up to the early 1960s when satellite and rocket observations in the upper atmosphere gave a more detailed picture of the true field and suggested modifications to the model.

Gauss and Weber initiated extremely precise measurements of the Earth’s magnetic field with relative accuracies of at least 10^{-5} . Small fluctuations could be detected, showing systematic variations in time and location as well as superimposed stochastic changes. These latter strongly indicated that the Earth is not an isolated object in the empty vastness of space. Instead, strong forces from outside are acting on spaceship Earth (cf. Chaps. 10 and 11).

In its heyday in the second half of the nineteenth century, solar–terrestrial relationships were an illustrative example of the development of science from correlations and apparently uncorrelated, sometimes even seemingly contradictory, observations into a consistent picture of a complex environment (cf. Sect. 11.3). The converging developments included, for example, the discovery of the 11-year sunspot cycle by H.S. Schwabe in 1843 and the correlation between sunspot numbers and the frequency of geomagnetic disturbances by

E. Sabine in 1852. In the same year, R. Wolf found a correlation between sunspots and aurorae. A relationship between individual aurorae and accompanying geomagnetic disturbances had already been noticed by A. Celsius and O.P. Hiorter in 1747 [1.3, 1.8] and by A. Humboldt in 1806 [1.9]. The spatial distribution of aurorae too suggested an involvement of the geomagnetic field. Aurorae always were known as a high-latitude phenomenon. But in the 1840s the ill-fated Arctic explorer John Franklin [1.5] noticed that the frequency of aurorae does not increase all the way towards the pole. In 1860, E. Loomis [1.10] showed that the highest incidence of aurora is seen inside an oval of 20° – 25° around the magnetic pole. In 1881 H. Fritz [1.6] published similar results with his famous map of isochasms.

During some ten years, out of these observations and statistical correlations a closed picture of solar–terrestrial relationships emerged with the Sun as a source of geomagnetic activity as well as aurorae. In the late 1870s, H. Bequerel [1.1] offered the first physical explanation: the sunspots are assumed to be a source of fast protons. On hitting the Earth’s magnetic field, these particles are guided towards the auroral oval by the magnetic field lines. Despite its simplicity, the model contains a revolutionary aspect: the Sun is not only a source of electromagnetic radiation but its blemishes, the sunspots, are also a source of energetic particles which could affect the Earth’s environment. Early this century, a similar idea led Birkeland to build the terrella, a model of the Earth which allows simulations of the aurora in the laboratory: a cathode-ray tube substitutes for the Sun as a source of energetic particles. A magnetic dipole inside a sphere covered by a fluorescent material simulates the Earth’s magnetic field surrounded by its atmosphere. With these ingredients, Birkeland showed that the geomagnetic field was responsible for the formation of the aurora ovals. From the correlation between aurorae and the number of active regions on the Sun, Birkeland [1.2] suggested that sunspots might be the source of a continuous stream of particles. This idea was an early introduction of the concept of a plasma flow from the Sun, which later evolved into the concept of the solar wind.

In the 1930s, Chapman and Ferraro [1.4] developed an idea about solar–terrestrial relationships which comes closer to our current understanding: sunspots are indicators of solar activity. Solar activity manifests itself in violent eruptions, called solar flares. Chapman and Ferraro suggested that flares not only emit electromagnetic radiation but also fling out clouds of ionized matter which in size dwarf the Earth. After a travel time of 1 to 3 days, such a cloud might hit the magnetosphere and compress it, leading to geomagnetic disturbances. Some of the cloud matter might penetrate the magnetosphere close to the poles, causing the aurora. The existence of these proposed clouds, today called coronal mass ejections, was first confirmed in the early 1970s by Skylab.

While their basic idea is accepted even today, Chapman and Ferraro’s ansatz still makes an insufficient assumption about the Earth’s magnetic

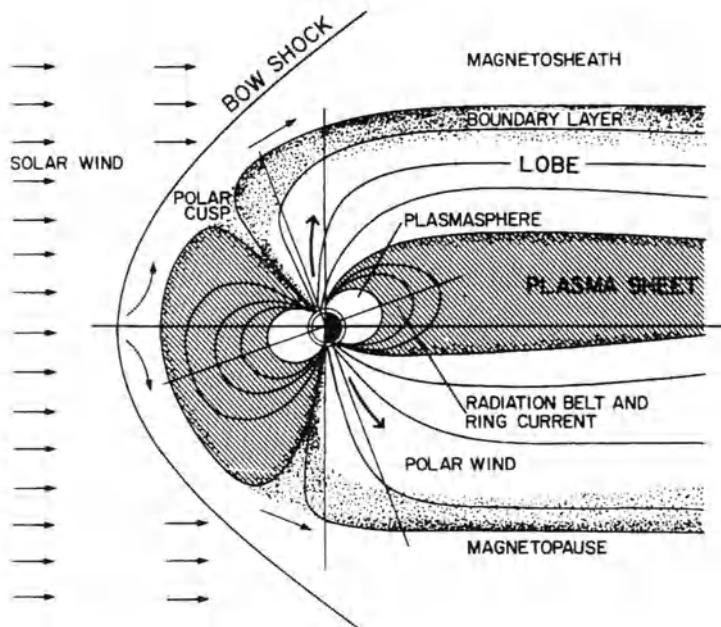


Fig. 1.1. Structure of the Earth's magnetosphere. Reprinted from H. Rosenbauer et al., [1.11], *J. Geophys. Res.* **80**, Copyright 1975, American Geophysical Union

field: it is described as a dipole. Only the first *in situ* measurements of the magnetosphere and of interplanetary space revealed the complexity of the magnetic field around the Earth. But again, indirect evidence had been found long before the first *in situ* measurements became available. In the 1920s, the existence of a region of charged particles in the atmosphere, the ionosphere, had been discovered by its effect on radio waves, allowing them to propagate far beyond the horizon: this layer can be used to explain the atmospheric contribution to the Earth's magnetic field which had been proposed 70 years earlier by Gauss. Thus early magnetosphere researchers knew of a conductive layer at a height of some tens of kilometers. In the early 1950s, in Arctica and Antarctica, van Allen and colleagues launched rockets to a height of about 110 km. Their instruments confirmed the existence of energetic electrons in this region, either directly or by observing the electron bremsstrahlung: the existence of the ionosphere had been confirmed by *in situ* measurements. In 1958, a Geiger counter on board the first US satellite, Explorer 1, detected the Earth's radiation belts, later named van Allen belts to honor their discoverer. In the same year, the Soviet lunar probe made the first measurements in interplanetary space, confirming the existence of the long-proposed solar wind. First detailed studies of the solar wind were made by Mariner 2 in 1962. The boundary between interplanetary space and the Earth system, the magnetopause, was first studied by Explorer 10 in 1961; the bow shock in

front of it was detected by Explorer 12 in 1962 and studied in detail by OGO (Orbiting Geophysical Observatory) in 1964.

These and subsequent observations led to the identification of the main components of the magnetosphere, as shown in Fig. 1.1. The magnetosphere is formed by the interaction of the solar wind with the Earth's magnetic field. A boundary sheet, the magnetopause, forms where the pressure of the solar wind equals the magnetic field pressure. The solar wind streams around the magnetopause but it does not penetrate into the magnetosphere. Where the supersonic solar wind is slowed down to subsonic speed, the bow shock develops. At high latitudes, polar cusps form, separating closed magnetic field lines in front of the magnetosphere from open field lines pulled away to the magnetotail by the solar wind. At these cusps, particles and plasmas can penetrate into the magnetosphere. These and many other details of the magnetosphere will be discussed in Chap. 10.

But morphology is only one property, functioning is another. To understand the dynamics of the complex system Sun, solar wind, magnetosphere, and ionosphere, we have to develop concepts of plasma physics and test them with the observations. This book is structured so as to first present a concise and consistent introduction into the most important concepts of plasma physics and interactions between plasmas and energetic particles (Chaps. 2–7). Later the phenomena and observations will be presented and discussed in the framework of these concepts (Chaps. 8–13).

Exercises and Problems

1.1 Define a plasma. Discuss the importance of the density. Is there a limit for the relative or the absolute size of the neutral component?

1.2 What do you need to explain a magnetosphere? What determines its spatial extent?

1.3 Describe the basic features of a magnetosphere. How do these properties change if the axis of the magnetic field changes with respect to the solar wind direction?

1.4 Why is space dominated by plasmas?

1.5 Where in the near-Earth environment do plasmas exist? Would we miss them if they were neutral matter instead?

2. Charged Particles in Electromagnetic Fields

I've gotten a rock, I've gotten a reel,
I've gotten a wee bit spinning-wheel;
An' by the whirling rim I've found
how the weary, weary warl goes round.
S. Blamire, *I've gotten a rock*

To discuss the basic motion of charged particles in electric and magnetic fields, they are described as test particles: their motion does not affect the field. It can be separated into two parts: the motion of the guiding center of the particle orbit and the particle's gyration around it. The motion of the guiding center can be interpreted as the effective motion of the particle, averaged over many gyrations. This concept will be applied to drifts of particles in static electromagnetic fields. The adiabatic invariants will allow simple estimates of the particle motion in temporally and spatially slowly varying fields. This concept will be applied to the motion of particles in the Earth's radiation belts. This chapter starts with a brief recapitulation of Maxwell's equations and a description of electromagnetic fields.

2.1 Electromagnetic Fields

Particle densities in interplanetary space and in the magnetosphere are low. Thus a description of the electromagnetic field in a vacuum is sufficient. Then the permeability μ and the permittivity ε both equal 1: the medium can be neither polarized nor magnetized. In principle, we can assign a permittivity to a plasma, but this is just another description of the existence of charged particles. A plasma can also be magnetized: for instance, an axial-symmetric ring current of charged particles in a dipole field leads to a reduction of the magnetic moment, which is a diamagnetic effect. An equivalent description is the magnetic induction \mathbf{B} in a vacuum, consisting of both the dipole field and the field disturbance, and the current arising from the particle motion.

In the cgs system, because of its simplicity for solving the problems under study still common in geophysics and spacephysics, the permittivity ε_0 and the permeability μ_0 in a vacuum both equal 1. Compared with the SI system,

this leads to a factor c in some equations because $c = 1/\sqrt{\epsilon_0 \mu_0}$. If quantities in SI units are inserted into equations given in the cgs system, the analysis of units automatically leads to the correct consideration of μ_0 and ϵ_0 .

2.1.1 Maxwell's Equations

In 1864 Maxwell introduced a unified theory of the electromagnetic field. The sources of the electric and magnetic fields are charges, magnetized bodies, and currents, which can be either discrete or continuous, and which can be either constant in time or time-dependent.

The electric field \mathbf{E} generated by a charge density ρ_c is described by Poisson's equation:

$$\nabla \cdot \mathbf{E} = 4\pi\rho_c . \quad (2.1)$$

Since the electric field is non-rotational, it can be expressed by the gradient of a scalar potential φ : $\mathbf{E} = -\nabla\varphi$. Integrating (2.1) over a volume V and using Gauss's theorem (A.23), Poisson's equation can be rewritten as Gauss's law for an electric field:

$$\oint_{O(V)} \mathbf{E} d\mathbf{S} = 4\pi \int_V \rho_c d^3r . \quad (2.2)$$

Gauss's law states: the electric flux through a surface S enclosing a volume V is determined by the total charge inside V . If there are no net charges enclosed in V , the flux through S is zero. But V is not necessarily field-free, e.g. if V is placed inside a dipole field either enclosing none of the charges or both of them. Using a spherical test volume V with radius r , Coulomb's law can be derived from Gauss's law.

Gauss's law for a magnetic field is formally strictly analogous, $\nabla \cdot \mathbf{B} = 0$, or in integral form

$$\oint_{O(V)} \mathbf{B} d\mathbf{S} = 0 . \quad (2.3)$$

It states that there are no magnetic monopoles.

Faraday's law describes the electric field (or electro-motoric force) generated by a changing magnetic field:

$$\nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} . \quad (2.4)$$

The magnetic flux Φ through a surface S in the magnetic field is defined as

$$\Phi = \int \mathbf{B} d\mathbf{S} . \quad (2.5)$$

Graphically, the magnetic flux can be interpreted as the number of field lines going through S . Using this definition and Stokes's theorem (A.29), Faraday's law can be rewritten as

$$\oint_C \mathbf{E} \cdot d\mathbf{l} = -\frac{1}{c} \frac{\partial}{\partial t} \int_O \mathbf{B} \cdot d\mathbf{S} = -\frac{1}{c} \frac{\partial \Phi}{\partial t} . \quad (2.6)$$

It states that a variation in the magnetic flux through a surface creates an electro-motoric force in its circumference. The minus sign indicates that the electric field, if it were to act on an electric charge, would give rise to an induced current that opposes the change in the magnetic flux (Lenz's rule). Faraday's law has two consequences: if the magnetic field is constant, the electric field vanishes. And if the electric field is zero, the magnetic field is static.

Ampere's law describes the relationship between a magnetic field on the one hand and a time-dependent electric field and a current:

$$\nabla \times \mathbf{B} = \frac{4\pi}{c} \mathbf{j} + \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} . \quad (2.7)$$

The last term on the right-hand side, the displacement current, arises from the equation of continuity for charges (cf. Sect. 3.1.2):

$$\frac{\partial \varrho_e}{\partial t} + \nabla (\varrho_c \mathbf{v}) = 0 \quad \text{or} \quad \frac{d\varrho_e}{dt} + \varrho_c \nabla \cdot \mathbf{v} = 0 . \quad (2.8)$$

Using Stokes' theorem, (2.7) can be written as

$$\oint_C \mathbf{B} \cdot d\mathbf{l} = \frac{4\pi}{c} \int j \cdot d\mathbf{S} + \frac{1}{c} \int \frac{\partial \mathbf{E}}{\partial t} \cdot d\mathbf{S} . \quad (2.9)$$

Ampere's law states that a changing electric field and/or a current creates a rotational magnetic field.

2.1.2 Transformation of Field Equations

In space physics, fields and plasmas move with respect to the observer: the solar wind is swept over a spacecraft in interplanetary space, another spacecraft crosses through the radiation belts of a planet. The fields \mathbf{E} , \mathbf{B} and \mathbf{E}' , \mathbf{B}' in two reference frames C and C' , with C' moving with velocity \mathbf{v} with respect to C , are connected by the transformations

$$\mathbf{E}' = \frac{1}{\gamma} \left[\mathbf{E} + \frac{\mathbf{v}}{v^2} (\mathbf{v} \cdot \mathbf{E}) (1 - \gamma) + \frac{1}{c} \mathbf{v} \times \mathbf{B} \right] \quad (2.10)$$

and

$$\mathbf{B}' = \frac{1}{\gamma} \left[\mathbf{B} + \frac{\mathbf{v}}{v^2} (\mathbf{v} \cdot \mathbf{B}) (1 - \gamma) - \frac{1}{c} \mathbf{v} \times \mathbf{E} \right] , \quad (2.11)$$

with $\gamma = \sqrt{1 - v^2/c^2}$. In the non-relativistic case, all terms in v^2/c^2 can be ignored ($\gamma \rightarrow 1$), reducing the equations to

$$\mathbf{E}' = \mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{B} \quad \text{and} \quad \mathbf{B}' = \mathbf{B} - \frac{1}{c} \mathbf{v} \times \mathbf{E} . \quad (2.12)$$

Within the framework of this book, the second set of transformations will be sufficient, considering effects of order v/c (e.g. the Doppler effect) but ignoring effects of order v^2/c^2 (e.g. the Lorentz contraction).

2.1.3 Generalized Ohm's Law

Ohm's law connects the current density \mathbf{j} and the electric field \mathbf{E} by a constant, the conductivity σ [s^{-1}]:

$$\mathbf{j} = \sigma \mathbf{E} . \quad (2.13)$$

Equation (2.13) is valid in the plasma's rest frame only. If an observer is moving with velocity \mathbf{v} with respect to the plasma's frame, a generalized form of Ohm's law is required. It can be obtained by applying (2.12). The plasma has a high conductivity and thus the current on the left-hand side of (2.13) transforms under consideration of (2.7) $\mathbf{j} = \mathbf{j}'$. The generalized form of Ohm's law than reads

$$\mathbf{j} = \sigma \left(\mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{B} \right) . \quad (2.14)$$

The second term on the right-hand side describes an electric induction field which gives rise to the Hall current.

2.1.4 Energy Equation of the Electromagnetic Field

From Faraday's law we can derive an energy equation for the electromagnetic field. Multiplication of (2.4) by \mathbf{B} and integration over a volume V gives

$$\int_V \mathbf{B} \frac{\partial \mathbf{B}}{\partial t} d^3r = -c \int_V \mathbf{B} \nabla \times \mathbf{E} d^3r . \quad (2.15)$$

The divergence of a vector product can be written as (cf. (A.20))

$$\nabla(\mathbf{E} \times \mathbf{B}) = \mathbf{B} \nabla \times \mathbf{E} - \mathbf{E} \nabla \times \mathbf{B} . \quad (2.16)$$

The first term on the right-hand side is the term on the right-hand side in (2.15). The second term can be rewritten using Ampere's law (2.7) as:

$$\mathbf{B} \nabla \times \mathbf{E} = \nabla(\mathbf{E} \times \mathbf{B}) + \mathbf{E} \frac{4\pi}{c} \mathbf{j} + \mathbf{E} \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} . \quad (2.17)$$

Inserting this term into (2.15), the energy equation reads

$$\begin{aligned} & \int_V \mathbf{B} \frac{\partial \mathbf{B}}{\partial t} d^3r \\ &= -c \int_V \nabla(\mathbf{E} \times \mathbf{B}) d^3r - c \int_V \mathbf{E} \frac{4\pi}{c} \mathbf{j} d^3r - c \int_V \mathbf{E} \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} d^3r . \end{aligned} \quad (2.18)$$

With Gauss's theorem (A.23) the volume integral in the first term on the right-hand side can be changed into a surface integral. The first and last terms on the right-hand side then contain the product of a vector and the temporal derivative of this vector. This is equal to half of the temporal derivative of the squared vector as can be seen by differentiating the middle term in

$$\mathbf{a} \frac{\partial \mathbf{a}}{\partial t} = \frac{1}{2} \frac{\partial(\mathbf{a} \cdot \mathbf{a})}{\partial t} = \frac{1}{2} \frac{\partial \mathbf{a}^2}{\partial t}. \quad (2.19)$$

The energy equation now reads

$$\frac{\partial}{\partial t} \int_V \left(\frac{B^2}{8\pi} + \frac{E^2}{8\pi} \right) d^3r = -\frac{1}{4\pi} \oint_{O(V)} (\mathbf{E} \times \mathbf{B}) dS - \int_V \mathbf{E} \mathbf{j} d^3r. \quad (2.20)$$

Here $B^2/8\pi$ and $E^2/8\pi$ are the energy densities of the magnetic and the electric fields, respectively. Equation (2.20) can be interpreted as an equation for the continuity of the electromagnetic field: the change in the energy density of the electromagnetic field inside a volume V is given by the energy flux (Poynting vector \mathbf{S}),

$$\mathbf{S} = \frac{\mathbf{E} \times \mathbf{B}}{4\pi}, \quad (2.21)$$

through the surface of the volume and ohmic losses $\mathbf{E} \mathbf{j}$ inside V . While a positive value of $\mathbf{E} \mathbf{j}$ indicates losses, a negative value describes a generator.

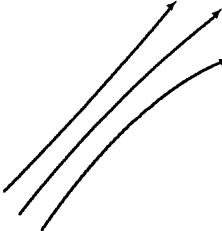
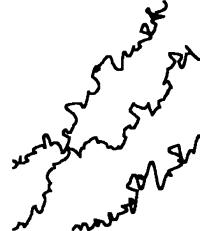
The energy density of the electromagnetic field is an important parameter when we consider the motion of an ionized gas instead of individual particles. A comparison of the energy densities of the field and the plasma shows whether the motion of the particles will be governed by the electromagnetic fields or by the gas laws: if the plasma's energy density is high and the conductivity infinite, the field is frozen-in (cf. Sect. 3.3.1) and carried away by the plasma (e.g. the interplanetary magnetic field frozen-into the solar wind). If the energy density of the field is high, the field determines the motion of the particles (e.g. energetic particles in interplanetary space or in the radiation belts).

2.2 Particle Motion in Electromagnetic Fields

Let us now turn to the motion of individual charged particles in a prescribed electromagnetic field. We will assume that the particle densities are small: there are no collisions between particles. In addition, the energy density of the particles is small; thus their motion does not modify the external field.

Although these limitations are strong, the resulting motion and their formal description are basic and instructive. Table 2.1 places them into the general framework of particle motion in electromagnetic fields. The basic types of fields are smooth fields, which sometimes can be described analytically, and

Table 2.1. Particle motion in different types of magnetic fields. The characteristic length scale L for changes in the magnetic field is defined as $1/L = 1/|B| \cdot \partial B / \partial x$

Field			
	Smooth field, analytical description possible	Turbulent, irregular fast fluctuations	
Scales	small variations in space and time $r_L \ll L$	large variations $B = \langle B \rangle + \delta B$, $r_L \geq L$	weak turbulence $\delta B \ll \langle B \rangle $ strong turbulence: $\delta B \approx \langle B \rangle $
Formalism	adiabatic invariants, guiding center motion, drifts	integration of the equation of motion	transport equations, pitch-angle scattering, resonance interaction
Occurrence	periodic motion in radiation belts, magnetic mirrors, and bottles	Størmer orbits	particle propagation in interplanetary space

turbulent fields. For smooth fields, two cases of particle motion can be distinguished: (a) The field varies only weakly on the temporal and spatial scales of the gyration and the particle motion can be described by the concepts of guiding center motion and adiabatic invariants (cf. Sect. 2.4). These concepts can be applied to particles in the radiation belts (cf. Sect. 10.4.1). (b) The field changes significantly during one gyration of the particle. The above concepts are no longer valid and the equation of motion has to be integrated. One example are the Størmer orbits of galactic cosmic rays in the magnetosphere (cf. Sect. 10.4.2). Turbulent fields require an entirely different approach. The particle motion is determined not only by the average magnetic field but also by scattering at field fluctuations, a stochastic process which requires the description of particle ensembles instead of single particles and transport equations instead of equations of motion. Particle propagation in turbulent fields can be understood as a diffusive process (cf. Chap. 6); applications are the interplanetary transport (cf. Sect. 9.2.2) or particle scattering into the loss cone in the radiation belts (cf. Sect. 10.4.1).

2.2.1 Lorentz Force and Gyration

The general equation of motion is Newton's second law $\mathbf{F} = d\mathbf{p}/dt$. The net force on a particle can consist of different components, e.g. gravitation, electromagnetic forces or a pressure gradient. In a pure electromagnetic field the Lorentz force

$$\mathbf{F}_L = m \frac{d\mathbf{v}}{dt} = q \cdot \left(\mathbf{E} + \frac{\mathbf{v} \times \mathbf{B}}{c} \right) \quad (2.22)$$

is the only force acting on a particle.

First Integral of Motion. In a pure magnetic field the electric field is zero. According to Faraday's law, the magnetic field then is static. The equation of motion reduces to

$$m \frac{d\mathbf{v}}{dt} = q \frac{\mathbf{v} \times \mathbf{B}}{c} . \quad (2.23)$$

Multiplication by the particle speed \mathbf{v} and application of (2.19) gives the first integral of motion:

$$\frac{1}{2} m \frac{d\mathbf{v}^2}{dt} = \frac{dE_{\text{kin}}}{dt} = q\mathbf{v} \cdot \frac{(\mathbf{v} \times \mathbf{B})}{c} = 0 . \quad (2.24)$$

It states that in a pure magnetic field the kinetic energy E_{kin} of a particle is constant. The kinetic energy of a particle is given in electronvolts (eV) with 1 eV being the energy gained by an electron after traversing a potential difference of 1 V. 1 eV equals 1.602×10^{-19} J.

Gyration. Let us now assume a homogeneous magnetic field along the z axis: $\mathbf{B} = B \cdot \mathbf{e}_z$. The equation of motion for the components then reads

$$m\dot{v}_x = \frac{qB}{c}v_y , \quad m\dot{v}_y = -\frac{qB}{c}v_x , \quad \text{and} \quad m\dot{v}_z = 0 . \quad (2.25)$$

Integration of the last equation gives $v_{||} = v_z = \text{const}$, the particle moves parallel to the field line with constant speed v_z . The other two equations give

$$\ddot{v}_x = \frac{qB}{mc}\dot{v}_y = -\left(\frac{qB}{mc}\right)^2 v_x \quad \text{and} \quad \ddot{v}_y = -\frac{qB}{mc}\dot{v}_x = -\left(\frac{qB}{mc}\right)^2 v_y , \quad (2.26)$$

describing a harmonic oscillator with a cyclotron frequency ω_c

$$\omega_c := \frac{|q|B}{mc} . \quad (2.27)$$

The solution of the equation of motion is a circular orbit around the magnetic field lines in the xy plane. The components of the trajectory are

$$x(t) = r_L \sin \omega_c t \quad \text{and} \quad y(t) = r_L \cos \omega_c t , \quad (2.28)$$

and the components of the particle velocity are

$$v_x(t) = r_L \omega_c \cos \omega_c t \quad \text{and} \quad v_y(t) = -r_L \omega_c \sin \omega_c t . \quad (2.29)$$

The particle speed v_{\perp} perpendicular to the magnetic field then is

$$v_{\perp}^2 = v_x^2 + v_y^2 = r_L^2 \omega_c^2 . \quad (2.30)$$

The radius of the particle orbit, the Larmor radius, r_L is given as

$$r_L := \frac{v_{\perp}}{\omega_c} = \frac{mv_{\perp}c}{|q|B} . \quad (2.31)$$

The direction of motion depends on the particle's charge and has to obey Lenz's rule: the ring current associated with the particle motion creates a magnetic field in a direction opposite to that of the external magnetic field. An electron obeys the right-hand rule (cf. Table 2.2): if the thumb is directed along the magnetic field line, the tips of the curved fingers give the direction of the electron motion. If the initial velocity has a component parallel to the magnetic field, the particle follows a helical path around the line of force.

Useful Definitions. The magnetic rigidity P describes the resistance of a particle to change its direction of motion in a magnetic field. It is defined as the ratio of the momentum p_{\perp} perpendicular to \mathbf{B} and the charge q :

$$P = \frac{p_{\perp}}{q} \cdot c . \quad (2.32)$$

The unit of P is the volt, V. The Larmor radius (2.31) then is as the ratio between the magnetic rigidity and the magnetic field strength $r_L = P/B$.

The gyration of the particle is determined by its speed v_{\perp} perpendicular to the magnetic field while the particle itself is characterized by its energy or velocity. The relative sizes of the velocity components parallel and perpendicular to the magnetic field can be described by the pitch angle α with

$$\tan \alpha = \frac{v_{\perp}}{v_{\parallel}} . \quad (2.33)$$

The velocity components perpendicular and parallel to the field therefore are $v_{\perp} = \sin \alpha \cdot v$ and $v_{\parallel} = \cos \alpha \cdot v$.

A particle with charge q moving in a circular orbit with radius r_L and gyration time T_c gives rise to the ring current $I = q/T_c = q\omega_c/(2\pi)$. The magnetic moment μ is defined as the product of the ring current and the enclosed area $A = \pi r_L^2$:

$$\mu := \frac{q}{c} \cdot I \cdot A = \frac{mv_{\perp}^2}{2B} = \frac{E_{\perp}}{B} . \quad (2.34)$$

The orientation of the magnetic moment is determined by the direction of the ring current:

$$\boldsymbol{\mu} = \frac{1}{2} qr_L \times \mathbf{v}_{\perp} = -\frac{mv_{\perp}^2}{2B^2} \mathbf{B} . \quad (2.35)$$

The magnetic moment does not depend on the charge of the particle; its direction is opposite to the external field. A plasma, therefore, is diamagnetic.

In a highly conductive plasma, no electric field exists and the particle energy is conserved (cf. (2.24)). Then v_{\parallel} is constant. Therefore v_{\perp} is constant, too. If the magnetic field is constant, the magnetic moment will also be constant. Even in the case of a slowly (adiabatically) varying field μ is constant. It is therefore called an adiabatic invariant (cf. Sect. 2.4.1).

2.3 Drifts of Particles in Electromagnetic Fields

Particle drifts in electromagnetic fields result from changes in the Larmor radius during one gyration, either because the particle speed (as in $\mathbf{E} \times \mathbf{B}$ and in gravitational drift) or because the magnetic field changes (as in gradient drift). For particles with pitch angle 0° the drifts vanish, except for the curvature drift, which arises from the field-parallel motion.

2.3.1 The Concept of the Guiding Center

The concept of the guiding center, introduced in the 1940s by H. Alfén, separates the motion \mathbf{v} of a particle into motions \mathbf{v}_{\parallel} parallel and \mathbf{v}_{\perp} perpendicular to the field. The latter can consist of a drift v_D and a gyration ω :

$$\mathbf{v} = \mathbf{v}_{\parallel} + \mathbf{v}_{\perp} = \mathbf{v}_{\parallel} + \mathbf{v}_D + \boldsymbol{\omega} = \mathbf{v}_{gc} + \boldsymbol{\omega}. \quad (2.36)$$

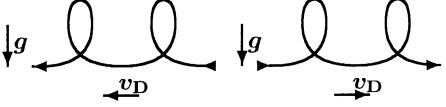
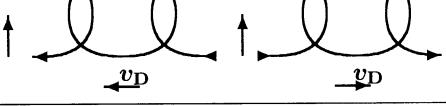
Thus the motions parallel and perpendicular to the field are decoupled. If we follow the particle for a long time, the gyration is of minor importance. In some sense it is averaged out, and the particle motion is described by the motion v_{gc} of the guiding center, consisting of a field-parallel motion and a drift. The particle is within a gyro-radius of this position.

2.3.2 Crossed Magnetic and Electric Fields: $\mathbf{E} \times \mathbf{B}$ Drift

Let us consider an electric field \mathbf{E} perpendicular to the magnetic field \mathbf{B} . Both fields are homogeneous and constant in time. What kind of motion do we expect? The magnetic field forces the particle into a gyration around the field line. During half of its orbit, the particle motion has a component parallel to the electric field, during the other half it is anti-parallel. Therefore the particle alternately is accelerated and decelerated.

Let us now look at this motion in detail. In the second row of Table 2.2 we show the motion of particles in a magnetic field pointing upwards through the paper and an electric field pointing downwards. The right column shows the motion of an electron. It starts in the upper right corner and is forced into a downward motion by the magnetic field. Since this motion is parallel to the electric field, the particle decelerates and its gyro-radius decreases. At the lowest point of its orbit the electron therefore is shifted towards the left with respect to its starting point. In the upward part of its gyro-orbit the

Table 2.2. Drift in different types of fields

\mathbf{B} upward through the paper	Charge	Equation
	Positive	Negative
\mathbf{B} homogeneous	 	(2.37)
\mathbf{E} homogeneous		(2.47)
Gravitation		(2.49)
Inhomogeneous \mathbf{B}		(2.50)

electron moves parallel to the field. Thus it is accelerated and its Larmor radius increases, shifting the electron farther to the left. This $\mathbf{E} \times \mathbf{B}$ drift results from a continuous change of the particle's gyro-radius and is perpendicular to both the electric and magnetic fields. Therefore, in spite of the presence of an electric field the particle does not gain energy. For a particle with positive charge, the direction of the drift is the same: its gyro-motion is opposite to that of an electron, as are the parts of its orbit with the smallest and largest gyro-radii.

To derive a quantitative description of drift we substitute

$$\mathbf{v} = \mathbf{w} + \frac{c}{B^2} \mathbf{E} \times \mathbf{B} . \quad (2.37)$$

This corresponds to the transformation into a frame of reference moving with velocity $c\mathbf{E} \times \mathbf{B}/B^2$. It is $\dot{\mathbf{v}} = \dot{\mathbf{w}}$ because both the electric and magnetic fields are constant. The equation of motion therefore reads

$$m\ddot{\mathbf{w}} = m\dot{\mathbf{v}} = q\mathbf{E} + \frac{q}{c}\mathbf{w} \times \mathbf{B} + \frac{q}{B^2}(\mathbf{E} \times \mathbf{B}) \times \mathbf{B} . \quad (2.38)$$

Because \mathbf{E} is perpendicular to \mathbf{B} , the last term on the right-hand side can be written as $-\mathbf{E}B^2$. Therefore, the first and last terms on the right-hand side cancel and the equation of motion reduces to

$$m\dot{\mathbf{w}} = q \frac{\mathbf{w} \times \mathbf{B}}{c} . \quad (2.39)$$

Now \mathbf{w} fulfills (2.23). In the new reference system the particle motion therefore is a gyration and the motion of this system gives the drift velocity:

$$\mathbf{v}_{\mathbf{E} \times \mathbf{B}} = c \frac{\mathbf{E} \times \mathbf{B}}{B^2} . \quad (2.40)$$

The direction and size of the drift depend on the fields alone: particle properties such as mass, charge or velocity do not enter into the drift.

With substitutions similar to the one in (2.37) the other drift velocities can be derived. The general form of the drift velocity in a force field \mathbf{F} is

$$\mathbf{v}_F = \frac{c}{q} \frac{\mathbf{F} \times \mathbf{B}}{B^2} . \quad (2.41)$$

In the above example $\mathbf{F} = \mathbf{E}/q$ has been used.

2.3.3 Magnetic and Gravitational Fields

Now consider a gravitational field perpendicular to a homogeneous magnetic field (third row in Table 2.2). As in the previous case, the Larmor radius changes continuously. But compared with the $\mathbf{E} \times \mathbf{B}$ drift the external force does not depend on the charge sign, and the points of the largest and smallest gyro-radii are the same for both positive and negative particles. Because particles with different charge signs have opposite directions of gyration, they drift in opposite directions. The direction of drift therefore depends on the charge q , and its size depends on the particle mass m , because the external force depends on the particle mass. With $\mathbf{F} = mg$, (2.41) yields for the drift velocity

$$\mathbf{v}_g = \frac{cm}{q} \frac{\mathbf{g} \times \mathbf{B}}{B^2} . \quad (2.42)$$

A gravitational field perpendicular to a magnetic field therefore allows the separation of particles with positive and negative charges. In addition, the magnetic field prevents the particles from “falling down”. There is no net acceleration along \mathbf{g} and the potential and kinetic energies are both constant.

2.3.4 Inhomogeneous Magnetic Fields

In an inhomogeneous magnetic field (bottom row in Table 2.2) particles with positive and negative charges behave in a similar way to the one in the previous case. While in the gravitational field or in an electric field the gyro-radius changes due to changes in particle speed, in an inhomogeneous magnetic field the particle speed is constant and the gyro-radius increases/decreases with decreasing/increasing field strength. This change in gyro-radius is independent of the charge sign, leading to drifts of electrons and protons in opposite directions. The resulting gradient drift is

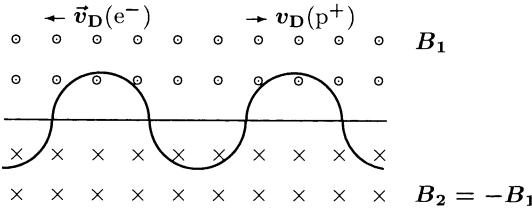


Fig. 2.1. An efficient gradient drift occurs in adjacent fields of opposite polarity, for instance along the heliospheric current sheet

$$\mathbf{v}_{\nabla B} = \frac{c\mu}{qB^2} \mathbf{B} \times \nabla B = \pm \frac{c}{2} \frac{v_{\perp} r_L}{B^2} \mathbf{B} \times \nabla B . \quad (2.43)$$

The \pm on the right-hand side reflects the charge dependence of the drift direction.

A very efficient drift develops in the configuration of two opposing magnetic fields. Imagine the guiding center of the particle orbit on the neutral line between the fields. The particle starts its gyration in the upper half of the field, but after crossing the neutral line, the direction of gyration is reversed. Therefore, instead of a closed circular orbit two semicircles result, leading to displacement by 4 Larmor radii during one gyration, cf. Fig. 2.1. In interplanetary space such a drift takes place along the heliospheric current sheet and contributes to the modulation of galactic cosmic rays (cf. Sect. 9.5).

2.3.5 Curvature Drift

Imagine a homogeneous magnetic field in which the lines of force are curved with a radius r_c . In a vacuum such a field would obey Maxwell's equations only if combined with a magnetic field gradient. The net drift therefore would be the sum of the curvature drift and the gradient drift. To derive the curvature drift alone, let us start from a simplified configuration without a magnetic field gradient. The drift arises from the centrifugal force \mathbf{F}_{cf} on the particle during its motion parallel to the field. Therefore, the drift is not determined by v_{\perp} as in the previous cases but by the speed v_{\parallel} parallel to the field. Inserting the centrifugal force $\mathbf{F}_{cf} = mv_{\parallel}^2 \mathbf{e}_r / r_c = mv_{\parallel}^2 \mathbf{r}_c / r_c^2$ into (2.41) yields

$$\mathbf{v}_R = \frac{c}{q} \frac{mv_{\parallel}^2}{B^2} \frac{\mathbf{r}_c \times \mathbf{B}}{r_c^2} . \quad (2.44)$$

The curvature drift depends on the charge, mass, and speed of the particle.

In a real field, a field gradient would exist in addition to the curvature. Ampere's law (2.7) in a vacuum without electric current gives $\nabla \times \mathbf{B} = 0$. In cylindrical coordinates (cf. Sect. A.2.3 and Fig. 2.2), \mathbf{B} has only one component in θ , ∇B only one in r , and $\nabla \times \mathbf{B}$ only one in z , given as

$$(\nabla \times \mathbf{B})_z = \frac{1}{r} \frac{\partial(rB_{\theta})}{\partial r} = 0 . \quad (2.45)$$

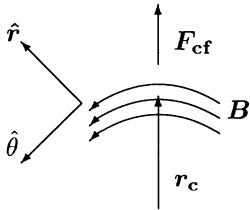


Fig. 2.2. Curved magnetic field and coordinate system

This is zero because B_θ is proportional to $1/r$. Then $|B|$ is proportional to $1/r_c$ and it is $\nabla|B|/|B| = -r_c/r_c^2$. The gradient drift (2.43) then reads

$$\mathbf{v}_{\nabla B} = \pm \frac{c}{2} \frac{m}{q} v_\perp^2 \frac{\mathbf{r}_c \times \mathbf{B}}{r_c^2 B^2}. \quad (2.46)$$

Combined with (2.44) the drift in a curved magnetic field is

$$\mathbf{v}_R + \mathbf{v}_{\nabla B} = \frac{cm}{q} \frac{\mathbf{r}_c \times \mathbf{B}}{r_c^2 B^2} \left(v_\parallel^2 + \frac{1}{2} v_\perp^2 \right). \quad (2.47)$$

This equation has an important implication: if particles are trapped inside a magnetic field torus, the best and finest adjustments in temperature and field cannot prevent the particles from drifting across the field lines and out of the torus sooner or later.

2.3.6 Drifts Combined with Changes in Particle Energy

The drifts discussed so far have been associated with acceleration in the sense of a change in the direction of motion but not in average speed. Certain combinations of fields, however, can lead to changes in average speed and therefore also in particle energy. Figure 2.3 shows the drift of a particle in a crossed electric (pointing upward in the drawing plane) and magnetic (pointing upward out of the drawing plane) field with a gradient perpendicular to both the magnetic and electric fields. The motion under the influence of two combined fields has been discussed above and is indicated by thin lines: for a particle with positive charge the $\mathbf{E} \times \mathbf{B}$ drift leads to horizontal motion towards the right while the $\text{grad}B$ drift leads to upward motion. A combination of both motions gives the $\mathbf{E}\text{grad}B$ drift oblique to the fields and to the magnetic field gradient. Owing to this drift, the particle moves from one potential to another, gaining energy. Note that the use of a gravitational field instead of the electric field would yield similar results.

2.3.7 Drift Currents in Plasmas

Some of the drifts discussed above lead to the separation of positive and negative charges, giving rise to a drift current. In a plasma, instead of single particles we have to consider k different particle species with number densities

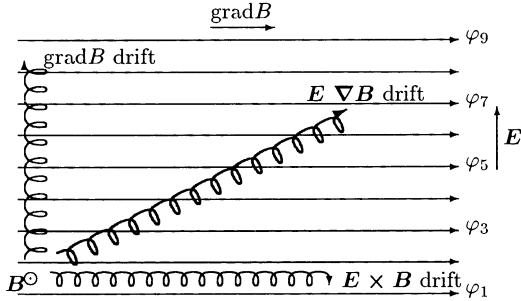


Fig. 2.3. One example of a configuration of fields in which the drift of a charged particle leads to changes in particle energy

n_k and masses m_k . With the general equation for the drift velocity (2.41) we can derive a drift current:

$$\mathbf{j}_D = c \frac{\sum_k n_k \cdot \mathbf{F}_k}{B^2} \times \mathbf{B} = \sum_k m_k q \cdot \mathbf{v}_{Dk} . \quad (2.48)$$

The drift current due to the gradient drift then is given as

$$\mathbf{j}_{D\nabla B} = c \sum_k \frac{m_k v_{k\perp}^2}{2B^2} (\mathbf{B} \times \nabla B) . \quad (2.49)$$

This current results from the inhomogeneity of the field and leads to a charge separation. An example of a natural drift current is the ring current in the magnetosphere (cf. Sect. 10.3.5). The centrifugal force leads to the drift current

$$\mathbf{j}_{Dc} = c \sum_k \frac{n_k m_k v_{k\parallel}^2}{B^4} \mathbf{B} \times (\mathbf{B} \nabla) \mathbf{B} . \quad (2.50)$$

2.4 Adiabatic Invariants

So far we have considered only temporally and spatially constant fields. In weakly and slowly varying fields, the concept of adiabatic invariants provides a powerful tool to describe the periodic motions of particles. A very simple analogy is a mathematical pendulum: if its length increases only weakly during one swing, the ratio of the pendulum's energy and frequency is a constant of the motion, called an adiabatic invariant. For charged particles in a magnetic field, such as in the radiation belts of a magnetosphere, an adiabatic invariant is associated with each of the following three types of motion, cf. Fig. 2.4:

1. the field changes only slowly during a gyration:

$$\frac{1}{B} \frac{\partial B}{\partial t} \ll \frac{\omega_c}{2\pi} ; \quad (2.51)$$

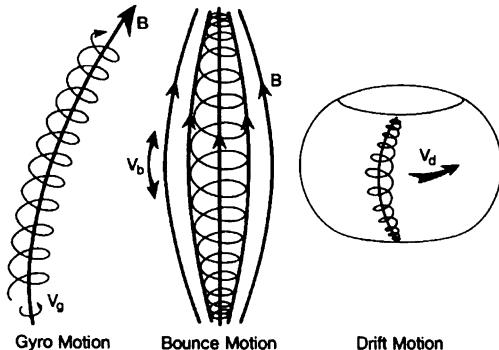


Fig. 2.4. Adiabatic invariants in the magnetosphere. Reprinted from M.G. Kivelson [2.6], in *Introduction to space physics* (eds. M.G. Kivelson and C.T. Russell), Copyright 1995, with kind permission from Cambridge University Press

2. the field varies only weakly on a scale comparable with the distance travelled along the field by the particle during one gyration (bounce motion, longitudinal oscillation):

$$\frac{\nabla B_{\parallel}}{B} \ll \frac{\omega_c}{2\pi v_{\parallel}} ; \quad (2.52)$$

3. the field varies only weakly in the area encircled by the particle during the drift motion:

$$\frac{\nabla B_{\perp}}{B} \ll \frac{\omega_c}{2\pi v_{\perp}} . \quad (2.53)$$

If the particle motion is described by a pair of variables (p_i, q_i) which are generalized momentums and coordinates, for each periodic coordinate q_i the action integral J_i ,

$$J_i = \oint p_i dq_i , \quad (2.54)$$

integrated over a complete cycle of motion, is an invariant or constant of the motion, provided changes in the variables occur slowly compared with the relevant periods of the system (for a proof of this statement, see, for example, [2.7, 2.8]). Thus a system can change from one state of motion into another and still have the same action integral.

2.4.1 First Adiabatic Invariant: The Magnetic Moment

The first adiabatic invariant states that in a slowly varying magnetic field the magnetic moment $\mu = E_{\perp}/B$ is constant. It finds applications in magnetic mirrors and bottles. We can derive it by inserting the momentum mv_{\perp} as p and the coordinate $q = r_L \psi$ with ψ as the azimuthal angle along the gyro-orbit into the action integral (2.54)

$$J_1 = \oint p dq = \oint mv_{\perp} r_L d\psi = 2\pi m r_L v_{\perp} . \quad (2.55)$$

With (2.31) and (2.34) this yields

$$J_1 = 4\pi c \frac{m}{|q|} \mu = \text{const} \quad (2.56)$$

under the tacit assumption that $\omega/\omega_c \ll 1$ with ω characterizing the change in \mathbf{B} . Thus (2.56) states: as long as $m/|q|$ is constant, the magnetic moment is constant, too.

A less formal but more illustrative proof, without using the action integral, starts from the motion of a particle in a magnetic field which slowly varies in space or time. Let us choose the latter attempt, i.e. $\partial\mathbf{B}/\partial t \neq 0$. Then, according to Faraday's law (2.4), a rotational electric field arises which in principle can lead to particle acceleration. If we further assume $v_{||}$ to be zero, the particle speed is $v = v_{\perp} = dl/dt$ with dl being a small element of the particle path. Multiplying the equation of motion (2.22) by the particle speed v_{\perp} , we derive the temporal change in kinetic energy:

$$\frac{d}{dt} \left(\frac{mv^2}{2} \right) = q\mathbf{E} \cdot \mathbf{v} = q\mathbf{E} \frac{dl}{dt}. \quad (2.57)$$

Integration over a gyro-orbit yields for the energy change during a single gyration

$$\delta E_{\text{kin}} = \int_0^{2\pi/\omega} q\mathbf{E} \frac{dl}{dt} dt. \quad (2.58)$$

According to our assumptions, the field changes only weakly during one gyration. The particle orbit therefore still is nearly circular and the integration in time can be substituted by a line integral along the particle path:

$$\delta E_{\text{kin}} = \oint q\mathbf{E} dl. \quad (2.59)$$

With Stokes' theorem (A.29) and Faraday's law (2.4) this gives

$$\delta E_{\text{kin}} = q \int_S (\nabla \times \mathbf{E}) \cdot d\mathbf{S} = -\frac{q}{c} \int_S \frac{\partial \mathbf{B}}{\partial t} d\mathbf{S}, \quad (2.60)$$

with S being the surface enclosed by the Larmor orbit, its direction given by the right-hand rule. The integral is positive for negatively charged particles and negative for positively charged ones. It then can be written as

$$\delta E_{\text{kin}} = \pm \frac{q}{c} \frac{\partial \mathbf{B}}{\partial t} \pi r_L^2. \quad (2.61)$$

With the Larmor radius expressed in terms of the cyclotron frequency this is

$$\delta E_{\text{kin}} = \pm \pi qc \frac{\partial \mathbf{B}}{\partial t} \frac{v^2}{\omega_c} \frac{m}{\pm qB} = \frac{E_{\text{kin}}}{B} \cdot \frac{2\pi \partial \mathbf{B} / \partial t}{\omega_c}. \quad (2.62)$$

The first part of the right-hand side gives the magnetic moment, the second part the variation $\delta\mathbf{B}$ of the magnetic field during one gyration: $\delta E_{\text{kin}} = \mu \delta B$.

Inserting the definition of the magnetic moment into the left-hand side, we get $\delta(\mu B) = \mu\delta B + B\delta\mu = \mu\delta B$, which implies

$$B\delta\mu = 0. \quad (2.63)$$

Because $B \neq 0$, this equation gives the first adiabatic invariant: in a slowly varying magnetic field the magnetic moment is constant, too.

Changes in the magnetic field result in variations in the Larmor radius. Thus the question arises: does the magnetic flux through a Larmor orbit change in a slowly varying field? With the definition of the magnetic flux (2.5) and the area S enclosed by the gyro-orbit we can write

$$\Phi = \int \mathbf{B} \cdot d\mathbf{o} = \pi B \frac{v_{\perp}^2}{\omega_c^2}. \quad (2.64)$$

Inserting (2.27) yields

$$\Phi = \frac{2\pi m}{q^2} \mu. \quad (2.65)$$

Thus as long as μ is an invariant of the motion, the magnetic flux Φ through a gyro-orbit is constant.

2.4.2 Magnetic Mirrors and Bottles

Magnetic mirrors and bottles are applications of the first adiabatic invariant. In its top left panel Fig. 2.5 shows the configuration of a magnetic mirror together with the path of a particle. The panel below shows the dependence of the components v_{\parallel} and v_{\perp} of the particle speed and the pitch angle α on the particle's location inside the bottle. The panel on the right shows details of the particle's orbit at the mirror point, in particular the restoring force.

Since the electric field is zero, the particle's kinetic energy is constant: $\frac{1}{2}mv^2 = \frac{1}{2}m(v_{\parallel}^2 + v_{\perp}^2) = \text{const}$. The kinetic energy perpendicular to the magnetic field can be expressed by the magnetic moment μ :

$$\frac{1}{2}mv^2 = \frac{1}{2}mv_{\parallel}^2 + \mu B. \quad (2.66)$$

The first term on the right-hand side can be interpreted as the drift energy or the kinetic energy of the guiding center, while the second term μB is the energy of the gyration. Since μ is invariant, an increase in B has to be compensated for by a decrease in the drift energy until v_{\parallel} becomes zero. At this mirror point the energy conservation yields

$$\mu B_{\text{mp}} = \frac{1}{2}mv^2 = \frac{1}{2}mv_{\perp\text{mp}}^2. \quad (2.67)$$

Thus at the mirror point the drift energy is entirely transformed into the energy of the gyration, which therefore is the particle's kinetic energy. But what happens at the mirror point? The guiding center has come to a standstill. But it will not stay there, instead it will be reflected back towards the diverging field. The right panel in Fig. 2.5 illustrates the origin of the restoring force:

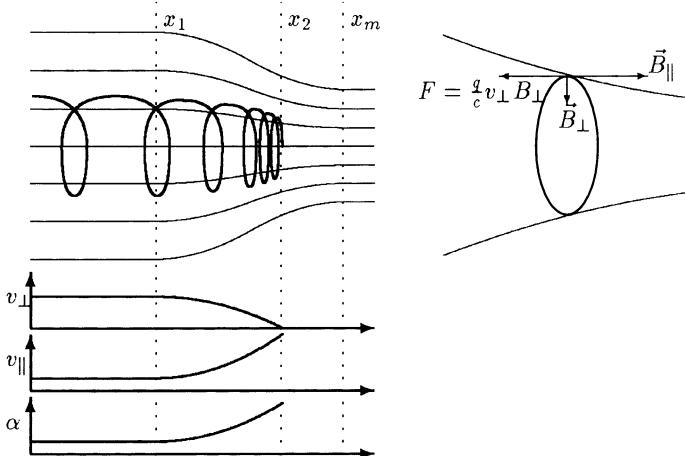


Fig. 2.5. Magnetic mirror: configuration (top left), variations in particle speed parallel and perpendicular to the magnetic field and pitch angle α with position (bottom left), and restoring force at the mirror point (right)

at the mirror point the magnetic field is inhomogeneous. Thus the plane of gyration is not perpendicular to the field and \mathbf{B} has a component \mathbf{B}_\perp in this plane, leading to the restoring force

$$\mathbf{F}_r = q \frac{\mathbf{v} \times \mathbf{B}}{c} = q \frac{v_\perp B_\perp}{c} \mathbf{e}_x . \quad (2.68)$$

The particle is pushed back into regions of decreasing field strength and its pitch angle decreases as energy from the gyration is transformed back into the drift energy.

The location of the mirror point B_{mp} depends on the initial pitch angle α_1 of the particle. If α_1 is zero, the magnetic moment is zero, too, and the particle's total kinetic energy is the drift energy. An increase in the magnetic field strength does not transform the energy of gyration into the drift energy and the particle traverses the magnetic mirror. If, on the other hand, α_1 is 90° , all the particle's energy is contained in the gyration and the guiding center already is at a standstill. For values of α_1 in between these two extremes either reflection occurs at a point $x_{mp}(\alpha_1)$ or the particle is transmitted if the increase in the magnetic field strength is not strong enough to convert all the drift energy. Let us now determine whether a particle will be reflected back or transmitted by the magnetic mirror. The constancy of the magnetic moment also means that the ratio of the energy of gyration and the magnetic field strength is constant. Thus for any two points in the magnetic field we have

$$\mu = \frac{mv_{\perp 1}^2}{2B_1} = \frac{mv_{\perp 2}^2}{2B_2} \quad (2.69)$$

or, taking the pitch angle into consideration,

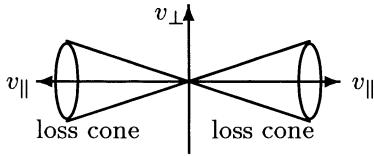


Fig. 2.6. Definition of the loss cone

$$\frac{v_1^2 \sin^2 \alpha_1}{B_1} = \frac{v_2^2 \sin^2 \alpha_2}{B_2} . \quad (2.70)$$

The kinetic energy is constant; thus it is $v_1 = v_2$ and

$$\frac{2\mu}{v^2} = \frac{\sin^2 \alpha_1}{B_1} = \frac{\sin^2 \alpha_2}{B_2} \quad (2.71)$$

becomes an invariant of the motion. At the mirror point, the particle's pitch angle α_{mp} is 90° . Thus to be reflected at the position B_m with maximal magnetic field strength, the initial pitch angle at B_1 has to be

$$\sin^2 \alpha_1^r = \frac{B_1}{B_m} = \frac{1}{R_m} . \quad (2.72)$$

Here R_m is the mirror ratio. Particles with an initial pitch angle α_1^r are reflected exactly at B_m ; particles with larger α_1 are reflected earlier and particles with smaller α_1 pass through the mirror point. Thus (2.72) defines the boundary of a region in velocity space in the shape of a cone, called the loss cone (cf. Fig. 2.6): particles inside this cone are not confined by the magnetic mirror. Note that the loss cone depends on pitch angle only, and does not depend on other particle parameters, such as mass, speed or charge.

2.4.3 Second Adiabatic Invariant: Longitudinal Invariant

Two magnetic mirrors can be combined to form a magnetic bottle (cf. Fig. 2.7): particles are confined inside this bottle due to repeated reflection between the two mirrors. The second adiabatic invariant is related to the drift motion inside such a bottle and the distance between the mirrors. The longitudinal invariant can be derived from (2.54) with the momentum $p = mv_\parallel$ as the generalized momentum and the distance s along the field as the spatial coordinate:

$$J_2 = \int_{s_1}^{s_2} mv_\parallel ds \text{ or briefly } \int v_\parallel ds = \text{const} . \quad (2.73)$$

With (2.66) this can be written as

$$J_2 = \int_{s_1}^{s_2} m \sqrt{v^2 - \frac{2\mu B}{m}} ds . \quad (2.74)$$

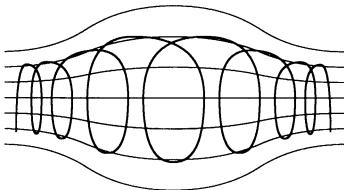


Fig. 2.7. Magnetic bottle

The second adiabatic invariant leads to an efficient mechanism for particle acceleration: for the acceleration of galactic cosmic rays, Fermi proposed a configuration of two converging magnetic mirrors (first-order Fermi effect [2.4, 2.5]). In terms of energy, the reflection of a particle at a fixed magnetic mirror is equivalent to a ball bouncing off a wall. Thus in a magnetic bottle with two fixed mirrors the particle oscillates between these mirrors without changes in total energy. The interaction with a moving magnetic mirror, however, finds its mechanical equivalent in the reflection off a moving wall. Depending on the relative speeds of the particle and the mirror, an energy gain or loss results: head-on collisions lead to an energy gain; if the particle and mirror move in the same direction, an energy loss results. According to (2.73) in each pair of collisions the total energy gain is determined by the shortening of the distance between the mirrors, independent of whether the particle meets both mirrors head-on or only one of them. A familiar equivalent of this adiabatic invariant is the warming of a gas during compression, e.g. inside a tyre pump.

The application of the second adiabatic invariant in the Earth's magnetosphere is not concerned with acceleration but with the asymmetry of the field. The dipole field in the inner magnetosphere forms magnetic bottles: particles moving from the equatorial regions towards the poles "see" a convergent magnetic field, they move into a magnetic mirror (cf. Sect. 10.4.1), at least as long as their pitch angle is sufficiently large to prevent them from entering the loss cone. Particles then are reflected back, cross the equator and travel towards the other pole. This bouncing motion, however, is not exactly the same as depicted in Fig. 2.7: the field lines are curved to follow the Earth's dipole field and the particle therefore faces a stronger magnetic field during that part of the Larmor orbit closest to Earth than in the other half of the orbit. The resulting gradient drift leads to a guiding center motion around the Earth, cf. Fig. 2.4. Electrons drift from west to east, protons from east to west, creating a ring current.

In a symmetric magnetic field, the particles should return to a certain field line after each drift period. But, as we will see in Chap. 10, the Earth's magnetosphere is neither symmetric nor constant in time. Why then should a particle return to a certain field line after one drift period? The particle's energy is conserved during the motion. The first adiabatic invariant then requires that at the mirror point $|B|$ be constant. If a particle has drifted back to a certain longitude, in principle it can be on a field line in an altitude

different from its initial one. This is ruled out by the second adiabatic invariant, which determines the length of the field line between the mirror points. Combination of both invariants requires finding field lines with the same $|\mathbf{B}|$ at the reflection point and the same length. But at a fixed longitude this is fulfilled by one field line only. Thus even in an asymmetric field the particle returns to its original field line after each drift period.

2.4.4 Third Adiabatic Invariant: Flux Invariant

The third adiabatic invariant, the flux invariant, is also related to the guiding center drift. It states that the magnetic flux enclosed by the drift orbit is constant. The particle moves on a surface which adjusts itself to variations in B so that the flux enclosed by this surface stays constant. Formally, the third adiabatic invariant can be derived from the action integral (2.54) by using the drift velocity v_D as the generalized momentum and ψ as the azimuthal angle of the particles orbit with radius r :

$$J_3 = \oint mv_D r d\psi . \quad (2.75)$$

Similar to (2.55) we then obtain

$$J_3 = \frac{4mc}{|q|} \cdot M = \text{const} , \quad (2.76)$$

with M being the magnetic moment of the axisymmetric field.

2.5 Summary

This chapter has introduced the motion of individual charged particles in electromagnetic fields. The particles are considered as test particles: they do not interact with other particles and their motion does not influence the fields. The main results are: (i) The elementary motion of a charged particle in a homogeneous magnetic field is the gyration around the field line, characterized by the Larmor radius (2.31) and the cyclotron frequency (2.27). If the particle has a velocity component parallel to the field, a helical orbit results. (ii) The pitch angle α describes the relation between the particle's motion parallel and perpendicular to the field. (iii) The motion of a charged particle can be separated into the motion of its guiding center and the gyration around the guiding center. The guiding center motion can consist of a field parallel motion and drifts. (iv) In slowly and weakly varying fields, three adiabatic invariants can be defined, each associated with a typical mode of motion: the first adiabatic invariant (constancy of the magnetic moment) is associated with the gyration, the second (longitudinal invariant) with the field parallel motion of the guiding center, and the third (flux invariant) with the drift of the guiding center. The adiabatic invariants can be applied to the motion of particles in the radiation belts.

Exercises and Problems

2.1 A particle gyrates in a homogeneous magnetic field. (a) Determine the size of a volume V which contains an amount of magnetic energy equal to the particle's kinetic energy. (b) Determine the height of a cylinder with this volume and a base given by the Larmor orbit. (c) Discuss this result.

2.2 In the equatorial plane, the Earth's magnetic field can be described as $B = B_o(R_E/r)^3$ with $B_o = 0.3$ G, R_E being the Earth's radius, and r being the geocentric distance. Determine the time a particle with pitch angle 90° needs to drift around the Earth in the equatorial plane. What is the meaning of this time? Determine the period for electrons and protons with an energy of 1 keV drifting in a height of $5r_E$ above the center of the Earth. Compare with the drift due to the gravitational field and the period of an uncharged particle (e.g. a satellite) in the same orbit.

2.3 A proton of cosmic radiation is trapped between two magnetic mirrors with $R_m = 5$. Initially, it has an energy of 1 keV and $v_\perp = v_\parallel$ in the meridional plane between the two mirrors. Each mirror moves with $v_m = 10$ km/s towards the other. Draw a sketch of the configuration. Determine the acceleration of the proton. (a) Does the acceleration continue until the mirrors are in contact with each other or does the particle escape? Determine the maximum energy acquired by the particle. Determine the maximum energy for other pitch angles, too. (b) How long does the particle need to acquire maximum energy? (Hint: assume the mirrors to be planes moving with speed v_m and show that the energy gain in each interaction is $2v_m$. How many interactions are required for the particle to reach maximum speed?)

2.4 The magnetic field of a magnetic mirror varies as $B_z = B_o(1+\alpha^2z^2)$ along the axis. (a) At $z = 0$ an electron has a speed of $v^2 = 3v_\parallel^2 = 1.5v_\perp^2$. Where does reflection occur? (b) Determine the motion of the guiding center. (c) Show that the motion is sinusoidal. Determine the frequency. (d) Determine the longitudinal invariant belonging to this motion.

2.5 A particle of mass m and charge q is at rest in a uniform magnetic field \mathbf{B} . At time $t = 0$, a uniform electric field perpendicular to \mathbf{B} is switched on. Show that the maximum energy gain is $2m(E/B)^2$.

2.6 Describe the concept of the guiding center. What is the reason for drifts?

2.7 Derive (2.41) and discuss the conditions under which it can be applied.

2.8 Derive an expression for the gyro-radius and frequency of a relativistic particle. The relativistic momentum is $p = \gamma m_o v$, where $\gamma = \sqrt{1 - v^2/c^2}$ and m_o is the rest mass. What is the expression for the magnetic moment of a relativistic particle? Show that the relativistic magnetic moment is conserved.

2.9 Show that $\mathbf{j}' = \mathbf{j}$ in the derivation of Ohm's generalized law (2.14).

2.10 Derive Coulomb's law from Eq. (2.2).

3. Magnetohydrodynamics

'Glorious stirring sight!' murmured Toad, never offering to move. 'The poetry of motion! The *real* way to travel! The *only* way to travel!'

K. Grahame, *The Wind in the Willows*

In the previous chapter we discussed the motion of individual charged particles in prescribed \mathbf{E} - and \mathbf{B} -fields. Magnetohydrodynamics is different for two reasons: (a) it considers many particles instead of just a single particle and (b) the \mathbf{E} - and \mathbf{B} -fields are not prescribed but determined by the positions and motions of the particles. Thus the field equations and the equation of motion have to be solved simultaneously and self-consistently: we are looking for a set of particle trajectories and field patterns such that the particles generate the field patterns as they move along their orbits and the field patterns force the particles to move in exactly these orbits. And all this has to be done in a time-varying situation.

While magnetohydrodynamics describes many useful and important concepts, it is only a simple approach to plasma physics because it describes the plasma as a fluid with all particles having the same speed, the bulk speed. The plasma therefore is considered as a cold plasma; the thermal motion of particles is neglected. In a more advanced description of plasma physics, in the kinetic theory (cf. Chap. 5), the velocity distribution of the particles will also be considered.

This chapter consists of four parts. It starts with an introduction to the basic equations. Subsequently, magnetohydrostatics will be concerned with the energetics of the field and the particles without allowing for the collective motion of the plasma. Concepts such as magnetic pressure and magnetic tension are introduced. Magnetohydrokinematics discusses the reaction of the field to a fluid with a given velocity field. Basic concepts to be introduced are frozen-in fields and the dissipation of fields. An application of these concepts is the merging of magnetic field lines, also called reconnection. In magnetohydrodynamics fields and particles can interact freely. In this chapter, the magnetohydrodynamic dynamo will be discussed; magnetohydrodynamic waves will be treated in the next chapter together with other types of plasma waves.

3.1 Basic MHD Equations

We will start with the one-fluid description of a plasma, i.e. the fluid consists of one particle species only. This is entirely sufficient to introduce the basic concepts. In a real plasma, quasi-neutrality suggests the existence of two fluids with positive and negative charges, respectively. For certain phenomena, such as ion waves, a description in the framework of a two-fluid theory will be required, which is briefly sketched in Sect. 3.1.5.

3.1.1 Momentum Balance

The motion of each individual particle is described by an equation of motion. If only electromagnetic forces act on a single particle, the equation of motion is given by (2.22). In a plasma we would have to solve the equations of motion for all particles, which might be billions inside a volume as small as 1 mm³. Such a task is impossible to archive. Instead, we can treat the plasma as a fluid: we are no longer interested in the motion of individual particles but only in the motion of a fluid element or the fluid as a whole.

Partial and Convective Derivatives. The equation of motion contains a total derivative of the momentum and the external forces acting at each given position on the particle. In a fluid, in principle we can do the same: single out a volume element and follow its path, always calculating the local forces acting on the moving volume. This corresponds to Lagrange's description of particle motion. Here we simply would have to multiply the transport equation by the number density n of the particles and get

$$nm \frac{d\mathbf{u}}{dt} = q \left(\mathbf{E} + \frac{\mathbf{u} \times \mathbf{B}}{c} \right), \quad (3.1)$$

with $\mathbf{u} = \langle \mathbf{v} \rangle$ being the bulk speed or average speed of the particles.

For practical purposes, we normally consider a volume fixed in space and measure the properties of the fluid streaming through the volume. A property ε of the fluid then is given as $\varepsilon = \varepsilon(x, y, z, t)$ with the spatial coordinates and the time being independent variables. This would be Euler's description of a fluid. In contrast, in Lagrange's description, the spatial coordinate depends on time, too, and a property ε of the fluid is given as $\varepsilon = \varepsilon(x(t), y(t), z(t), t)$. In the atmosphere, Euler's description could be applied to a stationary thermometer while Lagrange's description could be applied to a thermometer on a radio-sonde carried by the prevalent winds.

Let us now consider the derivative of the property ε . In Euler's description, the total derivative $d\varepsilon/dt$ and the partial derivative $\partial\varepsilon/\partial t$ are equal because all temporal derivatives of the spatial coordinates vanish. In Lagrange's description, the chain rule has to be applied:

$$\frac{d\varepsilon}{dt} = \frac{dx}{dt} \frac{\partial\varepsilon}{\partial x} + \frac{dy}{dt} \frac{\partial\varepsilon}{\partial y} + \frac{dz}{dt} \frac{\partial\varepsilon}{\partial z} + \frac{\partial\varepsilon}{\partial t} = (\mathbf{u} \cdot \nabla) \varepsilon + \frac{\partial\varepsilon}{\partial t}. \quad (3.2)$$

The change of a property ε in a moving fluid element therefore consists of two parts: (a) a change in ε at a fixed position in space (second term on the right-hand side) and (b) the relative motion between observer and medium (first term). Or, more formally: the total temporal derivative consists of a local temporal derivative and advection; it is also called convective derivative. Note that the product $(\mathbf{u} \cdot \nabla)$ is a scalar differential operator.

To understand the difference between a convective and a local temporal derivative let us take a look at the property of water as, for example, salinity or temperature. Let us first consider a closed volume, e.g. a fish-pond. The temperature then might change due to absorbed solar radiation, and the salinity might change due to evaporation. These changes are local temporal changes. Now think of this volume of water as a segment of a river. The local changes are still the same but there are also changes due to the advection of water from other sites: warmer water might be advected into the volume from a power station upstream or the salinity might increase as the incoming tide carries water with higher salinity into the volume.

Electromagnetic Forces. Taking into consideration the convective derivative, the equation of motion (3.1) can be written as

$$mn \left[\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} \right] = qn \left(\mathbf{E} + \frac{\mathbf{u} \times \mathbf{B}}{c} \right). \quad (3.3)$$

Because n is a number density, (3.3) can be written as

$$\mathbf{f}_{\text{elmag}} = \varrho \frac{d\mathbf{u}}{dt} = \varrho \left[\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} \right] = \varrho_c \mathbf{E} + \frac{\mathbf{j} \times \mathbf{B}}{c}, \quad (3.4)$$

with $\varrho = mn$ the density, ϱ_c the charge density, and $\mathbf{j} = nq\mathbf{u}$ the current density. Equation (3.4) gives the force density of the electromagnetic field. In a plasma, the electric field vanishes because the free-moving charges immediately rearrange to cancel out the field. The force density then reduces to

$$\mathbf{f}_{\text{elmag}} = \frac{\mathbf{j} \times \mathbf{B}}{c}. \quad (3.5)$$

Pressure–Gradient Force. Regions of different pressure in a gas exert forces: particles move from the high pressure towards the low pressure. This force is proportional to the pressure gradient $-\nabla p$ and is called the pressure-gradient force. Here we give its derivation, closely following Chen [3.3].

Pressure is related to the thermal motion of particles. The pressure-gradient force leads to a transport in momentum resulting from the motion of particles in and out of a fluid element $\Delta x \Delta y \Delta z$ at position x_o , cf. Fig. 3.1. If the motion is limited to the random thermal motion along the x -axis, particles enter and leave the volume through surfaces A and B only. During a time interval $\delta n = \Delta n_v \cdot v_x \Delta y \Delta z$ particles with speed v_x pass through A into the positive x -direction. Here $\Delta n_v = \Delta v_x \int \int f(v_x, v_y, v_z) dv_y dv_z$ is the number density of particles with speed v_x , with f being the distribution function of the plasma as discussed in Chap. 5.

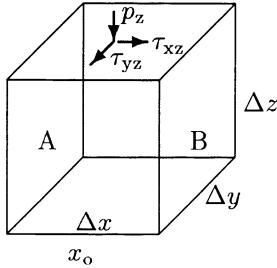


Fig. 3.1. Normal forces p and shear stresses τ acting on a cubic volume

Each particle carries a momentum mv_x . The total momentum P_A^+ transported through A into the positive x -direction then is

$$P_A^+ = \sum \Delta n_v mv_x^2 \Delta y \Delta z = \Delta y \Delta z \left[\frac{1}{2} m \langle v_x^2 \rangle n \right]_{x_o - \Delta x}. \quad (3.6)$$

Here the sum over Δn_v is expressed by the average $\langle v_x^2 \rangle$ of the distribution. The factor 1/2 is due to the fact that only half of the particles in the volume element at $x_o - \Delta x$ will have a speed in the positive x -direction and therefore will transport momentum through A into the volume element at x_o . But particles inside this latter volume element also have a momentum in the positive x -direction which is carried out of the volume through the surface B. Their number is given as

$$P_B^+ = \Delta y \Delta z \left[\frac{1}{2} m \langle v_x^2 \rangle n \right]_{x_o}. \quad (3.7)$$

The net gain of the positive x -momentum therefore is

$$P_A^+ - P_B^+ = \Delta y \Delta z \frac{1}{2} m (-\Delta x) \frac{\partial n \langle v_x^2 \rangle}{\partial x}. \quad (3.8)$$

Particles moving into the negative x -direction double the momentum gain in (3.8) because the negative x -momentum is transported into the negative x -direction:

$$\frac{\partial}{\partial t} (nmv_x) \Delta x \Delta y \Delta z = -m \frac{\partial}{\partial x} (n \langle v_x^2 \rangle) \Delta x \Delta y \Delta z. \quad (3.9)$$

The particle speed $v_x = u_x + v_{x,\text{th}}$ consists of two parts, the bulk speed u_x of the fluid element with $u_x = \langle v_x \rangle$ and the superimposed thermal speed $v_{x,\text{th}}$ with $\langle v_{x,\text{th}} \rangle = 0$. The latter is described by a one-dimensional Maxwell distribution (cf. Sect. 5.1.2), establishing a relationship between average thermal speed and temperature:

$$\frac{1}{2} m \langle v_{x,\text{th}}^2 \rangle = \frac{1}{2} k_B T. \quad (3.10)$$

With (3.9) we get

$$\frac{\partial}{\partial t} (nmu_x) = -m \frac{\partial}{\partial x} [n(\langle u_x^2 \rangle + 2\langle u_x v_{x,\text{th}} \rangle + \langle v_{x,\text{th}}^2 \rangle)]. \quad (3.11)$$

The last term on the right-hand side can be substituted by (3.10). The term in the middle is zero because u_x is constant and therefore $\langle u_x v_{x,\text{th}} \rangle = u_x \langle v_{x,\text{th}} \rangle = 0$, leading to

$$\frac{\partial}{\partial t}(nm u_x) = -m \frac{\partial}{\partial x} \left(n u_x^2 + \frac{nk_B T}{m} \right). \quad (3.12)$$

The partial differentiation on the right-hand side with $n u_x^2 = n u_x \cdot u_x$ gives

$$mn \frac{\partial u_x}{\partial t} + mu_x \frac{\partial n}{\partial t} = -mu_x \frac{\partial(n u_x)}{\partial x} - mn u_x \frac{\partial u_x}{\partial x} - \frac{\partial(nkT)}{\partial x}. \quad (3.13)$$

The second term on the left-hand side and the first term on the right-hand side cancel (cf. the equation of continuity (3.27)). With the pressure p defined as $p = nk_B/T$ rearrangement of (3.9) leads to

$$mn \left(\frac{\partial u_x}{\partial t} + u_x \frac{\partial u_x}{\partial x} \right) = -\frac{\partial p}{\partial x}. \quad (3.14)$$

Generalization to three dimensions gives the pressure-gradient force

$$\mathbf{f}_{\nabla p} = -\nabla p. \quad (3.15)$$

Stress Tensor. In the generalization of (3.14) we tacitly assumed that the x_i -momentum is transported in the x_i -direction only and that the fluid is isotropic. Such a description is entirely sufficient for an ideal gas or fluid. In a plasma or in a viscous fluid, however, the momentum can be transported from one direction to another and the momentum transport is not necessarily isotropic. Therefore, the scalar property p has to be replaced by a tensor \mathbf{P} and the pressure-gradient force ∇p has to be replaced by $\nabla \mathbf{P}$. \mathbf{P} not only considers the pressure, which is orthogonal to the surface of a volume element, but also shear stresses, which are forces parallel to the element's surface, cf. Fig. 3.1. The stress tensor \mathbf{P} has the dimension of a pressure or an energy density. It is symmetric with six independent components P_{ij} for each point: $P_{ij} = mn v_i v_j$ with i being the direction of the momentum transport and j the component of the momentum involved.

In the simplest case, the particle distribution is an isotropic Maxwellian and the stress tensor \mathbf{P} can be written as

$$\mathbf{P} = \begin{pmatrix} p & 0 & 0 \\ 0 & p & 0 \\ 0 & 0 & p \end{pmatrix}. \quad (3.16)$$

Here $\nabla \mathbf{P}$ equals ∇p . In the presence of a magnetic field, a plasma can have two different temperatures T_{\parallel} and T_{\perp} parallel and perpendicular to the magnetic field, leading to different pressures $p_{\parallel} = nk_B T_{\parallel}$ and $p_{\perp} = nk_B T_{\perp}$. In a coordinate system orientated with its z -axis parallel to B , the stress tensor can be written as

$$\mathbf{P} = \begin{pmatrix} p_{\perp} & 0 & 0 \\ 0 & p_{\perp} & 0 \\ 0 & 0 & p_{\parallel} \end{pmatrix}. \quad (3.17)$$

This tensor still is diagonal and it is also isotropic in a plane perpendicular to \mathbf{B} .

Stress Tensor and Viscosity. The off-diagonal elements of the stress tensor in an ordinary fluid are associated with viscosity. Viscosity results from collisions between particles and tends to make the flow more uniform. Quantitatively, the effect of viscosity is described by a kinematic viscosity coefficient $\nu = v_{\text{th}} \cdot \lambda$ with v_{th} being the thermal speed of the particles and λ the mean free path between particle collisions. Alternatively, a viscosity coefficient $\eta = \nu \rho$ can be used. In a fluid, friction is described by

$$\mathbf{f}_{\text{frict}} = \eta \nabla^2 \mathbf{u} + \frac{1}{3} \eta \nabla (\nabla \times \mathbf{u}) . \quad (3.18)$$

In an incompressible fluid, the second term on the right-hand side vanishes:

$$\mathbf{f}_{\text{frict}} = \eta \nabla^2 \mathbf{u} = \nu \rho \nabla^2 \mathbf{u} . \quad (3.19)$$

This can be interpreted as the collisional part of $\nabla P - \nabla p$. Note that the inclusion of viscosity into the momentum balance has two consequences: (a) in agreement with the irreversible character of the transport process, the transport equation is no longer time-reversible: if $\mathbf{u}(\mathbf{r}, t)$ is a solution of the transport equation, then $\mathbf{u}(\mathbf{r}, -t)$ is not. (b) Viscosity increases the order of the partial differential equation. Therefore, to determine solutions we need more boundary conditions than in the case of a non-viscous fluid.

In a plasma, off-diagonal elements also arise without collisions from gyration only: independent of the parallel component of their velocity, the gyro-orbit brings particles into different parts of the plasma, a process which tends to equalize the fluid speeds in different regions of the plasma. The scale of this ‘collisionless viscosity’ is given by the Larmor radius rather than by the particle mean free path.

Fictitious Forces in Rotating Systems. The forces discussed so far are sufficient to give the equation of motion for a plasma in the laboratory setting. In large-scale natural plasmas, such as the ionosphere or stellar atmospheres, additional forces act: the fictitious forces, i.e. the Coriolis force and the centrifugal force.

For a quantitative description let us consider two frames of reference C and C', with C rotating with an angular speed $\boldsymbol{\Omega}$ with respect to C'. A vector \mathbf{r} fixed in C, in C' moves with a speed $\boldsymbol{\Omega} \times \mathbf{r}$. The temporal derivative of \mathbf{r} in C' therefore is given as

$$\left(\frac{d\mathbf{r}}{dt} \right)_{C'} = \left(\frac{d\mathbf{r}}{dt} \right)_C + \boldsymbol{\Omega} \times \mathbf{r} \quad \text{or} \quad \mathbf{v}' = \mathbf{v} + \boldsymbol{\Omega} \times \mathbf{r} . \quad (3.20)$$

The temporal derivative gives the acceleration in the rotating frame:

$$\mathbf{a}' = \frac{d'\mathbf{v}'}{dt} = \frac{d\mathbf{v}'}{dt} + \boldsymbol{\Omega} \times \mathbf{v}' = \frac{d\mathbf{v}}{dt} + 2\boldsymbol{\Omega} \times \mathbf{v} + \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}) . \quad (3.21)$$

The fictitious forces in a rotating frame of reference therefore are

$$\mathbf{f}_{\text{rot}} = -\rho 2\boldsymbol{\Omega} \times \mathbf{v} - \rho \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}) \quad (3.22)$$

with the first term on the right-hand side describing the Coriolis force and the second term the centrifugal force.

In the near-Earth environment the Coriolis force has to be considered in the atmospheric motion and in the ionospheric and magnetospheric current systems; it is of vital importance in the dynamo process inside the Sun and the planets. The influence of the Coriolis force can be illustrated by its effect on the atmospheric motion. In the northern hemisphere, wind is deflected towards the right. On a global scale, this deflection leads to the break-up of the Headley cell driven by the temperature gradient between the equator and the pole into three separate cells, which determine the global atmospheric circulation and govern the energy transport from equator to pole. The Coriolis force, and therefore the size of the deflection, depends on the wind speed: with increasing speed, the distance traveled by a volume of air during a time interval increases. A longer trajectory also means a larger displacement. The Coriolis force becomes effective only if the scales of the system are large enough. Contrary to popular belief, the eddy at the outflow of a bath-tub is not due to the Coriolis force: its direction depends on residual motions in the water or the motion induced by pulling the plug.

Putting it all Together. Adding these forces gives the momentum balance:

$$\varrho \frac{du}{dt} = -\nabla P + \varrho \mathbf{E} + \frac{\mathbf{j} \times \mathbf{B}}{c} + \varrho \mathbf{g} - 2\boldsymbol{\Omega} \times \mathbf{v} - \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}) . \quad (3.23)$$

If we neglect the electric field and the fictitious forces and split the stress tensor into the pressure-gradient force and friction, (3.23) can be written as

$$\varrho \frac{du}{dt} = \varrho \left[\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} \right] = -\nabla p + \varrho \nu \nabla^2 \mathbf{u} + \frac{\mathbf{j} \times \mathbf{B}}{c} + \varrho \mathbf{g} . \quad (3.24)$$

This equation is the Navier–Stokes equation used in hydrodynamics complemented by the forces of the electromagnetic field.

The momentum balance (3.23) still is incomplete: (a) it does not consider sources and sinks, e.g. due to ionization or recombination, which might involve a net gain or loss of momentum; (b) it does not consider momentum transport due to Coulomb collisions between charged particles; and (c) it does not consider momentum transport arising from the forces exerted by a particle component of opposite charge inside the plasma. The latter will be discussed briefly in the two-fluid description of a plasma, cf. Sect. 3.1.5.

3.1.2 Equation of Continuity

An equation of continuity is concerned with the conservation of a property ε , such as mass or charge. A change in ε inside a volume V can result from the convergence of a flux $C(\varepsilon)$ into or out of the volume or sources and sinks $S(\varepsilon)$ inside the volume. The general form of an equation of continuity is

$$\frac{\partial \varepsilon}{\partial t} + \nabla \cdot \mathbf{C}(\varepsilon) = S(\varepsilon) . \quad (3.25)$$

The most common application is the conservation of mass:

$$\frac{\partial \varrho}{\partial t} = -\nabla(\varrho \mathbf{u}) = -\nabla \mathbf{j} , \quad (3.26)$$

with ϱ being the density and $\mathbf{j} = \varrho \mathbf{u}$ the mass current density. Using (3.2) the conservation of mass can be rewritten as

$$\frac{d\varrho}{dt} = \frac{\partial \varrho}{\partial t} + \mathbf{u} \cdot \nabla \varrho = -\varrho \nabla \cdot \mathbf{u} . \quad (3.27)$$

It states that a change of mass inside a volume is a consequence of the flow of matter into or out of the volume. Local sources and sinks of mass are not considered because, except for elementary particle physics, there are no sources or sinks for matter. Application of Gauss's theorem (A.23) leads to the integral form of the continuity equation:

$$\frac{\partial}{\partial t} \int_V \varrho dV = - \oint_{O(V)} \mathbf{j} d\sigma . \quad (3.28)$$

The equation of continuity for the electric charge is formally analogous:

$$\frac{\partial \varrho_c}{\partial t} + \nabla \cdot (\mathbf{u} \varrho_c) = 0 . \quad (3.29)$$

3.1.3 Equation of State

Finally, we need a relationship connecting the scalar pressure p and the density ϱ : $p = p(\varrho, T)$. The equation of state describes how the temperature changes during the motion or compression of a gas. In case of an isothermal ideal gas, the equation of state can be written as

$$p = C(T) \cdot \varrho , \quad (3.30)$$

with C being a constant proportional to temperature. If the compression is slow compared with thermal conduction (isothermal compression), the pressure increases as a result of the density increase but not because of a temperature increase. In many cases, because particles can freely flow along \mathbf{B} , conduction parallel to \mathbf{B} provides the possibility for the plasma to remain isothermal, especially if the compression is periodic or wave-like along \mathbf{B} .

A fast moving gas might not be able to exchange energy with its environment. The equation of state for such an adiabatic compression is

$$p = C \cdot \varrho^{\gamma_a} , \quad (3.31)$$

with $\gamma_a = c_p/c_v$ being the specific heat ratio or adiabatic exponent. For an ideal gas, γ_a equals $(N+2)/N$, with N being the number of degrees of freedom. For a three-dimensional ideal gas consisting of atoms, γ_a is 5/3.

A third important case arises if adiabatic compression is fast compared with heat conduction and also is anisotropic. Now the degrees of freedom parallel and perpendicular to the field are separated and $T_{||}$ ($N=1$, $\gamma_a=3$)

can be heated more efficiently than T_\perp ($N = 2, \gamma_a = 2$). The adiabatic invariants can be used to derive generalizations of this relationship.

The perpendicular pressure p_\perp can be expressed by the average magnetic moment μ : $p_\perp = \frac{1}{2}\varrho\langle v^2 \rangle = n\langle \mu \rangle B$. If the compression is fast compared with the heat conduction but slow compared with the gyration period, the magnetic moment is conserved, leading to the adiabatic relation

$$\frac{d}{dt} \left(\frac{p_\perp}{nB} \right) = \text{const.} \quad (3.32)$$

Pure perpendicular compression in general is equivalent to an increase in the magnetic field strength. For each area A , the conservation of particles implies $nA = \text{const}$ while conservation of the magnetic flux yields $BA = \text{const}$. Thus n/B is constant, too, and (3.32) reduces to $p = C\varrho^{\gamma_a}$ with γ_a being 2 as expected for two-dimensional adiabatic compression.

The pressure $p_\parallel = \varrho\langle v^2 \rangle$ parallel to the magnetic field is related to the second adiabatic invariant $J_2 \sim v_\parallel L = \text{const}$, with L being the scale length along the magnetic field. If the compression is slow compared with the particle's oscillation along the field line, then J_2 is conserved. Now length L , area A and volume $V = AL$ change. The conservation of particles and the magnetic flux yield $nV = \text{const}$ and $BA = \text{const}$. Thus L can be expressed as $L = V/A \sim B/n$ and we finally get

$$\frac{d}{dt} \left(\frac{p_\parallel B^2}{n^3} \right) = 0. \quad (3.33)$$

Pure parallel compression with $B = \text{const}$ then leads to $p = C\varrho^{\gamma_a}$ with $\gamma_a = 3$ as expected for one-dimensional compression.

The two adiabatic relations (3.32) and (3.33) are called the “double adiabatic” equations of state.

3.1.4 Basic Equations of MHD

Magnetohydrodynamics attempts to describe a plasma as a fluid. Its simplest approach is a one-fluid description, considering one particle species only. This is sufficient to derive the most important phenomena (cf. Sects. 3.2–4.2).

In magnetohydrodynamics some assumptions about the properties of the system are made, namely: (a) The medium can be neither polarized nor magnetized: $\varepsilon = \mu = 0$, cf. Sect. 2.1.1. (b) Flow speeds are small compared with the speed of light: $u/c = \gamma \ll 1$. (c) Speeds of changes in field properties are small compared with the speed of light: $v_{\text{ph}}/c = \beta \ll 1$. (d) Conductivity is high, and thus strong electric fields are immediately canceled and so $E/B = \alpha \ll 1$. (e) The displacement current $\partial E/\partial t$ can be ignored compared with the induction current. MHD is a theory linear in α , β , and γ and ignores all terms of second or higher order in these quantities. MHD considers the conservation laws of fluid mechanics which are concerned with mass, momentum, energy (mechanical and electromagnetic), and magnetic flux. The formal description is then based on the following set of equations:

- Maxwell's equations

$$\nabla \cdot \mathbf{E} = 4\pi\varrho_c , \quad (3.34)$$

$$\nabla \cdot \mathbf{B} = 0 , \quad (3.35)$$

$$\nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} , \quad (3.36)$$

$$\nabla \times \mathbf{B} = \frac{4\pi}{c} \mathbf{j} ; \quad (3.37)$$

- Ohm's law

$$\mathbf{j} = \sigma \left(\mathbf{E} + \frac{\mathbf{u} \times \mathbf{B}}{c} \right) ; \quad (3.38)$$

- equation of continuity

$$\frac{\partial \varrho_c}{\partial t} + \nabla(\mathbf{u} \varrho_c) = 0 ; \quad (3.39)$$

- equation of motion (momentum balance)

$$\varrho \frac{\partial \mathbf{u}}{\partial t} + \varrho(\mathbf{u} \cdot \nabla) \mathbf{u} = -\nabla p + \frac{\mathbf{j} \times \mathbf{B}}{c} + \varrho \mathbf{g} + \varrho \nu \nabla^2 \mathbf{u} ; \quad (3.40)$$

- equation of state

$$\frac{d}{dt} \left(\frac{p}{\varrho^{\gamma_a}} \right) = 0 . \quad (3.41)$$

This set of partial non-linear differential equations can be solved for given boundary conditions. For certain applications only a part of the equations is required, or some equations can be used in a simplified form: in magnetohydrostatics (cf. Sect. 3.2) the left-hand side of the momentum balance vanishes while in magnetohydrokinematics (cf. Sect. 3.3) an external velocity field is prescribed and therefore the momentum balance can be ignored completely.

The momentum balance gives us hints on the kind of motion: in certain slow motions the inertial term $\varrho \dot{\mathbf{u}}$ can be ignored while in weak magnetic fields the Lorentz force can be ignored. The relative strength of these two forces can be estimated by the ratio

$$S = \frac{B^2/8\pi}{\varrho u^2/2} \quad (3.42)$$

between the energy density of the magnetic field and the kinetic energy density. For $S \gg 1$ the magnetic field determines the motion of the particles and the single particle approach of Chap. 2 can be used. For $S \ll 1$ the magnetic field is swept away by the plasma motion, cf. the concept of the frozen-in field in Sect. 3.3.1. S is another expression for the plasma- β giving the ratio between the gas dynamic and the magnetic pressure: $\beta = 8\pi p/B^2$.



formal
→3.2

3.1.5 Two-Fluid Description

So far we have treated the plasma as a fluid consisting of one kind of charged particles only. A real plasma, however, contains electrons, ions and possibly also neutral particles. Each particle component has its own speed, temperature and partial pressure.

MHD Equations in Two-Fluid Theory. Since a plasma is expected to be quasi-neutral, the number of positive and negative charges has to be equal. The charge density is $\varrho_c = n_i q_i + n_e q_e = \varrho_i + \varrho_e$ with n_i and n_e being the number densities of ions and electrons with charges q_i and q_e . The current density is $\mathbf{j} = n_i q_i \mathbf{u}_i + n_e q_e \mathbf{u}_e = \mathbf{j}_i + \mathbf{j}_e$. If we limit ourselves to a two-fluid plasma, we have to deal with an electron and an ion component, the neutral component is ignored. In addition to the assumptions made in the one-fluid description we assume: (a) the fluid is in thermal equilibrium ($T_i = T_e$), and (b) the plasma is quasi-neutral ($\varrho_i = \varrho_e$). The basic equations in two-fluid MHD then are

- Maxwell's equations

$$\nabla \cdot \mathbf{E} = 4\pi(\varrho_i + \varrho_e) , \quad (3.43)$$

$$\nabla \cdot \mathbf{B} = 0 , \quad (3.44)$$

$$\nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} , \quad (3.45)$$

$$\nabla \times \mathbf{B} = -\frac{4\pi}{c} (\mathbf{j}_i + \mathbf{j}_e) + \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} ; \quad (3.46)$$

- Ohm's law

$$\frac{m_e}{e^2 n} \frac{\partial \mathbf{j}}{\partial t} = \mathbf{E} + \frac{\mathbf{u} \times \mathbf{B}}{c} - \frac{\mathbf{j} \times \mathbf{B}}{enc} + \frac{\nabla p_e}{en} - \frac{\mathbf{j}}{\sigma} ; \quad (3.47)$$

- equation of continuity

$$\frac{\partial n_j}{\partial t} + \nabla \cdot (n_j \mathbf{u}_j) = 0 , \quad j = i, e ; \quad (3.48)$$

- momentum balance

$$m_j n_j \frac{d\mathbf{u}_j}{dt} = q_j n_j \left(\mathbf{E} + \frac{\mathbf{u}_j \times \mathbf{B}}{c} \right) - \nabla p_j \pm \beta(\mathbf{u}_i - \mathbf{u}_e) , \quad j = i, e ; \quad (3.49)$$

- equation of state

$$p_j = p_j(\varrho_j, T_j) , \quad j = i, e . \quad (3.50)$$

Compared with the equations in one-fluid MHD we find the following differences: (a) The equations of state, motion and continuity are given for each component separately. (b) The equation of motion contains an additional term coupling the two components to consider momentum transfer arising from Coulomb collisions. The force between the two components depends on their relative speed, therefore $\mathbf{f}_i = -\mathbf{f}_e = \beta(\mathbf{u}_i - \mathbf{u}_e)$. (c) Gauss's law for the electric field contains both charge densities as Ampere's law contains both current densities. (d) Ohm's law has become unrecognizable.

Ohm's Law in Two-Fluid MHD. Following [3.11], the two-fluid form of Ohm's law can be derived from the equation of motion as follows. In (3.49) the momentum balance is given for both components separately. Adding both equations yields

$$n_i m_i \frac{d\mathbf{u}_i}{dt} + n_e m_e \frac{d\mathbf{u}_e}{dt} = e(Zn_i - n_e)\mathbf{E} + \frac{\mathbf{j} \times \mathbf{B}}{c} - \nabla p ; \quad (3.51)$$

with $p = p_i + p_e$. The friction terms, which describe the momentum transport between the two components, cancel each other. The first term on the right-hand side vanishes because the plasma is quasi-neutral ($Zn_i = n_e = n$):

$$n_i m_i \frac{d\mathbf{u}_i}{dt} + n_e m_e \frac{d\mathbf{u}_e}{dt} = \frac{\mathbf{j} \times \mathbf{B}}{c} - \nabla p . \quad (3.52)$$

Let us now multiply the equation of motion for electrons by m_i and that for ions by Zm_e . Subtraction of the new equations yields

$$\begin{aligned} & m_i m_e \left(Zn_i \frac{d\mathbf{u}_i}{dt} - n_e \frac{d\mathbf{u}_e}{dt} \right) \\ &= e(Z^2 n_i m_e + n_e m_i) \mathbf{E} + \frac{e}{c} (Z^2 n_i m_e \mathbf{v}_i + n_e m_i \mathbf{u}_e) \times \mathbf{B} \\ & \quad - Zm_e \nabla p_i + m_i \nabla p_e - \beta \mathbf{u}_i (Zm_e + m_i) + \beta \mathbf{u}_e (Zm_e + m_i) . \end{aligned} \quad (3.53)$$

Since m_e is much smaller than m_i , Zm_e is small compared with m_i . Thermal equilibrium requires $Zp_i = Zn_i kT = n_e kT = p_e$. Thus $Zm_e \nabla p_i = m_e \nabla Zp_i$ is small compared with $m_i \nabla p_e$. Equation (3.53) therefore can be simplified:

$$\begin{aligned} & m_i m_e \left(Zn_i \frac{d\mathbf{u}_i}{dt} - n_e \frac{d\mathbf{u}_e}{dt} \right) \\ &= enm_i \mathbf{E} + \frac{en}{c} (Zm_e \mathbf{u}_i + m_i \mathbf{u}_e) \times \mathbf{B} + m_i \nabla p_e - \beta m_i (\mathbf{u}_i - \mathbf{u}_e) . \end{aligned} \quad (3.54)$$

Using the total current

$$\mathbf{j} = en(\mathbf{u}_i - \mathbf{u}_e) \quad (3.55)$$

the last term on the right-hand side can be rewritten as

$$-\beta m_i (\mathbf{u}_i - \mathbf{u}_e) = -\frac{\beta m_i}{en} \mathbf{j} . \quad (3.56)$$

Multiplication of (3.55) by ϱ_i (ϱ_e) and subtraction (addition) of (3.56) gives

$$\mathbf{u}_i = \mathbf{u} + \frac{\varrho_e}{en\varrho} \mathbf{j} \quad \text{and} \quad \mathbf{u}_e = \mathbf{u} - \frac{\varrho_i}{en\varrho} \mathbf{j} . \quad (3.57)$$

With $\varrho_i \approx \varrho_e$ we get

$$Zm_e \mathbf{u}_i + m_i \mathbf{u}_e = (Zm_e + m_i) \mathbf{u} + \frac{Zm_e \varrho_e - m_i \varrho_i}{en\varrho} \mathbf{j} = m_i \mathbf{u} - \frac{m_i}{en} \mathbf{j} . \quad (3.58)$$

Combined with (3.55) and (3.58), (3.54) gives Ohm's law in two-fluid MHD:

$$\frac{m_e}{en} \left(Z n_i \frac{d\mathbf{u}_i}{dt} - n_e \frac{d\mathbf{u}_e}{dt} \right) = \mathbf{E} + \frac{\mathbf{u} \times \mathbf{B}}{c} - \frac{\mathbf{j} \times \mathbf{B}}{enc} + \frac{\nabla p_e}{en} - \frac{\beta}{e^2 n^2} \mathbf{j} . \quad (3.59)$$

The last term on the right-hand side is linear in \mathbf{j} with a proportionality constant $1/\sigma$. Thus from (3.59) the electric conductivity can be expressed by the friction β between the two plasma components: $\sigma = e^2 n^2 / \beta$.

As the equation of motion, Ohm's law on its right-hand side contains only macroscopic quantities. In both equations, the left-hand side still contains the total derivative. This can be simplified under certain assumptions. One of the assumptions in MHD is that speeds are small compared with the speed of light. Thus terms of second or higher order in \mathbf{u} can be ignored. The left-hand side of (3.54) therefore can be written as

$$\begin{aligned} n_i m_i \frac{d\mathbf{u}_i}{dt} + n_e m_e \frac{d\mathbf{u}_e}{dt} \\ = n_i m_i \frac{\partial \mathbf{u}_i}{\partial t} + n_i m_i (\mathbf{u}_i \cdot \nabla) \mathbf{u}_i + n_e m_e \frac{\partial \mathbf{u}_e}{\partial t} + n_e m_e (\mathbf{u}_e \cdot \nabla) \mathbf{u}_e \\ = n_i m_i \frac{\partial \mathbf{u}_i}{\partial t} + n_e m_e \frac{\partial \mathbf{u}_e}{\partial t} . \end{aligned} \quad (3.60)$$

The equation of continuity (3.48) can be simplified in a similar way:

$$\begin{aligned} n_j m_j \frac{\partial \mathbf{v}_j}{\partial t} &= \frac{\partial n_j m_j \mathbf{v}_j}{\partial t} - \mathbf{v}_j \frac{\partial n_j m_j}{\partial t} \\ &= \frac{\partial n_j m_j \mathbf{v}_j}{\partial t} + \mathbf{v}_j \nabla \cdot (n_j m_j \mathbf{v}_j) = \frac{\partial \varrho_j \mathbf{v}_j}{\partial t} . \end{aligned} \quad (3.61)$$

Here the index j indicates that for electrons and protons such an equation exists. With $\varrho \mathbf{u} = \varrho_e \mathbf{u}_e + \varrho_i \mathbf{u}_i = n_e m_e \mathbf{v}_e + n_i m_i \mathbf{v}_i$ we get

$$n_i m_i \frac{d\mathbf{u}_i}{dt} + n_e m_e \frac{d\mathbf{u}_e}{dt} = \frac{\partial \varrho_i \mathbf{u}_i}{\partial t} + \frac{\partial \varrho_e \mathbf{u}_e}{\partial t} = \frac{\partial \varrho \mathbf{u}}{\partial t} . \quad (3.62)$$

Neglecting terms of second order in \mathbf{u} yields

$$\frac{\partial \varrho \mathbf{u}}{\partial t} = \varrho \frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \frac{\partial \varrho}{\partial t} = \varrho \frac{\partial \mathbf{u}}{\partial t} - \mathbf{u} \cdot \nabla \cdot (\varrho \mathbf{u}) = \varrho \frac{\partial \mathbf{u}}{\partial t} . \quad (3.63)$$

Thus the momentum balance can be written to contain only macroscopic quantities:

$$\varrho \frac{\partial \mathbf{u}}{\partial t} = \frac{\mathbf{j} \times \mathbf{B}}{c} - \nabla p . \quad (3.64)$$

Note that the left-hand side of the momentum balance now only contains the partial derivative because terms of higher order have been ignored.

The left-hand side of Ohm's law (3.59) also can be rewritten with the help of the equation of continuity and the definition of the total current in (3.55):

$$\frac{m_e}{en} \left(Z n_i \frac{d\mathbf{u}_i}{dt} - n_e \frac{d\mathbf{u}_e}{dt} \right)$$

$$\begin{aligned}
&= \frac{m_e}{en} \left(Zn_i \frac{\partial \mathbf{u}_i}{\partial t} - n_e \frac{\partial \mathbf{u}_e}{\partial t} \right) + \frac{m_e}{en} [\mathbf{u}_i \nabla (Zn_i \mathbf{u}_i) - \mathbf{u}_e \nabla (n_e \mathbf{u}_e)] \\
&= \frac{m_e}{e^2 n} \cdot en \left(\frac{\partial \mathbf{u}_i}{\partial t} - \frac{\partial \mathbf{u}_e}{\partial t} \right) = \frac{m_e}{e^2 n} \frac{\partial \mathbf{j}}{\partial t} .
\end{aligned} \tag{3.65}$$

Ohm's law therefore reads

$$\frac{m_e}{e^2 n} \frac{\partial \mathbf{j}}{\partial t} = \mathbf{E} + \frac{\mathbf{u} \times \mathbf{B}}{c} - \frac{\mathbf{j} \times \mathbf{B}}{enc} + \frac{\nabla p_e}{en} - \frac{\mathbf{j}}{\sigma} . \tag{3.66}$$

The left-hand side gives the current acceleration. The first, second and final terms on the right-hand side are expressions already known from Ohm's law in one-fluid MHD. The $\mathbf{j} \times \mathbf{B}$ term is called the Hall term and describes the Hall effect: in a magnetic field the current created by the moving charges is deflected by the Lorentz force resulting in an additional electric field perpendicular to both \mathbf{j} and \mathbf{B} . The fourth term on the right-hand side gives the pressure diffusion: in the presence of a pressure gradient both particle species diffuse with respect to each other creating a current along ∇p .

3.2 Magnetohydrostatics

Magnetohydrostatics deals with the energetics of particles and fields. It does not require the entire set of MHD equations; instead the field equations and the momentum balance are sufficient. In addition, the inertial term in the equation of motion vanishes. We will use the one-fluid description of MHD because it is sufficient to derive and demonstrate the important concepts, namely magnetic pressure and magnetic tension.

In Chap. 2 we derived the energy equation for the electromagnetic field (2.20) considering the fields only. But the electromagnetic field exerts a force on matter, too, described by the $\mathbf{j} \times \mathbf{B}$ -term in the equation of motion. Thus the energy equation has to be supplemented by a term describing the work done by the field:

$$\begin{aligned}
&\frac{\partial}{\partial t} \int_V \left(\frac{B^2}{8\pi} + \frac{E^2}{8\pi} \right) d^3 r \\
&= - \oint_{O(V)} \frac{\mathbf{E} \times \mathbf{B}}{4\pi} dS - \int_V \mathbf{E} \mathbf{j} d^3 r - \int_V \frac{1}{c} \mathbf{u} \cdot (\mathbf{j} \times \mathbf{B}) d^3 r .
\end{aligned} \tag{3.67}$$

The total energy balance inside a volume therefore is determined by the energy flux into or out of the volume (Poynting vector), ohmic losses and the work exerted on the particles by the field.

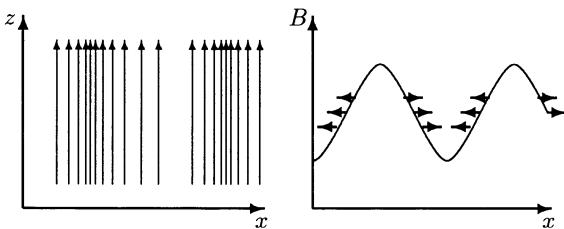


Fig. 3.2. Magnetic pressure: field gradient perpendicular to the field (left) and resulting spatial distribution of the field strength (right) with arrows indicating the direction of the magnetic pressure

3.2.1 Magnetic Pressure

Let us now take a closer look at a magnetic field such as the one shown in Fig. 3.2. The lines of force are parallel to the z -axis but the field strength varies along the x -axis: $\mathbf{B} = (0, 0, B(x))$. The force density exerted by the field is $\mathbf{f} = \mathbf{j} \times \mathbf{B}/c$. The current density can be expressed by Ampere's law (2.7):

$$\mathbf{f} = \frac{\mathbf{j} \times \mathbf{B}}{c} = \frac{1}{4\pi} (\nabla \times \mathbf{B}) \times \mathbf{B}. \quad (3.68)$$

Again, the electric field has been ignored because of the high conductivity. For the field defined above, (3.68) yields

$$\mathbf{f} = \frac{1}{4\pi} \left(-B \frac{\partial B}{\partial x}, 0, 0 \right), \quad (3.69)$$

and thus the force density only has a component along the x -axis which is

$$f_x = -\frac{1}{4\pi} B \frac{\partial B}{\partial x} = -\frac{\partial}{\partial x} \frac{B^2}{8\pi}. \quad (3.70)$$

An inhomogeneity in the magnetic field therefore gives rise to a force density pushing field lines back from regions of high density into low density areas. Such behavior is well known from an isothermal gas where a restoring force $f \sim \nabla p$ tries to cancel out pressure gradients. Therefore, (3.70) can be interpreted as the magnetic pressure:

$$p_M = \frac{B^2}{8\pi}. \quad (3.71)$$

Graphically, this magnetic pressure can be described as the tendency of neighboring field lines to repulse each other. Note that, in contrast to the gas-dynamic pressure, the magnetic pressure is not isotropic but is always perpendicular to the field.

The analogy with gas-dynamic pressure can be pushed even further if we invoke the concept of frozen-in magnetic fields (cf. Sect. 3.3.1). Imagine a magnetic field frozen-into a plasma: each plasma parcel contains a certain amount of magnetic flux which is tied to this plasma element and follows its path as the plasma parcels are shuffled around. Thus a field gradient always has to be combined with a gradient in gas-dynamic pressure. As the plasma attempts to reduce the pressure gradient, the field will be homogenized, too.

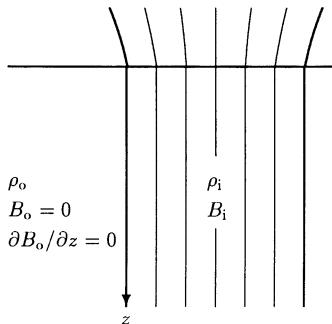


Fig. 3.3. Model of a sunspot: the gas-dynamic pressure from the outside is balanced by the magnetic pressure inside the sunspot

Sunspots. Sunspots are a prime example of the apparently paradoxical behavior of plasmas as well as a good illustration of the concept of magnetic pressure. Sunspots are cool and dark patches on the visible solar disk, their temperatures being 1000 K to 2000 K lower than the ambient photospheric temperature of 5700 K. Despite this temperature gradient, the sunspot does not mix with the ambient plasma. Instead, it is very stable and can survive for many solar rotations. This longevity results from the magnetic pressure: inside the sunspot the magnetic field is about 3000 gauss compared to a few gauss at the outside. The boundaries of a sunspot are sharp in both magnetic field strength and temperature.

Figure 3.3 shows a simple model of a sunspot. The indices ‘i’ and ‘o’ refer to plasma and field properties inside and outside of the sunspot, respectively. To prevent hot photospheric material from streaming into the sunspot, the gas-dynamic pressure p_o of the photospheric plasma has to be balanced by the combined magnetic and gas-dynamic pressure inside the spot: $p_i + B_i^2/(8\pi) = p_o$. Here we have assumed the magnetic field pressure outside the sunspot to be negligible and $\partial B/\partial z$ to be zero, as is suggested by observations. The variation of pressure with height is described by the hydrostatic equation (cf. Problem 3.3): $\partial p/\partial z = \rho g_{\text{Sun}}$. Since B is independent of height, this yields $\partial p_i/\partial z = \partial p_o/\partial z$ and $\rho_i = \rho_o$. On the other hand, the universal gas law yields for the pressure $p = \rho k_B T/m$. Because the densities inside and outside the sunspot are equal, the gas law requires T_i to be smaller than T_o to fulfill $p_i < p_o$. This is the desired result: for longevity, the higher magnetic field inside the sunspot has to be combined with a lower temperature and therefore less electromagnetic emission.

3.2.2 Magnetic Tension

The upper panel of Fig. 3.4 shows a homogeneous magnetic field parallel to the x -axis: $\mathbf{B} = (B_o, 0, 0)$. The field is assumed to be frozen-into a plasma with a velocity field $\mathbf{u} = (0, 0, u_z(x))$ as depicted in the middle panel, leading to the deformation of the field shown in the lower panel. The distorted field



Exam
→ 3.2

can be described as the superimposition of the undisturbed field \mathbf{B}_o and a disturbance $\delta\mathbf{B}$ (cf. (3.102) in Sect. 3.3.4):

$$\mathbf{B} = \mathbf{B}_o + \frac{\partial \mathbf{B}_o}{\partial t} dt = \mathbf{B}_o + \nabla \times (\mathbf{u} \times \mathbf{B}) dt . \quad (3.72)$$

With (3.68) and the field described above, the force is found to act parallel to the disturbing velocity field with a density

$$f_z = \frac{1}{4\pi} \frac{\partial u_z}{\partial x^2} B_o^2 dt . \quad (3.73)$$

The force always is a restoring force, called the magnetic tension: if the field lines have a convex curvature into the upward direction, $\partial^2 u_z / \partial x^2$ is less than zero, leading to a force directed downwards. If the curvature is opposite, the force also is in the opposite direction. The magnetic tension can also be interpreted graphically: magnetic field lines have a tendency to shorten.

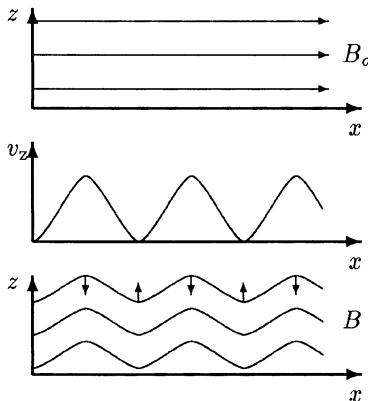


Fig. 3.4. Magnetic tension: the lines of force in a homogeneous magnetic field (*upper panel*) are distorted by a velocity field (*middle panel*), giving rise to a restoring force indicated by the arrows (*lower panel*)

Solar Filaments. Another example of the unusual behaviour of a plasma are solar filaments or protuberances which are cold and dense matter suspended from magnetic arcades high above the photosphere. Such a structure is called a filament as long as it is seen as a dark (because cold) stripe in front of the photosphere. As the Sun rotates, the filament becomes visible as a bright (because dense) arc extending high above the photosphere and is called a protuberance. The spatial structure clearly becomes visible: the filament extends vertically above the photosphere, covering angular distances of some 10 degrees. Filaments are stable and can last for a few solar rotations. Under certain conditions they become unstable and are blown out violently as coronal mass ejections (cf. Sect. 8.6). Typical numbers are: the temperature is about 7000 K compared with 10^6 K of the surrounding corona, the density is about 100 times larger than the ambient density, and a typical filament extends up to 30 Mm, which is about 100 times the scale height in the corona.



Exam ⊙
→ 3.3

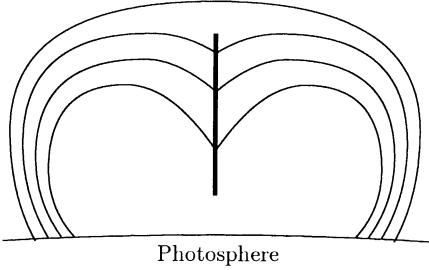


Fig. 3.5. Model of a solar filament (thick vertical line) suspended from magnetic arcades

But let us start with the stability of a filament. The first theoretical description, including the details inside the filament, goes back to Kippenhahn and Schlüter [3.12]. Here we will limit ourselves to a much shorter discussion, considering only the general structure of the filament. Filaments exist only where fields of opposite polarity meet: the filament is aligned along the neutral line. Therefore, the magnetic field appears to play an important role in the existence as well as stability of the field. Figure 3.5 sketches the situation: the filament (thick vertical line) is supported by magnetic arcs connecting opposite polarities in the photosphere. The magnetic field lines do not form perfect arcades, instead they are ditched-in at the position of the filament: gravity pulls down the filament which in turn pulls down the frozen-in magnetic field. The deformation of the magnetic field causes magnetic tension in the opposite direction. Thus the filament is held at a certain height by the equilibrium between gravity and magnetic tension.

We can obtain a more quantitative statement from the basic MHD equations. Here we need the equation of motion in the stationary case,

$$\nabla p = \frac{1}{4\pi} (\nabla \times \mathbf{B}) \times \mathbf{B} + \varrho \mathbf{g}, \quad (3.74)$$

and the equation of state $p = nk_B T$, with T being spatially constant.

Let us define a coordinate system with the xy -plane tangential to the surface of the photosphere and the y -direction extending along the filament into the drawing plane. The z -direction points upward; thus Fig. 3.5 is a cut through the filament in the xz -plane. All quantities are assumed to be independent of y , thus $\partial/\partial y$ is zero. The magnetic field is given as $\mathbf{B} = (B_x, 0, B_z)$ and \mathbf{g} is $(0, 0, -g)$. The double cross product in the equation of motion then can be written as

$$(\nabla \times \mathbf{B}) \times \mathbf{B} = \left(\frac{\partial B_x}{\partial z} - \frac{\partial B_z}{\partial x} \right) (B_z, 0, B_x). \quad (3.75)$$

Combining the equation of motion with the equation of state gives

$$k_B T \frac{\partial n}{\partial x} = \frac{1}{4\pi} B_z \left(\frac{\partial B_x}{\partial z} - \frac{\partial B_z}{\partial x} \right) \quad (3.76)$$

and

$$k_B T \frac{\partial n}{\partial z} = -\frac{1}{4\pi} B_x \left(\frac{\partial B_x}{\partial z} - \frac{\partial B_z}{\partial x} \right) - mng . \quad (3.77)$$

Differentiation of (3.76) to z and of (3.77) to x and subtraction of these equations yields

$$\begin{aligned} 0 &= \left(\frac{\partial B_z}{\partial z} + \frac{\partial B_x}{\partial x} \right) \left(\frac{\partial B_x}{\partial z} - \frac{\partial B_z}{\partial x} \right) \\ &\quad + B_z \left(\frac{\partial^2 B_x}{\partial z \partial x} - \frac{\partial^2 B_z}{\partial x^2} \right) + 4\pi \frac{\partial p}{\partial x} g . \end{aligned} \quad (3.78)$$

Gauss's law for a magnetic field in the two-dimensional case gives

$$\frac{\partial^2 B_z}{\partial x \partial z} = -\frac{\partial^2 B_x}{\partial x^2} \quad \text{and} \quad \frac{\partial^2 B_x}{\partial x \partial z} = -\frac{\partial^2 B_z}{\partial z^2} . \quad (3.79)$$

With $H_p = kT/mg$ being the scale height in the barometric height formula $p = p_0 \exp\{-\int dz/H_p\}$ and substitution of $\partial n/\partial x$ according to (3.76), Eq. (3.78) can be written as

$$B_z \nabla^2 B_x - B_x \nabla^2 B_z + \frac{1}{H_p} B_z \left(\frac{\partial B_x}{\partial z} - \frac{\partial B_z}{\partial x} \right) = 0 . \quad (3.80)$$

From this equation the details of the magnetic field as well as the density inside the filament can be determined and compared with observations. These results confirm the sharp bend in the magnetic field: inside the filament the lines of force are bent according to tanh. Thus there is still a steady field and not a discontinuity. In addition, the model predicts a decrease in density with increasing height, as is evident from the observations. This decrease leads to a flattening of the ditch in the field lines with increasing height, which also is indicated in Fig. 3.5.

3.3 Magnetohydrokinematics

Magnetohydrokinematics deals with the reaction of the electromagnetic field to a prescribed velocity field. The electromagnetic field is assumed not to influence the velocity field. Thus, we do not have to solve the equation of motion. Such a situation corresponds to a large plasma- β or a small value of S in (3.42). The basic equations to derive concepts such as frozen-in fields and the dissipation of fields are Maxwell's equations and Ohm's law.

3.3.1 Frozen-in Magnetic Fields

What happens to an electromagnetic field embedded in a moving medium with high conductivity? Let us assume a magnetic field $\mathbf{B}(\mathbf{r}, t_0)$ at a time t_0 and a prescribed velocity field $\mathbf{u}(\mathbf{r}, t)$. The magnetic flux through a surface

S enclosed by a curve C then is $\Phi = \int \mathbf{B} dS$. Let us follow the motion of C (cf. Fig. 3.6): as C moves, the magnetic flux through S changes because (a) the magnetic field varies in time and (b) the field lines move into or out of S . As C moves, it creates a cylinder with a mantle surface M . All changes in flux through S due to the field lines entering or leaving C is associated with a flux of the very same magnetic field lines through M . Thus the total change in magnetic flux through S can be described by

$$d\Phi = \Phi_2 - \Phi_1 = dt \int_{S_1} \frac{\partial \mathbf{B}}{\partial t} d\mathbf{S}_1 + \int_M \mathbf{B} d\mathbf{S}_M . \quad (3.81)$$

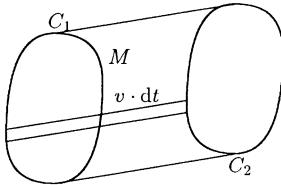


Fig. 3.6. Moving fluid line C to derive the concept of the frozen-in magnetic field

A surface element S_M on M is given as $d\mathbf{S}_M = \mathbf{u} \times d\mathbf{l} \cdot dt$ with $d\mathbf{l}$ being the path along C and $\mathbf{u} \cdot dt$ being the path along the direction of motion of C . Equation (3.81) therefore yields

$$\frac{d\Phi}{dt} = \int_{S_1} \frac{\partial \mathbf{B}}{\partial t} d\mathbf{S}_1 + \int_{C_1} \mathbf{B} \cdot \mathbf{u} \times d\mathbf{l}_1 . \quad (3.82)$$

With Stokes' theorem the last term can be written as

$$\int_{C_1} \mathbf{B} \cdot \mathbf{u} \times d\mathbf{l}_1 = - \int_{C_1} \mathbf{u} \times \mathbf{B} \cdot d\mathbf{l}_1 = - \int_{S_1} \nabla \times (\mathbf{u} \times \mathbf{B}) d\mathbf{S}_1 , \quad (3.83)$$

leading to

$$\frac{d\Phi}{dt} = \int_{S_1} \left[\frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{u} \times \mathbf{B}) \right] d\mathbf{S}_1 . \quad (3.84)$$

The $\mathbf{u} \times \mathbf{B}$ term can be expressed by Ohm's law while the $\partial \mathbf{B} / \partial t$ term can be expressed by Ampere's law:

$$\frac{d\Phi}{dt} = -c \int_{S_1} \nabla \times \mathbf{j} \frac{1}{\sigma} d\mathbf{S}_1 = -c \int_{C_1} \frac{1}{\sigma} \mathbf{j} \cdot d\mathbf{l}_1 . \quad (3.85)$$

The change in magnetic flux through a moving surface therefore is proportional to $1/\sigma$. If σ converges towards infinity, $d\Phi/dt$ converges towards zero: in a medium with infinite conductivity, the magnetic field is frozen-into the plasma and carried away by the matter as if glued to it (cf. left side of

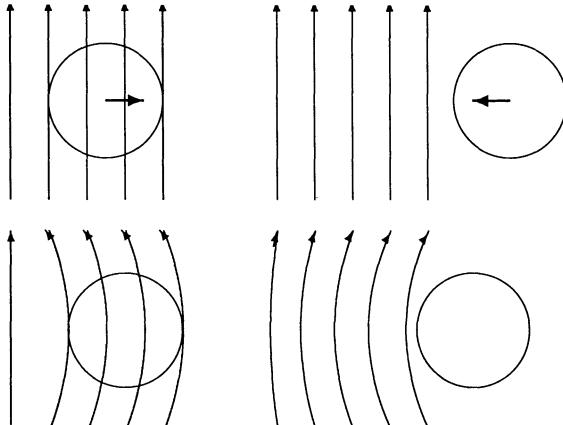


Fig. 3.7. Frozen-in (left) and frozen-out (right) magnetic fields

Fig. 3.7). A prime example of the application of this concept is the interplanetary magnetic field frozen-into the solar wind (cf. Sect. 8.3). A reversal of the concept, the frozen-out field, exists, too. In the right panel of Figure 3.7 a field-free plasma bubble moves towards a region filled with a magnetic field and pushes the field away until its kinetic energy is transferred to additional field energy as evidenced by an increase in magnetic pressure as well as magnetic tension. The field cannot enter into the bubble because in that case the magnetic flux inside the bubble would change. An example is the solar wind frozen-out of the Earth's magnetic field.

3.3.2 Dissipation of Fields

Frozen-in magnetic fields always are connected with an infinite conductivity. But what happens to a magnetic field embedded in a flow with finite conductivity? For simplicity let us assume σ to be spatially and temporally constant. The combination of Faraday's law (3.36) and Ohm's law (3.38) yields

$$\frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{u} \times \mathbf{B}) = -\frac{c}{\sigma} \nabla \times \mathbf{j}. \quad (3.86)$$

The current density can be expressed by Ampere's law (3.37), leading to

$$\frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{u} \times \mathbf{B}) = -\frac{c}{\sigma} \nabla \times \frac{\nabla \times \mathbf{B}}{4\pi/c}. \quad (3.87)$$

The double cross product can be simplified according to (A.13). With $\nabla \cdot \mathbf{B} = 0$ we get

$$\frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{u} \times \mathbf{B}) = -\nabla \times \left(\frac{c^2}{4\pi\sigma} \nabla \times \mathbf{B} \right) = \frac{c^2}{4\pi\sigma} \nabla^2 \mathbf{B}. \quad (3.88)$$

This equation allows us to determine how a given velocity field \mathbf{u} deforms a magnetic field \mathbf{B} .

Let us now consider the case of a vanishing external velocity field. Then the second term on the left-hand side vanishes. Let us align the x -axis of our coordinate system parallel to the magnetic field direction: $\mathbf{B} = (B_x(y, t), 0, 0)$. Equation (3.88) then can be simplified to

$$\frac{\partial B_x}{\partial t} = \frac{c^2}{4\pi\sigma} \frac{\partial^2 B_x}{\partial y^2}. \quad (3.89)$$

Formally, this is equivalent to the one-dimensional heat conduction equation

$$\frac{\partial T}{\partial t} = \chi \frac{\partial^2 T}{\partial y^2}, \quad (3.90)$$

where $T(y)$ is the one-dimensional distribution of temperature and χ is the thermal conductivity. Equation (3.90) gives the temporal change in temperature as the heat is transported away by conduction. Therefore, (3.89) gives the temporal change in magnetic field strength as the magnetic field lines are transported away by a process which depends on conductivity: the field dissolves. Note that while the magnetic flux inside the yz -plane stays constant during this process, the magnetic energy decreases because the field-generating currents are associated with ohmic losses.

If τ is a characteristic time scale for changes in the magnetic field (e.g. the dissipation time during which the field strength decreases to $1/e$ of its original value) and L is the characteristic scale length of the field, the change in B can be estimated from (3.89) to be

$$\frac{B}{\tau} \approx \frac{c^2}{4\pi\sigma} \frac{B}{L^2}. \quad (3.91)$$

Thus the dissipation time

$$\tau \approx \frac{4\pi\sigma}{c^2} L^2 \quad (3.92)$$

depends on the square of the characteristic scale length of the field: smaller fields dissipate faster than larger ones. In addition, the dissipation time increases with increasing conductivity: for infinite conductivity, the dissipation time becomes infinite, too, leading to the frozen-in field.

The Sun, for instance, has a linear dimension of about 7×10^8 m and an average conductivity of 10^{16} /s. This gives a dissipation time for the magnetic field of about 1.2×10^{10} years, nearly three times the age of the Sun. Thus if during its creation the Sun had received a magnetic field, this still would be present today. On the other hand, the solar magnetic field is highly variable on time scales of months to years (cf. Sect. 8.6), making the presence of a fossil field very unlikely. Instead, a MHD dynamo (cf. Sect. 3.4) appears to be responsible for the solar magnetic field.

According to (3.92) the dissipation time depends on the scale of the field. Thus if the field on the left-hand side of Fig. 3.8 is divided into smaller patches of length L/n instead of L , it dissipates n^2 times faster than the original field. Such a redistribution of field lines smearing out the boundary



Fig. 3.8. The dissipation time of the magnetic field decreases as the spatial scales decrease

between regions of opposite polarity and leading to structures on smaller scales can result from turbulent plasma motions. For instance, the stochastic motions in the photospheric and chromospheric network on the Sun might contribute to the dissipation of magnetic fields, in particular in the declining phase of the solar cycle. A vortex in the plasma flow might even create a field-free region inside an otherwise relatively undisturbed field. Numerical simulations of such a vortex first were made by Weiss [3.24].

3.3.3 Reconnection

The dissipation of magnetic field lines is important, e.g. in reconnection, which is assumed to take place in many locations in the solar system, such as solar flares, the tails of magnetospheres, and in the exchange of solar wind and magnetospheric plasma at the day-side magnetopause. Reconnection not only plays an important role in the rearrangement of magnetic fields but also in the formation of shock waves and the acceleration of energetic particles.

The concept of reconnection goes back to Petschek [3.17]. It is widely used in magnetospheric and solar physics, although the physics behind the process is still under debate; sometimes it is even questioned whether reconnection really exists. The basics of reconnection are outlined in Fig. 3.9. Reconnection requires a topology where two magnetic flux tubes of opposite polarity meet (a). According to Ampere's law, in the neutral line between these flux tubes a current flows perpendicular to the drawing plane with a current density

$$j = \frac{c}{4\pi} \frac{\Delta B}{d} . \quad (3.93)$$

The flux tubes are frozen into a plasma with infinite conductivity. As a plasma flow u pushes the flux tubes towards the neutral line, an X-point configuration arises where anti-parallel field lines meet (b). As the distance between the flux tubes decreases, the required current density increases according to (3.93) and may surpass the limit of the current density which the plasma can carry. Then the current becomes unstable, leading to a finite conductivity. Now the frozen-in approximation breaks down and magnetic field diffusion becomes important. Magnetic tension leads to a shortening of the newly merged field lines, pulling them away from the position of the former X-point (c). The energy of the terminated neutral line current is converted to high-speed tangential flows, indicated by v_A . The speed of this plasma flow might exceed the local Alfvén speed, forming two shock waves propagating away from the reconnection site. The shocks or the electric fields induced by them might lead to particle acceleration, cf. Chap. 7.



Space
→3.3.4

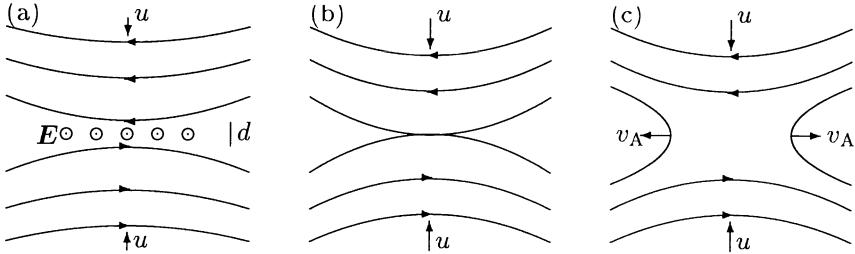


Fig. 3.9. Reconnection: merging of magnetic field lines leads to a rearrangement of fields. In addition, magnetic field energy is released, heating the plasma, creating a shock wave, and accelerating particles

If the onset of reconnection does not modify the general field and plasma configuration, stationary reconnection results as an equilibrium between inflowing mass and magnetic flux, magnetic diffusion, and outflowing mass and magnetic flux. In Sweet–Parker reconnection [3.22], a diffusion region of width d and length L is assumed with $L \gg d$, similar to the configuration in Fig. 3.9. The rate of reconnection and the properties of the outflow can be determined from the conservation of mass, momentum, energy and magnetic flux. The outflow speed of the plasma then is equal to the Alfvén speed $v_{A,\text{in}}$ in the incoming plasma flow, and the rate of reconnection R_{SP} equals the Mach number of the incident flow:

$$R_{\text{SP}} = \sqrt{\frac{c}{L\sigma v_{A,\text{in}} 4\pi}}. \quad (3.94)$$

Thus the reconnection process depends on the conductivity. For space plasmas, where the conductivity is high, a low rate of reconnection results. The Sweet–Parker reconnection therefore is a slow process in which about half of the incoming magnetic energy is converted into kinetic energy of the outflowing plasma. This acceleration leads to the two high-speed plasma flows indicated in panel (c) in Fig. 3.9.

Petschek reconnection occurs in more localized regions; the process is faster because the length scale L is smaller. In Petschek reconnection about three-fifth of the inflowing magnetic energy are converted into kinetic energy behind the shock waves, the remaining two-fifths is used to heat the plasma. The reconnection rate R_{P} is given as

$$R_{\text{P}} = \frac{\pi}{8} \ln \left(\sqrt{\frac{c}{L\sigma v_{A,\text{in}} 4\pi}} \right). \quad (3.95)$$

Petschek reconnection varies less with conductivity and therefore is much more efficient in mixing plasmas and fields. And, with a more efficient reconnection, the resulting acceleration becomes more violent, too.

Sweet–Parker reconnection appears to play an important role at the magnetopause where the high-speed flows streaming away from the reconnection

side can be detected *in situ*. Petschek reconnection probably does not play a role in magnetospheric plasmas but might be important in solar flares.

3.3.4 Deformation of the Field in a Plasma Flow

To determine the dissipation time of a magnetic field, we have ignored the plasma flow in (3.88). If an external flow $\mathbf{u}(\mathbf{r})$ is prescribed, the field will not dissipate but will be deformed by the flow. If we assume both magnetic field and plasma flow to be independent of time, a stationary solution exists with

$$-\nabla \times (\mathbf{u} \times \mathbf{B}) = \frac{c^2}{4\pi\sigma} \nabla^2 \mathbf{B}. \quad (3.96)$$

Again we can assume a characteristic time scale τ and a characteristic length scale. With (3.92) the flow speed u_\perp perpendicular to the field then can be estimated to be

$$u_\perp \approx \frac{L}{\tau} \approx \frac{c^2}{4\pi\sigma L}. \quad (3.97)$$

The physical interpretation is simple: if a plasma flows perpendicular to the magnetic field, it deforms the lines of force until their characteristic scale length is small enough to fulfill (3.97). Then the plasma starts to flow across the lines of force.

Dimensionless Variables and Dimensional Stability. To determine the deformation of a line of force quantitatively, we will use the technique of dimensionless variables. This technique is quite common in fluid dynamics (see e.g. [3.8]). It is helpful to determine not only one solution of the differential equation but the an entire manifold of solutions which can be scaled to the situation under study. This is particularly helpful in hydrodynamics when the solution for a certain size of syringe or nozzle is known and we are looking for a dynamically similar flow on a different scale.



The idea is quite simple: all equations representing scientific laws can be expressed such that both sides are dimensionless. In its simplest case, just divide one side of the equation by the other: the result, one, is dimensionless. To take advantage of dimensionless variables, we first have to identify the physical variables relevant to the problem and combine them into dimensionless groups A, B, C, \dots . These groups have to be independent of one another. If the groups are dimensionless, combinations of groups such as AB or A/B^2 are dimensionless, too. But they are not independent of either A or B , though any one of them might be included instead of A or B if this seems advantageous. If the groups are chosen in such a way that the quantity of interest occurs in only one of them, it can be expressed by the function $A = f(B, C, \dots)$. The nature of this unknown function can either be determined analytically (as will be demonstrated below) or by computational methods. In an analytical solution, the advantage of the use of dimensionless variables is small; it only shows which parameters are important in scaling. If the solution has to

be obtained by numerical simulations the advantage of this method is more obvious: the procedure to determine a solution for one particular set of parameters can be quite time consuming. Each other set of parameters would require a new run. If dimensionless variables are used instead, the nature of the solution becomes obvious and it can be scaled to suit different sets of parameters.

Let us now follow this principle and introduce new variables:

$$\mathbf{B} = b\tilde{\mathbf{B}}, \quad \mathbf{r} = r\tilde{\mathbf{r}}, \quad t = \tau\tilde{t}, \quad \mathbf{u} = U\tilde{\mathbf{u}}, \quad U = \frac{L}{\tau}. \quad (3.98)$$

The quantities with a tilde are dimensionless scalars or vectors. With the abbreviation

$$\eta = \frac{c^2}{4\pi\sigma}, \quad (3.99)$$

which can be interpreted as a magnetic viscosity, (3.88) yields

$$\frac{b}{\tau} \frac{\partial \tilde{\mathbf{B}}}{\partial \tilde{t}} - \frac{Ub}{L} \tilde{\nabla} \times (\tilde{\mathbf{u}} \times \tilde{\mathbf{B}}) = \frac{\eta b}{L^2} \tilde{\nabla}^2 \tilde{\mathbf{B}}. \quad (3.100)$$

Here a tilde above the differential operator indicates that the operator refers to a dimensionless variable.

In ordinary hydrodynamics, the Reynolds number is a measure of the ratio between inertial and viscous forces. If the Reynolds number exceeds a critical value, the flow becomes turbulent. In its definition, the Reynolds number contains typical scales which have to be adjusted to the problem under study. Here we will use a magnetic Reynolds number:

$$R_M = \frac{UL}{\eta} = \frac{4\pi\sigma UL}{c^2}. \quad (3.101)$$

It differs from the ordinary Reynolds number in so far as the viscous forces described by η do not depend on forces between particles but on the conductivity of the medium, cf. (3.99). Now we can rewrite (3.100) as

$$\frac{\partial \tilde{\mathbf{B}}}{\partial \tilde{t}} - \tilde{\nabla} \times (\tilde{\mathbf{u}} \times \tilde{\mathbf{B}}) = \frac{1}{R_M} \tilde{\nabla}^2 \tilde{\mathbf{B}}. \quad (3.102)$$

This dimensionless form has an advantage: it shows directly that a three-fold set of solutions exists. What does this mean? Assume we know a solution $\tilde{\mathbf{B}}(\tilde{\mathbf{r}}, \tilde{t})$ in dimensionless variables for a fixed Reynolds number R_M and a velocity field $\tilde{\mathbf{u}}$. In this case, $b \cdot \tilde{\mathbf{B}}(L\tilde{\mathbf{r}}, \tau\tilde{t})$ also is a solution of the same Reynolds number and the velocity field $U \cdot \tilde{\mathbf{u}}$ as long as the conditions $UL/\eta = R_M$ and $L/\tau = U$ are fulfilled. For instance, a free choice of U and L for a given R_M determines the values τ and η of the solution. But we still have a free choice for b . Thus one solution of (3.102) contains a three-fold infinite manifold of solutions, characterized, for instance, by U , η , and b .

Stationary Flow Perpendicular to the Field. Let us now determine the solution of (3.102) for the special case of a stationary parallel flow perpendicular to a homogeneous magnetic field. Since the flow is stationary, (3.102) reduces to

$$-\tilde{\nabla} \times (\tilde{u} \times \tilde{\mathbf{B}}) = \frac{1}{R_M} \tilde{\nabla}^2 \tilde{\mathbf{B}} . \quad (3.103)$$

Let us now orientate the x -axis of a cartesian coordinate system along the flow: $\tilde{u} = (\tilde{u}_x(z), 0, 0)$. The magnetic field $\tilde{\mathbf{B}} = (\tilde{B}_x(z), 0, \tilde{B}_z(z))$ has one component parallel to the flow and another perpendicular to it. Note that the flow varies along the perpendicular component. From $\nabla \times \mathbf{B} = 0$ we find that \tilde{B}_z is constant. Equation (3.103) is a second-order linear inhomogeneous partial differential equation for \tilde{B}_x as a function of z :

$$\frac{\partial^2 \tilde{B}_x}{\partial z^2} = -\tilde{B}_z R_M \frac{\partial \tilde{u}_x}{\partial z} . \quad (3.104)$$

Integrating twice we get

$$\tilde{B}_x = -\tilde{B}_z R_M \left(\int \tilde{u}_x dz + C\tilde{z} + D \right) , \quad (3.105)$$

with C and D to be determined to fulfill the boundary conditions.

Let us now assume the flow to have a cosine profile around $z = 0$, that is $\tilde{u}_x = \cos z$ for $|\tilde{z}| \leq \pi/2$ and $\tilde{u}_x = 0$ for $|\tilde{z}| \geq \pi/2$. Equation (3.105) then reads

$$\tilde{B}_x = -\tilde{B}_z R_M (\sin \tilde{z} + C\tilde{z} + D) . \quad (3.106)$$

The tangential component of $\tilde{\mathbf{B}}$ should be steady at the boundary of the flow to avoid currents; thus one boundary condition is $\tilde{B}_x(\tilde{z} = \pi/2) = 0$. In addition, the flow is assumed to be symmetric around $\tilde{z} = 0$; thus the second boundary condition is $\tilde{B}_x(0) = 0$. The integration constants therefore are $D = 0$ and $C = -\pi/2$ and (3.106) can be written as

$$\tilde{B}_x = -\tilde{B}_z R_M \left(\sin \tilde{z} - \frac{1}{\pi} \tilde{z} \right) . \quad (3.107)$$

We can now define a magnetic stream function ψ by

$$\frac{\partial \tilde{\psi}}{\partial \tilde{x}} = -\tilde{B}_z \quad \text{and} \quad \frac{\partial \tilde{\psi}}{\partial \tilde{z}} = \tilde{B}_x . \quad (3.108)$$

Then $\nabla \tilde{\psi} \times \nabla \tilde{\mathbf{B}}$ is zero and lines with $\tilde{\psi} = \text{const}$ are the field lines. Integration of the second part of (3.108) combined with (3.107) gives the line of force as

$$\tilde{\psi} = \tilde{B}_z R_M \left(\cos \tilde{z} + \frac{\tilde{z}^2}{\pi} \right) - \tilde{B}_z \tilde{x} . \quad (3.109)$$

Let us now determine the maximum displacement of a line of force, cf. Fig. 3.10. Because $\tilde{\psi}$ is constant along a field line, it is $\tilde{\psi}(\Delta \tilde{x}, 0) = \tilde{\psi}(0, \pi/2)$. Therefore the maximal displacement is given as

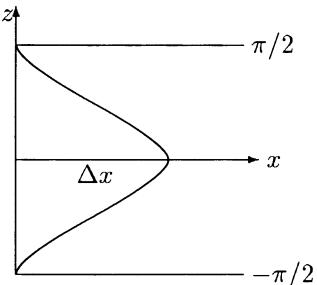


Fig. 3.10. Deformation of a magnetic field line by a plasma flow. The maximal displacement Δx is determined by the Reynolds number R_M

$$\Delta \tilde{x} = \left(1 - \frac{\pi}{4}\right) R_M . \quad (3.110)$$

Thus the deformation of the magnetic field line increases with increasing Reynolds number. This is not surprising because the Reynolds number depends linearly on the conductivity, cf. (3.101): if the conductivity and therefore the Reynolds number is infinite, the magnetic field is frozen into the fluid and deformation of the field lines becomes maximal.

Note that, in contrast to the frozen-in case, for finite conductivity matter starts to flow across the field after it is curved according to (3.109). The flow across the field is largest for small dimensions because on small scales the condition for frozen-in fields can be violated more easily. Thus we can confirm the suggestion made in connection with (3.97): at each point the flow curves the field such that the radius of curvature becomes small enough to allow for a small dissipation time and therefore a flow of matter across the field.



3.4 The Magnetohydrodynamic Dynamo

Space
→3.5

Magnetic fields can be found almost everywhere in space. The magnetosphere could not exist without the magnetic field of the Earth, interplanetary space is structured by the solar magnetic field carried by the solar wind, and the Sun itself would be an extremely boring star were it not for the magnetic field. But these fields are not permanent: the Sun reverses polarity in an 11-year cycle and polarity reversals of the Earth's magnetic field are known, too. Thus these fields cannot be remnants of a fossil field left from the time of the big bang. Instead, a mechanism is required that generates these fields and also allows for the (cyclic) variations. This mechanism is called the magnetohydrodynamic (MHD) dynamo: a residual seed field is amplified by a dynamo mechanism. The required energy is drawn from the rotational energy of the star or the planet. Thus the motion of the plasma drives a dynamo which amplifies a seed field and preserves it against losses. If we use the solar radius as the scale-length and a conductivity of $10^{16}/\text{s}$ in (3.97), a velocity of the order of 10^{-9} m/s results: thus very small flow speeds are sufficient to compensate for the diffusion of magnetic energy. But which topology do we have to choose?

In principle, a dynamo consists of a permanent magnet and a rotating circuit loop in which the current is induced. In the hot interior of the Sun and the planets, permanent magnets cannot exist. Thus the static magnetic field must be created by a current, too. Part of the current induced into the circuit loop then is fed back into the system to support the static field. Without such a feedback, the MHD dynamo would not work.

In the core of the Sun or the planets such well-defined parts as coils or rotating wires do not exist. Instead, we find a homogeneous and highly conductive fluid, rotating with the star or planet. Thus the dynamo also is called a homogeneous dynamo because all important processes take place inside one homogeneous medium. Since the matter inside the core is liquid, the question of how to create a magnetic field can be reduced to a simpler form: What is the nature of the plasma flow that allows to support the required currents?

Since we want to apply the dynamo to planets and stars, the model has to explain the most important features of their magnetic fields, such as: (a) the magnetic flux density increases with increasing rotation speed, (b) to first-order, the field is dipole like, (c) the dipole axis and the axis of rotation are nearly parallel, (d) the dynamo should allow for fluctuations in the magnetic field direction and flux density, and (e) polarity reversals with quasi-periodic but nonetheless stochastic character should be allowed. This latter point means that the reversal period can be identified (for instance 11 years for the Sun and about 500 000 years for the Earth), but that the individual cycle lengths are distributed stochastically around this average.

The fields are axial-symmetric; thus one might be tempted to try a configuration similar to the uni-polar inductor where a metal cylinder rotates parallel to a homogeneous magnetic field, leading to a potential difference between the center and the mantle of the cylinder. But in the uni-polar inductor the field cannot be amplified. In astrophysical plasmas this is expressed by Cowling's theorem, dating back to 1934: there is no finite velocity field that can maintain a stationary axial-symmetric magnetic field. The proof of this theorem is based on the induction equation (3.88) which, under the conditions cited in Cowling's theorem, would allow for decaying magnetic fields only.

The situation is different in a statistical magnetic field: on the Sun, for instance, the turbulent motion in the convection zone modifies the field. The average field $\mathbf{B}_o = \langle \mathbf{B} \rangle$ still is axial-symmetric but it is modified by fluctuations \mathbf{B}_1 with $\langle \mathbf{B}_1 \rangle = 0$. Thus the magnetic field is $\mathbf{B} = \mathbf{B}_o + \mathbf{B}_1$ and the velocity field is $\mathbf{u} = \mathbf{u}_o + \mathbf{u}_1$. The cross product of the speed and the magnetic field then reads

$$\langle \mathbf{u} \times \mathbf{B} \rangle = \mathbf{u}_o \times \mathbf{B}_o + \langle \mathbf{u}_1 \times \mathbf{B}_1 \rangle. \quad (3.111)$$

The products $\langle \mathbf{u}_1 \times \mathbf{B}_o \rangle$ and $\langle \mathbf{u}_o \times \mathbf{B}_1 \rangle$ vanish because the quantities with index 'o' are constant and the average of the other quantity equals zero. The product $\langle \mathbf{u}_1 \times \mathbf{B}_1 \rangle$, which is the correlation function, does not vanish

because the fluctuations are not independent: because the matter has a high conductivity, the magnetic field is frozen-into the plasma, and a change in the velocity field leads to a corresponding change in the magnetic field. To first order, the correlation function can be approximated as

$$\langle \mathbf{u}_1 \times \mathbf{B}_1 \rangle = \alpha \mathbf{B}_o - \beta \nabla \times \mathbf{B}_o . \quad (3.112)$$

Both α and β are determined by the properties of the turbulent velocity field. The β -term describes the increase in magnetic diffusion due to the turbulent motion, leading to a faster dissipation of the field. In the case of a mirror-symmetric velocity field, α would vanish. But in a rotating system, such as the stars or the planets, the velocity field is not symmetric. Taking the average of the induction equation (3.88) when considering the magnetic viscosity (3.99), we get

$$\frac{\partial \mathbf{B}_o}{\partial t} - \nabla \times (\mathbf{u}_o \times \mathbf{B}_o + \alpha \mathbf{B}_o) = -(\eta + \beta) \nabla \times (\nabla \times \mathbf{B}_o) . \quad (3.113)$$

While β modifies the viscosity, the α -term as an additional term contains the basic difference compared with (3.88): it allows for an electro-motoric force parallel to the average magnetic field; Cowling's theorem does not apply to this equation.

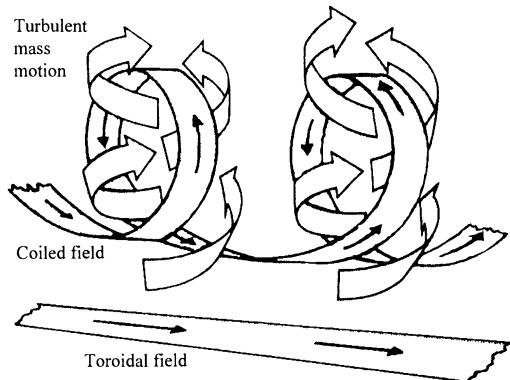


Fig. 3.11. A combination of the stochastic motion and the Coriolis force leads to turbulent motion (*short twisted arrows*) of the plasma which twists an originally toroidal field (*lower line*) into a coiled field which allows for a current parallel to the original field

The basic idea of the MHD dynamo can be applied to different geometries and to stationary as well as periodically changing magnetic fields. Because we are interested in axial-symmetric fields, it is reasonable to describe the magnetic field as consisting of a toroidal and a poloidal part:

$$\langle \mathbf{B} \rangle = \langle \mathbf{B}_{\text{tor}} \rangle + \langle \mathbf{B}_{\text{pol}} \rangle = B \cdot \mathbf{e}_\phi + \nabla \times A \cdot \mathbf{e}_\phi , \quad (3.114)$$

with \mathbf{e}_ϕ being the unit vector in the toroidal direction. Thus two scalar quantities, A and B , determine the three field components. With this ansatz, the induction equation gives two equations: one describing the ohmic dissipation of B and the generation of B out of A due to the α -effect and the differential

rotation $\nabla\Omega$, the other describing the ohmic dissipation of A combined with the generation of A out of B .

Differential rotation can occur for different reasons. The Sun, for instance, has a higher angular speed at the equator than at higher latitudes, thus the rotation depends on latitude. The differential rotation inside the Earth is due to the differences in angular speed between the faster inner and the slower outer core. In both cases, because the field is frozen-into the plasma, the deformation of the field line arises from the differential rotation.

The α -effect, on the other hand, is associated with the turbulent motion of the plasma, in particular the upward and downward motions associated with convection. Although this motion is stochastic, its combination with the Coriolis force leads to a turbulent motion which introduces a systematic twist into an originally toroidal field, as shown in Fig. 3.11. The resulting magnetic field coil allows for a current parallel to the undisturbed toroidal field.

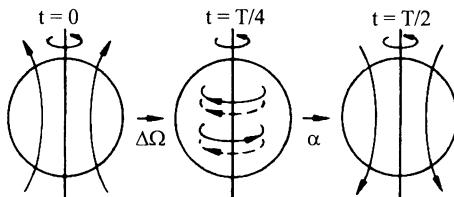


Fig. 3.12. Magnetohydrodynamic dynamo: differential rotation deforms a poloidal magnetic field into a toroidal one. The α -effect allows for field-parallel currents giving rise to a toroidal magnetic field in the opposite direction

The combination of both effects allows us to describe the MHD dynamo as sketched in Fig. 3.12. (a) We start with a poloidal field on the Sun. (b) The differential rotation deforms the magnetic field, leading to a toroidal field. The α -effect leads to electromagnetic forces parallel to the field, thus a toroidal current is flowing (dashed lines). Although the magnetic field directions are opposite on both hemispheres, the asymmetry of the Coriolis force leads to an asymmetric α -effect and therefore parallel currents in both hemispheres. (c) This current leads to a magnetic field directed opposite to the original field. Half a cycle is finished. This dynamo is called the $\alpha\Omega$ dynamo because both the α -effect and the differential rotation contribute to the dynamo process. The dynamos inside the Sun and the Earth are based on this principle; their details will be discussed in Sects. 8.6.2 and 10.1

If the α -effect was not at work, the differential rotation still would transform the poloidal magnetic field into a toroidal one. However, no polarity reversal would occur and with time the entire field would dissipate. The differential rotation, on the other hand, is not essential in the MHD dynamo. The α -effect can also work with turbulent motions which, for some reason, have a preferred direction of motion, often an upwelling of magnetic flux combined with a certain direction of rotation of the flux tubes (cf. e.g. [3.15]).

The MHD dynamo requires an initial magnetic field which is amplified by a suitable feedback mechanism. Thus at first glance the MHD dynamo vio-

lates Lenz's rule which states that all fields, currents and forces are directed so as to hinder the process that leads to their induction. For instance, an increase in the magnetic field leads to currents which create a magnetic field opposite to the original one. Lenz's rule thus stabilizes the system, allowing only for negative feedback but not for the positive feedback required in the MHD dynamo. Were we to build such a dynamo on the basis of one process only, Lenz's rule would be violated. But the MHD dynamo has the remarkable feature that although all individual processes obey Lenz's rule, the sum of them still allows for positive feedback.

3.5 Debye Shielding

We will now, though in a simple formalism, make use of the stochastic, thermal motion of particles in a two-component plasma consisting of electrons and protons. A local deviation from quasi-neutrality arises from the random thermal motion of the particles. Quasi-neutrality depends on the size of the volume under consideration. If the volume is very small, housing only one particle, quasi-neutrality cannot be obtained. But even if we increase the size of the volume, the thermal motion of the particles might lead to an excess of particles with one charge sign. Then the shielding of a certain particle with one polarity due to particles of the opposite polarity becomes important. The typical spatial scale for such shielding is the Debye length. This parameter is important, e.g. to estimate the influence of a spacecraft or the instrument on the plasma parameters to be measured.

The region depleted of electrons due to their random thermal motion is limited in extent because the displaced electrons create an electric field which acts as a restoring force. Consider a sheath of width D depleted of all electrons. Because of their larger mass, the ions are less mobile and stay within this sheath. Within D therefore a positive charge exists while the electrons can be regarded as a surface charge collected at the boundary of the sheath. The electric field is different from zero within the sheath; outside the sheath the field of the positive ions is screened by the surface charge.

The energy in the electric field stems from the kinetic energy of the thermal electron motion. With n_e as the electron density in the undisturbed plasma, the kinetic energy of the electrons in a sheath of thickness D is $n_e k_B T D / 2$. If all electrons are removed from this sheath, a restoring force proportional to D acts on them. The energy contained in the electric field created by the charge separation depends on D^2 . Thus there is a certain width λ_D at which the energy contained in the field equals the kinetic energy of the electrons originally present within this region: the kinetic energy of the thermal motion of the electrons is converted entirely into field energy.¹

¹ Note that this is different from the discussion of the plasma oscillations in Sect. 4.3.1 because here the sheath depleted of electrons results from the thermal

The electric field created by the ions inside the sheath is $\nabla \cdot \mathbf{E} = 4\pi e Z n_i$. Thus the Coloumb potential φ , related to the electric field by $\mathbf{E} = -\nabla \varphi$, can be written as $\nabla^2 \varphi = -4\pi e Z n_i$. In the one-dimensional case, the potential is

$$\frac{\partial^2 \varphi}{\partial x^2} = -4\pi e Z n_i = \text{const.} \quad (3.115)$$

The general solution of this equation is $\varphi(x) = -2\pi Z e n_i x^2 + C_1 x + C_2$, with the constants determined by the boundary conditions. The coordinate system can be fixed such that the potential vanishes at $x = 0$ and C_2 is zero. If the field is symmetric around $x = 0$, then C_1 is zero, too. In addition, the transition between the potential inside and outside the sheath has to be steady. Thus the potential can be written as

$$\varphi = \begin{cases} -2\pi Z e n_i x^2 & (|x| \leq D/2) \\ -2\pi Z e n_i (D/2)^2 & (|x| \geq D/2) \end{cases}. \quad (3.116)$$

The electric field then is

$$E(x) = \begin{cases} 4\pi Z e n_i x & (|x| \leq D/2) \\ 0 & (|x| \geq D/2) \end{cases}. \quad (3.117)$$

The energy contained in the electric field can be determined as

$$\int_{-D/2}^{+D/2} \frac{E^2}{8\pi} dx = 2\pi e^2 Z^2 n_i^2 \cdot \int_{-D/2}^{+D/2} x^2 dx = \frac{\pi e^2 Z^2 n_i^2}{6} D^3. \quad (3.118)$$

We are now looking for the thickness λ_D of a sheath where the field energy equals the kinetic energy of the electrons originally present inside the sheath:

$$\frac{n_e k_B T \lambda_D}{2} = \frac{\pi e^2 Z^2 n_i^2 \lambda_D^3}{6}. \quad (3.119)$$

Note that according to the one-dimensional geometry considered so far only one degree of freedom is considered in the kinetic energy. Quasi-neutrality requires $n_e = Z n_i$, and thus the Debye length λ_D is given as

$$\lambda_D = \sqrt{\frac{3k_B T}{\pi e^2 n_e}} = \sqrt{\frac{k_B T}{m}} \cdot \frac{1}{\omega_{pe}}, \quad (3.120)$$

with ω_{pe} being the angular frequency of electron plasma oscillations.

Though derived for the one-dimensional case, this equation also holds in the three-dimensional case where the Debye length can be interpreted as the maximal radius of a sphere which, due to thermal motion, might be depleted of electrons.

In spatial regions small compared with the Debye length, quasi-neutrality is likely to be violated, while on larger scales the plasma is quasi-neutral.

motion of the electrons. In plasma oscillations, instead, a cold plasma is considered and the charge separation results from an external force, e.g. a beam of electrons traveling through the plasma.

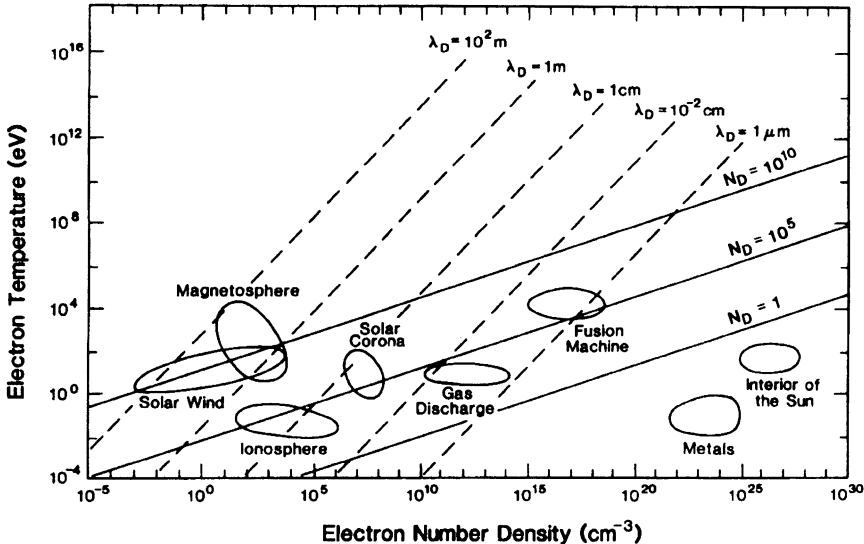


Fig. 3.13. Temperature, electron density and Debye length for different types of plasmas. Reprinted from M.G. Kivelson [3.14], in *Introduction to Space Physics* (eds. M.G. Kivelson and C.T. Russell), Copyright 1995, with kind permission of Cambridge University Press

In the latter case the kinetic energy contained in the thermal motion is not large enough to disturb the particle distribution over the entire region. Figure 3.13 shows typical values of the Debye length and numbers of particles inside a sphere with the Debye radius for different electron densities and temperatures. Some typical plasmas also are indicated.

Another, more illustrative motivation for the Debye length is related to the influence of the instrument on measuring plasma parameters. Let us start from an initially cold plasma, i.e. the motion of the electrons can be ignored. Let us now insert a small positive charge q into this plasma. Immediately, this will be surrounded by a cloud of electrons while the ions are repulsed by the charge. The electron cloud screens the additional charge; thus outside the electron cloud the electric field vanishes. If we now increase the temperature, the electrons gain thermal velocity. Deep inside the electron cloud this velocity will be too small to overcome the attraction of the positive charge. At the outer edges of the cloud, on the other hand, the thermal energy might be large enough to exceed the electrostatic potential of the partly screened charge, allowing an escape from the cloud. The Debye length than can be interpreted as the spatial scale over which the potential of the point charge q has decreased by a characteristic value, cf. Fig. 3.14: within the Debye length, electrons are influenced by the test charge, while at larger distances the test charge goes unnoticed.

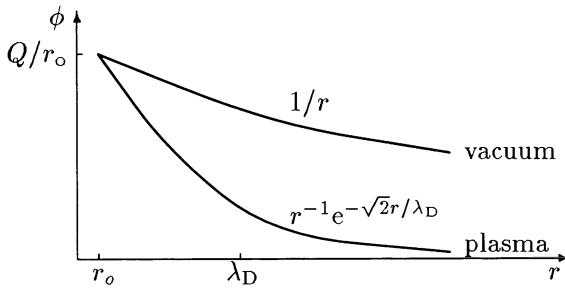


Fig. 3.14. The electric potential of a test charge is reduced by the surrounding plasma

A more general definition would read: only charged particles within a distance of λ_D exert an electrostatic force on each other. This is different from bodies which interact gravitationally, like, for example, interacting stars: gravitation cannot be screened by repulsing forces, it has an indefinite range.

Debye screening also is important in preventing local clusters of charges. For $e\varphi \ll k_B T$ the influence of the electric field on the particle energy is small; thus the particle motion is determined by the thermal speed. The Debye shield arises from small differences in the particle motion: some particles with opposite charge stay close to the test charge slightly longer, while particles with equal charge move away a little bit faster. Charge inhomogeneities in a plasma therefore are balanced on the scale of the thermal propagation time.

3.6 Summary

Magnetohydrostatics is concerned with the energetics of the electromagnetic field without allowing for the collective motion of the plasma. Basic concepts derived from this approximation are the magnetic pressure and the magnetic tension. Magnetic pressure, graphically interpreted as mutual repulsion of field lines, prevents a magnetic field from being compressed by external forces. If part of a magnetic line of force is displaced from its original position, the magnetic tension creates a restoring force, which graphically can be interpreted as the tendency of a line of force to shorten itself.

Magnetohydrokinematics assumes the energy density of the plasma to be much larger than the energy density of the field, which allows us to ignore the influence of the field on particle motion. The basic concepts are as follows: (a) If the conductivity of the plasma is infinite, the magnetic field lines are frozen-into the plasma. Thus the plasma flow carries away the magnetic field. (b) If the conductivity is finite, the magnetic field is deformed by the plasma flow until diffusion allows the plasma to flow across the field lines. (c) In a stationary plasma, the magnetic field dissipates, with the dissipation time depending on the square of the linear dimensions (small fields dissipate faster) and linearly on the conductivity (if the latter is infinite, the dissipation time is infinite, too, and the field is frozen-in). These concepts are important, e.g. in

our understanding of the merging of magnetic field lines, called reconnection, and in dynamo theory.

Exercises and Problems

- 3.1** Explain the difference between convective and partial derivatives. Find examples to illustrate the differences.
- 3.2** Recall simple hydrodynamics and give other examples of the momentum balance. Discuss the different forms and compare with the Navier–Stokes equation.
- 3.3** Derive the hydrostatic equation from the Navier–Stokes equation. Which terms do you need?
- 3.4** Give a quantitative discussion of the stability of a sunspot (all important numbers are given in Tab. 8.1).
- 3.5** Is the filament sketched in Fig. 3.5 realistic? Why does it not dissolve towards the sides (remember, it is a plasma, not a solid body)?
- 3.6** What is the meaning of viscosity and Reynolds number? What are the formal differences between hydrodynamics and magnetohydrodynamics? What are the differences in substance?
- 3.7** Explain the consequences of stationary parallel flows parallel and oblique to the magnetic field.
- 3.8** Show that in an ideal, non-relativistic magnetohydrodynamic plasma the ratio between the electric and the magnetic energies is $(v_{\perp}/c)^2$.
- 3.9** Determine the dissipation times for a copper block (side length 10 cm, conductivity $10^{18}/\text{s}$) and the interstellar medium (linear dimension 10^{21} m, conductivity $10^{10}/\text{s}$). Compare with the age of the universe (about 10^{18} s).
- 3.10** Determine the gyro radii, gyro frequencies, Debye length and the number of particles inside a Debye sphere for electrons and protons moving with thermal speeds (cf. Sect. 5.1.2) in the following fields: (a) the Earth's magnetosphere with $n = 10^4 \text{ cm}^{-3}$, $T = 10^3 \text{ K}$, $B = 10^{-2} \text{ G}$; (b) the core of the Sun with $n = 10^{26} \text{ cm}^{-3}$, $T = 10^{7.2} \text{ K}$, $B = 10^6 \text{ G}$; (c) the solar corona with $n = 10^8 \text{ cm}^{-3}$, $T = 10^6 \text{ K}$, $B = 1 \text{ G}$; and (d) the solar wind with $n = 10 \text{ cm}^{-3}$, $T = 10^5 \text{ K}$, $B = 10^{-5} \text{ G}$.

4. Plasma Waves

Roll on, thou deep and dark blue Ocean – roll!
Ten thousand fleets sweep over thee in vain;
Man marks the earth with ruin – his control
Stops with the shore.
Lord Byron, *Childe Harold's Pilgrimage*

In this chapter we will catch a glimpse on the vast zoo of plasma waves. The formalism to derive these waves, perturbation theory, briefly will be introduced. We will not derive all types of waves formally; instead, we will limit ourselves to the magnetohydrodynamic waves, which are Alfvén waves and ion acoustic waves. The derivation of the dispersion relations for other types of waves follows the same scheme, it only differs in the terms considered in the equation of motion, in the assumptions made in the equation of state, and in whether a one-fluid description of the plasma is sufficient or if a two-fluid description is required. The detailed derivations can be found e.g. in [4.1, 4.3, 4.4, 4.5, 4.7, 4.8, 4.9, 4.6]. For the purpose of this book, however, it is more important to grasp the nature of the waves than to fiddle around with the mathematical tricks involved in solving the equation of motion.

This chapter is limited to elementary types of waves. First, the geometry always is simple: in a magnetized plasma the waves either propagate parallel or perpendicular to the undisturbed magnetic field. Physically more important is the limitation to small disturbances, i.e. small amplitude waves: the basic set of magnetohydrodynamic equations is a set of coupled non-linear partial differential equations. Thus in principle we can expect non-linear couplings between different fluctuating quantities of the wave. If we limit our discussion to small amplitude waves, the equations can be linearized: whenever two oscillating quantities are multiplied, since both are small, we consider this a higher order term and ignore it. If we apply these results to a real situation, we always have to take one step back and justify whether the amplitudes calculated in our real situation are small enough so that the non-linear terms actually are negligible compared with the linear ones.

In a plasma, a large variety of waves can exist. A simple phenomenological classification in transversal and longitudinal waves is insufficient. Instead, we have to consider the conditions under which certain types of waves can

exist. For instance, in a cold plasma the thermal motion of the particles and therefore the pressure vanishes. Thus elastic waves cannot exist; they can form only in a warm plasma where a pressure gradient can build up. In an isotropic plasma, no magnetic field exists, which allows other modes of propagation than an anisotropic plasma. As a third criterion, we also have to consider which particle species is in motion.

4.1 What is a Wave?

To start, let us briefly repeat the basic descriptions and properties of waves. A wave is a disturbance propagating through a continuous medium, leading to a periodic motion of the fluid. Even a complex oscillation, at least as long as the amplitude of the disturbance is small, can be decomposed into different sinusoidal waves by the technique of Fourier analysis. Small-amplitude disturbances lead to plane waves with the direction of propagation and the amplitude being the same everywhere.

A sinusoidal wave is described by its frequency ω and wave vector \mathbf{k} :

$$\mathbf{B}(\mathbf{r}, t) = \mathbf{B}_0 \exp\{\mathrm{i}(\mathbf{k}\mathbf{r} - \omega t)\} . \quad (4.1)$$

The measurable quantity is the real part of this complex expression. The exponent in (4.1) is the phase of the disturbance. The temporal derivative of the phase gives the frequency ω , and the spatial derivative gives the wave vector \mathbf{k} that specifies the direction of wave propagation. A surface of constant phase, also called a wave surface, is displaced by the phase velocity v_{ph} , which can be determined from $d(\mathbf{k}\mathbf{r} - \omega t)/dt$ as $v_{\text{ph}} = \omega/k$ or, in vector form,

$$\mathbf{v}_{\text{ph}} = \frac{\omega}{k^2} \mathbf{k} . \quad (4.2)$$

If ω/k is positive, the wave moves to the right; for negative ω/k it moves to the left. For electromagnetic waves the refraction index n is defined as the ratio between the speed of light and the phase speed of the wave: $n = c/v_{\text{ph}} = ck/\omega$.

The phase velocity of a plasma wave often can exceed the speed of light. This is not in contradiction to the theory of relativity because an indefinitely long wave train of constant amplitude does not carry information. Information can be carried by a wave if it is modulated, i.e. variations in frequency or amplitude are superimposed. We can regard each bit of information in this modulated signal as a wave packet which then moves with the group velocity

$$\mathbf{v}_g = \frac{\partial \omega}{\partial \mathbf{k}} . \quad (4.3)$$

The group velocity v_g given by the slope of the dispersion relation (cf. dotted line in Fig. 4.1). It is always less than the speed of light and determines the velocity with which the energy of the wave is transported.

The most important tool in the description of waves is the dispersion relation $\omega = \omega(k)$, relating frequency and wave vector. From this relation

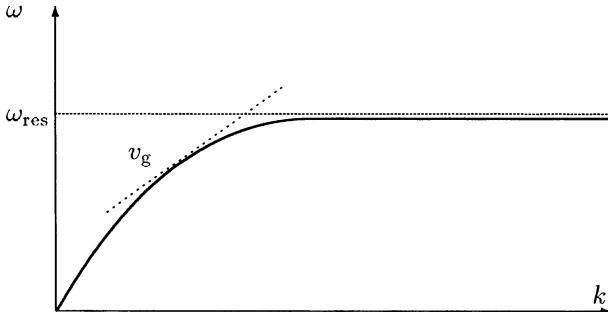


Fig. 4.1. Dispersion relation with a resonance at the frequency ω_{res} . The slope of the curve gives the group velocity v_g , as indicated by the dashed line

the group and phase velocities of a wave can be determined. The dispersion relation contains the physical parameters of the medium under consideration. If the plot of the dispersion relation shows an asymptotic behavior towards a certain frequency ω_{res} , as depicted in Fig. 4.1, there is a resonance at this frequency: as $\partial\omega/\partial k$ converges towards zero, the wave no longer propagates and all the wave energy is fed into stationary oscillations.

4.2 Magnetohydrodynamic Waves

Magnetohydrodynamic (MHD) waves are low-frequency waves related to the motion of the plasma's ion component. They can be understood intuitively from the concepts of magnetic pressure (Sect. 3.2.1) and magnetic tension (Sect. 3.2.2): in a magneto-sonic wave, compression of the field lines creates a magnetic pressure pulse which propagates perpendicular to the field in the same way as an ordinary pressure pulse propagates through a gas to form a sound wave. Magneto-sonic waves therefore are longitudinal waves: the disturbance is parallel to the propagation direction of the wave. And the displacement of part of a field line in an Alfvén wave can be likened to plucking a string: magnetic tension, like the tension in a string, acts as a restoring force and a transversal wave propagates along the field line.

Despite the simplicity and graphic quality of these concepts, we will treat these waves formally in a concept which is useful for small disturbances: the linearization of the equations. The MHD waves provide just one example for this concept; it is applied in a more elaborate way also in the quasi-linear theory of wave-particle interaction (Sect. 6.2.1).

4.2.1 Linearization of the Equations: Perturbation Theory

A wave is a disturbance of the medium with a certain speed, amplitude, and frequency. Thus the parameters of the medium, such as pressure, density, and the electromagnetic field, can be described by an average state, indicated by the index '0', and a superimposed disturbance, indexed as '1':

$$\mathbf{B} = \mathbf{B}_o + \mathbf{B}_1 , \quad \mathbf{E} = \mathbf{E}_o + \mathbf{E}_1 , \quad \mathbf{u} = \mathbf{u}_o + \mathbf{u}_1 , \quad (4.4)$$

$$\mathbf{j} = \mathbf{j}_o + \mathbf{j}_1 , \quad \varrho = \varrho_o + \varrho_1 , \quad p = p_o + p_1 , \quad (4.5)$$

with $\langle \mathbf{u}_1 \rangle = \langle \mathbf{j}_1 \rangle = \langle \mathbf{E}_1 \rangle = \langle \mathbf{B}_1 \rangle = \langle \varrho_1 \rangle = \langle p_1 \rangle = 0$. The resulting MHD equations (3.34)–(3.41) are difficult to solve. If we limit ourselves to small disturbances, i.e. $u_1 < u_o$, $B_1 < B_o$, $E_1 < E_o$, $j_1 < j_o$, $\varrho_1 < \varrho_o$, and $p_1 < p_o$, we can derive two sets of equations which are more convenient: the set describing the equilibrium state of the undisturbed medium contains quantities with index ‘0’ only, and a second set for the fluctuating quantities contains fluctuating quantities and products of undisturbed and fluctuating quantities. Products of fluctuating quantities are ignored because they are small in second order.

The basic MHD equations (3.34)–(3.41) for the undisturbed quantities are

$$\nabla \times \mathbf{B}_o = \frac{4\pi}{c} \mathbf{j}_o , \quad (4.6)$$

$$\nabla \times \mathbf{E}_o = 0 , \quad (4.7)$$

$$\nabla \cdot \mathbf{B}_o = 0 , \quad (4.8)$$

$$\frac{\mathbf{j}_o}{\sigma} = \mathbf{E}_o + \frac{\mathbf{u}_o \times \mathbf{B}_o}{c} , \quad (4.9)$$

$$\varrho_o \cdot (\mathbf{u}_o \cdot \nabla) \mathbf{u}_o = -\nabla p_o + \frac{\mathbf{j}_o \times \mathbf{B}_o}{c} , \quad (4.10)$$

$$\nabla \cdot (\varrho_o \mathbf{u}_o) = 0 , \quad (4.11)$$

$$p_o = C \varrho_o^{\gamma_a} . \quad (4.12)$$

Note that in the equation of motion friction and gravity have been ignored. The equation of state describes the plasma as an ideal gas with all changes occurring adiabatically.

The undisturbed medium is assumed to be homogeneous in pressure, density, and magnetic field. Since it is assumed to be at rest ($\mathbf{u}_o = 0$), the current \mathbf{j}_o vanishes (4.6) and the undisturbed electric field vanishes, too (4.9). Note that this is just another expression for the high conductivity of the plasma ($\sigma \rightarrow \infty$). The equations for the fluctuations then read

$$\nabla \times \mathbf{B}_1 = \frac{4\pi}{c} \mathbf{j}_1 , \quad (4.13)$$

$$\nabla \times \mathbf{E}_1 = -\frac{1}{c} \frac{\partial \mathbf{B}_1}{\partial t} , \quad (4.14)$$

$$\nabla \cdot \mathbf{B}_1 = 0 , \quad (4.15)$$

$$\mathbf{E}_1 = -\frac{\mathbf{u}_1 \times \mathbf{B}_o}{c} , \quad (4.16)$$

$$\varrho_o \frac{\partial \mathbf{u}_1}{\partial t} = -\nabla p_1 + \frac{\mathbf{j}_1 \times \mathbf{B}_o}{c}, \quad (4.17)$$

$$\frac{\partial \varrho_1}{\partial t} = -\nabla \cdot (\varrho_o \mathbf{u}_1). \quad (4.18)$$

$$\frac{p_1}{p_o} = \gamma_a \frac{\varrho_1}{\varrho_o}. \quad (4.19)$$

Now (4.13)–(4.19) is a homogeneous linear system of equations for the fluctuating quantities. Since neither time nor the spatial coordinate are explicit in one of the equations, the system can be solved by an exponential ansatz.

4.2.2 Alfvén Waves

Alfvén waves are transversal waves propagating parallel to the magnetic field, cf. Fig. 4.2. The magnetic tension acts as the restoring force. The fluctuating quantities are the electromagnetic field and the current density.

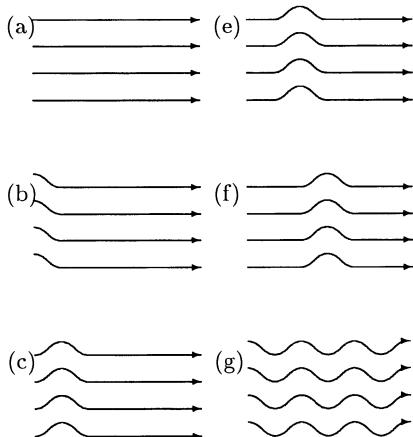


Fig. 4.2. Alfvén waves depicted as an oscillating string or an elastic rope

To derive the properties of Alfvén waves, we have to solve the equations for the fluctuating quantities. Let us start with the equation of motion (4.17). The pressure gradient force vanishes because the Alfvén wave is limited to fluctuations in the magnetic field but not in the gas-dynamic pressure. If we express the current \mathbf{j}_1 by Ampere's law (4.13), the momentum balance reads

$$\varrho_o \frac{\partial \mathbf{u}_1}{\partial t} = \frac{\mathbf{j}_1 \times \mathbf{B}_1}{c} = \frac{1}{4\pi} (\nabla \times \mathbf{B}_1) \times \mathbf{B}_1. \quad (4.20)$$

Combining Ohm's law (4.16) with Faraday's law (4.14) yields

$$\frac{\partial \mathbf{B}_1}{\partial t} = \nabla \times (\mathbf{u}_1 \times \mathbf{B}_o). \quad (4.21)$$

The remaining equation is Gauss's law for the magnetic field (4.15):

$$\nabla \cdot \mathbf{B}_1 = 0 . \quad (4.22)$$

The equation of state (4.18) is not considered here because we are not concerned with fluctuations in pressure and density, only in field.

Equations (4.20)–(4.22) can be solved by means of a Fourier transformation. If we assume that the solutions are plane waves, temporal and spatial derivatives can be substituted according to

$$\partial/\partial t \rightarrow -i\omega , \quad \nabla \rightarrow ik , \quad \nabla \cdot \rightarrow ik \cdot , \quad \nabla \times \rightarrow ik \times . \quad (4.23)$$

Equations (4.20)–(4.22) then read

$$-i\omega \varrho_o \mathbf{u}_1 = \frac{i}{4\pi} (\mathbf{k} \times \mathbf{B}_1) \times \mathbf{B}_o , \quad (4.24)$$

$$i\omega \mathbf{B}_1 = ik \times (\mathbf{u}_1 \times \mathbf{B}_o) , \quad (4.25)$$

$$\mathbf{k} \cdot \mathbf{B}_1 = 0 . \quad (4.26)$$

Alfvén waves propagate along the field, thus $\mathbf{k} \parallel \mathbf{B}_o$. With $\nabla \cdot \mathbf{B}_o = 0$ the above equations can be reduced to

$$\omega \mathbf{u}_1 = \frac{1}{4\pi \varrho_o} (\mathbf{B}_o \cdot \mathbf{B}_1) \mathbf{k} , \quad (4.27)$$

$$\omega \mathbf{B}_1 = (\mathbf{k} \cdot \mathbf{u}_1) \mathbf{B}_o , \quad (4.28)$$

$$\mathbf{k} \cdot \mathbf{B}_1 = 0 . \quad (4.29)$$

Let us now multiply (4.27) by \mathbf{k} and (4.28) by \mathbf{B}_o . Adding both equations gives the dispersion relation for the Alfvén wave:

$$\omega^2 = \frac{B_o^2}{4\pi \varrho_o} k^2 . \quad (4.30)$$

Alfvén waves are non-dispersive waves. Thus group and phase speed are the same. The Alfvén speed

$$v_A = \frac{B_o}{\sqrt{4\pi \varrho_o}} \quad (4.31)$$

is an important characteristic of a plasma: it is the maximum speed of a disturbance propagating along the magnetic field and can be compared with the speed of sound in a gas: if a disturbance propagates faster than the Alfvén speed, a shock wave develops (cf. Chap. 7). Typical Alfvén speeds are some tens of kilometers per second in the interplanetary medium and some 100 km/s in the solar corona.

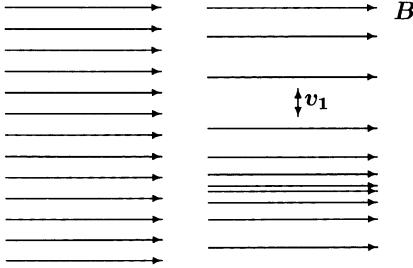


Fig. 4.3. Sketch of a magneto-sonic wave: undisturbed field (*left*) and fluctuating field (*right*)

4.2.3 Magneto-Sonic Waves

A magneto-sonic wave is similar to a sound wave: it is a longitudinal wave parallel to the magnetic field with alternating regions of compression and rarefaction in the plasma as well as in the magnetic field, cf. Fig 4.3.

Since we allow for a compression of the plasma, we also have to consider the equation of state (4.19) which can be written as $p_1 = \gamma_a \varrho_1 p_0 / \varrho_0 = v_s^2 \varrho_1$, with

$$v_s = \sqrt{\frac{\gamma_a p_0}{\varrho_0}} \quad (4.32)$$

being the (adiabatic) sound speed. To solve the equation of motion (4.17), we have to express p_1 and $\mathbf{j}_1 \times \mathbf{B}_1$ by \mathbf{u}_1 .

Fourier transformation of the equation of continuity gives $i\omega \varrho_1 = ik \varrho_0 u_1$ or, combined with the equation of state and (4.32),

$$p_1 = \frac{v_s^2}{i\omega} ik \varrho_0 u_1. \quad (4.33)$$

Thus the fluctuating pressure p_1 in the momentum balance can be expressed by \mathbf{u}_1 .

Ohm's law (4.16) yields the dependence of \mathbf{E}_1 on \mathbf{u}_1 . Combined with Faraday's law (4.14), \mathbf{B}_1 can be expressed as a function of \mathbf{u}_1 :

$$\nabla \times \mathbf{E}_1 = -\nabla \times \left(\frac{\mathbf{u}_1 \times \mathbf{B}_0}{c} \right) = -\frac{1}{c} \frac{\partial \mathbf{B}_1}{\partial t}. \quad (4.34)$$

Transformation yields

$$ik \times \left(\frac{\mathbf{u}_1 \times \mathbf{B}_0}{c} \right) = \frac{i\omega}{c} \mathbf{B}_1. \quad (4.35)$$

This expression for \mathbf{B}_1 can be inserted into Ampere's law (4.13):

$$\frac{4\pi}{c} \mathbf{j}_1 = -ik \times \left(\frac{ck}{\omega} \times \frac{\mathbf{u}_1 \times \mathbf{B}_0}{c} \right). \quad (4.36)$$

With (A.13), this can be simplified to

$$-\frac{4\pi\omega i}{c} \mathbf{j}_1 = \mathbf{k} \times \mathbf{u}_1 \cdot \mathbf{k} \times \mathbf{B}_0. \quad (4.37)$$

Here we have used $\mathbf{k} \perp \mathbf{B}_o$ and therefore $\mathbf{k} \cdot \mathbf{B}_o = 0$ and $\mathbf{k} \times \mathbf{B}_o = kB_o$. The vector product of (4.37) and \mathbf{B}_o yields

$$\mathbf{j}_1 \times \mathbf{B}_o = \frac{c}{4\pi i\omega} k^2 B_o^2 \mathbf{u}_1 . \quad (4.38)$$

Thus the equation of motion (4.17) combined with (4.33) and (4.38) can be written as

$$-i\omega \varrho_o \mathbf{u}_1 = -v_s^2 \frac{\varrho_o ik^2}{\omega} \mathbf{u}_1 - \frac{B_o^2 ik^2}{4\pi\omega} \mathbf{u}_1 , \quad (4.39)$$

giving the dispersion relation for the magneto-sonic wave. The phase speed then is determined by the squared sound and Alfvén speeds:

$$v_{ms}^2 = \frac{\omega^2}{k^2} = v_s^2 + v_A^2 . \quad (4.40)$$

The phase speed is independent of frequency or wave number: the wave is dispersion-free. If the magnetic field vanishes, the Alfvén speed goes to zero and the phase speed of the magneto-sonic wave is the speed of sound. If the magnetic field is strong, the phase speed of the magneto-sonic wave becomes the Alfvén speed. But the wave still behaves differently because it propagates perpendicular to the field instead of parallel to it. This latter kind of wave is also called the compressive Alfvén wave. Since the phase speed of the magneto-sonic wave exceeds the Alfvén speed, it often is called the fast magneto-sonic wave or just the fast MHD wave.

MHD Waves Oblique to the Field. So far, we have considered two special geometries: a transversal wave propagating parallel to the field and a longitudinal wave propagating perpendicular to it. But MHD waves can propagate at any angle relative to the field. For this general case, we have to solve the whole set of (4.6)–(4.12) for the undisturbed and (4.13)–(4.19) for the fluctuating quantities, neither one of the equations nor a term in one of the equations can be ignored. Again, a solution can be obtained by combining the equations into one equation for the desired quantity, the speed of the fluctuations. The dispersion relation obtained from this equation yields solutions for the above wave types with phase speeds depending on the angle θ between the wave vector and the undisturbed field. The speed of the MHD wave is determined by

$$u^4 - (v_A^2 + v_s^2)u^2 + v_A^2 v_s^2 \cos^2 \theta = 0 . \quad (4.41)$$

For $\theta = 90^\circ$, the fast magneto-sonic wave (4.40) results. For $\theta = 0^\circ$, two solutions exist: the Alfvén wave with $u = v_A$ and the sound wave with $u = v_s$.

The different solutions can be represented in a hodograph or Friedrichs diagram, cf. Fig. 4.4. The hodograph is a polar diagram of velocities with the polar angle θ relative to the undisturbed magnetic field direction and the wave's phase speed as distance from the origin. The wave front then propagates perpendicular to the velocity vector. Depending on whether the Alfvén speed is greater or less than the sound speed, two sets of solutions

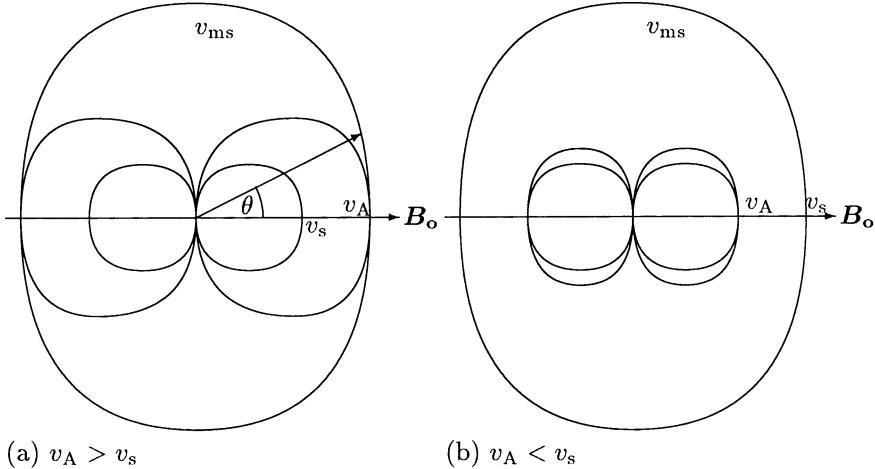


Fig. 4.4. Friedrichs diagram (hodograph) representing the different types of MHD waves for Alfvén speeds greater than the speed of sound (a) or less than (b)

arise, cf. Fig. 4.4. For almost any direction we find three different phase speeds, corresponding to an Alfvén wave and a slow and a fast mode magneto-sonic wave. For $\theta = 90^\circ$ or 270° only the fast magneto-sonic wave exists. As can be seen from the hodographs, the Alfvén speed $u_A = v_A \cdot \cos \theta$ always is between the phase speeds of the fast (outer curve) and the slow (inner curve) magneto-sonic wave. For waves propagating parallel to the field, the Alfvén speed becomes identical to the fast magneto-sonic speed for $v_A > v_s$. This does not contradict (4.40) because the latter had been derived for waves propagating perpendicular to the field, the direction in which the magneto-sonic speed becomes maximal. The speeds of the Alfvén wave and the slow mode magneto-acoustic wave, on the other hand, are maximal in the direction parallel to the field. The Alfvén speed of waves parallel to the field equals the slow mode speed for $v_A < v_s$.

4.3 Electrostatic Waves in Non-Magnetic Plasmas

Electrostatic waves start from a charge imbalance in an initially quasi-neutral fluid element. This charge imbalance accelerates the electrons and ions in its neighborhood, resulting in charges oscillating back and forth. Since these oscillations only involve the electric field, they are defined as electrostatic waves. The oscillating magnetic field is zero. Electrostatic waves can occur in non-magnetic plasmas (this section) or in magnetized plasma (Sect. 4.4). Fourier transformation of Faraday's law (4.14) for the fluctuating quantities of the wave yields $i\mathbf{k} \times \mathbf{E}_1 = i\omega \mathbf{B}_1 = 0$, thus the fluctuating electric field is parallel to the wave vector \mathbf{k} .

4.3.1 Plasma Oscillations

Plasma oscillations, also called Langmuir oscillations, are a prime example of a plasma phenomenon that requires consideration of both types of charges. Nonetheless, this does not automatically imply that plasma oscillations can be described only in the framework of two-fluid theory. Since the ions are assumed to be stationary, we do not have to consider their equation of motion. Thus a one-fluid description of the electrons is sufficient as long as the electric field created by the ions is considered.

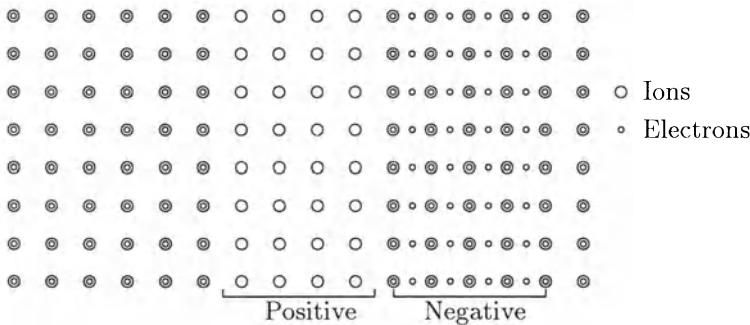


Fig. 4.5. The displacement of electrons in a cold plasma leads to plasma oscillations

Let us assume ions and electrons are distributed equally in space. Thus quasi-neutrality is fulfilled even in rather small volumes. In addition, we assume that the thermal motion of the particles vanishes; the fluid is treated as a cold plasma. Such a fluid can be disturbed by the displacement of part of the electrons, as indicated in Fig. 4.5. This displacement creates an electric field that pulls the electrons back to their initial rest positions while the heavier ions stay in their positions. Since the electrons are accelerated along the field, they gain kinetic energy which in turn drives them behind their initial position, creating an electric field in the opposite direction. This field slows down the electrons, eventually driving them back. The period of the resulting electron oscillation around their rest position is the electron plasma frequency ω_{ep} . It can be derived from the equation of motion

$$m_e \frac{\partial^2 x}{\partial t^2} = -eE, \quad (4.42)$$

with x the direction parallel to the electric field. E can be determined by applying Gauss's law to a closed surface along the boundary between the positive and negative charges in Fig. 4.5 and extending at least for the displacement x of the charges: $E = 4\pi n_e \cdot e \cdot x$. The equation of motion then is

$$\frac{\partial^2 x}{\partial t^2} = -\frac{4\pi n_e e^2}{m_e} x = -\omega_{\text{pe}}^2 x. \quad (4.43)$$

It describes a harmonic oscillator with the electron plasma frequency

$$\omega_{pe} = \sqrt{\frac{4\pi n_e e^2}{m_e}} . \quad (4.44)$$

Plasma oscillations are an important tool in plasma diagnostics because they allow us to measure the electron density n_e . If the latter is given in m^{-3} the relation between the electron plasma frequency and the electron density can be written as

$$\omega_{pe} = 0.564 \cdot \sqrt{n_e} \quad (\text{rad/s}) . \quad (4.45)$$

The electron displacement responsible for a plasma oscillation, for instance, can be created by a stream of fast charged particles. Examples are solar radio bursts (cf. Sect. 8.7.1).

4.3.2 Electron Plasma Waves (Langmuir Waves)

Plasma oscillations have been derived for a cold plasma. There the group velocity equals zero: the disturbance does not propagate. In a warm plasma, the situation is different. The thermal motion of the electrons allows them to carry the information about a disturbance into the undisturbed neighboring plasma, giving rise to a wave propagating through the plasma. Formally, one can derive this kind of wave by adding the pressure gradient force $-\nabla p$ to the equation of motion (4.42). If the plasma behaves adiabatically, the dispersion relation of the plasma wave in the one-dimensional case ($\gamma_a = 3$) reads

$$\omega^2 = \omega_{pe}^2 + \frac{3}{2} k^2 v_{th}^2 , \quad (4.46)$$

or in the three-dimensional case

$$\omega^2 = \omega_{pe}^2 + \frac{5}{3} k^2 v_{th}^2 , \quad (4.47)$$

with the thermal velocity $v_{th} = \sqrt{2k_B T/m_e}$, cf. Fig. 4.6. Both equations are based on the assumption that a local Maxwell equilibrium is established at all times. Thus the plasma must allow for frequent collisions. In space plasmas, in general this is not the case. Here the correct dispersion relation is

$$\omega^2 = \omega_{pe}^2 + 3k^2 v_{th}^2 . \quad (4.48)$$

This equation is called the Bohm–Gross equation [4.2]. From (4.48) the group velocity of plasma waves can be determined as

$$v_g = \frac{d\omega}{dk} = \frac{3k}{\omega} v_{th}^2 = \frac{3v_{th}^2}{v_{ph}} . \quad (4.49)$$

As expected, the group velocity vanishes as the thermal energy of the plasma goes to zero. The group velocity of the wave always is significantly smaller than the thermal speed and thus also much smaller than the speed of light. For large wave numbers k , the information propagates with the thermal velocity.

For small wave numbers, the information travels slower than v_{th} because the density gradient becomes small for large wavelengths and therefore the net flux of momentum into adjacent layers becomes small.

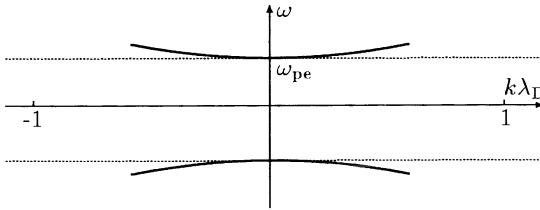


Fig. 4.6. Dispersion relation for electrostatic electron waves in a warm unmagnetized plasma (*Langmuir waves*)

4.3.3 Ion-Acoustic Waves (Ion Waves)

In Langmuir waves, as in plasma oscillations, the ions are assumed to be indefinitely massive and therefore stay fixed at their position. Therefore Langmuir waves are high-frequency waves. If we allow the ions to move, the properties of the wave change. The inertia of the ions requires rather slow oscillations. Therefore, ion waves are low-frequency waves. In ordinary fluids, sound waves are the counterpart of the ion wave. They can be derived from the Navier–Stokes equation with the pressure gradient force being the only term on the right-hand side. The dispersion relation then reads $v_s = \omega/k = \sqrt{\gamma_a p_0 / \rho_0} = \sqrt{\gamma_a k_B T / m}$, with v_s being the speed of sound.

Sound waves are pressure waves that transport momentum from one layer to the next due to collisions between molecules or atoms. Despite its often low density, in a plasma a similar phenomenon exists. Here the momentum is transported by Coulomb collisions, and the information therefore is contained in the charges and the fields. Since we have to consider the motion of both electrons and ions, the ion wave can be derived in the framework of two-fluid theory only. In the equation of motion, we have to consider the pressure gradient force and the force exerted by the electromagnetic field. The dispersion relation then can be derived as

$$\frac{\omega^2}{k^2} = \frac{\gamma_e k_B T_e + \gamma_i k_B T_i}{m} := v_s . \quad (4.50)$$

Note that here $\gamma_e \approx 1$ because the electrons are fast compared with the waves; thus an isothermal distribution can be established easily. The ions, on the other hand, experience a one-dimensional compression in the plane wave and then $\gamma_i = 3$. The group velocity of the ion wave is independent of the wave number k .

If the ion population can be considered as cold, $T_i \rightarrow 0$, and the wave length is small, $k\lambda_D \gg 1$, the ions can oscillate with the ion plasma frequency

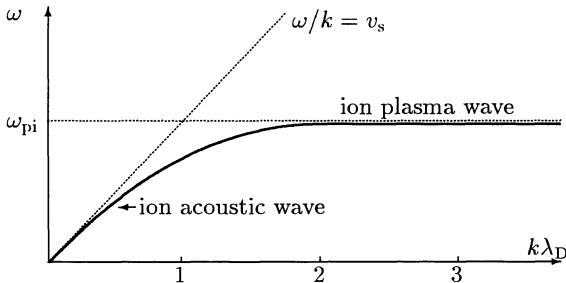


Fig. 4.7. Dispersion relation for ion waves in a warm unmagnetized plasma. The dashed lines are asymptotes for small and large wave numbers

$$\omega^2 = \omega_{\text{pi}}^2 = \frac{n_0 e^2}{m_i} . \quad (4.51)$$

The dispersion relation for ion waves is shown in Fig. 4.7. The ion plasma frequency as an asymptote for the short wavelength is indicated in the right part of the figure. There are fundamental differences in the dispersion relation for electron and ion waves. Electron plasma waves are waves with constant frequency (cf. Fig. 4.6), the thermal motion only adds a small correction. In contrast, ion waves are waves with a constant speed and can exist only if thermal motion is allowed. The phase speed and group speed are identical. The difference between these two types of waves is due to the behavior of the second particle component: in the electron plasma wave, the electrons oscillate while the ions stay fixed. In the ion wave, the ions oscillate but the electrons are in motion, too. In particular, they can be carried along by the ions, screening part of the electric field resulting from the ion oscillation.

In space plasmas, ion acoustic waves are observed upstream of planetary bow shocks where they are generated by suprathermal particles streaming away from the shock front, cf. Sect. 9.4.

4.4 Electrostatic Waves in Magnetized Plasmas

Let us now consider electrostatic waves in a magnetized plasma. They can be divided into waves with \mathbf{k} parallel or perpendicular to \mathbf{B}_o . The terms longitudinal or transversal refer to the direction of the wave vector \mathbf{k} relative to the fluctuating electric field \mathbf{E}_1 . Only longitudinal waves are electrostatic because $\mathbf{k} \times \mathbf{E}_1 = \omega \mathbf{B}_1$ vanishes. In a transverse wave, \mathbf{B}_1 is finite and the wave therefore is an electromagnetic wave. Waves propagating oblique to the field can be regarded as the superimposition of longitudinal and transversal wave modes.

4.4.1 Electron Oscillations Perpendicular to B (Upper Hybrid Frequency)

As for electron oscillations in an unmagnetized plasma, in the upper hybrid oscillations the ions are assumed to be fixed in space, creating a uniform

background of positive charges. The plasma is cold, and so the thermal motion of the electrons can be ignored. The equation of motion then has to consider only the forces exerted by the electric and magnetic fields.

For longitudinal waves, the dispersion relation reads

$$\omega_{\text{uh}}^2 = \omega_{\text{pe}}^2 + \omega_{\text{ce}}^2 , \quad (4.52)$$

with ω_{pe} being the frequency of electron plasma oscillations and ω_{ce} being the electron cyclotron frequency. Only electrostatic waves perpendicular to \mathbf{B} have this upper hybrid frequency. Disturbances parallel to \mathbf{B} oscillate with the plasma frequency ω_{pe} : particles moving parallel to the magnetic field do not gyrate and therefore ω_{ce} vanishes.

We can understand these waves as the superimposition of two motions. In the plane wave, the electrons exhibit regions of compression and rarefaction as in ordinary plasma oscillations. The magnetic field, which is perpendicular to the direction of electron motion, forces the electrons to move around an ellipse instead of oscillating back and forth along a straight line. As in simple plasma oscillations, the electric field accelerates the electrons displaced from their rest position. But as the electron speed increases, the Lorentz force exerted by the magnetic field increases, too, reversing the direction of electron motion. The electrons therefore move against the electric field, losing energy. Thus two restoring forces act on the electrons: the electrostatic force resulting from the electron displacement and the Lorentz force. This additional restoring force leads to a higher frequency as in simple plasma oscillations. If the magnetic field vanishes, the cyclotron frequency vanishes, too, leaving us with ordinary plasma oscillations. If the plasma density decreases, the plasma frequency decreases, too. For vanishing plasma density, the remaining motion is a gyration around the magnetic field line.

4.4.2 Electrostatic Ion Waves Perpendicular to \mathbf{B} (Ion Cyclotron Waves)

The upper hybrid wave is a high-frequency wave with ω much larger than both the plasma and cyclotron frequencies. It results from the motion of electrons in a magnetized plasma. Electrostatic ion waves, like ion acoustic waves, are low-frequency waves. Let us now consider an ion acoustic wave with \mathbf{k} nearly perpendicular to \mathbf{B} . In the equation of motion only the forces of the electromagnetic field have to be considered. The dispersion relation then is

$$\omega_{\text{pi}}^2 = \omega_{\text{ci}}^2 + k^2 v_s^2 , \quad (4.53)$$

with ω_{ci} being the ion cyclotron frequency. An electrostatic ion wave parallel to the magnetic field has the same frequency $\omega^2 = k^2 v_s^2$ as an ion acoustic wave because the ions do not gyrate around the field and therefore ω_{ci} vanishes.

The physical explanation is the same as in the upper hybrid wave: the Lorentz force provides an additional restoring force leading to an elliptical path and a higher frequency, described by the additional term ω_{ci} in (4.53).

4.4.3 Lower Hybrid Frequency

In the ion cyclotron wave we have assumed \mathbf{k} to be nearly perpendicular to \mathbf{B} . The small remaining component of the wave vector and therefore the particle motion parallel to \mathbf{B} is essential because it allows the electrons to travel freely along \mathbf{B} to obtain thermal equilibrium as was required in the derivation of the ion acoustic wave. If this cannot be archived, the equation of motion for the electrons must be solved differently, though that for the ions still is valid. In this case, the lower hybrid wave results with

$$\omega_{lh} = \sqrt{\omega_{ce} \cdot \omega_{ci}} . \quad (4.54)$$

The motion is maintained in the perpendicular direction only. Thus charge neutrality can be maintained in that direction. From this charge neutrality we can understand the frequency given in (4.54). The fluctuating electric field \mathbf{E}_1 of the wave is perpendicular to \mathbf{B}_o . The ions move along \mathbf{E}_1 while the Lorentz force acting on them is rather small. The ion displacement along \mathbf{E}_1 is limited because the electric field oscillates; the maximum displacement is about $\Delta x_i = eE_1/m_i\omega^2$. The electrons gyrate around the magnetic field. In addition, they experience an $\mathbf{E} \times \mathbf{B}$ drift perpendicular to both fields. Their displacement parallel to \mathbf{E}_1 is given roughly as $\Delta x_e = E_1/B_o\omega_{ce}$. Charge neutrality requires $\Delta x_e = \Delta x_i$ and therefore $\omega = \omega_{lh}$.

Lower hybrid waves are of great importance in the auroral regions where they may be responsible for ion heating.

4.5 Electromagnetic Waves in Non-Magnetized Plasmas

Electromagnetic waves consist of both a fluctuating electric field and a fluctuating magnetic field. In this section we will consider electromagnetic fields in an unmagnetized plasma; thus the background field \mathbf{B}_o vanishes and then $\mathbf{B} = \mathbf{B}_1$. Electromagnetic waves are high-frequency waves, and because of their large inertia the ions do not follow the fluctuating field. Electromagnetic waves therefore can be treated within the framework of one-fluid theory. In the equation of motion we have to consider the pressure gradient force and the forces exerted by the electromagnetic field. The dispersion relation for electromagnetic waves then reads

$$\omega^2 = \omega_{pe}^2 + k^2 c^2 , \quad (4.55)$$

with $c = 1/\sqrt{\epsilon_0 \mu_0}$ being the speed of light in a vacuum.

The dispersion relation (4.55) is shown in Fig. 4.8. For plasma frequencies much smaller than the frequency of the waves we get light waves with $\omega = kc$.

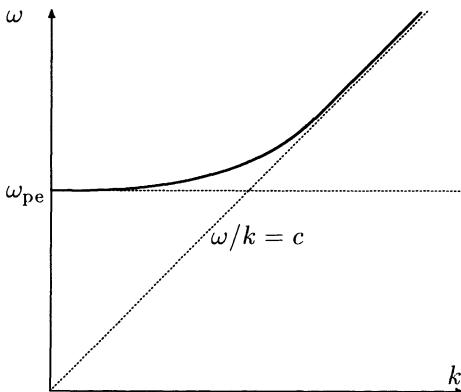


Fig. 4.8. Dispersion relation for electromagnetic waves in a cold unmagnetized plasma. For small wave numbers the group velocity approaches zero and a plasma oscillation results. For large wave numbers both the group and phase speeds converge towards the speed of light

The index of refraction for these waves is $n = c/v_{\text{ph}} = ck/\omega$ or, taking into consideration the electron plasma frequency,

$$n = \sqrt{1 - \frac{\omega_{\text{pe}}^2}{\omega^2}}. \quad (4.56)$$

Waves can propagate through a medium only if n^2 is larger than zero. Thus electromagnetic waves can exist only if $\omega > \omega_{\text{pe}}$. For transverse electromagnetic waves, the ordinary light waves, the cutoff frequency is the electron plasma frequency, the plasma is opaque at lower frequencies.

For $\omega < \omega_{\text{pe}}$ an imaginary refraction index results. For a real frequency an imaginary wave vector k would result. Such a wave would not propagate but decay. The second solution, $-ik$, formally could be interpreted as wave growth. Physically, however, this cannot happen: because we are considering a cold, stationary plasma, the energy required for wave growth cannot be drawn from the plasma. The cold plasma only can absorb the energy of a decaying wave and cannot support wave growth.



Electromagnetic waves can be used for plasma diagnostics in space or in the ionosphere. For instance, the density of a space plasma can be determined by detecting the radio signal of a satellite from another satellite or a ground station. As the frequency of the radio signal is varied, absorption sets in at the plasma frequency of the medium from which the density can be determined.

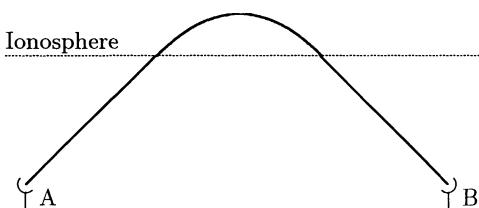


Fig. 4.9. The electron density in the ionosphere can be determined from the reflection of electromagnetic waves

Another application is the measurement of the electron density in the Earth's ionosphere. Radio waves of variable frequency are sent from a transmitter A to a receiver B, cf. Fig. 4.9. As the wave enters the ionosphere, it is refracted according to Snell's law from optics. Thus from the travel time of a radio pulse between transmitter and receiver we can determine the height at which the ray path is bent down towards Earth again. From the geometry, we also can determine the angle at which the electromagnetic wave enters the ionosphere and, from Snell's law, the refraction index. The latter gives us the plasma frequency and also the electron density. Ionospheric reflection is used in short-wave communications because it allows us to send signals around the Earth. With a maximum ionospheric electron density of about 10^{12} m^{-3} , the critical frequency for reflection is about 10 MHz.

The experiment sketched in Fig. 4.9 can be simplified to a vertical geometry in which the transmitter and receiver are located at the same position. Total reflection of the pulse then occurs when the emitted frequency equals the local plasma frequency, i.e. $n = 0$. The travel time of the signal gives the height of reflection. Emitting at different frequencies over a broad frequency band, the height profile of the electrons in the ionosphere can be determined. In principle, this sounding experiment is very simple because reflection occurs exactly at the point where the signal frequency is equal to the local plasma frequency. But a small error remains: the electron density varies continuously with height, leading to changes in the propagation speed of the wave. Therefore, the estimate of the reflection height from the propagation time is not exact. In reality, the situation becomes even more complex because the electromagnetic wave propagates into a magnetized plasma and is split into ordinary and extraordinary modes, both having different propagation speeds.

4.6 Electromagnetic Waves in Magnetized Plasmas

As in the case of electrostatic waves, electromagnetic waves in magnetized plasmas are characterized by the direction of the wave vector relative to the background magnetic field \mathbf{B}_o and to the fluctuating electric field \mathbf{E}_1 .

4.6.1 Electromagnetic Waves Perpendicular to \mathbf{B}_o

Let us start with an electromagnetic wave perpendicular to the undisturbed magnetic field. The wave is assumed to be transversal, and thus $\mathbf{k} \perp \mathbf{E}_1$. Two modes are possible: either the fluctuating electric field is parallel to \mathbf{B}_o (ordinary wave) or it is perpendicular to \mathbf{B}_o (extraordinary wave).

Ordinary Waves ($\mathbf{E}_1 \parallel \mathbf{B}_o$). For $\mathbf{E}_1 \parallel \mathbf{B}_o$ the wave equation takes the same form as in an unmagnetized plasma, leading to the dispersion relation

$$\omega^2 = \omega_{pe}^2 + c^2 k^2 . \quad (4.57)$$

The propagation of ordinary or O-waves therefore is not influenced by the magnetic field: in an ordinary wave, the fluctuating electric field is parallel to \mathbf{B}_o and therefore the magnetic field does not influence the wave dynamics.

Extraordinary Waves ($\mathbf{E}_1 \perp \mathbf{B}_o$). If the fluctuating electric field is perpendicular to \mathbf{B}_o , the electron motion is influenced by the magnetic field and the dispersion relation has to be modified accordingly. As the wave propagates perpendicular to \mathbf{B}_o and \mathbf{E}_1 oscillates perpendicular to \mathbf{B}_o , the wave vector \mathbf{k} has components parallel and perpendicular to \mathbf{E}_1 . The wave therefore is a mixed mode of both a transversal and a longitudinal component. While the ordinary wave is linearly polarized (\mathbf{E}_1 only along \mathbf{B}_o), the extraordinary wave is elliptically polarized. The dispersion relation for the extraordinary wave can be written as

$$\frac{c^2 k^2}{\omega^2} = 1 - \frac{\omega_{pe}^2}{\omega^2} \frac{\omega^2 - \omega_{pe}^2}{\omega^2 - \omega_{uh}^2} = n^2. \quad (4.58)$$

The influence of the magnetic field is contained in the upper hybrid frequency (4.52). The dispersion relation for both ordinary and extraordinary electromagnetic waves is shown in Fig. 4.10. The hatched area indicates the stop band separating two different modes of the extraordinary or X-wave.

A closer look at the refraction index (4.58) helps us to understand these two modes. The refraction index becomes zero for

$$\omega_{(L,R)}^X = \pm \frac{\omega_{ce}}{2} + \sqrt{\omega_{pe}^2 + \frac{\omega_{ce}^2}{4}}, \quad (4.59)$$

with the indices ‘R’ and ‘L’ indicating right-hand and left-hand polarized waves.

The frequencies defined by (4.59) are cutoff frequencies: for lower frequencies the refraction index becomes imaginary and the wave vanishes. The ordinary wave has its cutoff at the electron plasma frequency. The refraction index also can go towards infinity. This happens for very large wave numbers (or very small wavelengths). The corresponding frequency can be found by letting k go towards infinity. This occurs when ω goes towards the upper hybrid frequency ω_{uh} in the denominator of (4.58). As the extraordinary wave approaches this resonance, its phase and group velocities approach zero and the electromagnetic energy is converted into electrostatic oscillations. Therefore, in Fig. 4.10 between the upper hybrid frequency ω_{uh} and the cutoff frequency ω_R a stop band results where, because $n < 0$, the extraordinary wave cannot propagate.

The existence of a stop band has an interesting application. Magnetized plasmas are emitters of radio waves. Therefore, all planets emit radio signals. In the solar system, the largest sources of radio waves is Jupiter. While it is easy to detect these waves, their interpretation in terms of sources and excitation mechanisms is difficult. The stop band, however, allows us to exclude one possible mechanism. The radio waves are excited close to the planet where both the electron plasma frequency as well as the electron cyclotron

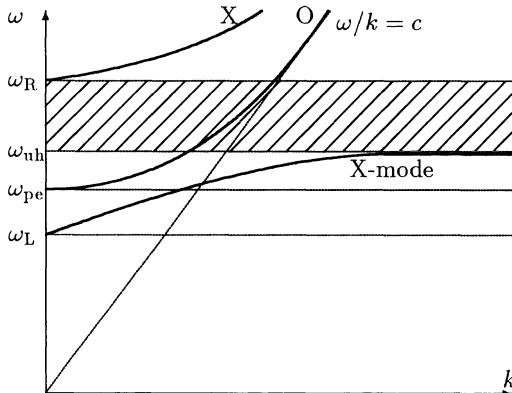


Fig. 4.10. Dispersion relation for ordinary and extraordinary electromagnetic waves in a cold magnetized plasma. The hatched region between ω_{uh} and ω_R separates the two modes of the extraordinary electromagnetic wave

frequency are large. As the waves propagate outward, their frequency decreases. Extraordinary waves excited at a frequency below the upper hybrid frequency will encounter a stop band where ω_{uh} has decreased below the frequency of the wave. These waves cannot escape from the vicinity of the planet. Planetary radio waves observed at large distances therefore have not been excited as extraordinary waves below the local upper hybrid frequency.

4.6.2 Waves Parallel to the Magnetic Field: Whistler (R-Waves) and L-Waves

Let us now consider electromagnetic waves propagating parallel to B_o . These waves are of particular importance in the magnetosphere because they can propagate along B towards the ground. Again, two cases can be distinguished: the right-hand and the left-hand waves. The dispersion relation becomes

$$\frac{\omega^2 - k^2 c^2}{\omega_{pe}^2} \left(1 \pm \frac{\omega_{ce}}{\omega} \right) = 1 \quad (4.60)$$

and the refraction index is

$$n^2 = \frac{c^2 k^2}{\omega^2} = 1 - \frac{\omega_e^2 / \omega^2}{1 \pm \omega_{ce} / \omega} . \quad (4.61)$$

In both equations the '+' sign refers to the right-hand or R-wave. The electric field of an R-wave rotates in the same way as an ordinary screw rotates: if the thumb of the right-hand points in the direction of k , the curved fingers point in the rotation direction of the electric field. Thus the R-wave rotates in the same direction as an electron gyrates. The L-wave rotates in the opposite direction, following the ion gyration. Thus both waves have a resonance. The R-wave has a resonance at $\omega = \omega_{ce}$. The electron gyrating around the magnetic field then experiences a constant electric field which, depending on the phase between the electron and the field, either accelerates or decelerates it continuously. This resonance is called cyclotron resonance.

The phase velocity of the wave goes to zero at the resonance and the wave no longer propagates. The L-wave does not resonate with the electrons because the wave field rotates in the direction opposite to the electron gyration. Thus the L-wave has no resonance at high frequencies, and its resonance would be at $\omega = \omega_{ci}$ if ion motion were considered in addition to electron motion. The cutoff frequencies for L- and R-waves are the same as for the extraordinary waves:

$$\omega_{(L,R)}^{\text{cutoff}} = \pm \frac{\omega_{ce}}{2} \pm \sqrt{\omega_{ce}^2 + \frac{\omega_{pe}^2}{4}} . \quad (4.62)$$

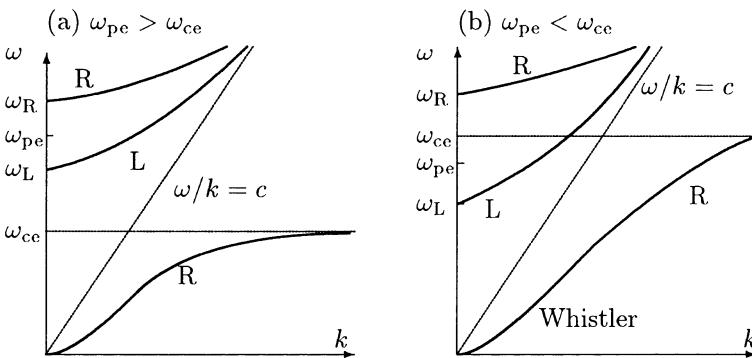


Fig. 4.11. Dispersion relation for electromagnetic waves propagating parallel to the magnetic field in a cold magnetized plasma

The dispersion relation for electromagnetic waves propagating parallel to the magnetic field is shown in Fig. 4.11. Two cases can be distinguished: $\omega_{pe} < \omega_{ce}$ (right panel) and $\omega_{pe} > \omega_{ce}$ (left panel). The main difference between these cases is the cutoff for the left-hand mode: for $\omega_{pe} < \omega_{ce}$ the cutoff is below the electron cyclotron frequency, while in the opposite case it is above. For a constant cyclotron frequency these waves approach the $n^2 = 1$ line for high frequencies. The R-wave has a stop band between ω_R and ω_{ce} . The R-waves in the lower frequency range also are called Whistler waves and are important for propagation studies in the magnetosphere. Whistler waves can be excited by lightning discharge. Thus the source has a short duration but creates a wide spectrum of different frequencies. Since the propagation time depends on the group speed, the first waves arriving at a distant observer have higher frequencies than those arriving later, rather like a whistle with decreasing pitch. The rate of frequency change contains information about the change in plasma density along the propagation path.

4.7 Summary

Table 4.1 summarizes the waves discussed in this chapter. In electrostatic waves only the electric field fluctuates while the magnetic field either is static or zero. In electromagnetic waves, both the electric and magnetic fields oscillate. The waves are further divided into electron and ion waves. In electron waves the electrons oscillate while the ions create a uniform background. Therefore, electron waves are high-frequency waves. In ion waves, both ions and electrons oscillate. Because of the large inertia of the ions, ion waves are low-frequency waves. The propagation of the wave depends on the orientation of the wave vector \mathbf{k} relative to the background magnetic field and relative to the fluctuating electric field.

We should note that this set of dispersion relations is greatly simplified in so far as it considers only waves propagating into the principal directions perpendicular or parallel to the field. Waves propagating oblique to the field are more difficult; however, often they can be understood as the superimposition of the two modes parallel and perpendicular to the field.

Table 4.1. Different types of plasma waves

Wave	Geometry	Dispersion relation	Equation
Electron Waves (Electrostatic)			
Langmuir waves	$\mathbf{B}_o = 0$ or $\mathbf{k} \parallel \mathbf{B}_o$	$\omega^2 = \omega_{pe}^2 + 3k^2 v_{th}^2 / 2$	(4.48)
Upper hybrid waves	$\mathbf{k} \perp \mathbf{B}_o$	$\omega_{uh} = \omega_{pe}^2 + \omega_{ce}^2$	(4.52)
Ion Waves (Electrostatic)			
Ion acoustic waves	$\mathbf{B}_o = 0$ or $\mathbf{k} \parallel \mathbf{B}$	$\omega^2 = k^2 v_s^2$	(4.50)
Ion cyclotron waves	$\mathbf{k} \perp \mathbf{B}_o$	$\omega^2 = \omega_{ci}^2 + k^2 v_s^2$	(4.53)
Lower hybrid waves	$\mathbf{K} \perp \mathbf{B}_o$	$\omega_{lh}^2 = \omega_{ci} \cdot \omega_{ce}$	(4.54)
Electron Waves (Electromagnetic)			
Light waves	$\mathbf{B}_o = 0$	$\omega^2 = \omega_{pe}^2 + k^2 c^2$	(4.55)
O-waves	$\mathbf{k} \perp \mathbf{B}_o$ $\mathbf{E}_1 \parallel \mathbf{B}_o$	$\omega^2 = c^2 k^2 + \omega_{pe}^2$	(4.57)
X-waves	$\mathbf{k} \perp \mathbf{B}_o$ $\mathbf{E}_1 \perp \mathbf{B}_o$	$\omega^2 = c^2 k^2 + \omega_{pe}^2 \cdot (\omega^2 - \omega_{pe}^2) / (\omega^2 - \omega_h^2)$	(4.58)
Whistler (R-waves)	$\mathbf{k} \parallel \mathbf{B}_o$	$\omega^2 = c^2 k^2 - \omega_{pe}^2 / [1 - (\omega_{ce}/\omega)]$	(4.60)
L-waves	$\mathbf{k} \parallel \mathbf{B}_o$	$\omega^2 = c^2 k^2 + \omega_{pe}^2 / [1 + (\omega_{ce}/\omega)]$	(4.60)
Ion Waves (Electromagnetic)			
Alfvén waves	$\mathbf{k} \parallel \mathbf{B}_o$	$\omega^2 = k^2 v_A^2$	(4.30)
Magneto-sonic waves	$\mathbf{k} \perp \mathbf{B}_o$	$\omega^2 = c^2 k^2 \cdot (v_s^2 + v_A^2) / (c^2 + v_A^2)$	(4.40)

Some of these waves can be used for plasma diagnostics, with the sources of the waves either being natural, such as in Whistler waves to study propagation in the magnetospheric plasma, or artificial, such as radio waves to probe the ionosphere.

Exercises and Problems

4.1 Explain, in your own words, the important quantities characterizing a wave. What are group and phase speeds?

4.2 Show that in an electron plasma wave the energy contained in the electron oscillation exceeds the energy in the ions by the mass ratio $m_i/(Z_i m_e)$.

4.3 On re-entry into the Earth's atmosphere, a spacecraft experiences a radio blackout due to the shock developing in front of the spacecraft. Determine the electron density inside the shock if the transmitter works at 300 MHz.

4.4 Show that the maximum phase speed of a Whistler wave is at a frequency $\omega = \omega_{ce}/2$. Prove that this is below the speed of light.

4.5 Show that if a packet of Whistler waves with a spread in frequency is generated at a given instant, a distant observer will receive the higher frequencies earlier than the lower ones.

4.6 Show that if the finite mass of the ions is included, the frequency of Langmuir waves in a cold plasma is given by $\omega^2 = \omega_{pe}^2 + \omega_{pi}^2$.

4.7 How would you use pulse delay as a function of frequency to measure the average plasma density between the Earth and a distant pulsar?

4.8 Determine the Alfvén speeds and the electron plasma frequencies for the situations described in Problem 3.11.

4.9 Use Fig. 4.4 to describe the properties of magnetohydrodynamic waves propagating parallel to \mathbf{B}_o for $v_A > v_s$ and $v_A < v_s$.

4.10 Show that in an Alfvén wave the average kinetic energy equals the average magnetic energy.

4.11 Discuss an Alfvén wave with $\mathbf{k} \parallel \mathbf{B}_o$. (a) Determine the dispersion relation under the assumption of a high but finite conductivity (the displacement current nevertheless can be ignored). (b) Determine the real and the imaginary parts of the wave vector for a real frequency.

5. Kinetic Theory

All animals are equal
but some animals
are more equal
than other.

G. Orwell, *Animal Farm*

Magnetohydrodynamics has treated the plasma as a fluid with all particles moving at the same speed, the bulk speed. The thermal motion is ignored and the plasma is described as a cold plasma. But space plasmas often are hot plasmas with the thermal speed by far exceeding the flow speed. Thus thermal motion has to be taken into account and the plasma has to be described by a distribution function which considers the positions and velocities of the individual particles. The Boltzmann equation gives the equation of motion for this phase space density.

5.1 The Distribution Function

Kinetic theory starts from the physics of individual particles (the microscopic approach). The macroscopic phenomena then can be described by averaging over a sufficiently large number of particles, an approach which also is used in statistical mechanics. This formalism therefore is also called the statistical description of a plasma.

5.1.1 Phase Space and Distribution Function

The mechanical properties of each particle are described completely by its position and momentum. The phase space is a six-dimensional space defined by the three spatial coordinates q_1, q_2, q_3 and the three generalized momenta p_1, p_2, p_3 . Each particle is related unambiguously to one point in phase space:

$$\mathbf{Q} = (q_1, q_2, q_3; p_1, p_2, p_3) = (\mathbf{q}, \mathbf{p}) . \quad (5.1)$$

The speed of the particle in phase space, i.e. the combined change in its position and momentum in ordinary three-dimensional space, then is

$$\mathbf{C} = \frac{d\mathbf{Q}}{dt} = \left(\frac{dq_1}{dt}, \frac{dq_2}{dt}, \frac{dq_3}{dt}; \frac{dp_1}{dt}, \frac{dp_2}{dt}, \frac{dp_3}{dt} \right) = \left(\frac{d\mathbf{q}}{dt}, \frac{d\mathbf{p}}{dt} \right). \quad (5.2)$$

For a plasma consisting of N particles, in phase space N such points exist. The phase space density $f(\mathbf{q}, \mathbf{p}, t)$ gives the number of particles inside a volume element $[(q_i, q_i + dq_i), (p_i, p_i + dp_i)]$. This density function is also called the distribution function of the plasma. The particle density in ordinary three-dimensional space is related to it by

$$n(\mathbf{q}, t) = \int_{-\infty}^{+\infty} f(\mathbf{q}, \mathbf{p}, t) d^3\mathbf{p}. \quad (5.3)$$

In contrast to the phase space density, the number density still depends on space and time. It is used in the definition of averages, which describe the macroscopic properties of the plasma. If $a(\mathbf{q}, \mathbf{p}, t)$ is a function in phase space, its average is defined as

$$\langle a(\mathbf{q}, t) \rangle = \frac{1}{n(\mathbf{q}, t)} \int_{-\infty}^{\infty} a(\mathbf{q}, \mathbf{p}, t) \cdot f(\mathbf{q}, \mathbf{p}, t) d^3\mathbf{p}. \quad (5.4)$$

The average speed of the plasma, for instance, is given as

$$\mathbf{u}(\mathbf{q}, t) = \langle \mathbf{v}(\mathbf{q}, t) \rangle = \frac{1}{n} \int_{-\infty}^{\infty} \mathbf{v}(\mathbf{p}, \mathbf{q}, t) \cdot f(\mathbf{q}, \mathbf{p}, t) d^3\mathbf{p}. \quad (5.5)$$

Application of this averaging scheme to the equation of motion yields the Vlasov equation (5.22) which is the basic equation of statistical plasma physics and can be used to derive the MHD equations of two-fluid theory.

5.1.2 Maxwell's Velocity Distribution

The average speed $\langle \mathbf{v} \rangle$ of the particles in a plasma, as defined by (5.5), is also the plasma flow speed \mathbf{u} . The speeds \mathbf{v}_i of individual particles, however, can be substantially different; in particular, speeds of individual particles can exceed the flow speed by orders of magnitude. A plasma contains different particle species s which all have their own average speed \mathbf{u}_s . If in an electron-proton plasma the average speeds \mathbf{u}_e and \mathbf{u}_p are different, a current results.

To derive the velocity distribution of the particles, let us first determine the kinetic energy contained in a volume element of the plasma. The kinetic energy of the plasma flow is determined by the average speed \mathbf{u} , while the entire kinetic energy is the sum of the kinetic energies of all particles. Since the latter is larger, there is also a kinetic energy contained in the stochastic motion of the particles, which can be described by

$$\left\langle \frac{m}{2} (\mathbf{v} - \mathbf{u})^2 \right\rangle = \frac{\int m(\mathbf{v} - \mathbf{u})^2 f(\mathbf{r}, \mathbf{v}, t) d\mathbf{v}}{2 \cdot \int f(\mathbf{r}, \mathbf{v}, t) d\mathbf{v}}. \quad (5.6)$$

This kinetic energy is related to the hydrostatic pressure by

$$\frac{p}{n} = \left(\frac{2}{N} \right) \left\langle \frac{m}{2} (\mathbf{v} - \mathbf{u}_s)^2 \right\rangle , \quad (5.7)$$

with N being the number of degrees of freedom, normally three.

If the system is in thermal equilibrium, which especially in a hot plasma is not necessarily the case because times between collisions can be quite long, the velocity distribution is given by the Maxwell distribution:

$$f(\mathbf{r}, \mathbf{v}) = n \sqrt{\left(\frac{m}{2\pi k_B T} \right)^3} \exp \left\{ -\frac{m_2(\mathbf{v} - \mathbf{u})^2}{2k_B T} \right\} . \quad (5.8)$$

Here T is the plasma temperature and k_B is Boltzmann's constant. According to (5.8), the relative number of particles with large stochastic speeds $|\mathbf{v} - \mathbf{u}|$ increases with T . The maximum of the distribution is at the most probable thermal speed v_{th} :

$$v_{\text{th}} = \sqrt{\frac{2k_B T}{m}} . \quad (5.9)$$

In a one-atomic gas in equilibrium, the temperature is related to the kinetic energy of the stochastic motion by

$$\left\langle \frac{1}{2} m (\mathbf{v} - \mathbf{u})^2 \right\rangle = \frac{1}{2} N k_B T . \quad (5.10)$$

In a plasma, N normally is 3. Even if particles are moving in a magnetic field, N is 3 because the particle speed is described completely only by giving one of the following triples: either $v_x - u_x$, $v_y - u_y$, and $v_z - u_z$ or v_{\parallel} , $|\mathbf{v}_{\perp} - \mathbf{u}|$, and ψ_v , the latter describing the direction of the perpendicular speed relative to the gyro-center. Note that the combination of (5.7) and (5.10) yields the ideal gas law $p = nk_B T$.

Occasionally, we are concerned only with particle speeds and not with the direction of motion. This might be the case if the plasma is at rest, i.e. \mathbf{u} equals zero. The distribution function (5.8) then is

$$\iiint f(\mathbf{r}, \mathbf{v}, t) d\Omega_v v^2 dv = (4\pi f(\mathbf{r}, |v|, t) v^2) dv = g(\mathbf{r}, v) dv . \quad (5.11)$$

This gives the number of particles inside a volume element with speeds between v and $v + dv$. The function $g(\mathbf{r}, v)$ gives the number of particles per velocity unit, again with speeds between v and $v + dv$. For small speeds, this function increases with the square of v while for large speeds it decreases exponentially (cf. upper panel in Fig. 5.1). If not the speed, but only one component of the speed is considered, the distribution is symmetric around zero if the plasma is at rest (middle) or symmetric around the flow speed in that particular direction if the plasma is in motion (lower panel).

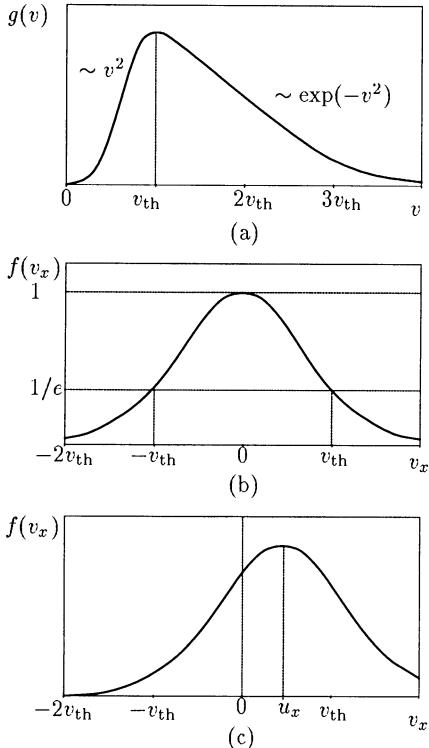


Fig. 5.1. Maxwell distribution. (a) Distribution $g(v, r)$ for the particle speeds in a plasma at rest. (b) Maxwell distribution of one velocity component for a plasma at rest, the other velocity components have been removed by integration. (c) Same as above but for a plasma moving with average speed u_x in the x -direction. Note that an increase in temperature would not affect the position of the maximum in (b) and (c) but would shift the maximum towards the right in (a), combined with an increase of the maximum

5.1.3 Other Distributions

Particle distributions always cannot be described by a Maxwellian. If the plasma is not in equilibrium, normally no analytical distribution function can be given. But on many occasions, distribution functions can at least be good approximations on the observations.

For instance, if in a plasma there is a pronounced difference in the speeds parallel and perpendicular to the magnetic field, the distribution function often can be approximated by a product of two Maxwell distributions which take into account the different speeds and temperatures of the two motions:

$$f(\mathbf{r}, \mathbf{v}) = n \sqrt{\left(\frac{m}{2\pi k_B T} \right)^2 \frac{T^{3/2}}{T_\perp \sqrt{T_\parallel}}} \times \exp \left\{ -\frac{(v_\parallel - u_\parallel)^2}{k_B 2T_\parallel} \right\} \exp \left\{ -\frac{m(v_\perp - u_\perp)^2}{2k_B T_\perp} \right\}. \quad (5.12)$$

The average kinetic energies parallel and perpendicular to the field are

$$\left\langle \frac{1}{2}m(v_\parallel - u_\parallel)^2 \right\rangle = \frac{1}{2}k_B T_{\parallel s} \quad \text{and} \quad \left\langle \frac{1}{2}m(v_\perp - u_\perp)^2 \right\rangle = k_B T_\perp. \quad (5.13)$$

Here the difference in the number of degrees of freedom for the parallel and the perpendicular motion becomes obvious: there is one degree of freedom parallel to the field while there are two perpendicular to it. Equation (5.13) can also be used to define the pressure perpendicular and parallel to \mathbf{B} .

Occasionally, the particle distribution can be described by a Maxwellian up to a certain energy. At higher energies the particle distribution can be fitted much better by a power law than by the exponential decay of the Maxwell distribution. Here the kappa distribution,

$$f(\mathbf{r}, \mathbf{v}) = A_\kappa \left[1 + \frac{m(\mathbf{v} - \mathbf{u})^2}{2\kappa E_T} \right]^{-\kappa-1}, \quad (5.14)$$

can be used as an approximation. Instead of the temperature T , the parameters κ and E_T are characteristics of the distribution, with E_T being closely associated with the temperature and κ describing the deviation of the distribution from a Maxwellian. For energies $E \gg \kappa E_T$, the distribution decays more slowly than a Maxwellian. The decay can then be described by a power law in kinetic energy with κ being the power law index: $f(E) = f_0 \cdot E^{-\kappa}$. If κ converges towards infinity, the distribution is a Maxwellian with temperature $kT = E_T$. Note that for the kappa distribution, as well as for all other non-Maxwellian distributions, (5.10) cannot be applied.

5.1.4 Distribution Function and Measured Quantities

While the distribution function is important for our theoretical treatment of (space) plasmas, it is a quantity which cannot be measured directly. Instead, observations give the differential flux $J(E, \Omega, \mathbf{r})$ of particles within a solid angle $d\Omega$ and an energy interval $(E, E + dE)$. Thus the quantity

$$J(E, \Omega, \mathbf{r}) dA d\Omega dE dt$$

is the number of particles in the energy band from E to $E + dE$ coming from the direction Ω within a solid angle $d\Omega$, going through a surface dA perpendicular to $d\Omega$ during the time interval dt . The differential flux therefore can be measured in units of particles per $(\text{m}^2 \text{ sr s MeV})$. Since J depends on Ω , it can also be interpreted as the angular distribution of the particles. On a rotating spacecraft, often the omnidirectional intensity is measured. The latter can be obtained by averaging over all directions:

$$J_{\text{omni}}(E, \mathbf{r}) = \frac{1}{4\pi} \int J(E, \Omega, \mathbf{r}) \cdot d\Omega. \quad (5.15)$$

The number density of particles with velocity v in a phase space element is given as $dn = fv^2 dv d\Omega$. Multiplication by v gives the differential flux of particles with velocity v as $f(\mathbf{v}, \mathbf{p})v^3 dv d\Omega$. Comparison with the same quantity expressed by the differential flux yields

$$J(E, \Omega, \mathbf{r}) dE d\Omega = f(\mathbf{r}, \mathbf{p})v^3 dv d\Omega. \quad (5.16)$$

Since the energy is related to speed, dE is related to dv by $dE = mvdv$. The relation between the differential flux and the distribution function therefore can be written as

$$J(E, \Omega, \mathbf{r}) = \frac{v^2}{m} f(\mathbf{r}, \mathbf{p}) . \quad (5.17)$$

5.2 Basic Equations of Kinetic Theory

As mentioned above, the equation of motion in kinetic theory can be derived by applying the averaging scheme (5.4). The most important equations are the Boltzmann equation, the fundamental equation of motion in kinetic theory, the Vlasov equation, which can be applied if the forces are purely electromagnetic, and the Fokker–Planck equation considering scattering.

5.2.1 The Boltzmann Equation

The Boltzmann equation is the fundamental equation of motion in phase space. It is not limited to plasmas, and the only assumption inherent in the Boltzmann equation is that only external forces \mathbf{F} might act on the particles while internal forces vanish: there are no collisions between the particles.

The Boltzmann equation is a direct consequence of the equation of continuity in phase space:

$$\frac{\partial f}{\partial t} + \nabla_6(f\mathbf{C}) = 0 \quad \text{or} \quad \frac{\partial f}{\partial t} + \nabla_{\mathbf{r}} \cdot (\mathbf{v}f) + \nabla_{\mathbf{v}} \cdot (\mathbf{a}f) = 0 , \quad (5.18)$$

where ∇_6 , $\nabla_{\mathbf{r}}$, and $\nabla_{\mathbf{v}}$ are the divergence in phase space, in ordinary space, and in momentum space, respectively, and \mathbf{v} and \mathbf{a} are velocity and acceleration. In phase space, \mathbf{r} and \mathbf{v} are independent variables. If we further assume that the acceleration \mathbf{a} , and therefore also the force \mathbf{F} , is independent of \mathbf{v} , (5.18) can be simplified:

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{r}} f + \mathbf{a} \cdot \nabla_{\mathbf{v}} f = 0 \quad \text{or} \quad \frac{\partial f}{\partial t} + \mathbf{v} \nabla f + \frac{\mathbf{F}}{m} \cdot \frac{\partial f}{\partial \mathbf{v}} = 0 . \quad (5.19)$$

Equation (5.19) is called the collisionless Boltzmann equation. It also can be written as

$$\frac{df}{dt} = 0 , \quad (5.20)$$

which states that the convective derivative of the phase space density is always zero for a collisionless assembly of particles. Thus for an observer moving with the flow, the phase space density is constant. Or, in other words: the substrate of points in phase space behaves like an incompressible fluid. This is also called Liouville's theorem.

The general form of the Boltzmann equation can be written as

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla f + \frac{\mathbf{F}}{m} \cdot \frac{\partial f}{\partial \mathbf{v}} = \left(\frac{\partial f}{\partial t} \right)_{\text{coll}}, \quad (5.21)$$

where the term on the right-hand side is the rate of change in phase space density due to collisions.

5.2.2 The Vlasov Equation

The Vlasov equation is the application of the Boltzmann equation to a plasma on which only electromagnetic forces act. These forces are described by the Lorentz force (2.22). In the derivation of (5.19) we have made the assumption of a force independent on \mathbf{v} . At first glance the Lorentz force violates this assumption and therefore should not be considered in (5.19). Closer inspection, however, shows that this is not true. Since the Lorentz force contains the cross product of speed and magnetic field, the resulting force is perpendicular to the speed. Thus each individual component of the force does not depend on the same component of the velocity. Since in (5.18) a scalar product is considered, the only derivatives of \mathbf{a} are of the form $\partial a_x / \partial v_x$ and therefore vanish. Thus the Lorentz force can be inserted into (5.19):

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla f + \frac{q}{m} \left(\mathbf{E} + \frac{\mathbf{v} \times \mathbf{B}}{c} \right) \cdot \frac{\partial f}{\partial \mathbf{v}} = 0. \quad (5.22)$$

Equation (5.22) is called the Vlasov equation. Because of its simplicity, this is the equation most commonly studied in kinetic theory.

The Vlasov equation is derived under the assumption of non-interacting particles. On the other hand, interactions are the very essence of a plasma. Thus we have to discuss whether the Vlasov equation can be applied as often as it is. As we will see, the Vlasov equation is a valid approach. It does not consider collisions in the sense of short-range, local interactions, such as collisions between two billiard balls or Coulomb collisions between two charged particles. Nonetheless, that kind of interaction, which is essential in a plasma, is considered: each particle moves in the average Coulomb field created by thousands of other particles. Thus the fields in the Vlasov equation are due to the rest of the plasma and describe the interaction of the particles. These fields often are called self-consistent fields. External fields can be included in (5.22), too. Since the fields \mathbf{E} and \mathbf{B} are determined by the rest of the plasma, they depend on the distribution function f .

The Vlasov equation thus is non-linear; analytical solutions in general are not possible. But Jeans' theorem identifies some solutions. It states: any function of the constants of motion is a solution of the Vlasov equation. For instance, if there are no electric fields, the kinetic energy is a constant of motion. Thus any function of $mv^2/2$ is a solution of the Vlasov equation. In particular, the Maxwell distribution (5.8) is a solution.

Jeans' theorem therefore shows the equivalence of kinetic theory and orbit theory. Following an approach given by Boyd and Sanderson [5.2], this

equivalence can be shown easily. The basic equation of orbit theory is Newton's second law $\mathbf{F} = md^2\mathbf{r}/dt^2$. This is a second-order differential equation in three dimensions and therefore the general solution must contain six constants of integration, $\gamma_1, \dots, \gamma_6$. Thus the solutions of Newton's second law can be written as $\mathbf{r} = \mathbf{r}(\gamma_1, \dots, \gamma_6, t)$ and $\mathbf{v} = \mathbf{v}(\gamma_1, \dots, \gamma_6, t)$. These six scalar equations can be solved in principle to give the γ_i : $\gamma_i = \gamma_i(\mathbf{v}, \mathbf{r}, t)$. Jeans' theorem then states that each function $f = f(\gamma_1, \dots, \gamma_6)$ is a solution of the fundamental equation of kinetic theory, the Boltzmann equation (5.19). This can be seen easily by inserting the γ_i into the Boltzmann equation:

$$\sum_i \left(\frac{\partial \gamma_i}{\partial t} + \mathbf{v} \cdot \nabla \gamma_i + \frac{\mathbf{F}}{m} \cdot \frac{\partial \gamma_i}{\partial \mathbf{v}} \right) = \sum_i \frac{\partial f}{\partial \gamma_i} \frac{d\gamma_i}{dt} = 0. \quad (5.23)$$

The result is identically zero because the γ_i are constants.

5.2.3 The Fokker–Planck Equation

In contrast to the Vlasov equation, the Fokker–Planck equation considers the short-range, local interactions between plasma particles. Collisions arise from many small Coulomb interactions between charged particles (see Sect. 5.3.2). The collision term has its mechanical analogy in the Brownian motion of particles in a gas; however, both are not equivalent as will be discussed below.

Collisions are not a deterministic but a stochastic process. Thus for a given particle, although we might know its momentary position and velocity, we cannot determine its future motion. Only for an assembly of particles can we determine the collective behavior. This can be done by means of probabilities. Let $\psi(\mathbf{v}, \Delta\mathbf{v})$ be the probability that a particle with velocity \mathbf{v} after many small collisions during a time interval dt has changed its velocity to $\mathbf{v} + \Delta\mathbf{v}$. The phase space density $f(\mathbf{r}, \mathbf{v}, t)$ also is a probability function. At a time t it can be written as the product of the phase space density at an earlier time $t - \Delta t$ multiplied by the probability of changes during this time interval and integrated over all possible velocities \mathbf{v} :

$$f(\mathbf{r}, \mathbf{v}, t) = \int f(\mathbf{r}, \mathbf{v} - \Delta\mathbf{v}, t - \Delta t) \psi(\mathbf{v} - \Delta\mathbf{v}, \Delta\mathbf{v}) d(\Delta\mathbf{v}). \quad (5.24)$$

Since we only consider scattering by small angles, i.e. $|\Delta\mathbf{v}| \ll |\mathbf{v}|$, Taylor expansion to second order of the product $f\psi$ yields

$$\begin{aligned} f(\mathbf{r}, \mathbf{v}, t) &= \int \left[f(\mathbf{r}, \mathbf{v}, t - \Delta t) \psi(\mathbf{v}, \Delta\mathbf{v}) - \Delta\mathbf{v} \cdot \frac{\partial(f\psi)}{\partial\mathbf{v}} \right] d(\Delta\mathbf{v}) \\ &\quad + \int \left[\frac{\Delta\mathbf{v}\Delta\mathbf{v}}{2} \odot \frac{\partial^2(f\psi)}{\partial\mathbf{v}\partial\mathbf{v}} \right] d(\Delta\mathbf{v}). \end{aligned} \quad (5.25)$$

Note that the \odot in the last term indicates a product between two tensors.¹

Because some interactions always take place, the probability can be normalized to $\int \psi d(\Delta v) = 1$. Equation (5.25) than can be simplified to

$$f(\mathbf{r}, \mathbf{v}, t) = f(\mathbf{r}, \mathbf{v}, t - \Delta t) - \frac{\partial}{\partial \mathbf{v}}(f\langle \Delta \mathbf{v} \rangle) + \frac{1}{2} \frac{\partial}{\partial \mathbf{v} \partial \mathbf{v}} \odot (f\langle \Delta \mathbf{v} \Delta \mathbf{v} \rangle), \quad (5.26)$$

with

$$\langle \Delta \mathbf{v} \rangle = \int \psi \Delta \mathbf{v} d(\Delta \mathbf{v}) \quad \text{and} \quad \langle \Delta \mathbf{v} \Delta \mathbf{v} \rangle = \int \psi \Delta \mathbf{v} \Delta \mathbf{v} d(\Delta \mathbf{v}). \quad (5.27)$$

By definition, the collision term as written down in (5.21) is

$$\left(\frac{\partial f}{\partial t} \right)_{\text{coll}} = \frac{f(\mathbf{r}, \mathbf{v}, t) - f(\mathbf{r}, \mathbf{v}, t - \Delta t)}{\Delta t}. \quad (5.28)$$

Thus the Fokker–Planck equation can be written as

$$\left(\frac{\partial f}{\partial t} \right)_{\text{coll}} \Delta t = -\frac{\partial}{\partial \mathbf{v}} \cdot (f\langle \Delta \mathbf{v} \rangle) + \frac{1}{2} \frac{\partial^2}{\partial \mathbf{v} \partial \mathbf{v}} \odot (f\langle \Delta \mathbf{v} \Delta \mathbf{v} \rangle). \quad (5.29)$$

The first term on the right-hand side basically contains $\langle \Delta \mathbf{v} \rangle / \Delta t$, which is an acceleration. Thus the term describes the frictional forces leading to an acceleration of the slower particles and a deceleration of the faster particles, which tends to equalize the speeds of the particles. The negative divergence in velocity space describes this narrowing of the distribution function. The second term, $\langle \Delta \mathbf{v} \Delta \mathbf{v} \rangle / \Delta t$, is a diffusion in velocity space. This term describes the broadening of a narrow velocity distribution, e.g. a beam, as a result of the collisions. The two terms therefore operate in the opposite sense. They are in balance in an equilibrium distribution, e.g. the Maxwell distribution. The physics of the collision processes is contained in the probability function ψ .

Figure 5.2 shows the evolution of a suprathermal distribution of particles. At time t_0 the particle distribution is a parallel beam of uniform speed. Since the particles are much faster than the plasma, at early times (t_1 and t_2) pitch-angle scattering dominates, the distribution spreads towards larger pitch angles, leading finally to an isotropic ring distribution around the origin (t_3). With increasing time, friction and energy losses become important and the ring distribution contracts towards the origin. This is an example of a more general rule of thumb: the faster a particle moves through a plasma, the smaller is its frictional drag. This longevity property allows the existence of suprathermal particles as a distinct population in some cosmic plasmas.

¹ Tensors stem from tensor products (dyads) \mathbf{ab} of two vectors \mathbf{a} and \mathbf{b} :

$$\mathbf{T} = \mathbf{ab} = \begin{pmatrix} a_x b_x & a_x b_y \\ a_y b_x & a_y b_y \end{pmatrix}.$$

The product $\mathbf{S} \odot \mathbf{T}$ of two tensors \mathbf{S} and \mathbf{T} itself is a tensor and can be obtained by application of the rules of matrix multiplication.

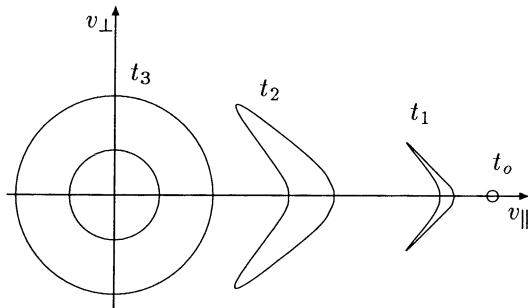


Fig. 5.2. Evolution of a suprathermal distribution of particles from a monoenergetic beam at time t_o towards an isotropic ring distribution at later times t_3

5.3 Collisions



We have mentioned collisions twice in this chapter. In Sect. 5.1.2 we introduced the Maxwell distribution. The basic requirement for such a distribution is thermal equilibrium, which in turn requires collisions between the particles. If we have a distribution with a suprathermal tail, such as the kappa distribution, in time collisions will transform it into a Maxwell distribution. This time scale depends, of course, on the time scales of the collisions. We have also mentioned collisions in connection with the basic equations of kinetic theory, in particular the Fokker–Planck equation. We have even mentioned that the collisions should lead to small changes in speed only. But we did not talk explicitly about the nature of these collisions. This section is a supplementary, briefly introducing some of the basics of collisions.

Collisions are also important in the energy transfer between different components in a plasma: imagine a plasma which also contains neutral particles. The charged particles might be accelerated by an electric field. In time, collisions between charged and neutral particles will equalize the two distributions, leading to an acceleration of the neutrals.

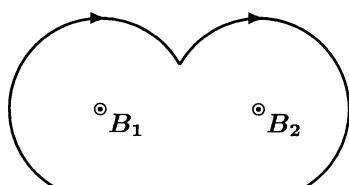


Fig. 5.3. Collisions in a plasma can shift the gyro-center of a particle onto another field line, here illustrated for a collision of a charge particle with a neutral

As in a neutral gas, collisions in a plasma change the path of the individual particles. In a magnetized plasma, collisions between a charged particle and a neutral can shift the gyro-center of a particle onto another field line, cf. Fig. 5.3. Collisions between charged particles can also lead to a shift in the gyro-center and/or changes in pitch angle.

5.3.1 Collisions Between Neutrals

Collisions between neutral particles give rise to the Brownian motion in a gas. The individual process is a collision between two hard spheres. The basic equations are the conservation of momentum and the conservation of energy. The changes in momentum and direction depend on the masses and speeds of the particles and on the angle between their velocities. The change in momentum is largest in a head-on collision: here the particle loses twice its initial momentum as its velocity is reversed. Thus scattering by a large angle up to 180° is possible in collisions between neutrals.

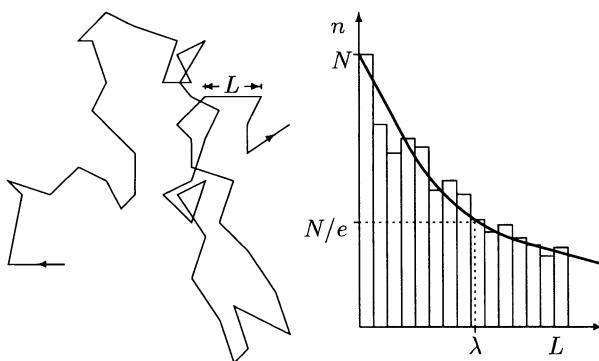


Fig. 5.4. Statistical path of a particle under the influence of collisions (left) with other particles (Brownian motion) and distribution of the path length between two collisions (right)

The relevant parameters are the mean free path and the scattering cross-section. The particle mean free path λ is defined as the average distance traveled by a particle between two subsequent collisions. If we could follow a smoke particle in air under a microscope, we would detect a path similar to the one depicted on the left-hand side of Fig. 5.4. The statistical motion is composed of many straight lines with different length L . The right-hand side of Fig. 5.4 shows the distribution of these L . This probability distribution can be described by a function $p(L) = a \cdot \exp(-L/\lambda)$ with a being a constant depending on the total number of collisions and λ being the particle mean free path.

The number of collisions, and therefore the mean free path, depends on the number density of particles and on their ‘size’ as described by the scattering cross-section. Consider a fast particle with radius r_1 moving in a gas of slow particles with radii r_2 . A collision happens if the distance between the two particles has decreased below $r_1 + r_2$. Alternatively, we can assume the fast particle to be a mass point. Then we have to attribute a radius of $r_1 + r_2$ to the gas molecules. Thus for the fast particle, a gas molecule is equivalent to a disk with the scattering cross-section $\sigma = \pi(r_1 + r_2)^2$.

Now consider a beam of particles incident on a slab of area A and thickness dx . The number density of molecules in this slab is n_s , the total number of molecules in the slab is $n_s A dx$. The fraction of the slab blocked by atoms

therefore is $\sigma n_s A dx / A = \sigma n_s dx$. Out of N particles incident on the slab, $\Delta N = N n_s \sigma dx$ will experience a collision, leading to a reduction in N according to $dN/N = -\sigma n_s dx$. Integration yields

$$N(x) = N_0 \cdot \exp(-\sigma n_s x) = N_0 \cdot \exp(-x/\lambda), \quad (5.30)$$

with the mean free path λ defined as

$$\lambda = \frac{1}{n_s \sigma}. \quad (5.31)$$

Thus, the mean free path can also be interpreted as the distance over which the number of particles decreases to $1/e$ of its initial value. After traveling a distance λ , the particle will have a high probability of colliding. The average time between two collisions is

$$\tau = \frac{\lambda}{\langle v \rangle} = \frac{1}{n_s \sigma \langle v \rangle}. \quad (5.32)$$

This also can be written as a collision frequency ν_c :

$$\nu_c = n_s \sigma \langle v \rangle. \quad (5.33)$$

The formalism for interactions between a charged particle and a neutral is the same as for a collision between two neutral particles.

5.3.2 Collisions Between Charged Particles

Collision between charged particles do not require a direct contact, instead the interaction takes place as each particle is deflected in the electric field of the other charge. This Coulomb force has a long range and therefore leads to a gradual deflection. Nonetheless, one can derive a kind of cross-section for this process. Following the attempt given in Chen [5.3], we will only make an order-of-magnitude estimate.

The geometry of a Coulomb collision is shown in Fig. 5.5: an electron with velocity v approaches an ion with charge e . If no Coulomb force acts, the electron will pass the ion at a distance r_o , the impact parameter. But the Coulomb force $F = -e^2/(4\pi r^2)$ leads to a deflection of the electron from its original direction by an angle φ . The force acts on the electron for a time $T \approx r_o/v$ when it is in the vicinity of the ion. The change in electron momentum then can be approximated by $\Delta p = |FT| \approx e^2/(4\pi r_o v)$. For a 90° collision, the change in momentum is of the order of mv . Thus it is $\Delta p \cong mv \cong e^2/(4\pi r_o v)$. To get a deflection by 90° , the impact parameter therefore must be $r_{90^\circ} = e^2/(4\pi m v^2)$. The cross-section for a deflection of at least 90° therefore can be written as

$$\sigma_{>90^\circ} = \pi r_o^2 = \frac{e^4}{16\pi m^2 v^4}, \quad (5.34)$$

leading to a collision frequency of

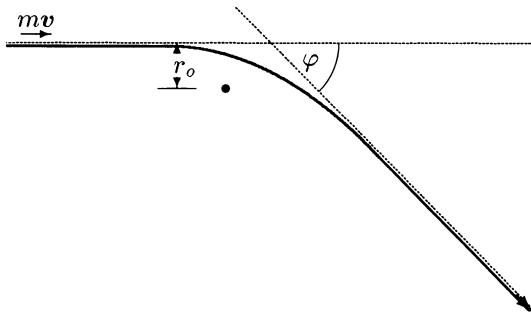


Fig. 5.5. Coulomb scattering: orbit of an electron in the Coulomb field of an ion

$$\nu_{ei,>90^\circ} = n\sigma v = \frac{ne^4}{16\pi m^2 v^3}. \quad (5.35)$$

In a real plasma the situation is more complex. Let us consider the motion of one particle, a test particle, in the field created by the other particles, the field particles. The fields of these particles add to a stochastic field changing continuously in time and space. The test particle therefore will not move in a hyperbolic orbit as in the interaction between two charged particles, instead it basically follows its original direction of motion, though not on a straight line but on a jittery trajectory. Because of the stochastic nature of the collisions, test particles with nearly identical start conditions will diverge in space and velocity. Most of these collisions result in small changes in the particle path only. Occasionally, also large deviations of the original direction result. These are called large-angle collisions.

To understand the different types of collisions, we have to consider the typical spatial scales. One characteristic scale is the Debye length (3.120): the test particle is screened from the electric field of the charges outside the Debye sphere. The Debye length can be interpreted as the range of microscopic electric fields and separates the field particles into two groups: (i) particles at distances larger than the Debye length can influence the test particle by the macroscopic fields only, leading to gyration, drift or oscillations; and (ii) particles inside the Debye sphere create a microscopic field leading to the stochastic motion of the test particle.

The other spatial scale is related to the scattering angle, which in turn is related to the impact parameter. In the derivation of the Fokker–Planck equation we have assumed scattering by small angles only. The deflection angle will be small if the kinetic energy $m_T v_T^2/2$ of the test particle is large compared with the electrostatic potential $Z_T Z_F e^2/r_o$, with r_o being the impact parameter as in Fig. 5.5. The test particle will be deflected by a small angle only if the impact parameter r_o fulfills the condition $r_{90^\circ} < r_o < \lambda_D$, with r_{90° being the impact parameter for a deflection by 90° defined as $r_{90^\circ} = Z_T Z_F e^2 / (m_T v_T^2)$. The ratio for deflections by small and large angles can be determined from the ratio of cross-sections for both processes:

$$\frac{\lambda_D^2 - r_{90^\circ}^2}{r_{90^\circ}^2} \leq \frac{\lambda_D^2}{r_{90^\circ}^2} \approx \left(\frac{9}{Z_T Z_F} \right)^2 \left(\frac{4\pi}{3} n \lambda_D^3 \right)^2 =: A^2. \quad (5.36)$$

The expression inside the second parentheses is the number of particles inside a Debye sphere. If we assume this number to be large, (5.36) states that collisions leading to deflections by a small angle by far outnumber the collisions with large-angle deflection. A careful calculation shows that small-angle interactions are about two orders of magnitude more efficient in the deflection of test particles than the few large-angle interactions [5.1]. Thus the Fokker–Planck formalism can be applied to the Coulomb collisions in a plasma, too.

5.4 Summary

Kinetic theory describes a plasma as an assembly of particles with statistically distributed properties. The basic quantity is the phase space density. In thermal equilibrium it is described by the Maxwell distribution. The equation of continuity for the phase space density allows the derivation of the basic equations of kinetic theory: the Boltzmann equation in its most general form or as collisionless Boltzmann equation. For a pure electromagnetic field, the collisionless Boltzmann equation becomes the Vlasov equation. If collisions lead to small-angle deflections only, the collision term in the Boltzmann equation can be described by the Fokker–Planck equation. Despite the limitation to small-angle collisions, the Fokker–Planck equation can be applied to a plasma: while Coulomb collisions can lead to large-angle deflection, these are outnumbered by the small-angle deflections.

Exercises and Problems

5.1 Describe the meaning of the mean free path. What are the physical and formal differences in a neutral gas and in a plasma?

5.2 The solar wind is a proton gas with a temperature of about 1 million K. Plot the distribution function and determine the most probable speed and energy. Compare with the flow speed of 400 km/s and the kinetic energy contained in the flow.

5.3 A spacecraft measures the proton distribution in the solar wind. Above an energy of about 20 keV, the distribution can be described as a power law in E with E^{-4} . Plot the distribution and compare with the results of Problem 5.2. What kind of distribution is this?

6. Diffusive Transport

They reel to and fro, and stagger like a
drunken man: and are at their wit's end.
Psalm 107, v. 27

This chapter introduces the basic concepts of diffusion. It starts from simple spatial diffusion as in the Brownian motion. As we saw in the previous chapter, collisions of charged particles in a plasma are different: they preferentially lead to small-angle interactions, deflecting the particle only slightly from its original path. In a magnetized plasma, small-angle interactions can be described as scattering in pitch angle. Pitch-angle diffusion plays an important role in particle propagation in interplanetary space (cf. Sect. 9.2.2) and in the scattering of radiation belt particles into the loss cone. The basic scattering process, however, is different from the one sketched in Sect. 5.3.2: in the rarefied space plasmas, scattering does not result from Coulomb collisions between charged particles but from wave-particle interactions where energy is transferred from waves to particles and vice versa.

6.1 Diffusion

Diffusion is the consequence of frequent, stochastically distributed collisions and is a stochastic process. Therefore, it is not reasonable to discuss the motion of individual particles instead one has to consider an assembly of particles, described by the distribution function.

But diffusion is not only spatial diffusion. If we carefully drip a drop of ink into a glass of water, in time the drop will spread and eventually ink and water will be mixed completely: the thermal motion leads to collisions between ink and water molecules, distributing both species uniformly. This is spatial diffusion. If we carry out the same experiment with a good drop of cold cream and a cup of steaming hot tea, we find a second consequence of the collisions: after some time, tea and milk have the same temperature, thus thermal energy is transferred from the faster molecules to the slower ones, leading to diffusion in momentum space. The existence of both processes has already been mentioned in connection with Fig. 5.2.

6.1.1 Spatial Diffusion

Let us start with spatial diffusion alone. All particles have the same speed and collisions lead to changes only in the direction of motion. To describe the effect of diffusion, we have to keep track of a larger number of small spatial steps for a large number of particles. Because the stochastic aspect is important, we can borrow some concepts from probability calculus and use simple games with coins as illustrations.

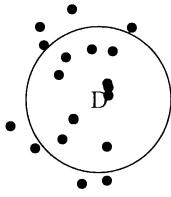


Fig. 6.1. Distribution of drunkards staggering away from a distiller D. The circle indicates the average expected distance $\lambda\sqrt{N}$

Tumbling Drunkards and Tossed Coins. Spatial diffusion, or more correctly, the motion of a particle in spatial diffusion, occasionally is called “drunkards walk”. To get the picture, imagine a couple of drunkards, happily lingering around a distiller. As they hear a police siren in the distance, they start to stagger away, everyone in his own direction. They all make steps of equal length but random direction. As a police helicopter arrives at the scene, every drunkard had made N steps of length λ . The spatial distribution of the drunkards, as seen from the helicopter, is shown in Fig. 6.1. None of the drunkards has covered the maximum possible distance $N\lambda$. Instead, they are still relatively close to the distiller. How close, compared with the maximum distance, can be described by a quantity called the expected distance or, mathematically more correct, the average squared distance. This average displacement is $\lambda\sqrt{N}$ as indicated by the circle.

Let us now reduce the problem to the one-dimensional case: the test objects can only move along a straight line, again with constant step length. We can simulate the resulting motion by flipping a coin: a head leads to a step in the positive direction, a tail to a step in the negative one. Let us consider one particle only. At first glance one might expect the expected distance to be close to the starting point. In particular, after a large number of tosses, we would expect the number of heads and tails to be roughly equal and therefore the net displacement to be small. This, however, is a faulty reversion of the law of large numbers, which is often observed in people gambling only occasionally: if the coin has shown tails 9 times in succession, the chance of heads in the next toss is exactly the same as in all previous tosses, 50%. Thus in a long series of tosses, there can be quite a large deviation from a deadlock between heads and tails. This has been known since the middle of the seventeenth century when game theory was quite popular. Thus if for a

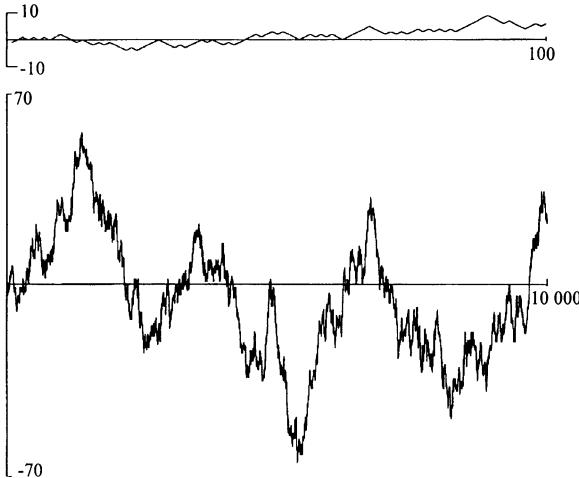


Fig. 6.2. Gain and loss chart for tossing coins. In the upper panel, the result of 100 tosses is shown, in the lower panel 10 000 tosses have been considered

long time one side of the coin can be dominant, as indicated in Fig. 6.2, then a large net gain for the one and a large loss for the other gambler results. Or, in case of one-dimensional motion, the displacement from the starting position can become quite large.

The average squared distance $\langle \Delta x \rangle^2$, or the expected distance for short, can be determined easily. The total squared displacement of the particle is the sum of the displacements dx_i in each individual step:

$$\langle \Delta x \rangle^2 = \left(\sum_{i=1}^N dx_i \right)^2 = (dx_1 + dx_2 + \dots + dx_N)^2 = \sum_{i=1}^N \sum_{j=1}^N dx_i dx_j . \quad (6.1)$$

The individual displacements dx_i are either $+\lambda$ or $-\lambda$, both with a probability of 0.5. Thus the product $dx_i dx_j$ is either λ^2 or $-\lambda^2$. For $i \neq j$, dx_i and dx_j are independent and both positive and negative values of the product have the probability 0.5. In the sum (6.1) these terms cancel and only products with $i = j$ remain. They are always $+\lambda^2$ and there are N such products. Thus (6.1) becomes

$$\langle \Delta x \rangle^2 = N\lambda^2 . \quad (6.2)$$

Thus with increasing number N of steps, the average displacement from the starting point increases as \sqrt{N} .

If the particle has a speed v , the total distance s traveled during a time t is $s = vt$. If N is the number of direction reversals during this time interval, the distance also can be written as $s = N\lambda$. Therefore in (6.2) we can substitute N by vt/λ :

$$\langle \Delta x \rangle^2 = N\lambda^2 = v\lambda t = 2Dt . \quad (6.3)$$

Here D is the diffusion coefficient:

$$D = \frac{1}{2}v\lambda . \quad (6.4)$$

Note that this diffusion coefficient has been defined for one-dimensional motion. For three-dimensional motion, the diffusion coefficient becomes

$$D = \frac{1}{3}v\lambda . \quad (6.5)$$

Galton Board and Bell Curve (Gauss's Distribution). The average distance is a statistical term which refers to a large assembly of particles. The individual particles scatter around the starting point. Their distribution can be described by the bell curve (Gauss's distribution).

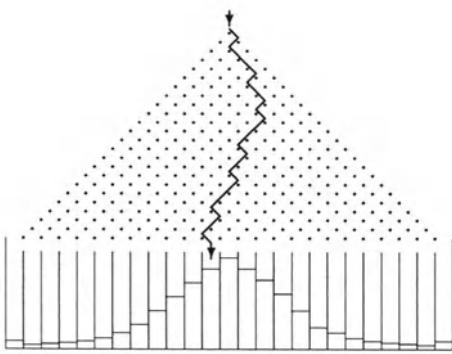


Fig. 6.3. Galton board: many small scatters work together stochastically, forming a bell curve (Gauss's distribution)

The Galton board is a graphical way to derive this distribution. It consists of rows of pins, indicated by dots in Fig. 6.3, and models the scattering the particles experience: as a ball hits a pin, it is deflected either to the left or to the right. Then it hits a pin in the next row, which leads to another deflection and so on. The solid line indicates a sample path. Below the lowest row, the particles are collected in slots. The slot in which a ball finally comes to rest, results from a large number (equal to the number of rows) of stochastic interactions of comparable strength. If we use a large number of balls, the distribution in the slots will be a bell curve or Gauss's distribution:

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - x_o)^2}{2\sigma^2}\right) . \quad (6.6)$$

Here x_o is the average and σ is the standard deviation. $P(x)$ describes the probability of a ball to be found in the slot at position x . The standard deviation σ defines the width of the distribution: 68.3% of all values will be inside the interval $[x_o - \sigma, x_o + \sigma]$ and 95.4% inside $[x_o - 2\sigma, x_o + 2\sigma]$. It is given as

$$\sigma^2 = \frac{1}{n} \sum (x - x_o)^2 =: \langle \Delta x \rangle^2 , \quad (6.7)$$

and therefore describes the widening of the particle distribution or the expected displacement from the origin.

We can rewrite (6.6) and (6.7) to find an expression depending on the diffusion coefficient. With (6.4) we find for the standard deviation

$$\sigma = \sqrt{\langle x \rangle^2} = \sqrt{2Dt} = \sqrt{v\lambda t}, \quad (6.8)$$

and therefore for the bell curve

$$P(x) = \frac{1}{\sqrt{2\pi v\lambda t}} \exp\left(-\frac{(x - x_0)^2}{2v\lambda t}\right). \quad (6.9)$$

Note that the maximum of the distribution stays fixed while it broadens with time, as described by (6.3) and shown in Fig. 6.4.

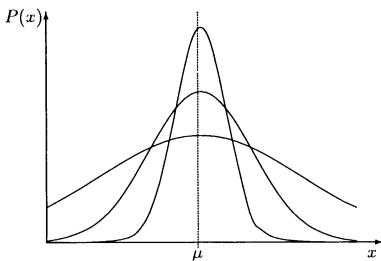


Fig. 6.4. Broadening of a Gauss distribution with increasing standard deviation. Physically, this is equivalent to the diffusive broadening of a distribution with time

The Diffusion Equation. If collisions happen in a homogeneous gas enclosed in a fixed volume, the relevant quantity to describe the diffusive process is the mean free path. It does not make sense to talk about a diffusion coefficient or an expected displacement because, viewed from the outside, the collisions do not change the properties of the gas, only the individual molecules change positions. This could be depicted by something similar to the Galton board: while in the Galton board the pins are arranged to form a triangle with the input only at the tip of the triangle, the modified board would consist of pins arranged in a rectangle with the input all over the top line. For each input slot, the spatial distribution is the same as for a Galton board. But the superimposition of all the different input slots leads to the same number of particles in each output slot. For a gas this implies: for each particle leaving a volume element another one enters.

The situation is different if there is a gradient. Then there are more particles of the species under study in one part of the volume than in the other. Accordingly, a random walk carries more particles out of the volume with high density than particles are carried in from the lower density region. Thus a net transport results, reducing the gradient and eventually leading to an equalized distribution. The streaming \mathbf{S} of particles can be described as

$$\mathbf{S} = -\mathbf{D}\nabla N, \quad (6.10)$$

with \mathbf{D} being the diffusion tensor for anisotropic diffusion. The gradient is the driving force for the flow, a larger gradient leading to a larger flow. The

flow also depends on the mobility of the particles, described by the diffusion tensor. If diffusion is isotropic, the diffusion tensor reduces to the diffusion coefficient and the streaming becomes $\mathbf{S} = -D\nabla N$. Since the diffusion coefficient depends on the particle speed and on the mean free path, for a given gradient the flow as well as the average displacement are largest for fast particles undergoing only few collisions (thus having a large mean free path) and smallest for slow particles undergoing many collisions.

With the equation of continuity (3.27), Eq. (6.10) can be written as

$$\frac{\partial N}{\partial t} + \oint_{O(S)} \mathbf{S} d\mathbf{o} = 0 . \quad (6.11)$$

Here N is the number of particles and \mathbf{S} is the flux of particles through the surface \mathbf{o} of the volume element V . If U is the particle density, (6.11) yields

$$\frac{\partial}{\partial t} \int_V U d^3x + \oint_{O(V)} \mathbf{S} d\mathbf{o} = 0 . \quad (6.12)$$

With Gauss' theorem (A.23) this is

$$\frac{\partial U}{\partial t} + \nabla \cdot \mathbf{S} = 0 . \quad (6.13)$$

With (6.10) we can write the diffusion equation as

$$\frac{\partial U}{\partial t} = \nabla \cdot (D \nabla U) . \quad (6.14)$$

If the diffusion is independent of the direction (isotropic diffusion), we can use the diffusion coefficient (6.4) instead of the diffusion tensor. If the diffusion coefficient is also independent of the spatial coordinate, as for example in a homogeneous medium, the equation can be reduced further:

$$\frac{\partial U}{\partial t} = D \Delta U . \quad (6.15)$$

We have already encountered the one-dimensional form of this equation in Sect. 3.3.2 in Eqs. (3.89) and (3.90).

Solutions of the Diffusion Equation. The solution of the diffusion equation depends on the boundary conditions. In the general case we will consider propagation from the source at a position r_o . Thus we have to consider a source Q in the diffusion equation:

$$\frac{\partial U}{\partial t} - D \Delta U = Q(r_o, t) . \quad (6.16)$$

For a spherical symmetric geometry this can be written as

$$\frac{\partial U}{\partial t} - \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 D \frac{\partial U}{\partial r} \right) = Q(r_o, t) . \quad (6.17)$$

The simplest case is a pulse-like injection of N_0 particles at the position $r_0 = 0$ at time $t_0 = 0$. In space plasmas, a typical example would be the injection of solar flare particles into the interplanetary medium. The solution of the diffusion equation for a radial-symmetric geometry then reads

$$U(r, t) = \frac{N_0}{\sqrt{(4\pi Dt)^3}} \exp\left(-\frac{r^2}{4Dt}\right). \quad (6.18)$$

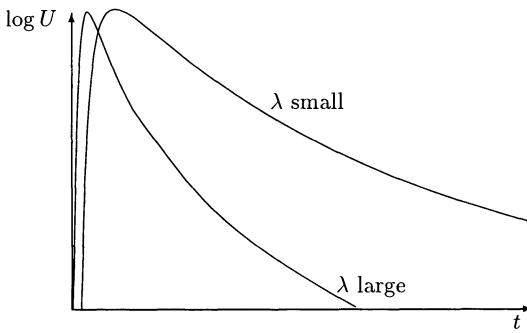


Fig. 6.5. Typical diffusive profiles for small and large λ s

Two typical diffusive profiles are shown in Fig. 6.5. The intensity rises fast to a maximum and then decays slowly $\sim t^{-3/2}$. The time of maximum t_m can be determined by setting the first temporal derivative to zero:

$$t_m(r) = \frac{r^2}{6D}. \quad (6.19)$$

The time of the maximum decreases with increasing mean free path and increasing particle speed. That is what we expect from our experience with gases and liquids: the diffusion of a minor constituent is faster with increasing temperature (corresponding to a higher particle speed) and decreasing density (corresponding to an increase in particle mean free path). The time of the maximum increases quadratically with increasing distance. This can be understood easily from (6.8): the average distance increases with \sqrt{t} .

Wibberenz [6.11] offers an interesting interpretation of the time to the maximum: if we write (6.19) in the form $t_m = (r/2\lambda) \cdot (r/v)$, we have r/v as the direct travel time for the distance r and can interpret $r/2\lambda$ as a measure of the number of mean free paths between the origin and the observer at r . The quantity $r/2\lambda$ therefore characterizes the delay due to diffusion.

Inserting (6.19) into (6.18) gives the density at the time of the maximum:

$$U(r, t_m) = \frac{N_0}{\sqrt{(4\pi r^2/6)^3}} \exp\left(-\frac{3}{2}\right) \sim \frac{N_0}{r^3}. \quad (6.20)$$

The intensity at the time of the maximum thus decreases fast with increasing radial distance but it is independent of the diffusion coefficient.

Equation (6.19) is used frequently as a simple estimate of the mean free path from the time of maximum of a particle profile observed in interplanetary space. Rewriting (6.19) we get

$$\lambda = \frac{r^2}{2vt_m} . \quad (6.21)$$

Diffusion–Convection Equation. So far, we have considered particles being scattered in a medium at rest. A different situation arises if the medium is also moving, for instance an oil spill in a river: the oil is distributed by diffusion but is also carried along with the moving fluid. In interplanetary space, the convection is due to the solar wind: the particles are scattered at inhomogeneities frozen into the solar wind and therefore propagating together with the solar wind. Thus the streaming in (6.10) has to be supplemented by the convective streaming $\mathbf{S}_{\text{conv}} = \mathbf{U} \cdot \mathbf{u}$, giving $\mathbf{S} = \mathbf{U}\mathbf{u} - D\nabla\mathbf{U}$. Here \mathbf{u} is the velocity of the convective flow. This streaming can be inserted into the equation of continuity (3.27), giving the diffusion–convection equation

$$\frac{\partial U}{\partial t} + \nabla(U\mathbf{u}) = \nabla(D\nabla U) . \quad (6.22)$$

If \mathbf{u} and D are independent of the spatial coordinate, (6.22) reads

$$\frac{\partial U}{\partial t} + \mathbf{u}\nabla U = D\Delta U . \quad (6.23)$$

In the radial-symmetric case, the solution for a δ -injection is

$$U(r, t) = \frac{N_0}{\sqrt{(4\pi Dt)^3}} \exp\left\{-\frac{(r - ut)^2}{4Dt}\right\} . \quad (6.24)$$

For small bulk speeds \mathbf{u} of the medium the transport equation as well as the solution converge towards the simple diffusion equation.

6.1.2 Pitch-Angle Diffusion

Thus far we have not discussed the physical process of scattering. In the graphical description we have tacitly assumed that we are dealing with large-angle interactions: either the particle continues to propagate in its original direction of motion or it is turned around by 180° . In the previous chapter, however, we learned that fast particles in a plasma are more likely to encounter small-angle interactions. Thus to turn a particle around, a large number of interactions is required.

In space plasmas, a different process leads to small-angle interactions: the energetic particles are not scattered in the electric field of the background plasma but by plasma waves. The physical processes will be briefly sketched in Sect. 6.2. The formal description, however, is similar to the one discussed for spatial diffusion.

Let us assume a magnetized plasma. Thus a pitch angle can be assigned to each particle, often written in the form $\mu = \cos \alpha$. Each interaction leads

to a small change in μ , i.e. a diffusion in pitch angle space. We can derive a scattering term strictly analogous to spatial diffusion. Let us start from (6.14). This can be rewritten easily: while the driving force for spatial diffusion is a spatial gradient, the driving force for pitch angle diffusion is a gradient in pitch angle space. Therefore the spatial derivatives have to be replaced by a derivative to μ and the scattering term reads

$$\frac{\partial}{\partial \mu} \left(\kappa(\mu) \frac{\partial f}{\partial \mu} \right),$$

with κ being the pitch-angle diffusion coefficient and f the phase space density. Note that the scattering depends on μ , and thus the scattering can be different for different pitch angles, depending on the waves available for wave-particle interaction. The pitch angle diffusion coefficient can be related to the particle mean free path [6.3, 6.4] by

$$\lambda := \frac{3}{8} v \int_{-1}^{+1} \frac{(1 - \mu^2)^2}{\kappa(\mu)} d\mu. \quad (6.25)$$

Here λ does not describe the average distance that a particle traveled between two consecutive small-angle scatterings, but the distance traveled before its pitch angle had been changed by 90° , i.e. the direction of motion has been reversed. Thus for the overall motion, λ has a meaning comparable to the mean free path in spatial diffusion.

This term also can describe spatial scattering if we also consider the field-parallel motion μv of the particles. Thus as in the diffusion-convection equation we also have to consider the streaming of particles with respect to the scattering centers. The transport equation then can be written as

$$\frac{\partial f}{\partial t} + \mu v \frac{\partial f}{\partial s} = \frac{\partial}{\partial \mu} \left(\kappa(\mu) \frac{\partial f}{\partial \mu} \right). \quad (6.26)$$

Here $\partial f / \partial s$ is the spatial gradient along the magnetic field line (the motion of the guiding center is one-dimensional along the magnetic field line and the particle gyrates around it). We will encounter this equation again as part of the focused transport equation (9.1) for particles in interplanetary space.

6.1.3 Diffusion in Momentum Space

Collisions not only change a particle's direction of motion but also its energy. We had already mentioned this as a basic requirement in the establishment of a thermal distribution. Momentum transfer can happen by collisions between particles as well as by wave-particle interaction. If the energy gain in each interaction is small compared with the particle energy, this can be described as diffusion in momentum space [6.2, 6.9]. Instead of particle flow, streaming S_p in momentum results in

$$S_p = -D_{pp} \frac{\partial f}{\partial p} + \frac{dp}{dt} f . \quad (6.27)$$

Here D_{pp} is the diffusion coefficient in momentum space. The second term describes non-diffusive changes in momentum, such as ionization, and corresponds to the convective term in the spatial diffusion equation. It therefore can also be described as convection in momentum space. Again, the physics of the scattering process is hidden in D_{pp} .



6.2 Wave–Particle Interactions

Suppl.
→6.3

In this section we will briefly introduce some of the basic processes of wave–particle interactions. While in all previous sections the plasmas were regarded as well-behaved, wave–particle interactions are an example of the non-linearity of plasma physics. While for the linear aspects treated before a well-developed mathematical description is available, in the non-linear theory no general algorithms exist. Only few analytical methods are known, most of them relying on approximations. One of them is the limitation to lowest-order perturbations, similar to the approximation described in Sect. 4.2. Let us start this section with a brief introduction to quasi-linear theory.

6.2.1 Quasi-Linear Theory

Quasi-linear theory is based on perturbation theory; interactions between waves and particles are considered to first order only. As in perturbation theory, all terms of second order in the disturbance should be small enough to be ignored. Only weakly turbulent wave–particle interactions can be treated this way: the particle distribution is only weakly affected by the self-excited waves in a random-phase uncorrelated way. This requirement not only corresponds to small disturbances in perturbation theory but even directly results from it as the waves are described in the framework of perturbation theory. The waves generated by the particles will affect the particles in a way which will tend to reduce the waves. Thus we assume the plasma to be a self-stabilizing system: neither indefinite wave growth happens nor are the particles trapped in a wave well.

The basic equation is the Vlasov equation (5.22). We split all quantities into a slowly evolving average part, such as f_0 , $\mathbf{E}_0 = 0$, and \mathbf{B}_0 , and a fluctuating part f_1 , \mathbf{E}_1 , and \mathbf{B}_1 . The long-term averages of the fluctuating quantities vanish: $\langle f_1 \rangle = \langle \mathbf{E}_1 \rangle = \langle \mathbf{B}_1 \rangle = 0$. Note that in contrast to the ansatz in Sect. 4.2.1 here the quantities with index ‘0’ are not constant background quantities but slowly evolving average properties of the system. These are the quantities we are interested in. The fluctuating quantities, on the other hand, are of interest only in so far as they give rise to the evolution of the phase space density, in contrast to Sect. 4.2.1 where we were interested

in the fluctuating quantities because they gave rise to a new phenomenon, the waves.

With the above ansatz, the average Vlasov equation reads

$$\frac{\partial f_o}{\partial t} + \mathbf{v} \cdot \nabla f_o + \frac{q}{m} \frac{\mathbf{v} \times \mathbf{B}_o}{c} \cdot \frac{\partial F_o}{\partial \mathbf{v}} = -\frac{q}{m} \left\langle \left(\mathbf{E}_1 + \frac{\mathbf{v} \times \mathbf{B}_1}{c} \right) \cdot \frac{\partial f_1}{\partial \mathbf{v}} \right\rangle. \quad (6.28)$$

The term on the right-hand side contains the non-vanishing averages of the fluctuations and describes the interactions between the fluctuating fields and the fluctuating part of the particle distribution. These interactions combined with the slowly evolving fields on the left-hand side of (6.28) lead to the phase space evolution of the slowly varying part of the distribution. Note that we have not made any assumptions about the smallness of the fluctuations, the only limitation is a clear separation between the fluctuating part and the average behavior of the plasma.

Equation (6.28) is the fundamental equation in non-linear plasma physics. Solutions, however, are difficult to obtain because they require an a priori knowledge of the fluctuating fields to calculate the average term on the right-hand side. This term has the nature of a Boltzmann collision term. Note that these collisions are not particle–particle interactions but result from the non-linear coupling between the particles and the fluctuating wave fields.

If the particles and the fluctuating fields are known, the term on the right-hand side can be calculated. It can then be used to derive an expression for the scattering coefficients mentioned above which depends on particle properties, in general the rigidity, and the properties of the waves, in particular their power density spectrum.

6.2.2 Resonance Scattering

The scattering of particles by waves can be described as a random walk process if the individual interactions lead to small-angle scattering only. Thus a reversal of the direction of motion requires a large number of these small-angle scatters. If we assume a particle to be in resonance with the wave, the scattering is more efficient because the small-angle changes all work together into one direction instead of trying to cancel each other. Thus pitch-angle scattering will mainly occur from interactions with field fluctuations with wavelengths in resonance with the particle motion along the field. Wibberenz [6.11] describes how such a resonance interaction can formally be understood from a simple mechanical or electrical analogy, such as a light torsion pendulum in a turbulent gas or a resonant circuit excited by noise. A full treatment of the theory with application to the scattering of particles in interplanetary space was first given by Jokipii [6.4] and, with a somewhat different approach, also by Hasselmann and Wibberenz [6.3].

The idea is sketched in Fig. 6.6. Let us assume a simple model of magnetic field fluctuations: the (relevant) waves propagate only along the magnetic field ($\mathbf{k} \parallel \mathbf{B}_o$) and the fluctuating quantities are symmetric around the wave vector.

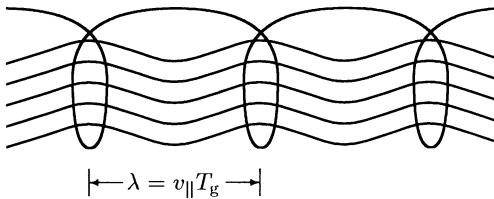


Fig. 6.6. Resonance scattering: scattering is more efficient if the particle interacts only with waves in resonance with its motion parallel to the magnetic field

This assumption is called the slab model. Let us single out a certain wave number k . A particle is in resonance with this wave if it propagates a wave length λ_{\parallel} along the magnetic field during one gyration: $\lambda_{\parallel} = v_{\parallel} \cdot T_c = \mu v \cdot T_c$. With $k_{\parallel} = 2\pi/\lambda_{\parallel}$ and $T_c = 2\pi/\omega_c$ the resonance condition can be written as

$$k_{\parallel} = \frac{\omega_c}{v_{\parallel}} = \frac{\omega_c}{\mu v}. \quad (6.29)$$

The amount of scattering a particle with pitch angle α experiences basically depends on the power density $f(k_{\parallel})$ of the waves at the resonance frequency. This dependence clearly shows why the pitch-angle diffusion coefficient κ depends on μ : particles with different pitch angles are scattered by different wave numbers with different power densities.

6.2.3 Alfvén Waves and Interplanetary Propagation

Magnetohydrodynamic waves are Alfvén waves with wave vectors parallel to the undisturbed field and magneto-sonic waves propagating perpendicular to the field. In interplanetary space, the resonance scattering of energetic particles at Alfvén waves plays an important role in particle propagation. The formalism describing this process is derived in [6.3, 6.4] or in the review [6.1].

The basic process is the interaction of the particle with the fluctuating component of the magnetic field. Since the fluctuations are assumed to be small, $B_1 \ll B_0$, the change in pitch angle during a single gyration is small, too. Scattering at waves in resonance with the particle motion parallel to the field is essential for efficient scattering. The relation between the power density of the magnetic field fluctuations and the pitch angle diffusion coefficient is relatively simple for the limitation to waves with wave vectors parallel to the field and axially symmetric transverse fluctuating components (the slab model).

The magnetic field density power spectrum in interplanetary space can be described by the power law (cf. Sect. 8.4, in particular Fig. 8.14)

$$f(k_{\parallel}) = C \cdot k_{\parallel}^{-q}. \quad (6.30)$$

Here q is the spectral shape, k_{\parallel} the wave number parallel to the field, and C the power at a certain frequency. The pitch-angle diffusion coefficient is related to this spectrum by

$$\kappa(\mu) = A(1 - \mu^2)|\mu|^{q-1}, \quad (6.31)$$

with A being a constant related to the level C of the turbulence. The particle mean free path can then be determined by (6.25). The mean free path depends on particle rigidity as $\lambda_{\parallel} \sim P^{2-q}$, as long as $q < 2$.

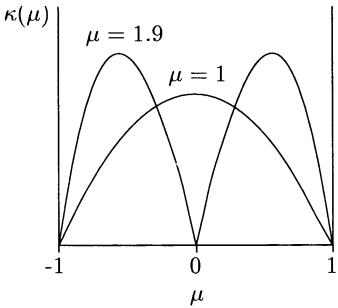


Fig. 6.7. Shape of the pitch-angle scattering coefficient for different slopes q of the power density spectrum

Figure 6.7 shows the shape of the pitch-angle scattering coefficient for different slopes q of the magnetic field power density spectrum. The case $q = 1$ corresponds to isotropic scattering, i.e. the strength of the scattering is independent of the particle's pitch angle and the diffusion coefficient is given as $\kappa_{\text{iso}} = A \cdot (1 - \mu^2)$. With increasing q , a gap develops around $\mu = 0$, the width of the gap increasing with q . Thus for particles with pitch angles close to 90° , scattering decreases as the power spectrum becomes steeper and it becomes increasingly difficult to scatter the particle through 90° , i.e. to turn around its motion along the field line. The physical basis is simple: particles with large pitch angles are in resonance with waves with small wavelength or high wave numbers. Since the power density spectrum decreases towards higher wave numbers, there is not enough power left for efficient scattering. For $q \geq 2$ the gap around $\mu = 0$ becomes too wide to scatter particles through 90° : although pitch-angle scattering still occurs, the direction of the particle motion along the field is not reversed. Or, in other words: the small-angle interactions do not sum up to a large-angle interaction. In this case the particle transport into the two hemispheres parallel and anti-parallel to the magnetic field is totally decoupled.

6.2.4 Electromagnetic Waves

Pitch-angle scattering also occurs in the magnetosphere. Here electrons trapped in the radiation belts are scattered into the loss cone by Whistler waves. Since Whistlers are electromagnetic waves, both a fluctuating electric and a fluctuating magnetic field component exist. Whistlers propagate parallel to the field, and they are circularly polarized with the electric field rotating in the same direction as an electron gyrates, cf. Sect. 4.6.2. Whistler waves therefore have a resonance at the electron cyclotron frequency.

Particles trapped in the radiation belt perform a bounce motion along the field line. Since they are trapped between two magnetic mirrors, they show a typical pitch-angle distribution: particles with small pitch angles are missing because their mirror point requires a strong increase in the magnetic field strength and therefore lies deep inside the atmosphere. Thus particles with small pitch angles are far more likely to make interaction with the atmosphere than to be reflected. The resulting pitch-angle distribution is often called a loss-cone distribution. The population of particles with large pitch angles, on the other hand, will bounce back and forth between north and south. This distribution would be stable if sufficiently large electron fluxes did not excite Whistlers. These Whistlers propagate on the background plasma. Like the electrons, Whistlers are trapped in the geomagnetic field, bouncing back and forth between the lower hybrid resonance points. These Whistlers lead to pitch-angle scattering, scattering the electrons into the loss cone, and providing a mechanism for the depletion of the radiation belts. This scattering is particularly strong if the electron component in the radiation belts is large, e.g. during a substorm, because large electron fluxes are necessary for efficient wave generation. The electrons scattered into the loss cone then precipitate into the ionosphere.

Under these conditions the quasi-linear diffusion equation is a pure pitch angle diffusion equation:

$$\frac{\partial f}{\partial t} = \frac{1}{\sin \alpha} \frac{\partial}{\partial \alpha} \left(D \sin \alpha \frac{\partial f}{\partial \alpha} \right). \quad (6.32)$$

The diffusion coefficient written in terms of the fluctuating magnetic field then is

$$D = \pi^2 \omega_{\text{thresh}} \sum_k \frac{1}{|k_{\parallel}|} \left| \frac{B_1(k)}{B_o} \right|^2 \delta \left(v \cos \alpha - \frac{\omega - \omega_{\text{thresh}}}{k_{\parallel}} \right), \quad (6.33)$$

with ω_{thresh} being the threshold frequency up to which Whistler waves can be excited by electrons.

6.2.5 Diffusion in Momentum Space

The above-mentioned types of wave-particle interactions were related to pitch-angle scattering and its consequences for spatial propagation, either in the interplanetary magnetic field or in the inner magnetosphere due to loss cone scattering. But we have not been concerned with another aspect of pitch-angle scattering, namely the diffusion in momentum space.

A change in pitch angle for constant particle speed is not related to diffusion in momentum space but only to diffusion in the pitch angle and the resulting spatial diffusion. In this case the particle interacts with the fluctuating magnetic field of the waves. A different kind of interaction can take place if the particle interacts with the fluctuating electric field of a wave, in particular if this field rotates around the magnetic field line as in a circularly

polarized wave. A particle is in resonance with the wave if its gyration frequency equals the frequency of the wave. In that case the particle is either accelerated or decelerated continuously. Figure 6.8 illustrates this principle. In the upper panel the gyration of a particle around B_o is shown together with the fluctuating electric field. To make the figure readable, the fluctuating magnetic field is not shown. In the lower panel, the two extreme cases are illustrated: depending on the phase between the wave and gyro-orbit, the particle either moves parallel or anti-parallel to the electric field. Thus either deceleration or acceleration results. In the latter case, the wave energy is converted into particle energy, and in the former case the particle energy is converted into wave energy. The resulting acceleration is called stochastic acceleration.

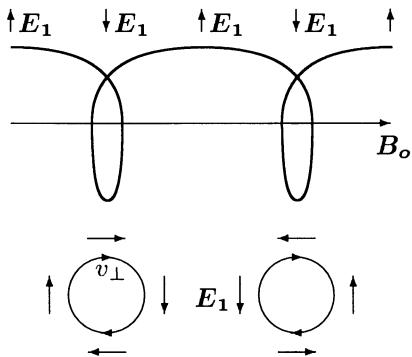


Fig. 6.8. Resonant wave-particle interaction: the particle interacts with the fluctuating electric field of a circularly polarized wave. The particle is in resonance with the wave if the wave frequency equals the gyration frequency. Depending on the phase between wave and particle gyration, the particle is either constantly accelerated or decelerated

Since the acceleration or deceleration changes the particle speed perpendicular to the average magnetic field, the particle's pitch angle changes, too. Stochastic acceleration can be described as diffusion in momentum space as long as the energy gain is small compared with the particle's energy (cf. Sect. 6.1.3). Then the accompanying change in pitch angle is small, too.

6.3 Summary

Collisions in a gas lead to particle scattering as evident in Brownian motion. In the presence of a gradient, diffusive transport results. In the rarefied space plasmas the interactions are different: here particles are scattered at waves. These wave-particle interactions lead to small changes in pitch angle and/or momentum. The basic requirement for efficient scattering is a resonance between the particle and the wave. The combined action of many such small scatterings can be described as pitch-angle diffusion and diffusion in momentum space. Applications of pitch-angle scattering are particle propagation in interplanetary space and the scattering of radiation belt particles into the loss

cone. Diffusion in momentum space can be found in stochastic acceleration in the turbulent plasmas behind shock fronts.

Wave-particle interactions are examples of non-linear plasma physics. Quasi-linear theory provides a framework for a simple formal description. The starting point is the Vlasov equation. If the relevant quantities, i.e. the phase space density and the electromagnetic field, are separated into a slowly evolving part and a fluctuating part, the resulting additional term in the Vlasov equation has the form of a collision term and describes the interaction between waves and particles. Depending on the assumptions about waves and particles, this collision term finally leads to the diffusion coefficients.

Exercises and Problems

6.1 Assume a Galton board with n rows of pins. For each pin, the possibility of a deflection to the left or to the right is 0.5. (a) Give the probability distribution in the n th layer. (b) Show that for large n this distribution converges towards the bell curve. Give the standard deviation. (c) Write a small computer program to simulate a Galton board. Compare the runs of your simulation with the expected result for different numbers of rows. Alternatively, simulate the results for a Galton board with 5 rows and 100 balls by tossing a coin. Compare with the expected results.

6.2 Get an idea about changes of times scales in diffusion. Imagine a horde of ants released at time $t_0 = 0$ onto a track in the woods. The speed of the ants is 10 m/min, their mean free path 10 cm. How long would you have to wait until the number of ants passing your observation point at 1 m (10 m, 100 m) is largest. How does your result change if you get faster ants (10 m/min, 50 m/min) or ants moving more erratically (mean free path reduced to 5 cm, 1 cm). Can you imagine different populations of ants characterized by different speeds and different mean free paths reaching their maximum at the same time at the same place? (More realistic numbers for interplanetary space: particle speeds of 0.1 AU/h, 1 AU/h, and 6 AU/h, distances of 0.3 AU, 1 AU, 5 AU, and mean free paths of 0.01 AU and 0.1 AU.)

6.3 In interplanetary space, propagation should be described by the diffusion-convection equation instead of the simple diffusion equation. The flow speed of the solar wind is about 400 km/s. Calculate profiles with the diffusion-convection equation with the numbers given in the parentheses to Problem 6.2. Compare with solutions of the simple diffusion model. Discuss the differences (how do they change with particle speed and mean free path and why?).

6.4 Explain the shape of $\kappa(\mu)$ in Fig. 6.7 for isotropic scattering. Why is it not a straight line?

7. Shock Waves

‘Will you move a little faster?’ said a whiting to a snail,
‘There’s a porpoise close behind us, and he’s treading on my tail.’
L. Carroll, *Alice’s Adventures in Wonderland*

A shock is a discontinuity separating two different regimes in an otherwise continuous medium. It is associated with something moving faster than the signal speed in the medium: a shock front separates the Mach cone of a supersonic jet from the ambient, undisturbed air. If a duck is paddling faster through a pond than a water wave can propagate, a shock separates the wake from the undisturbed water. In both cases the disturbance and with it the shock is moving, and thus the shock is called a traveling shock. Standing shocks also form: in a river, a shock forms in front of the bridge pier where the fast stream suddenly is slowed down. In space plasmas, both kinds of shocks exist: mass ejections propagating from the Sun through interplanetary space drive traveling shocks. The supersonic solar wind is slowed down at planetary magnetospheres, forming the bow shock, a standing shock wave. At these discontinuities the properties of the medium change dramatically. Combined with different wave modes in plasmas and the collisionless nature of the rarefied space plasmas, different types of shocks can develop. Shocks also play an important role in the acceleration of energetic particles. Shock waves in supernovae accelerate the galactic cosmic radiation, planetary bow shocks accelerate particles, as do shocks in front of coronal mass ejections.

7.1 What is a Shock Wave?

In our daily experience, shock waves are related to the supersonic bang of aircraft or blast waves from explosions. Thus a shock is related to abrupt changes in the properties of the medium, in particular an abrupt increase in the gas-dynamic pressure. More exactly, we can define a shock as follows:

1. The disturbance propagates with a speed faster than the signal speed. In a gas, the signal speed is the speed of sound; in space plasmas, it depends on the Alfvén speed and the sound speed.

2. At the shock front, the properties of the medium change abruptly. In a hydrodynamic shock, pressure and density increase; in a magnetohydrodynamic shock, plasma density and magnetic field strength increase.
3. Behind the shock front a transition back to the properties of the undisturbed medium must occur. Behind a gas-dynamic shock, density and pressure decrease; behind a magnetohydrodynamic shock, plasma density and magnetic field strength decrease. If this decrease is fast, a reverse shock develops.

Shock waves can be stable for long times (in the solar system up to some months) and can propagate even out to the boundary of the solar system.

While the study of gas-dynamic shocks started in the late nineteenth century and had its heyday in the 1940s, the study of plasma shocks started only in the fifties as an interest in fusion plasmas and the consequences of nuclear explosions in the atmosphere awoke. At that time, also a certain kind of shocks in a plasma has been detected that differed strongly from the gas-dynamic shock: in these collisionless shocks, densities are so low that the elementary transport process is no longer the collision between individual atoms or molecules. Instead, the collective effects of the electrical and magnetic properties of the plasma allow for frequent interactions and the formation of a shock wave.

Collisionless shocks therefore are different from gas-dynamic shocks. Nonetheless, the concepts about the fundamental nature of shocks are the same as in a gas-dynamic shock, as are the basic conservation laws.

7.1.1 Information, Dissipation, and Non-Linearity

A shock is a non-linear wave of “permanent” form propagating faster than the signal speed. Thus an understanding of a shock has to invoke the concepts of information, dissipation, and non-linearity.

Information can be transferred by a propagating disturbance; the sound wave is the simplest example. It can transfer information either as a continuous stream of different waves (as in language or music) or as a rather sharp pulse like the “bang” following an explosion or the “clap” in hand-clapping. Though the information contained in these latter signals might be more difficult to decipher, they are more useful for the explanation of shock formation because they are wave parcels with a well-defined onset. These sounds travel as pressure pulses through the air. Their distance from the source defines the information horizon: inside the information horizon, the signal has already been received, outside it has not yet been detected.

A sound wave is a compressional wave: the density increases with increasing pressure. Sound waves are “simple” waves. The compression is assumed to be adiabatic: the gas is compressed such that on expansion it returns to its original state. This is possible because the compression is fast enough to prevent thermal conduction from removing heat. Thus the compression is

isentropic, the entropy does not change. Furthermore, we assume that the disturbance is small and therefore viscosity, friction, and heat conduction are negligible, and that the gas behaves according to the ideal gas law. We then can calculate the sound speed according to (4.32). It is independent of frequency but depends on the gas parameters, for instance the temperature.

The creation of a supersonic disturbance can then be viewed from different perspectives: we can ask ourselves whether information can travel faster than the speed of sound, we can study a disturbance moving faster than sound, or we can study a large-amplitude disturbance.

Let us start with the latter case. A large-amplitude pressure pulse can be created by an explosion. In air, close to the site of the blast, the sound speed can increase up to about 1000 km/s. But the signal does not propagate as a harmonic wave: in the compressional phase the pressure amplitude can exceed the atmospheric pressure. During decompression, however, the pressure cannot drop below zero. Thus the amplitudes of the positive and negative half-waves are different. In addition, in a large-amplitude wave the change in sound speed during compression and decompression becomes significant: during compression the temperature increases and the wave can propagate faster, while during the other half-wave the temperature decreases, leading to a slower propagation. Thus the wave front steepens in time, similar to a water wave running into shallow water. If the steepening eventually leads to a jump in density and pressure, a shock wave has formed. This kind of shock is called a blast wave shock. Type II radio burst in the solar corona probably indicate blast waves (cf. Sect. 8.7.1). Blast wave shocks are often used in the numerical simulation of interplanetary shocks [7.20, 7.23, 7.40], although today it is realized that interplanetary shocks are driven shocks.

A driven shock is associated with an object moving faster than a sound wave. For a driven shock to form, the disturbance need not necessarily have a large amplitude. A small-amplitude disturbance also can lead to the pressure and density jump that defines a shock. An object moving through air transfers momentum and energy to the ambient molecules. The motion of the object also requires motion of the air: molecules in front of the object must give way to it by streaming around the object and again collecting behind it. But to do this, the molecules must first receive information about the approaching object. The pressure pulse in front of a subsonic vehicle provides this information. For a supersonic vehicle, the information also has to be provided, even if only for a short distance ahead. Thus the medium has to be changed to allow for the faster propagation of information. As the motion starts to become supersonic, the initially small-amplitude pressure pulses pile up in front of the vehicle, leading to a large-amplitude disturbance. This large-amplitude disturbance is able to change the properties of the medium irreversibly: after the disturbance has passed by, the medium will be in a different state. Since the disturbance will change the temperature of the medium it also will change the sound speed.

7.1.2 The Shock's Rest Frame

The simplest description of a gas-dynamic shock uses the shock's rest frame (cf. Fig. 7.1): gas with a speed larger than the signal speed is flowing into the shock from the upstream medium which so far has not received any information about the approaching shock. Since this side is not modified by the shock, it is also called the low-entropy side. At the shock front, irreversible processes lead to the compression of the gas and a change in speed: across the shock front, mass conservation is required, and thus the amount of matter flowing through each element of the shock surface has to be constant. The flow out of the shock front into the downstream medium is subsonic and the density is increased. Thus we can define a shock as an entropy-increasing or irreversible wave that causes a transition from subsonic to supersonic flow.

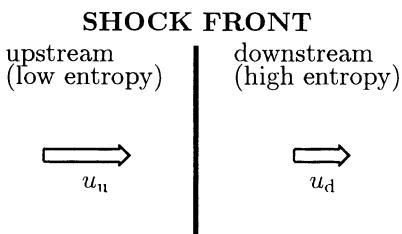


Fig. 7.1. Frame of reference for the description of a shock. The shock front is at rest, plasma flows with a high speed u_u from the upstream medium into the shock front and leaves it with a lower speed u_d into the downstream medium

Theoretically, the disturbance driving the shock can propagate at any speed. Therefore, the shock's propagation speed can increase without any limit. The Mach number M is defined as the ratio between the shock speed in the upstream medium and the sound speed. It is always determined in the rest frame of the shock. In the upstream medium, M is larger than 1, in the downstream medium it is smaller. Thus in the downstream medium the plasma leaves the shock with a speed smaller than the sound speed: any disturbance in the downstream medium can propagate away from the shock.

We can now easily understand that a traveling shock and a standing shock are identical: in a traveling shock a supersonic disturbance propagates through the medium while in a standing shock the object is at rest and the flow is supersonic. Thus the difference between standing and traveling shocks depends on the frame of reference only, the systems are Galilean invariant.

7.1.3 Collisionless Shock Waves

In a gas-dynamic shock, the important process is the collision between molecules: they establish a temperature distribution, temperatures of different species are equalized, density and temperature fluctuations can propagate, and the viscous forces associated with them lead to dissipation.

Space plasmas are rarefied, and here collisions are rare. These shocks are called collisionless shocks. The lack of collisions has some implications, for instance: electrons and protons can have different temperatures, their distributions can be very different from a Maxwellian making the classical concept of temperature obsolete, the presence of a magnetic field might even lead to highly anisotropic particle distributions, and processes of dissipation involve complex interactions between particles and fields.

Nonetheless, shocks are frequently observed in space plasmas. While the coupling between the particles due to collisions is negligible, the magnetic field acts as a coupling device, binding the particles together. We have already used this assumption in magnetohydrodynamics (MHD) where we have described a magnetized plasma by concepts such as pressure, density, and bulk velocity. These concepts also prove helpful in the description of the plasmas upstream and downstream of the shock. The details of the shock front and the plasma immediately around it, however, cannot be covered within this framework because MHD does not consider the motions of and the kinetic effects due to individual particles. While we are still far from understanding the details of these processes, observations indicate that the collective behavior of the plasma is mainly due to wave-particle interactions. Thus collisionless shocks are an example of a macroscopic flow phenomenon regulated by microscopic kinetic processes. A popular account can be found in [7.35].

7.2 Shock Conservation Laws

Plasma properties in the upstream and the downstream media are different in parameters such as bulk flow speed \mathbf{u} , magnetic field \mathbf{B} , plasma density ϱ , and pressure p . The relationship between these two sets of parameters is established by basic conservation laws, the Rankine-Hugoniot equations.

7.2.1 Rankine-Hugoniot Equations in Ordinary Shocks

A very clear description of the Rankine-Hugoniot equations and their application to hydro- and aerodynamic shocks can be found in [7.10]. These equations describe the conservation of mass, energy, and momentum through the shock front, and can even be applied to simple shocks in space plasmas as long as the distribution functions are isotropic Maxwellians and the magnetic field is roughly parallel to the flow.

In these conservation laws the shock is assumed to be an infinitesimally thin discontinuity. In optics, a boundary is thin with respect to the wavelength and thick with respect to the spacing of the molecules in the crystal structure. Analogously, in MHD a boundary can be regarded as thin with respect to the scale length of the fluid parameters (if waves are involved in the boundary, as is the case in a shock, it is thin with respect to the wavelength) but thick with

respect to the Debye length and the ion gyro-radius, both being characteristic for the collective behavior of the plasma.

In the remainder of this section the abbreviation $[X] = X_u - X_d$ gives the difference of a quantity X in the upstream and the downstream media. The Rankine–Hugoniot equations for a gas-dynamic shock then are:

- conservation of mass:

$$[\varrho u_n] = [\varrho u_n] = 0 ; \quad (7.1)$$

- conservation of momentum normal to the shock:

$$[\varrho u_n^2 + p] = 0 ; \quad (7.2)$$

- conservation of momentum tangential to the shock:

$$[\varrho u_n u_t] = 0 ; \quad (7.3)$$

- conservation of energy normal to the shock:

$$\left[\left(\frac{\varrho u^2}{2} + \frac{\gamma_a}{\gamma_a - 1} p \right) u_n \right] = 0 . \quad (7.4)$$

Here u is the flow speed, u_n (u_t) the flow speed normal (tangential) to the shock, ϱ the density, p the pressure, and γ_a the specific heat ratio, all measured in the shock's rest frame. The conservation of energy considers both kinetic flow energy and internal energy. Combining (7.1) and (7.2) yields $[u_t = 0]$: the tangential component of the flow is continuous. Therefore, we can choose a coordinate system moving along the shock front with the speed u_t . In this normal incidence frame (cf. left panel in Fig. 7.2), u equals u_n .

The mass conservation (7.1) can be used to estimate the local shock speed. Making a Galilean transformation into the laboratory system, it can be written as $[\varrho(v_s - u_n)] = 0$. Rearrangement gives the shock speed

$$v_s = \frac{\varrho_d u_{n,d} - \varrho_u u_{n,u}}{\varrho_d - \varrho_u} . \quad (7.5)$$

In applying (7.5) to shocks in space plasmas, in particular traveling interplanetary shocks, we should be aware of its limitations. First, the magnetic field is neglected; the shock is a simple gas-dynamic one. Second, the shock is assumed to be spherically symmetric with the flow perpendicular to the shock surface. For an interplanetary shock, this is an oversimplification [7.22, 7.39] and the speed estimated from (7.5) only gives the radial component of the shock speed and therefore can be used as a lower limit only. However, observations suggest that the deviation of the shock normal from the radial direction is often less than 20° [7.8], and thus (7.5) is a reasonable approximation.

Note that the shock speed alone is not indicative of the energetics of the shock. The shock speed becomes large if the denominator in (7.5) is small. Thus a small increase in density across the shock can often be associated with a large shock speed, while the total energy in terms of compression and mass motion is rather small.

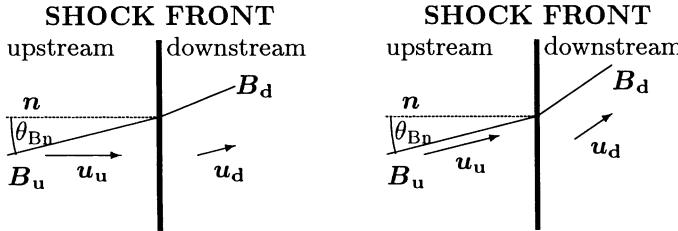


Fig. 7.2. Frames of reference for MHD shocks: normal incidence frame (*left*) and de Hoffmann–Teller frame (*right*)

7.2.2 Rankine–Hugoniot Equations in MHD Shocks

The crucial difference between a MHD shock and an ordinary shock is the magnetic field. Thus we have to expand the conservation laws to also accommodate the field. In addition, the geometry becomes more complex because the flow is not necessarily parallel to the field.

In addition to the normal incidence frame, often another rest frame is used, namely the de Hoffmann–Teller frame (right panel in Fig. 7.2). In the normal incidence frame, the upstream plasma flow is normal to the shock and oblique to the magnetic field. The downstream flow is oblique to both the magnetic field and shock normal. In the de Hoffmann–Teller frame [7.17], the plasma flow is parallel to the magnetic field on both sides of the shock. Therefore, the $\mathbf{u} \times \mathbf{B}$ induction field in the shock front vanishes: the reference frame moves parallel to the shock front with the de Hoffmann–Teller speed $v_{HT} \times \mathbf{B}/c = -E$.

For a MHD shock, the Rankine–Hugoniot relations can be inferred in the same way as for a gas-dynamic shock (see, for example, [7.6, 7.7, 7.11]). With \mathbf{n} being the unit vector along the shock normal, the Rankine–Hugoniot equations are:

- the mass balance, which is the same as for the ordinary shock,

$$[\varrho \mathbf{u} \cdot \mathbf{n}] = 0 ; \quad (7.6)$$

- momentum balance, where the additional terms describe the magnetic pressure perpendicular and normal to the shock front,

$$\left[\varrho \mathbf{u} \cdot (\mathbf{u} \cdot \mathbf{n}) + \left(p + \frac{\mathbf{B}^2}{8\pi} \right) \cdot \mathbf{n} - \frac{(\mathbf{B} \cdot \mathbf{n}) \mathbf{B}}{4\pi} \right] = 0 ; \quad (7.7)$$

- energy balance, where the additional terms describe the electromagnetic energy flux $\mathbf{E} \times \mathbf{B}/4\pi$ with the electric field expressed by $\mathbf{E} = -\mathbf{v} \times \mathbf{B}/c$,

$$\left[\mathbf{u} \cdot \mathbf{n} \cdot \left(\frac{\varrho \mathbf{u}}{2} + \frac{\gamma}{\gamma-1} p + \frac{\mathbf{B}^2}{4\pi} \right) - \frac{(\mathbf{B} \cdot \mathbf{n}) \cdot (\mathbf{B} \cdot \mathbf{u})}{4\pi} \right] = 0 ; \quad (7.8)$$

- Maxwell's equations

$$[\mathbf{B} \cdot \mathbf{n}] = 0 , \quad (7.9)$$

which follows from $\nabla \cdot \mathbf{B} = 0$, and states that the normal component of the magnetic field is continuous ($B_n = \text{const}$), and

$$[\mathbf{n} \times (\mathbf{u} \times \mathbf{B})] = 0 , \quad (7.10)$$

which states that the tangential component of the electric field must be continuous.

The Rankine–Hugoniot equations are a set of five equations for the unknown quantities ϱ , \mathbf{u} , p , B_n , and B_t .

7.2.3 Jump Conditions and Discontinuities

The Rankine–Hugoniot equations allow the calculation of the downstream plasma parameters from the knowledge of the upstream parameters of a MHD shock. But these conservation equations are more general; the solutions of (7.6)–(7.10) are not necessarily shocks, instead a multitude of different discontinuities can be described, too.

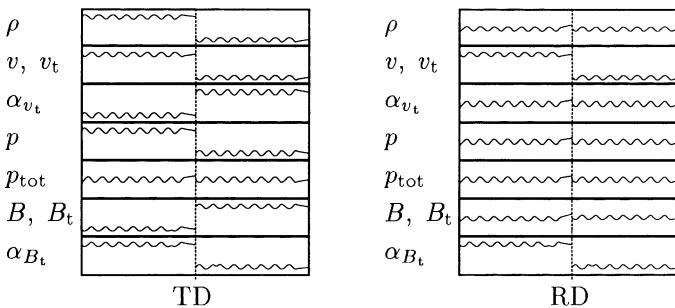


Fig. 7.3. Changes in magnetic field and plasma parameters across a tangential discontinuity (left) and a rotational discontinuity (right). Here α_x indicates a change in the direction of the quantity x . p_{tot} gives the total pressure, the sum of kinetic, plasma, and magnetic pressures

A contact discontinuity does not allow for a plasma flow across it, and thus it is $u_n = 0$. It is associated with an arbitrary density jump while all other quantities remain unchanged. The magnetic field has a component normal to the discontinuity ($B_n \neq 0$), and thus the two sides of the discontinuity are not completely decoupled but tied together by the field such that they flow together at the same tangential speed u_t .

A tangential discontinuity separates two plasma regions completely from each other. There is no flux across the boundary ($u_n = 0$ and $B_n = 0$) and

the tangential components of both quantities change ($[u_t] \neq 0$ and $[B_t] \neq 0$). Plasma and field change arbitrarily across the boundary but a static pressure balance is maintained: $[p + B^2/8\pi] = 0$. Typical changes in plasma and field parameters are sketched in the left panel in Fig. 7.3.

A rotational discontinuity can be regarded as a large-amplitude wave. In an isotropic plasma, the field and the flow change direction but not magnitude. The rotational discontinuity requires pressure equilibrium according to (7.7). Because there is a flux across the boundary, then $u_n \neq 0$ and $B_n \neq 0$. The normal flow speed is $u_n = B_n/\sqrt{4\pi\rho}$ and the change in tangential flow speed is related to the change in tangential magnetic field: $[u_t] = [B_t/\sqrt{4\pi\rho}]$. Thus normal flow speed and the change in tangential flow speed are directly related to the Alfvén speed and the change in tangential Alfvén speed. The rotational discontinuity therefore is closely related to the transport of magnetic signals across the boundary. The jump conditions for a rotational discontinuity also apply at boundaries suitable for reconnection. Typical changes in plasma and field parameters are sketched in the right panel in Fig. 7.3.

7.3 Shock Parameters

7.3.1 Shock Geometry

One important parameter in the description of a MHD shock is the local geometry, i.e. the angle θ_{Bn} between the magnetic field direction and the shock normal. Shocks can be classified according to θ_{Bn} :

- a perpendicular shock propagates perpendicular to the magnetic field: $\theta_{Bn} = 90^\circ$;
- a parallel shock propagates parallel to the magnetic field: $\theta_{Bn} = 0^\circ$;
- an oblique shock propagates at any θ_{Bn} between 0° and 90° . Oblique shocks can be subdivided into
 - quasi-parallel shocks with $0^\circ < \theta_{Bn} < 45^\circ$ and
 - quasi-perpendicular shocks with $45^\circ < \theta_{Bn} < 90^\circ$.

The shock shown in Fig. 7.2 therefore is a quasi-parallel shock.

7.3.2 Fast and Slow Shocks

A shock differs from the discontinuities described in Sect. 7.2.3 in so far as there is a flow of plasma through the surface ($u_n \neq 0$) combined with compression and changes in flow speed. Note that in a parallel shock B_t equals 0 and the magnetic field is unchanged by the shock, while in a perpendicular shock the normal component of the magnetic field vanishes, $B_n = 0$, and both plasma pressure and field strength increase at the shock. The parallel shock therefore behaves like a gas-dynamic shock – except for the fact that

the magnetic field is the only way to allow for the collective behavior of the plasma.

In a plasma, different modes of MHD waves exist which can steepen to form a shock: fast, slow, and intermediate waves, cf. Fig. 4.4 in Sect. 4.2. Of these waves, only the fast and the slow waves are compressive. The intermediate wave is a purely transverse wave with the velocity perturbation perpendicular to both \mathbf{k}_o and \mathbf{B}_o . The intermediate shock, sometimes also called an Alfvén shock, only exists in an anisotropic medium. In an isotropic plasma, such as the solar wind, it is not a shock but a rotational discontinuity: there is a rotation of the magnetic field by 180° in the plane of the shock but no density jump across the shock. Thus there is a flow across the boundary, but without compression or dissipation. In addition, the planes defined by the magnetic field and the plasma flow direction in the upstream and downstream media are not parallel but oblique. In an intermediate shock the propagation speed parallel to the magnetic field equals the Alfvén speed, i.e. $v_{\text{int}} = v_A \cos \theta_{Bn}$.

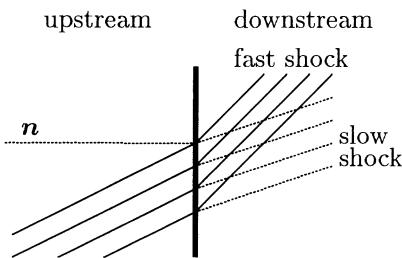


Fig. 7.4. Change in magnetic field direction across a fast and a slow MHD shock

Real shocks are formed by fast and slow magneto-sonic waves only. In both modes, the plasma density and pressure change across the shock. The phase speed of these modes is (cf. (4.41))

$$2v_{\text{fast,slow}}^2 = (v_s^2 + v_A^2) \pm \sqrt{(v_s^2 + v_A^2)^2 - 4v_s^2 v_A^2 \cos^2 \theta} , \quad (7.11)$$

with the + sign referring to the fast and the - sign to the slow mode. If these waves propagate perpendicular to \mathbf{B} , then $v_{\text{inter}} = v_{\text{slow}} = 0$ and $v_{\text{fast}} = \sqrt{v_A^2 + v_s^2}$. For propagation parallel to \mathbf{B} either v_{fast} equals v_{inter} for $v_A > v_s$ or v_{inter} equals v_{slow} for $v_A < v_s$, cf. Fig. 4.4. For the different modes, different Mach numbers can be introduced: M_A is the Alfvén Mach number, M_{cs} the sonic Mach number (the same as the Mach number in a simple gas-dynamic shock), and M_{sl} and M_f the slow and fast Mach numbers, respectively.

The change in the magnetic field is different in fast and slow shocks: in a fast shock the magnetic field increases and is bent away from shock normal because the normal component of the field is constant. The normal component of the upstream (downstream) flow speed is larger (smaller) than the propagation speed of a fast MHD wave and both upstream and downstream

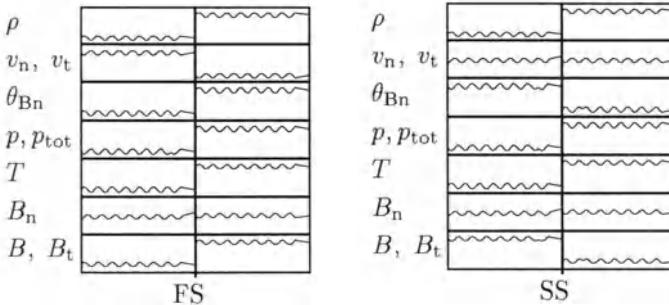


Fig. 7.5. Changes in magnetic field and plasma parameters across a fast (*left*) and a slow (*right*) shock

flow speeds exceed the Alfvén speed. In a slow shock, the upstream speed exceeds the sound speed but not the Alfvén speed. In addition, the magnetic field strength decreases across the shock and the field therefore is bent towards the shock normal (cf. Figs. 7.4 and 7.5).

Traveling interplanetary shocks in general and planetary bow shocks in particular always are fast MHD shocks. So far, only a few slow shocks have been observed *in situ* in the solar system [7.9, 7.33]. In the solar corona, however, slow [7.24] and intermediate [7.41] shocks might be more common because both the Alfvén and sound speed are much higher.

For many aspects, in particular shock formation and particle acceleration at the shock, the crucial quantity is not the shock speed but its component $v_{s\parallel} = v_s \sec \theta_{Bn}$ parallel to the magnetic field. A rather slow disturbance can still have a large propagation speed parallel to \mathbf{B} if θ_{Bn} is large enough. To form a shock, a disturbance must propagate with $v_{s\parallel} > v_A$ to catch up with the waves propagating along the field, but it must not necessarily propagate with $v_s > v_A$. This problem becomes evident in the definition of the Alfvén Mach number. In a stationary frame of reference, the Alfvén Mach number is defined as

$$M_A = \frac{v_s - u_u}{v_A}, \quad (7.12)$$

which is the shock speed in an upstream reference system relative to the Alfvén speed. With this definition, even fast MHD shocks occasionally can have Mach numbers smaller than 1. A better definition, such as the critical Mach number

$$M_c = \frac{v_s - u_u}{v_A \cos \theta_{Bn}}, \quad (7.13)$$

is formally more helpful because $M_c = 1$ exactly gives the intermediate shocks while fast shocks always have $M_c > 1$. But to determine the critical Mach number, the local geometry θ_{Bn} must be known.

7.3.3 The Coplanarity Theorem

In Fig. 7.2 we have tacitly assumed that the shock normal and the magnetic field directions in the upstream and the downstream media all lie in the same plane. This assumption is called the coplanarity theorem and is a consequence of the jump conditions at the shock. It can be expressed as

$$\mathbf{n} \cdot (\mathbf{B}_d \times \mathbf{B}_u) = 0. \quad (7.14)$$

If we only consider the transverse component, the momentum balance (7.7) for an isotropic pressure p can be written as

$$\left[\varrho u_n \mathbf{u}_t - \frac{B_n}{8\pi} \mathbf{B}_t \right] = 0. \quad (7.15)$$

Equation (7.10) can be written as $[u_n \mathbf{B}_t - B_n \mathbf{u}_t] = 0$. Therefore both $[\mathbf{B}_t]$ and $[u_n \mathbf{B}_t]$ are parallel to $[\mathbf{u}_t]$ and thus also parallel to each other. Then we have $[u_n \mathbf{B}_t] \times [u_n \mathbf{B}_t] = 0$. Resolving the parentheses gives

$$(u_{n,u} - u_{n,d})(\mathbf{B}_{t,u} \times \mathbf{B}_{t,d}) = 0. \quad (7.16)$$

Since $[u_n]$ does not vanish, the upstream and downstream tangential magnetic components must be parallel to each other. Thus the upstream and downstream magnetic field vectors are coplanar with the shock normal vector and the magnetic field across the shock has a two-dimensional geometry. The bulk velocity is coplanar with the shock, too.

7.3.4 The Shock Normal Direction

An application of the coplanarity theorem is the calculation of the shock normal in observational data. If the shock normal is known, the angle θ_{Bn} , which is crucial for shock formation and particle acceleration, can be calculated.

If only magnetic field measurements are available, the coplanarity theorem (7.14) for the magnetic field can be used. Since the magnetic field is divergenceless, we have $(\mathbf{B}_u - \mathbf{B}_d) \cdot \mathbf{n} = 0$. Thus together with (7.14) we have defined two vectors perpendicular to the shock normal. These vectors can be used to calculate the shock normal:

$$\mathbf{n} = \frac{(\mathbf{B}_u \times \mathbf{B}_d) \times (\mathbf{B}_u - \mathbf{B}_d)}{|(\mathbf{B}_u \times \mathbf{B}_d) \times (\mathbf{B}_u - \mathbf{B}_d)|}. \quad (7.17)$$

This method does not work if \mathbf{B}_u is parallel to \mathbf{B}_d . The shock normal derived according to (7.17) is called the coplanarity normal. If also three-dimensional plasma measurements are available, other constraints, in particular the coplanarity theorem for the bulk velocity, can be used to determine the shock normal. In addition, different methods can be combined into one overdetermined solution and solved for a best-fit shock normal [7.34].

With the known shock normal we are also able to determine the shock speed more accurately as suggested by (7.5):

$$v_s = \frac{\varrho_d \mathbf{u}_d - \varrho_u \mathbf{u}_u}{\varrho_d - \varrho_u} \cdot \mathbf{n}. \quad (7.18)$$

7.4 Particle Acceleration at Shocks

Particle acceleration at collisionless shocks can be observed best at planetary bow shocks and traveling interplanetary shocks. Compared with shock waves in other astrophysical objects, such as supernovae remnants or quasars, in these shocks both plasmas and energetic particles can be observed *in situ*. In this section the fundamentals of shock acceleration will be reviewed. More detailed reviews can be found in [7.25, 7.32, 7.42, 7.43, 7.44].

There are different physical mechanisms involved in the particle acceleration at interplanetary shocks:

- the shock drift acceleration (SDA), sometimes also called scatter-free acceleration, in the electric induction field in the shock front;
- the diffusive shock acceleration due to repeated reflections in the plasmas converging at the shock front; and
- the stochastic acceleration in the turbulence behind the shock front.

The relative contributions of these mechanisms depend on the properties of the shock. For instance, shock drift acceleration is important in perpendicular shocks ($\theta_{Bn} = 90^\circ$) where the electric induction field is maximal, but vanishes in parallel shocks. Stochastic acceleration requires a strong enhancement in downstream turbulence to become effective, while diffusive acceleration requires a sufficient amount of scattering in both upstream and downstream media. In addition, shock parameters such as speed, compression ratio (ratio of the densities in the upstream and downstream media), or Mach number, determine the efficiency of the acceleration mechanism.

If we discuss particle acceleration at shocks, we treat the particles as test particles, which do not affect the shock. The shock is thin compared with the Larmor radius, and thus the adiabatic invariants still hold. The jump in field and plasma may be arbitrarily large but both are homogenous in both upstream and downstream media, and the shock front is planar. Thus effects due to the curvature, in particular drifts, are neglected. Since the total extension of a shock is large compared with the Larmor, this approximation is reasonable for a start.

7.4.1 Shock Drift Acceleration (SDA)

Shock drift acceleration (SDA) takes advantage of the electric induction field in the shock front [7.1, 7.12, 7.18, 7.21]. Since scattering is assumed to be negligible, SDA often is also called scatter-free shock acceleration. The calculations by Schatzmann [7.36] were the first theoretical studies containing also an order-of-magnitude estimate of the efficiency of this mechanism. They showed that a very efficient acceleration on short temporal and spatial scales is only possible if there is additional scattering which feeds the particles back into the shock for further acceleration. As we will see below, this problem of feeding the particles back into the acceleration mechanism is a recurrent

problem in the study of shock acceleration, independent of the actual acceleration mechanism.

In shock drift acceleration, a charged particle drifts in the electric induction field in the shock front. In the shock's rest frame, this is

$$\mathbf{E} = -\frac{\mathbf{u}_u \times \mathbf{B}_u}{c} = -\frac{\mathbf{u}_d \times \mathbf{B}_d}{c}. \quad (7.19)$$

This field is directed along the shock front and perpendicular to both magnetic field and bulk flow. It is maximal at a perpendicular shock and vanishes at a parallel shock. But how does a particle move relative to this electric field? The shock is a discontinuity in magnetic field strength. Thus a particle can drift along the shock front according to (2.43). The direction of the drift depends on the charge of the particle and is always such that the particle gains energy.

Figure 7.6 shows sample trajectories for particles in the rest frame of a quasi-perpendicular shock with $\theta_{Bn} = 80^\circ$. The abscissa shows the distance from the shock in gyro-radii, and the ordinate gives the particle energy in units of its initial energy E_0 . The dashed lines indicate the shock with the upstream medium to the left and the downstream medium to the right. The electric induction field is parallel to the shock front in the upward direction. In the left panel, the motion of the particle starts in the upstream medium and the particle is reflected back into the upstream medium after the interaction. The other two panels show particles transmitted through the shock, either from the upstream medium to the downstream medium (middle panel) or vice versa (right panel).

Let us now consider the left panel. As the particle hits the shock front, it starts to drift along it, gaining energy. With increasing energy its velocity component perpendicular to the shock front increases, too, eventually becoming larger than the shock speed. At this point the particle escapes from the shock front into the upstream medium. The details of the particle trajectory, in particular the question whether the particle is transmitted or reflected, strongly depend on its initial energy and pitch angle (for additional examples of trajectories see, for example, [7.1, 7.13]), leading to a characteristic angular distribution of the accelerated particles: an initially isotropic distribution of particles in the upstream and downstream media is converted to a very strong field-parallel beam in the upstream medium and to a smaller beam perpendicular to the field in the downstream medium.

The energy gain of a particle is largest if the particle can interact with the shock front for a long time. This time depends on the particle's speed perpendicular to the shock. If this is small, the particle sticks to the shock, if it is large, the particle escapes before it has gained a large amount of energy. This particle speed relative to the shock is determined by particle speed, shock speed, pitch angle, and θ_{Bn} . With increasing particle speed the range of the other three parameters, which allow for efficient acceleration, narrows, reducing the chance of efficient acceleration.

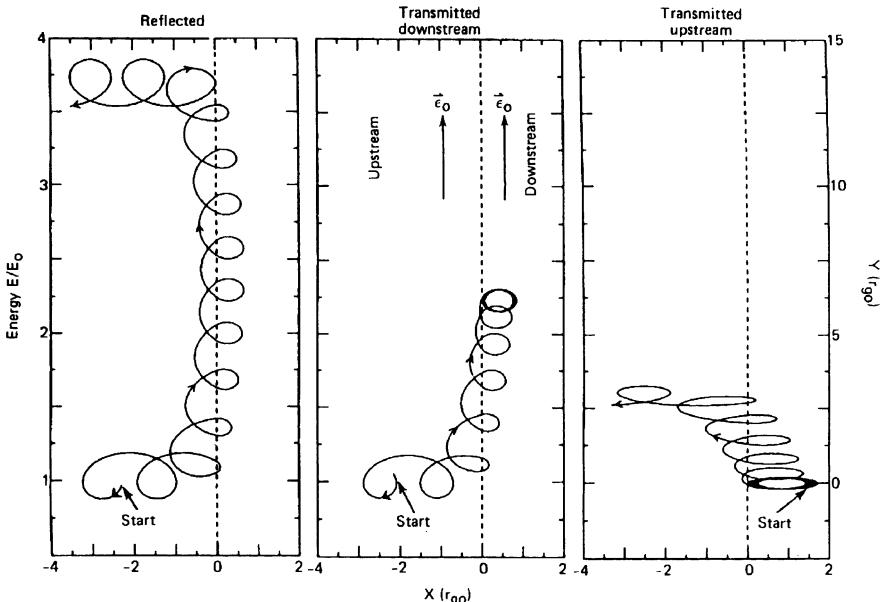


Fig. 7.6. Shock drift acceleration: sample trajectories in the shock rest frame of a quasi-perpendicular shock with $\theta_{Bn} = 80^\circ$. Reprinted from R.B. Decker [7.13], *Space Science Reviews* 48, Copyright 1988, with kind permission from Kluwer Academic Publishers

Since the magnetic moment is conserved, the momentum perpendicular to both shock and field after an interaction between particle and shock is

$$\frac{p_{2\perp}}{p_{1\perp}} = \frac{B_d}{B_u} = r_B . \quad (7.20)$$

The normal component of the momentum is unchanged by the interaction. Thus the change in momentum is determined by the magnetic compression r_B , the ratio between the upstream and downstream magnetic field strengths. Equation (7.20) is valid only for perpendicular shocks. For oblique shocks, Toptygin [7.43] gives the approximation $\Delta E \sim p_{u\perp}/\theta_{Bn}$.

The average energy gain is a factor 1.5–5. Evidently, such an energy gain is much too small to accelerate particles out of the solar wind to energies of some tens of kiloelectronvolts or even some tens of megaelectronvolts. Acceleration up to higher energies by shock drift acceleration would require repeated interactions between particles and shock. If a particle is in the downstream medium, further acceleration is not difficult because the turbulence created by the shock leads to stochastic acceleration or scatters the particle back towards the shock front. But once a particle has escaped into the upstream medium, the shock will not be able to catch up with it again, and thus no further acceleration is possible. This statement is only true for scatter-free conditions as originally assumed in shock drift acceleration. Space plasmas,

however, are turbulent plasmas, and particles are scattered depending on the level of turbulence. This allows for repeated interactions between particles and shock and thus a higher energy gain [7.14, 7.15, 7.16]. Note that in the presence of scattering the energy gain in each individual interaction between particle and shock in general is smaller because the particle is easily scattered off from its drift path along the shock front. The larger net effect results from the larger number of interactions.

Shock drift acceleration out of the solar wind up to energies of a few kiloelectronvolts to some tens of kiloelectronvolts can be observed at the Earth's bow shock and at interplanetary traveling shocks, cf. Chap. 9.

7.4.2 Diffusive Shock Acceleration

Diffusive shock acceleration is the dominant mechanism at quasi-parallel shocks. Here the electric induction field in the shock front is small and shock drift acceleration becomes negligible. In diffusive shock acceleration, the particle scattering on both sides of the shock is crucial. Since the scattering centers are frozen into the plasma, particle scattering back and forth across the shock can be understood as repeated reflection between converging scattering centers. The concept was developed in the 1970s [7.3, 7.4, 7.5, 7.19]; for reviews, see, for example, [7.2, 7.21, 7.29, 7.37, 7.38, 7.45].

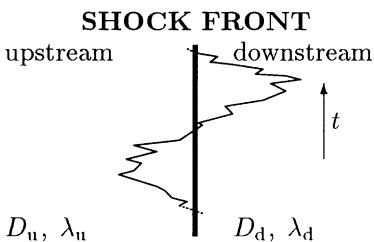


Fig. 7.7. Diffusive shock acceleration: motion of a particle in the shock rest frame. One cycle of motion consists of a crossing from one side of the shock to the other and back again

Figure 7.7 shows the motion of a particle in the rest frame of a quasi-parallel shock. The magnetic fields on both sides of the shock are turbulent; the resulting pitch-angle scattering is quantified by the diffusion coefficients D_u and D_d or the mean free paths λ_u and λ_d . Assume that the particle has just traversed the shock front in the upstream direction. The particle then follows a zigzag path and eventually is scattered back to the shock front, traversing it into the downstream medium. Here the same process is repeated, only with the diffusion coefficient characteristic of the downstream medium. As the particle is finally scattered back towards the shock front, traversing it into the upstream medium, a new cycle of motion begins.

But where in this cycle does the acceleration occur? Whereas in shock drift acceleration the location of acceleration is well defined, in diffusive shock acceleration the acceleration is given by the sum of all pitch-angle scatters

within the cycle. The energy gain can be determined from a simplification. Let us reduce the cycle to two isolated collisions, one in the upstream medium and the other in the downstream medium: in the upstream medium the particle gains energy due to a head-on collision with a scattering center; in the downstream medium it loses energy because the scattering center moves in the same direction as the particle. Since the flow speed, and therefore the speed of the scattering center, is larger upstream than downstream, a net gain in particle energy results. The amount of energy gained in each interaction depends on the relative speed between particle and scattering centers. The simplified picture sketched here assumes particle motion parallel to the field. In reality, the particle will have a pitch angle. Then the energy gain depends on the particle velocity parallel to the magnetic field. Thus for a given particle speed the energy gain depends on pitch angle, too.

Diffusive shock acceleration is based on scattering, which is a stochastic process. Therefore we cannot calculate the path of a certain particle. However, from the initial particle distribution, the shock parameters, and the scattering conditions in the upstream and downstream media we can calculate the average behavior of a large number of particles. Diffusive shock acceleration, therefore, must be described in terms of a phase space density and a transport equation. The latter can be written as

$$\frac{\partial f}{\partial t} + \mathbf{U} \nabla f - \nabla \cdot (\mathbf{D} \cdot \nabla f) - \frac{\nabla \cdot \mathbf{U}}{3} p \frac{\partial f}{\partial p} + \frac{f}{T} + \frac{1}{p^2} \frac{\partial}{\partial p} \left(p^2 \left(\frac{dp}{dt} \right) f \right) = Q(p, r, t). \quad (7.21)$$

Here f is the phase space density, \mathbf{U} the plasma speed, \mathbf{D} the diffusion tensor, and T the loss time. The terms, from left to right, present: convection of particles with the plasma flow, spatial diffusion, diffusion in momentum space (acceleration), losses due to particle escape from the acceleration region, and convection in momentum space due to processes which affect all particles, such as ionization or Coulomb losses. The term on the right-hand side is a source term describing the injection of particles into the acceleration process. In a first-order approximation losses and convection in momentum can be ignored, reducing the transport equation to

$$\frac{\partial f}{\partial t} + \mathbf{U} \nabla f - \nabla \cdot (\mathbf{D} \cdot \nabla f) - \frac{\nabla \cdot \mathbf{U}}{3} p \frac{\partial f}{\partial p} = Q(p, r, t). \quad (7.22)$$

Here the terms containing the plasma speed \mathbf{U} and the diffusion tensor \mathbf{D} describe the acceleration of the particles across the shock front.

Often we are not interested in the evolution of the particle distribution at the shock but only want to know the particle spectrum and the acceleration time in steady state. In this case, $\partial f / \partial t$ equals zero and the transport equation can be solved for suitable boundary conditions, such as steadiness in particle density and in the normal component of the particle flow $S = -4\pi p^2 [Up(\partial f / \partial p) + \mathbf{D} \cdot \nabla f]$ across the shock front. This steady-state

equation leads to predictions of the acceleration time, the particle energy spectrum, and the intensity increase upstream of the shock.

Characteristic Acceleration Time. From the transport equation (7.22) the time required to accelerate particles from momentum p_o to momentum p can be inferred to be

$$t = \frac{3}{u_u - u_d} \int_{p_o}^p \frac{dp}{p} \cdot \left(\frac{D_u}{u_u} + \frac{D_d}{u_d} \right) . \quad (7.23)$$

For all momenta below p the particle distribution at the shock is in steady state. Note that here isotropic scattering in a homogeneous medium is assumed, reducing the diffusion tensor D to a diffusion coefficient D .

If we assume the diffusion coefficient D to be independent of momentum p , then (7.23) can be integrated to give a characteristic acceleration time

$$\tau_a = \frac{p}{dp/dt} = \frac{3}{u_u - u_d} \left(\frac{D_u}{u_u} + \frac{D_d}{u_d} \right) \quad (7.24)$$

in $p(t) = p_o \exp(t/\tau_s)$. Equation (7.24) is often written as

$$\tau_a = \frac{3r}{r-1} \frac{D_u}{u_u^2} , \quad (7.25)$$

with $r = u_u/u_d$ being the ratio of the flow speeds in the shock rest frame. For a parallel shock, this ratio equals the compression ratio. In (7.25) D_d/u_d is assumed to be small compared with D_u/u_u because of enhanced downstream turbulence.

The Energy Spectrum. The energy spectrum expected from diffusive shock acceleration is a power law:

$$J(E) = J_o \cdot E^{-\gamma} . \quad (7.26)$$

The spectral index γ depends only on r . In the non-relativistic case this is

$$\gamma = \frac{1}{2} \frac{r+2}{r-1} . \quad (7.27)$$

In the relativistic case, the spectral index becomes $\gamma_{\text{rel}} = 2\gamma$. Again, this holds only for the steady state, i.e. $t \gg \tau_a$.

Why do we get an energy spectrum and not, for instance, a monoenergetic distribution? The energy gain for each particle is determined by its pitch angle and the number of shock crossings. The latter is determined by the stochastic process of scattering. For a high energy gain, the particle must be “lucky” to be scattered back towards the shock again and again. Most particles make a few shock crossings and then escape into the upstream medium. They are still scattered there, but they have escaped too far to be scattered back to the shock. Thus the stochastic nature of diffusion allows high gains for a few particles, while most particles make only small gains.

Intensity Increase Upstream of the Shock. With an infinitely thin shock at $x = 0$ and a continuous particle injection $Q = q\delta(x)\delta(p - p_0)/(4\pi p_0^2)$ at the shock front, the spatial intensity variation around the shock is

$$f(x, p) = f(0, p) \exp\{-\beta_i|x|\}, \quad (7.28)$$

with $\beta_i = u_i/D_i$, the index i indicating the upstream and the downstream medium, respectively. If particle escape is considered (the f/T term in (7.21)), the scale length β_i for the upstream intensity increase is

$$\beta_i = \frac{u_i + \sqrt{u_i^2 + 4D_i/T}}{2D_i}. \quad (7.29)$$

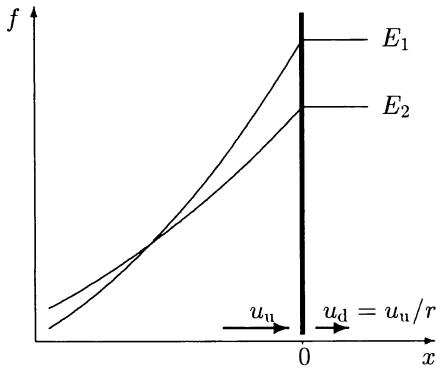


Fig. 7.8. Intensity increase upstream of the shock front for two energies $E_1 < E_2$

Figure 7.8 shows a typical upstream intensity increase and its dependence on particle energy. If β is spatially constant, the intensity increase is exponential, with its slope depending on scattering strength. Because λ increases with energy (cf. Sect. 9.2.2), the ramp is steeper for lower energies E_1 . In addition, the intensity at the shock is higher for lower energies, reflecting the power-law spectrum (7.26). In interplanetary space, the slope of the intensity increase is often used to determine the scattering conditions upstream of the shock.

7.4.3 Diffusive Shock Acceleration and Self-Generated Turbulence

Occasionally, (7.25) is used to show that proton acceleration up to megaelectronvolt energies at traveling interplanetary shocks is difficult: “Yet the Fermi process develops so slowly that the protons accelerated by quasi-parallel interplanetary shocks only reach energies of a few hundred thousand electron volts in the one day it takes the shock to travel from the Sun to the Earth” [7.35]. Note that this conclusion depends strongly on the assumptions about the particle mean free path: the particle mean free paths inferred from solar

energetic particle events, cf. Sect. 9.2.2, lead to acceleration times too high to explain the observations. For higher energies (e.g. in the megaelectronvolt or tens of megaelectronvolt range), the mean free path inferred from solar energetic particle events is so large that a particle would interact with the shock only a few times during its travel time from the Sun to the Earth. This even violates the assumption of frequent interactions inherent in the transport equation.

However, if the turbulence upstream of the shock could be increased, the acceleration would be more efficient. But how can we increase the amount of upstream scattering? As we saw in Chap. 6, scattering is a consequence of wave–particle interactions. But waves cannot propagate away from the shock into the upstream medium. Thus the upstream medium does not have any knowledge of the approaching shock (which, of course, is inherent in the definition of a shock). This statement is true from the viewpoint of the plasma. But as we saw in the previous section, the energetic particles do escape from the shock front, as evident in the upstream intensity increase. Whenever energetic particles stream faster than the Alfvén speed, they generate and amplify MHD waves with wavelengths in resonance with the field parallel motion of the particles, i.e. the same resonance condition as in Fig. 6.6. These waves grow in response to the intensity gradient of the energetic particles. A hydrodynamic approximation on the wave growth, treating the energetic particles as a fluid, can be found in [7.30]. Lee has introduced a model in which the diffusion equations describing the particle transport and the wave kinetic equations describing the wave transport are solved self-consistently for the Earth's bow shock [7.27] and traveling interplanetary shocks [7.28]. Lee's model allows some predictions regarding the particles and waves upstream of the shock: (i) At or just upstream of the shock the particle spectrum can be described by a power law in momentum $f \sim p^{-\sigma}$ with $\sigma = (1 - u_2/u_1)/3$. The scale-length of the intensity increase just upstream of the shock increases with increasing particle momentum as $p^{\sigma-3}$. (ii) The upstream waves generated by the particles propagate into the upstream direction. The growth rate is largest for waves with wave vectors parallel to the ambient magnetic field. The scale-length for the decay of the upstream waves is about the scale-length for the intensity increase of particles in resonance with these waves. Thus the scale-length of wave decay depends on the wave number. The power density spectrum of the excited upstream waves can be described as $f(k) \sim k^{\sigma-6}$. (iii) The ratio between the particle energy density ε_p and the magnetic field density of the fluctuating waves $\langle |\delta \mathbf{B}|^2 \rangle / 8\pi$ is proportional to the ratio between upstream flow speed and Alfvén speed:

$$\frac{\varepsilon_p}{\langle |\delta \mathbf{B}|^2 \rangle / 8\pi} \propto \frac{u_u}{v_A}. \quad (7.30)$$

These predictions have been found to be in fair agreement with the observations in the 11 November 1978 event [7.26, 7.29], although some differences in the details between the observed and the predicted turbulence have been

found. The energy up to which the accelerated particle population is in steady state is a few hundred kiloelectronvolts.

Thus in sum, Lee's model suggests acceleration at the shock to occur as follows. First, accelerated particles stream away from the shock. As they propagate upstream, the particles amplify low-frequency MHD waves in resonance with them. Particles escaping from the shock at a later time are scattered by these waves and are partly reflected back towards the shock. These latter particles again interact with the shock, gaining additional energy. Thus even more energetic particles escape from the shock, amplifying waves in resonance with these higher energies. The net effect therefore is an equilibrium between particles and waves which in time shifts to higher energies and larger wavelengths.

7.4.4 Stochastic Acceleration

In the previous section we saw that turbulence generated (or amplified) by particles streaming into the upstream medium is important for efficient shock acceleration. In the downstream medium, the situation is different. As the shock passes by, waves are generated with wave vectors in both the upstream and downstream directions. Thus a particle can either gain energy from the wave or supply energy to it, cf. Sect 6.2.5. This is a stochastic process, and the resulting particle acceleration is called stochastic acceleration or second-order Fermi acceleration. Fermi originally had suggested particle scattering at uncorrelated magnetic inhomogeneities moving in arbitrary directions. A particle then gains or loses energy, depending on whether it hits the inhomogeneity head-on or not. Because of the Doppler effect, head-on collisions are slightly more frequent, leading to a net gain in energy. The first considerations of the possibility of second-order Fermi acceleration in the turbulence in the downstream medium suggested that this effect is of minor importance [7.31] compared with the first-order Fermi acceleration described above. Thus applications of stochastic acceleration in general are limited to solar flares or astrophysical objects. At traveling interplanetary shocks stochastic acceleration occasionally seems to lead to intensity increases in the upstream medium which become evident as jumps in the intensity as the shock passes over the observer (cf. Sect. 9.3).

7.5 The Shock as a Non-Linear System

Shocks are highly non-linear systems with complex interactions between plasmas, waves, and energetic particles. We often use linear or quasi-linear approximations to describe some aspects of the entire phenomenon, such as shock formation or particle acceleration, but we have to be aware that these are approximations. For instance, particle acceleration and wave generation

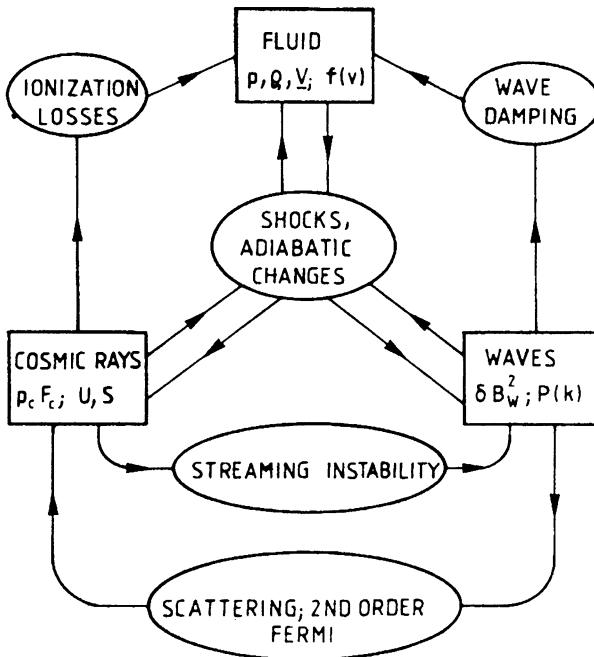


Fig. 7.9. Sketch of the non-linear interactions at a shock front. Arrows indicate the direction of energy flows. Reprinted from W.I. Axford [7.2], in *Plasma astrophysics* (eds. T.D. Tuyenen and T. Levy), Copyright 1982, European Space Agency

require energy which, of course, has to be taken from the plasma flow, thus altering the shock.

Figure 7.9 sketches the complexity of the phenomenon shock. The system can be divided into three parts: the plasma, the wave field, and the energetic particles (all shown as rectangles). The plasma is characterized by its distribution function f which also serves as the injection function into the acceleration process, and by its hydromagnetic parameters pressure p , density ϱ , bulk speed \mathbf{u} , and magnetic field \mathbf{B} . The energetic particles are characterized by their density U and their streaming S or by the corresponding hydromagnetic quantities, pressure p_c and energy flux \mathbf{F}_c . The wave field can be described by the power spectrum $P(k)$ or the average squared magnetic fluctuations $\langle (\delta \mathbf{B})^2 \rangle$. The wave field determines the diffusion coefficient and therefore the coupling between background plasma and energetic particles.

In a linear system, we could determine the relevant quantities for the waves and fields independently and use the resulting diffusion coefficient and velocity field to calculate the acceleration of energetic particles. In a non-linear system, we also have to consider mutual interactions. This can be illustrated by any pair of shock components in Fig. 7.9. Let us start with the wave field and the background plasma: waves can be amplified by the

interaction with the shock, they can be damped by the plasma, feeding wave energy into the plasma, or the wave pressure and energy flux can change the dynamics of the plasma. The non-linear interaction between another pair of shock constituents, the wave field and the energetic particles, has been partially discussed before: energetic particles can excite or amplify waves by streaming instabilities while these waves in turn scatter the particles, leading either to second-order Fermi acceleration in the downstream medium or more efficient first-order Fermi acceleration. Part of the particle's energy therefore can be transferred to the plasma by first converting it to wave energy which is then converted into plasma energy by damping. There is also the possibility of a direct interaction between energetic particles and plasmas by ionization losses.

A shock therefore is a complex, highly non-linear system. Even if all relevant equations can be written down, it is not possible to solve them self-consistently. Therefore, either approximations can be used, such as the ones presented in this chapter, or the system or part of the system can be studied by means of Monte-Carlo simulations.

7.6 Summary

Shocks in space plasmas are disturbances that propagate faster than the signal speed in the medium. They differ from ordinary hydrodynamic shocks in two respects. First, in plasmas there is a large number of different waves and correspondingly different types of MHD shocks exist, in particular fast and slow shocks related to the fast and slow magneto-sonic waves. Second, space plasmas are too thin to support frequent collisions between particles as is required in the formation of hydrodynamic shocks. Instead, the electromagnetic field is responsible for the collective behavior of the plasma. Shocks in space plasmas are therefore called collisionless shocks.

A shock is described by the Rankine–Hugoniot equations, the conservation laws for mass, momentum, and energy. Compared with hydrodynamics, in MHD these equations have to be complemented by terms that take into account the effects of the fields and by equations that describe the conservation of field properties across the shock.

Shocks in space plasmas are not only of interest from the viewpoint of plasma physics. They also provide an important acceleration mechanism for energetic particles of planetary, solar, and even galactic origin. Three acceleration mechanisms can be distinguished which are not mutually exclusive but work together with different relative contributions depending on the properties of the shock:

- In shock drift acceleration (SDA, also called scatter-free shock acceleration), particles gain energy by drifting along the electric induction field in the shock front. The energy gain is between a factor of 1.5 and 5, and the

particle interacts with the shock only once. If particles are scattered in the ambient plasma, the energy gain during one interaction is reduced because the drift path is shortened, but particles can interact repeatedly with the shock, leading to a higher total energy gain.

- In diffusive shock acceleration, the particles gain energy due to repeated scattering in the turbulent plasmas converging at the shock front (first-order Fermi acceleration). The time scales of acceleration and the energies acquired depend on the amount of scattering. Diffusive acceleration is a slow process, the maximum energy acquired depends on scattering conditions, shock parameters, and the time available for acceleration.
- Diffusive acceleration becomes more efficient if the turbulence created by the accelerated particles is considered. This self-generated turbulence in the upstream medium scatters the particles back towards the shock front, leading to further acceleration. This process is an example of the non-linearity of processes at the shock.
- Stochastic acceleration is a second-order Fermi process where the particle gains energy due to scattering in the downstream turbulence.

Exercises and Problems

7.1 Explain the basic conservation laws across a shock front. What are the differences between a gas-dynamic and a magnetohydrodynamic shock?

7.2 Explain the differences between a fast and a slow shock.

7.3 The plasma instrument on an interplanetary spacecraft detects a sudden increase in plasma density. No other changes in plasma or field are observed. Is this a shock?

7.4 The plasma instrument on an interplanetary spacecraft detects a discontinuity with a jump in plasma density from 4 cm^{-3} to 8 cm^{-3} and a jump in plasma flow speed from 400 km/s to 700 km/s (all quantities in the spacecraft frame). Determine the shock speed. What is the meaning of this speed?

7.5 An interplanetary shock propagates with a speed of 800 km/s in the spacecraft frame into a solar wind with speed 400 km/s. The ratio of upstream to downstream flow speed in the shock rest frame is 3, and the upstream diffusion coefficient is $10^{21} \text{ cm}^2/\text{s}$. Determine the characteristic acceleration time. Determine the power-law spectral index for times larger than the acceleration time.

7.6 A shock propagates with a speed of 1000 km/s through interplanetary space. The solar wind speed is 400 km/s. The particle instrument on a spacecraft observes an exponential intensity increase by two orders of magnitude, starting 3 hours prior to shock arrival. Determine the diffusion coefficient in the upstream medium (losses from the shock front can be ignored, the shock is assumed to be quasi-parallel).

8. Sun and Solar Wind: Plasmas in the Heliosphere

... but it is reasonable to hope that in not too distant a future we shall be competent to understand so simple a thing as a star.

A.S. Eddington, *The Internal Constitution of the Stars*¹

Plasmas in interplanetary space originate from the Sun, as do most of the disturbances and waves embedded in them. The solar atmosphere, the corona, extends as solar wind far beyond the orbit of the outermost planet, Pluto, filling a cavity in the interstellar medium called the heliosphere. The solar magnetic field, frozen into the solar wind, is carried out and is wound up to Archimedian spirals by the Sun's rotation. Fluctuations and waves on different scales are superimposed, sometimes steepening to form collisionless shock waves. The solar wind and the frozen-in magnetic field change during the solar cycle because of systematic changes in solar properties and transient disturbances related to solar activity.

8.1 The Sun

For an astrophysicist, the Sun is an ordinary star of spectral class 2, also called yellow dwarf, and luminosity V. It consists mainly of hydrogen (about 92%) as the fuel for solar energy production and helium (about 8%) as its waste product. But the Sun is also more interesting than other stars: owing to its close proximity, we are able to study not only its electromagnetic radiation but also solar emissions of a different kind – plasmas and energetic particles. These not only can be studied for the quiet or average Sun but also for their dependence on the solar cycle, their relation to certain features on the Sun, such as sunspots and filaments, and their association with violent processes of the active Sun. Many of these emissions are of interest for the layman, too, because they shape and influence our environment, from the atmosphere and the weather down to the realms of biology and physiology (cf. Chap. 13). The basic properties of the Sun are summarized in Table 8.1.

¹ copyright 1926, reprinted with kind permission from Cambridge University Press

Table 8.1. Properties of the Sun

Radius	$r_{\odot} = 696\,000 \text{ km}$
Mass	$M_{\odot} = 1.99 \times 10^{31} \text{ kg}$
Average density	$\varrho_{\odot} = 1.91 \text{ g/cm}^3$
Gravity at the surface	$g_{\odot} = 274 \text{ m/s}^2$
Escape velocity at the surface	$v_{\text{esc}} = 618 \text{ km/s}$
Luminosity	$L_{\odot} = 3.86 \times 10^{23} \text{ kW}$
Magnetic field	
polar	1 G
general	some G
protuberance	10–100 G
sunspot	3 000 G
Temperature	
core	15 million K
photosphere	5780 K
sunspot (typical)	4200 K
chromosphere	800 000 K
corona	2 million K
Sidereal rotation	
equator	26.8 d
30° latitude	28.2 d
60° latitude	30.8 d
75° latitude	31.8 d

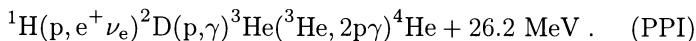
Most of the Sun's emission is electromagnetic radiation, amounting to $3.86 \times 10^{23} \text{ kW}$ integrated over its surface or $6.3 \times 10^4 \text{ kW/m}^2$. With Stefan–Boltzmann's Law, an effective temperature² T_{eff} of about 5780 K can be obtained. At Earth's orbit, the solar constant, which is the solar power received per unit area, is 1380 W/m^2 . The solar radiation can be divided into five frequency bands: (a) X-rays and extreme ultra-violet (EUV) with $\lambda < 1800 \text{ \AA}$ contributes to about 10^{-3} of the total energy output. It is emitted from the lower corona and the chromosphere, and varies during the solar cycle with enhancements up to orders of magnitude during solar flares. (b) Ultra-violet with wavelength between 1800 and 3500 \AA contributes to about 9% of the solar flux. It is radiated from the photosphere and the corona, its variations are similar to the ones in X-rays, although they are smaller. (c) The visible light between 3500 and 2500 \AA contributes to 40% of the energy flux and does not vary with the solar cycle. Only in extremely strong flares can a local brightening be observed. (d) The maximum energy flux of 51% is in the infrared between 7400 \AA and 10^7\AA , showing no variation with solar activity. As the visible light, it is emitted from the photosphere. (e) Radio-emission above 1 mm, originating from the solar corona, contributes to only $10^{-10}\%$ of the solar energy flux, but can be enhanced significantly during solar flares.

² The effective temperature of a body is the temperature a black body must have in order to emit the same total radiation q . Stefan–Boltzmann's law states that for a black body q is proportional to the fourth power of the temperature: $q = \sigma T_{\text{eff}}^4$ with $\sigma = 8.26 \times 10^{-11} \text{ cal cm}^{-2} \text{ min}^{-1} \text{ K}^{-4}$.

8.1.1 Nuclear Fusion

The source of the Sun's energy is nuclear fusion. This idea goes back to Eddington in 1926 [8.30] and replaced Lane's concept of 1869 which saw the energy source in the Sun's gravitational contraction.

Inside the Sun the basic process of nuclear fusion is the proton–proton cycle: as the net effect four protons merge to one ${}^4\text{He}$ -nuclei (α -particle). In the first step, two protons merge to a deuteron ${}^2\text{H}$, emitting a positron and a neutrino. In the second step, a proton collides and merges with the deuteron, forming a ${}^3\text{He}$ -nuclei under emission of a γ -quant. When two ${}^3\text{He}$ -nuclei collide, they merge to an α -particle, emitting two protons and a γ -quant. The mass difference between the four protons and the α -particle corresponds to an energy of $4.3 \times 10^{-12} \text{ W s}$ or 26.2 MeV. Formally, the fusion reaction can be written as



Under solar conditions, half of the hydrogen initially present is converted into deuteron within 10^{10} years: for two protons to merge, their distance must decrease below a proton radius and one of the protons has to undergo spontaneous β -emission. The life-time of ${}^2\text{H}$ is only a few seconds because it immediately captures another proton. The time scale for the fusion of the resulting ${}^3\text{He}$ -nuclei is about 10^6 years. Thus the time scale of the proton–proton cycle basically is determined by the first step, the fusion of two protons.

The last step in the PPI cycle can be replaced by one of the two reactions:



or



In both cases the ${}^3\text{He}$ -nucleus merges with an α -particle. In the PPII chain, the resulting ${}^7\text{Be}$ is converted into ${}^7\text{Li}$ by electron capture. The ${}^7\text{Li}$ then is converted by proton capture into the unstable isotope ${}^8\text{Be}$ which decays into two α -particles. In the PPIII chain, the ${}^7\text{Be}$ -nucleus captures a proton. The resulting ${}^8\text{B}$ -nucleus emits a neutrino and a positron, leading to an excited ${}^8\text{Be}$ -nucleus which decays into two α -particles. With increasing temperature, the latter two reactions become more important compared with the PPI cycle.

Independent of the details of the reaction, energy is liberated in the form of electromagnetic radiation, positrons, neutrinos and, to a smaller extent, the kinetic energy of the protons. The energy of γ s, positrons, and protons immediately is converted into thermal energy while the neutrinos escape from the Sun's core: their scattering mean free path is about 7000 AU, compared with the few cm of a photon. Thus although a photon travels with the speed of light, it needs about 10 000 years to diffuse from the Sun's core to its surface.

Neutrinos emitted in the different chains of the proton–proton cycle have characteristic energies. Thus their energy spectrum gives information about the processes inside the Sun, allowing for a test of our standard model of the Sun. Neutrinos are one of the current problems in solar astrophysics. Compared with the first observations, today a larger number of solar neutrinos is detected; however, it still is smaller than expected. Two possible explanations are offered: either our model of the Sun or our understanding of the neutrino is wrong [8.4, 8.62, 8.63]. Fine-tuning of the solar standard model to fit the neutrino observations appears possible without changing the energy flux on the solar surface. Helioseismology will show whether this approach leads to consistent results. The other explanation would have a strong influence on elementary particle physics: it suggests that neutrinos, contrary to our current understanding, have a mass and can oscillate between the different kinds. The electron neutrinos created in the proton–proton chains are partly converted into τ - or μ -neutrinos during their travel time from Sun to Earth. Since only the electron neutrinos can be detected by experiments, a lower than expected neutrino flux would result.

8.1.2 Structure of the Sun

Nuclear fusion takes place in the Sun’s core, which is about $0.3 r_{\odot}$ in radius, cf. Fig. 8.1. It is surrounded by a $0.44 r_{\odot}$ wide shell, the radiative core or radiation zone, where energy is transported by radiation. In the surrounding hydrogen convection zone energy is transported by convection. The top of the convection zone is the photosphere, the visible surface of the Sun. It is called the visible surface because here most of the visible light is emitted. The convection cells can be seen as granulation of the photosphere. From the center of the core to the photosphere, the density decreases by more than ten orders of magnitude, and the temperature decreases by a factor of 3000.

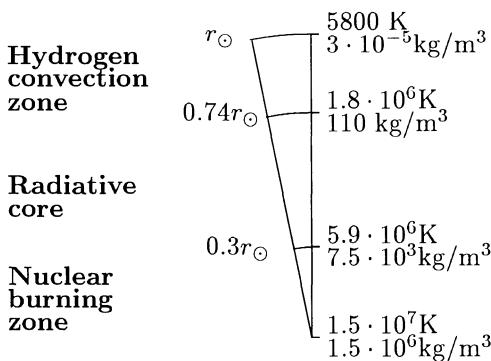


Fig. 8.1. Internal structure of the Sun with typical densities and temperatures. Based on [8.27]

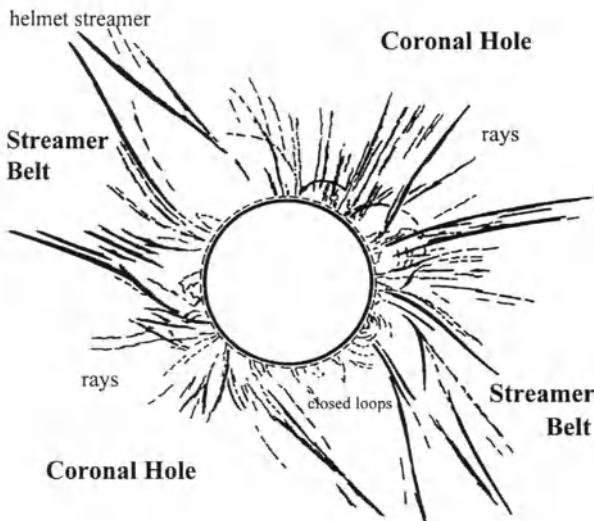


Fig. 8.2. Coronal structure during the total eclipse of 11 July 1991. Based on a sketch by S. Koutchmy in K. Lang [8.53], *Sun, earth and sky*, Copyright 1995, Springer Verlag

8.1.3 The Solar Atmosphere

Above the photosphere the solar atmosphere consists of three layers: the chromosphere, the transition region, and the corona. The corona can be seen during a solar eclipse as a structured, irregular ring of rays around the solar disk. Its structure and extent vary with the solar cycle. Figure 8.2 shows a sketch of the coronal structure during the total eclipse of 11 July 1991. Since this sketch has been taken during the solar maximum, the corona is highly structured and extends far outwards. During the solar minimum, only few structures are visible and the corona appears smaller. However, the corona does not have a sharp outer boundary, but instead shows structures which extend into different heights and then fade into the background.

The charge states of heavier elements such as O, Si, Mg, and Fe indicate coronal temperatures of about 1 million K. But the corona does not radiate like a black body because it is too thin. Its temperature is roughly independent of height, but it is about a factor of 200 higher than the photospheric temperature. This fast increase in temperature occurs in the transition region, which is only a few hundred kilometers thick and separates the corona from the chromosphere below. The latter can be seen during a total eclipse as a thin red ring around the solar disk, giving the appearance of small flames. The chromospheric emission is weak compared with the photospheric one.

Since the coronal “emission” is photospheric light scattered at the electrons in the corona, the coronal electron density can be inferred from its intensity. Thus Fig. 8.2 also can be interpreted as an electron density distribution. Typical coronal features include bright arcades, which can be interpreted as closed magnetic loops where electrons are stored, or helmet streamers, which are arcades from which a thin beam extends upwards. Ray-like structures

suggest electrons streaming away from the Sun. Rather dark regions are depleted of electrons. These coronal holes can also be seen as dark patches in soft X-ray images and are the source regions of the fast solar wind.

Depending on the underlying structures, the electron density at a certain height can vary by more than three orders of magnitude. In the lower corona the intensity gradient is very steep with a scale height of about $0.1 r_{\odot}$. Thus even with an extremely good radiometric resolution it is not possible to create images of the corona from just above the photosphere out to ten or twenty solar radii. Instead, different instruments have to be combined, as is done, for example, on the SOHO spacecraft [8.33].

8.1.4 The Coronal Magnetic Field

Figure 8.3 is a very simplified schematic of the corona, showing only the most important features, namely coronal holes and helmet streamers. The latter develop over active regions, the legs of the helmet streamer connecting regions of opposite magnetic field polarity. Electrons are captured inside these loops, therefore the helmet streamer is a bright feature in the corona. The coronal holes on the other hand are regions with open field lines, allowing for a fast electron escape. Thus they appear as rather dark features, often with rays indicating the direction of the field.

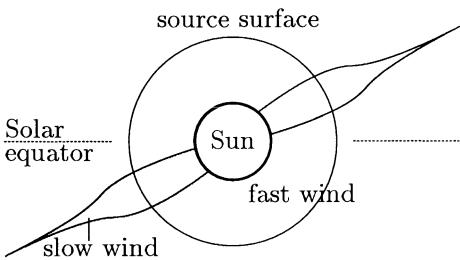


Fig. 8.3. Sketch of the solar corona. Two helmet streamers are shown together with the coronal holes, the solar equator, and the nominal location of the source surface

Over the poles of the Sun, coronal holes are the dominant feature. Occasionally, they can extend down to the solar equator or even into the opposite hemisphere. Streamer-like configurations are confined to the streamer belt, which is associated with the active regions. Thus the streamer belt's extension and its inclination relative to the solar equator varies over the solar cycle.

While the coronal magnetic field as described by these two distinct features appears to be rather orderly, spectral line observations of the photosphere reveal a complex magnetic field pattern associated with sunspots, active regions, and other visible features. Coronal loops, structures such as filaments and prominences, and *in situ* observations in interplanetary space suggest that many of the small-scale photospheric structures form closed loops within less than two solar radii (cf. Fig. 8.9). Thus a solar source surface can be defined: below it, the small-scale structures are closed, while the

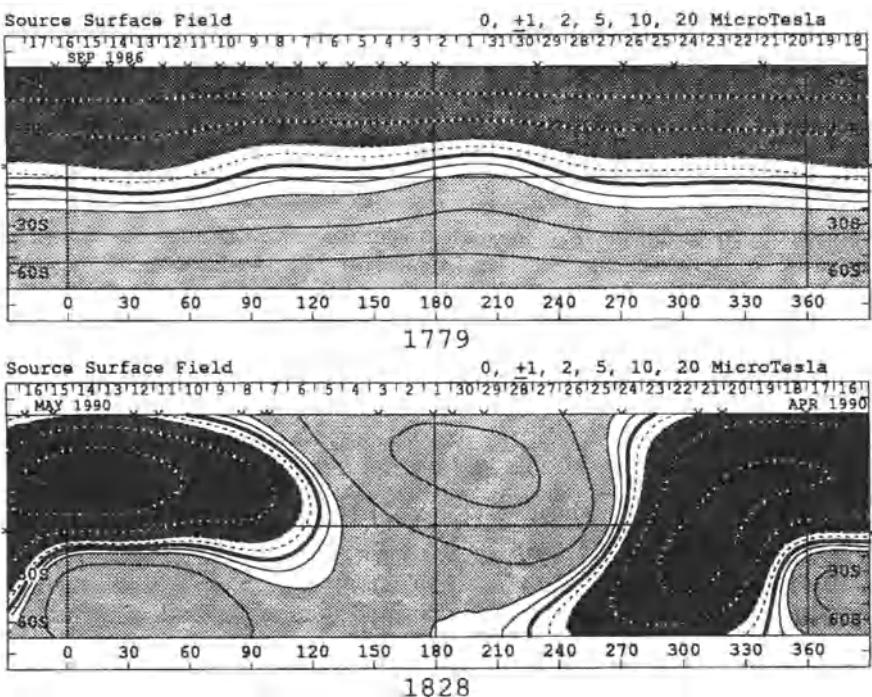


Fig. 8.4. Magnetic field pattern on the source surface map for solar minimum (upper panel) and maximum conditions (lower panel). The thick line is the neutral line separating the fields of opposite polarity. Reprinted from J.T. Hoeksema [8.40], *Advances Space Research* 11, Copyright 1991, with kind permission from Elsevier Science Ltd

overall field pattern is carried outwards by the solar wind. The source surface lies at a height of about 2.5 solar radii and can be determined from the photospheric field pattern using potential theory [8.75, 8.94]. Constraints on the field are as follows: (a) the magnetic field at the source surface is directed radially and (b) currents either vanish or are horizontal in the corona.

Figure 8.4 shows source surface maps for solar minimum and maximum conditions. A distinctive feature is the neutral line (thick line), separating the two magnetic field polarities. At the neutral line, the radial magnetic field vanishes. During solar minimum (upper panel) it is roughly aligned with the solar equator while with increasing solar activity (lower panel) the neutral line becomes wavy and extends to higher solar latitudes. During solar maximum, when the Sun changes polarity, the inclination of the neutral line is maximal. Since the neutral line separates magnetic fields of opposite polarity, a current must flow inside it, and thus the neutral line is a current sheet. In Fig. 8.3 it would extend outwards through the tips of the helmet streamers: its extension into interplanetary space is called the heliospheric current sheet (HCS).

8.2 The Solar Wind

The corona does not show a well-defined outer boundary but ragged structures blending into the background. Thus how far does it extend? Is there an outer boundary?

The Earth's atmosphere is stationary, shaped by an equilibrium between incoming solar radiation and outgoing terrestrial radiation. On the Sun, the situation is different. Here the temperature is much higher and the solar atmosphere is not stable but blown away as solar wind, filling the entire heliosphere. The first direct measurements of the solar wind [8.38] started in 1960. However, a particle flow from the Sun towards the Earth had already been suggested at the beginning of the twentieth century. To explain the relationship between aurorae and sunspots, in 1908 Birkeland [8.11] proposed a continuous particle emission out of these spots. Chapman [8.20] and Chapman and Ferraro [8.23], on the other hand, suggested the emission of clouds of ionized particles during flares only. Except for these plasma clouds, interplanetary space was assumed to be empty. Evidence to the contrary came from an entirely different source, i.e. the observation of comet tails. The tail of a comet neither follows the path of the comet nor is directed exactly radially away from the Sun. Instead, its direction deviates several degrees from the radial direction. Hoffmeister [8.41] suggested that solar particles and not the solar light pressure shape the comet tails. Biermann [8.10] noted that the fainter dust tails of the comets are indeed directed radially and most probably shaped by light pressure, especially since their spectra resemble the solar spectrum. To explain the shape of the main tail, he too invoked a solar particle radiation.

8.2.1 Properties

The high variability of the solar wind in space and time reflects the underlying coronal structures. The most important features can be summarized as follows [8.34, 8.76]: the solar wind is a continuous flow of charged particles. It is supersonic with a speed of about 400 km/s, which is 40 times the sound speed in the solar wind. A plasma parcel travels from the Sun to the Earth within roughly four days. The solar wind carries the solar magnetic field out into the heliosphere, the magnetic field strength amounting to some nanoteslas at the Earth's orbit. Two distinct types of plasma flow are observed – the fast and the slow wind, see, for example, [8.76].

The fast solar wind originates in the coronal holes, the dark parts of the corona dominated by open field lines. Fast solar wind streams are often stable over a long time period (some solar rotations) and variations from one stream to another are small. The fast solar wind has flow speeds between 400 km/s and 800 km/s, the average density is low, about 3 ions/cm³ at 1 AU. About 4% of the ions are helium. This ratio is very stable over different fast streams. The average particle flux is about $2 \times 10^{12} \text{ m}^{-2} \text{ s}^{-1}$, implying a total particle

loss from the Sun of about 1.3×10^{31} /s. The proton temperature is about 2×10^5 K and the electron temperature about 1×10^5 K.

The slow solar wind has lower speeds between 250 km/s and 400 km/s. Its density is about 8 ions/cm³ at 1 AU, and the flux density is about twice as large as that in the fast solar wind. During solar minimum the slow solar wind originates from regions close to the current sheet at the heliomagnetic equator. The relative amount of helium is highly variable, its average is about 2%. During solar maximum the slow solar wind originates above the active regions in the streamer belt, and its helium content is about 4%. Compared with the fast solar wind, it is highly variable and turbulent, often containing large-scale structures such as magnetic clouds or shocks. The proton temperatures are markedly lower, about 3×10^4 K, while the ion temperatures are similar. As in the fast wind, the temperature is always higher parallel to the magnetic field than perpendicular to it. On the average it is $T_{\parallel} \approx 2T_{\perp}$.

Despite their differences, fast and slow solar wind streams also have similarities. For instance, the momentum flux $M = n_p m_p v_p^2$ on average is similar. The same is true for the total energy flux, despite the fact that its individual components, such as kinetic energy, potential energy, thermal energy, electron and proton heat fluxes, and wave energy flux, are different.

A special feature of the solar wind are coronal mass ejections. Here the bulk speed is in the range 400 km/s to 2000 km/s and the composition is significantly different, in particular, up to 30% of the ions can be α -particles, and even Fe¹⁶⁺ or He⁺ can be observed occasionally.

8.2.2 Solar Wind Models

The charge states of heavy ions indicate temperatures of about 10⁶ K in the corona, independent of height. Thus hydrogen is completely ionized and the corona basically can be described as an electron–proton gas with small admixtures of heavier elements. In the lower corona, the electron density is about 10^8 to 10^9 cm⁻³ and it decreases with a scale height of about $0.1 r_{\odot}$. One of the basic questions in understanding the corona and the solar wind is related to heating: because the photosphere only has a temperature of about 5800 K, how can the corona be heated up to a million kelvins?

Chapman's Hydrostatic Corona. One of the first models of the corona, introduced by Chapman in 1957 [8.21], avoided this question and simply described the corona as a static atmosphere:

$$\frac{dp}{dt} = -\varrho \frac{GM_{\odot}}{r^2}, \quad (8.1)$$

with G being the universal constant of gravitation and M_{\odot} the Sun's mass.

Owing to some unknown mechanism, heat is supplied continuously from the photosphere to the corona. Thus thermal energy has to be transported outward through the corona by heat conduction with a heat flow $Q_H = -4\pi r^2 \chi dT/dr$. Because the electrons are the more mobile part in

the electron–proton gas, the heat basically is transported by the electron motion. Thus from the electron density distribution one expects a weakly height-dependent heat conduction coefficient χ . Chapman assumed a completely ionized corona, the variations of density and temperature with height are described as

$$\frac{n}{n_o} = \frac{r^{2/7}}{r_o} \exp \left\{ -\frac{7r_o}{5H_o} \left[1 - \left(\frac{r}{r_o} \right)^{-5/7} \right] \right\} \quad (8.2)$$

and

$$T = T_o \cdot \left(\frac{r}{r_o} \right)^{2/7} \quad (8.3)$$

with the scale height

$$H_o = \frac{2k_B T_o}{m M_\odot G / r^2}. \quad (8.4)$$

Under coronal conditions, heat conduction is about a factor of 20 more efficient than in copper. Thus the temperature decreases only weakly with height, cf. (8.3), with the consequence that “... the coronal gas surrounding the Earth may be expected to have a temperature of order of 100 000 K. This is ... consistent with my main inference – that the Earth is surrounded by very hot coronal gas, which greatly distends our outer atmosphere and that heat must flow from it by conduction into our atmosphere” [8.21, p. 477].

Parker’s Hydrodynamic Corona. But the continuous particle flow inferred from the comet tails is in contrast to a static atmosphere extending far behind Earth’s orbit. Parker [8.66] argued that, because of the high temperatures, the corona is not stationary but that heat is transported by particle streaming. Thus the solar atmosphere and the continuous particle radiation from the Sun are both the same. Parker used a hydrodynamic approach. Thus the hydrostatic equation has to be complemented by a term describing the fluid motion, leading to Bernoulli’s equation. However, Parker did not solve the heating and heat transport problems.

In a simple approach, only protons are considered because they are the dominant ion species and carry virtually all of the mass of the solar wind. The momentum balance then is $\varrho(\mathbf{u} \cdot \nabla)\mathbf{u} = -\nabla p - \varrho M_\odot G / r^2$ or, in the one-dimensional case for a spherically symmetric corona,

$$u_r \frac{du_r}{dr} = \frac{1}{nm} \frac{d}{dr} (2nk_B T) - \frac{GM_\odot}{r^2}. \quad (8.5)$$

The factor 2 in the pressure term $nk_B T$ considers that both electrons and protons contribute a factor of $nk_B T$ to the pressure. With the equation of continuity $n(r) \cdot u_r(r) \cdot r^2 = n_o \cdot u_{r_o} \cdot r_o^2$, (8.5) can be written as

$$\frac{du_r}{dr} \left[u_r - \frac{2k_B T}{mv_r} \right] = \frac{2k_B r^2}{m} \frac{d}{dr} \frac{T}{r^2} - \frac{GM_\odot}{r^2}. \quad (8.6)$$

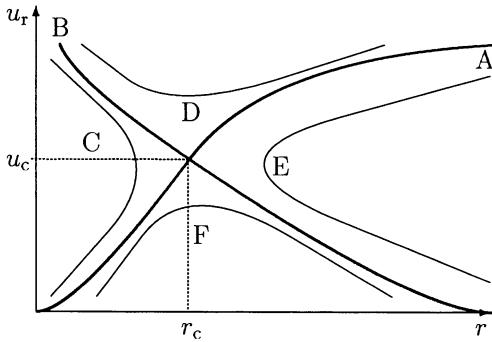


Fig. 8.5. Topology of different solutions for the solar wind equation (8.6)

To describe the temperature gradient, Parker assumed an isothermal corona above about $1.4r_\odot$ as suggested by the charge states of the heavier ions.

Solutions of (8.6) are shown in Fig. 8.5. There are two curves representing special solutions. These curves intersect at the critical point (u_c, r_c) with

$$r_c = \frac{GM_\odot m}{4k_B T} \quad \text{and} \quad u_c = \sqrt{\frac{2k_B T_o}{m}}. \quad (8.7)$$

The solution ‘A’ is the observed solar wind: it starts as a subsonic flow in the lower corona and accelerates with increasing radius. At the critical point r_c , the solar wind becomes supersonic. The solution than takes the form

$$\frac{u_r^2}{u_c^2} - \ln \frac{u_r^2}{u_c^2} = -3 + 4 \ln \frac{2u_c^2 r}{w r_o} + \frac{2w^2 r_o}{u_c^2 r}, \quad (8.8)$$

with $w = \sqrt{GM_\odot/r}$. For large distances, (8.8) can be approximated as $u_r \sim 2u_c \sqrt{\ln(r/r_o)}$. Figure 8.6 shows the radial dependence of the solar wind speed for different coronal temperatures. For a coronal density of about $2 \times 10^8 \text{ cm}^{-3}$ and a temperature of one million kelvins, the critical point is at about $6r_\odot$. The solar wind than accelerates up to about $40r_\odot$, afterwards propagating at a nearly constant speed of 500 km/s. The supersonic flow does not extend indefinitely, its density decreases during expansion. At a certain radial distance, probably 70 AU or beyond, the solar wind pressure will become too small to further support a supersonic flow. Where the flow is slowed down to subsonic speed, a termination shock forms. The solar wind then continues as a subsonic flow until the pressure of the interstellar gas becomes larger than the combined pressure of the solar wind and the frozen-in magnetic field. This is the heliopause, the boundary of the heliosphere which is expected at about 100 AU or beyond. In front of the heliopause, a bow shock might develop where the interstellar gas is slowed down by the obstacle heliosphere.

Solutions ‘C’ and ‘F’ of the solar wind equation also start as subsonic flows in the lower corona but develop differently. In solution ‘F’ the speed increases only weakly with height and the critical velocity is not acquired at

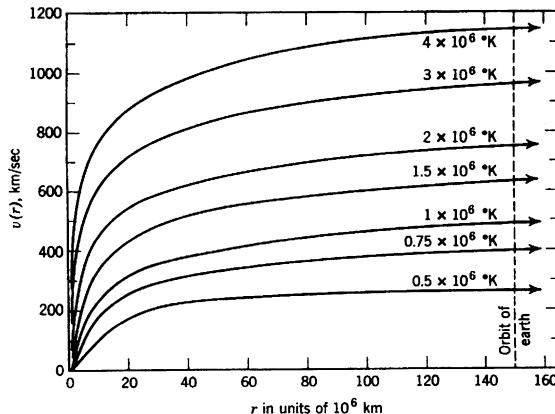


Fig. 8.6. Radial dependence of the solar wind speed for different coronal temperatures. Reprinted from E.N. Parker [8.67], *Interplanetary dynamical processes*, Copyright 1963, with kind permission from J. Wiley & Sons, Inc.

the critical radius. The flow then continues to propagate radially outward, but then slows down and can be regarded as a solar breeze only. In 'C' the flow has accelerated too fast and has become supersonic before reaching the critical height. It then turns around and flows back towards the Sun as a supersonic flow.

The other curve going through the critical point, 'B', starts as a supersonic flow in the lower corona and becomes subsonic at the critical point. This flow would continue to propagate outwards at subsonic speed. If the flow had decelerated less, as in curve 'D', it still would be supersonic at the critical point, where it would accelerate again, leaving the Sun as a supersonic flow. Solution 'E' is entirely different. It starts as an inward flow blowing subsonically towards the Sun from infinity. The flow accelerates as it approaches the Sun, becoming supersonic at some distance larger than v_c . At that point the flow turns back and propagates outward as a supersonic flow.

Only one of the mathematically possible solutions of (8.6), curve 'A' in Fig. 8.5, is an approximation of the solar wind. But how good is this model? The assumption of an isotropic pressure p is valid near the Sun where isotropy can be maintained by collisions. As the plasma moves farther out, collisions become less frequent and the plasma pressure parallel to the magnetic field exceeds the perpendicular one by a factor of about 2. The temperature then enters the equation of state. Here it is assumed to be isotropic, too. Again, this might be true close to the Sun but it is not true at the orbit of Earth. In addition, electron and proton temperatures are not the same, as is assumed in the model. These differences do not change the general character of the solution but modify the numbers. Another limitation of the model is the consideration of only one particle species, namely protons. Since α -particles are four times heavier than protons, even the 2% to 4% of He in the solar wind contribute significantly to the momentum transport. Thus an additional set of equations should be considered, leading to a reduction of the flow speed.

A more severe limitation concerns the fields. In the derivation of the hydrodynamic flow, no effects of the magnetic and electric fields were considered: the electromagnetic forces in the momentum balance were ignored. A more elaborate model should consider fields, too. In such a magnetohydrodynamic (MHD) model the critical point is lower in the corona, for average conditions at a height of about two solar radii, which is close to the height of the (fictitious) source surface. The general character of the solution nonetheless is the same as in the hydrodynamic model.

Parker's hydrodynamic solar wind model nevertheless is a valid approximation of the solar wind in quiet conditions. A comparison of model results with the solar wind observations, however, reveals a rather puzzling fact: the hydrodynamic model is more appropriate to describe the slow wind originating in the streamer belt with its complex magnetic field structures than the fast solar wind blowing directly out of the coronal holes, a situation which is much closer to the assumptions inherent in the model.

8.2.3 The Problem: Coronal Heating and Solar Wind Acceleration

Although the hydrodynamic description of the solar wind is a reasonable and valuable approach, one fundamental problem has been neglected: the heating of the corona. Basically, two lines of thought have evolved: heating by MHD waves and turbulence or, alternatively, small-scale impulsive energy releases due to reconnection, sometimes called nano-flares.

Waves and Turbulence. Of the basic MHD waves, the fast and slow magnetoacoustic waves are compressive, while the Alfvén wave is a non-compressive propagation of fluctuations along the field. In a collisionless plasma, the Alfvén wave propagates undamped, whereas the magnetoacoustic waves undergo Landau damping. Wave energy is then converted into thermal energy, mainly of the ion component in the plasma. Under solar conditions, the slow mode is damped very strongly while the fast mode can propagate up to a distance of about $20r_{\odot}$. On the basis of this, Barnes et al. [8.7] developed a model of the solar wind with coronal heating by fast magnetoacoustic waves. While the Alfvén wave is not damped, it nonetheless contributes to momentum transport and can be interpreted as a radiation pressure, accelerating the plasma. Although non-thermal broadening of some spectral lines indicates the existence of waves or turbulence in the lower corona, it is not completely understood which kind of waves these are, how they propagate outward, and whether the observations really are indicative of wave fields or, rather, of turbulence. A brief review can be found in [8.6, 8.56].

Impulsive Energy Release. So far, the corona has been treated only hydrodynamically and the magnetic field has been ignored. Even in the attempts to explain coronal heating by MHD waves, the field has been considered only



Suppl.
→8.3

as an agent that makes these waves possible. But the field also contains magnetic energy. The conversion of field energy into thermal energy therefore could provide a heating mechanism. On a much larger scale, this concept is widely applied in space physics: acceleration models of particles in the magnetosphere's tail or solar flares often involve reconnection because the magnetic field is the only source of energy available.

As we saw in Sect. 3.3.3, reconnection requires a certain magnetic field configuration, namely fields of opposite polarity. The photosphere, as the top of the convection zone, is in continuous motion with bubbles rising and falling and plasma flowing in and out. With this plasma the magnetic field is shuffled around, too. Thus on a small scale, magnetic field configurations suitable for reconnection might form, thereby converting field energy into thermal energy. Observational evidence for such small-scale impulsive energy releases is found in electromagnetic radiation, in particular in so-called bright X-ray points and small-scale exploding EUV events. Reconnection as a mechanism for coronal heating is utilized in many slightly different models, occasionally not only involving the heating of the plasma but also a direct expulsion of plasma bubbles.

8.3 The Interplanetary Magnetic Field

The photospheric magnetic field was discovered by Hale in 1902. The splitting of spectral lines due to the Zeeman effect suggests a photospheric field in the order of 10^{-4} T or 1 G (G: gauss) outside of sunspots and 3000 G to 4000 G inside it. Within less than 2 solar radii, this complex and highly variable field is reduced to a rather simple, radially directed one. Since the conductivity of the solar wind is high, the magnetic field is frozen into it and carried out into interplanetary space. The Sun's rotation winds up these field lines to Archimedean spirals. Thus with increasing radial distance, the originally radial magnetic field becomes more and more toroidal.

8.3.1 Spiral Structure

The Sun rotates with a sidereal rotation period of 27 days. The solar wind flows radially away from the Sun, carrying the frozen-in magnetic field. While the solar wind propagates outward, the base of the field line frozen into the plasma parcel is carried westward, forming an Archimedean spiral, as shown in Fig. 8.7. A similar effect can be observed with a rotating sprinkler; thus the deformation of the field lines is also called the garden-hose effect.

The equation of the Archimedean spiral can be derived from the displacements Δr and $\Delta\varphi$. If we assume as initial conditions of the plasma parcel on the Sun a source longitude φ_o and a source radius r_o , at a time t the parcel can be found at the position $\varphi(t) = \omega_\odot \cdot t + \varphi_o$ and $r(t) = u_{\text{sowi}} \cdot t + r_o$. Eliminating the time yields the equation for the Archimedean spiral:

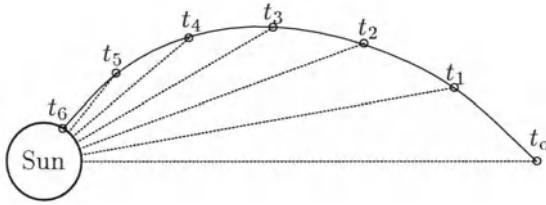


Fig. 8.7. Deformation of a magnetic field line due to the combination of a radial plasma flow and the Sun's rotation. The numbers indicate consecutive times after a plasma parcel has left the Sun at time t_o .

$$r = u_{\text{sowi}} \cdot \frac{\varphi - \varphi_o}{\omega_\odot} + r_o . \quad (8.9)$$

With $\psi = \omega_\odot r / u_{\text{sowi}}$, the path length s along the spiral is given as

$$s = \frac{1}{2} \frac{u_{\text{sowi}}}{\omega_\odot} \left(\psi \cdot \sqrt{\psi^2 + 1} + \ln \left\{ \psi + \sqrt{\psi^2 + 1} \right\} \right) . \quad (8.10)$$

The magnetic field in the equatorial plane can be expressed in polar coordinates $\mathbf{B} = (B_r, B_\varphi)$. The magnitude of \mathbf{B} depends on the radial distance only, thus it is $|\mathbf{B}| = B(r)$. Gauss's law in spherical coordinates (cf. Appendix Sect. A.2.4) yields

$$\nabla \cdot \mathbf{B} = \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 B_r) = 0 \quad (8.11)$$

or $r^2 B_r = r_o^2 B_{r_o}$. Thus the magnetic flux through spherical shells is conserved and the radial component of the field decreases as

$$B_r = B_o \left(\frac{r_o}{r} \right)^2 . \quad (8.12)$$

Since the magnetic field is constant, it is $\partial B / \partial t = 0$. From the frozen-in condition (3.88) we then get $\nabla \times (\mathbf{u} \times \mathbf{B}) = 0$, or in spherical coordinates

$$\frac{1}{r} \frac{\partial}{\partial r} (r (u_\varphi B_r - u_r B_\varphi)) = 0 . \quad (8.13)$$

Thus we have $r(u_\varphi B_r - u_r B_\varphi) = \text{const}$. Let us assume r_o to be at the source surface. There \mathbf{B} is radial and we get

$$r u_\varphi B_r - r u_r B_\varphi = r_o u_{\varphi o} B_o = r_o^2 \omega_\odot B_o . \quad (8.14)$$

In the second step, the rotation speed of the Sun was used to describe the azimuthal component of the solar wind speed at the source surface. From (8.14) the azimuthal component of the magnetic field is

$$B_\varphi = \frac{r u_\varphi B_o - r_o^2 \omega_\odot B_o}{r u_r} = \frac{u_\varphi - r \omega_\odot}{u_r} B_r . \quad (8.15)$$

For large distances, $r \omega_\odot > v_\varphi$, (8.15) can be approximated as $B_\varphi = -r \omega_\odot B_r / u_r$. The azimuthal component therefore decreases with $1/r$ while the radial component decreases as $1/r^2$. The field strength decreases with r as

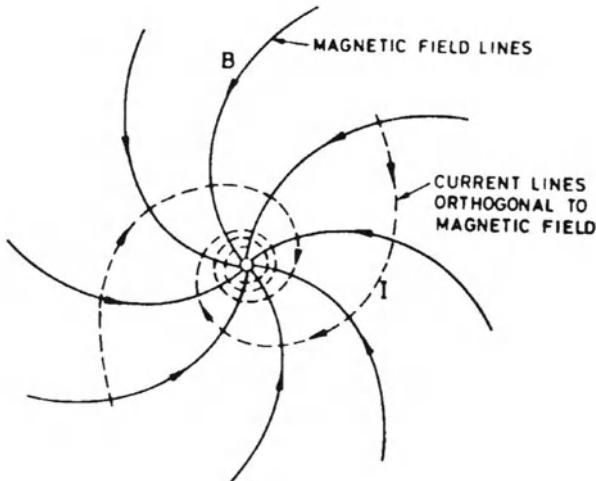


Fig. 8.8. A sketch of the Archimedean magnetic field spirals and the spiraling electric field in the current sheet. Magnetic field lines are shown slightly above the equatorial plane, close below they have opposite polarity. Reprinted from H. Alfvén [8.1], *Cosmic plasmas*, Copyright 1981, with kind permission from Kluwer Academic Publishers

$$B(r) = \frac{B_0 r_0}{r^2} \sqrt{1 + \left(\frac{\omega_\odot r}{u_r}\right)^2}. \quad (8.16)$$

The angle ψ between the magnetic field direction and the radius vector from the Sun is $\tan \psi = B_\varphi / B_r$. For large distances this reduces to $\tan \psi = \omega_\odot r / u_r$. At the Earth's orbit, $\tan \psi$ is about 1 for typical solar wind conditions, and thus the field line is inclined by 45° with respect to the radial direction.

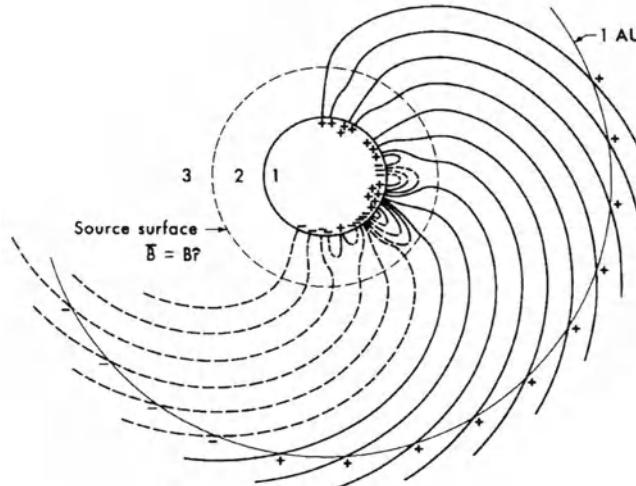
The current in the heliospheric current sheet is related to the magnetic field by Ampere's law (2.7). In spherical coordinates the current density in the plane of the ecliptic then is

$$j_r = j_0 \cdot \frac{r_0}{r} \quad \text{and} \quad j_\varphi = \frac{B_\varphi}{B_r} j_r = j_0 \frac{B_\varphi}{B_r} \frac{r_0}{r}. \quad (8.17)$$

Figure 8.8 shows a sketch of the magnetic and the spiraling electric fields.

8.3.2 Sector Structure

So far we have considered only the shape of the interplanetary magnetic field but not its direction. First, long-term observations by IMP 1 in 1963 over a couple of solar rotations revealed a sector pattern as shown in Fig. 8.9: the magnetic field polarity is uniform over large angular regions and then abruptly changes polarity. These magnetic field sectors are stable over many solar rotations. At most times either two or four sectors can be observed: if the neutral line is tilted without any wiggles, a pattern of two magnetic field sectors arises; a wavy neutral line leads to four or more sectors. During solar maximum the sector structure is complex and distorted by a large number of transient disturbances.



1. Photospheric magnetic field
2. Magnetic field calculated from potential theory $\nabla^2\phi = 0$
3. Magnetic field transported by solar wind $\frac{d\vec{B}}{dt} = -\vec{B}(\nabla \cdot \nabla) + (\vec{B} \cdot \nabla)\nabla$
(observed by spacecraft at 1 AU)

Fig. 8.9. Relationship between the photospheric magnetic field, the source surface field, and the interplanetary magnetic field. Dashed and solid field lines indicate negative and positive magnetic field polarities, respectively. Reprinted from K.-H. Schatten et al. [8.75], *Solar Physics* 6, Copyright 1969, with kind permission from Kluwer Academic Publishers

8.3.3 The Ballerina Model

The sector boundaries are the extension of the neutral line into the interplanetary medium, the so-called heliospheric current sheet (HCS). Figure 8.10 shows the HCS extending far into the interplanetary medium. The inclination of the neutral line defines the width of a cone inside which an observer in space alternately sees different polarities of the coronal/interplanetary magnetic field. The maximum inclination of the neutral line at each time is called the tilt angle. It can be used as an alternative measure of solar activity and is an indicator of the range over which both field polarities can be observed.

Figure 8.11 shows a three-dimensional sketch of the wavy current sheet, where the solid lines indicate magnetic field lines. The figure depicts conditions close to the solar minimum. During solar maximum a more complex and more wavy structure would be observed and the neutral line would be bent towards high latitudes, as also suggested in the cross-section shown in Fig. 8.10.

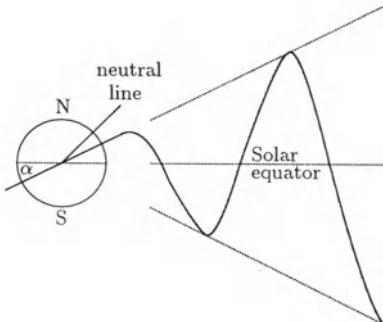


Fig. 8.10. Heliospheric current sheet and definition of the tilt angle α as inclination of the neutral line in the tilted dipole model proposed in [8.69]. The width of the cone is twice the tilt angle α

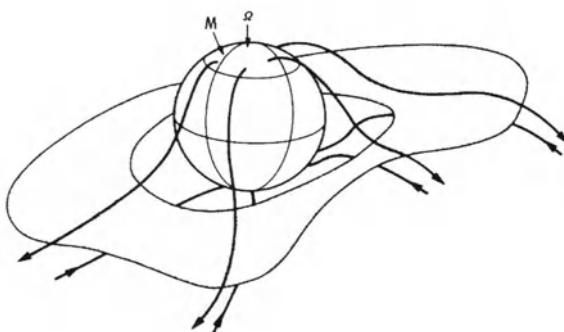


Fig. 8.11. Current sheet in the inner heliosphere in the Ballerina model. The thick lines indicate the magnetic field lines. Reprinted from Smith et al. [8.82], *J. Geophys. Res.* **83**, Copyright 1978, American Geophysical Union

8.3.4 Corotating Interaction Regions

Fast and slow solar wind streams originate on the Sun. While these streams propagate outward, the frozen-in magnetic field is wound up to Archimedean spirals. In a slow stream, the field line is curved more strongly than in a fast stream. Since field lines are not allowed to intersect, at a certain distance from the Sun an interaction region develops between the two streams. Because both rotate with the Sun, this is called a corotating interaction region (CIR). Often the source locations of the fast and slow solar winds are rather stable and an observer in space sees the CIR again during the following solar rotations. In this case, it is called a recurrent corotating interaction region.

Figure 8.12 shows an idealized sketch of the evolution of a CIR in the inner heliosphere. On the Sun, there is an abrupt change in solar wind speeds from fast to slow. As these streams propagate outward, flow compression and deflection on both sides of the interface smooth the jump, leading to a continuous increase in plasma speed. The region of compressed plasma at the transition between the fast and slow stream at 1 AU typically extends over about 30° while the plasma might originate from a coronal region as wide as 90° or more. Thus magnetic field sector boundaries often found close to the compression region are not necessarily related to the interface but originate in a coronal region far from the boundary between the fast and slow streams. This is also evident from observations of the corona and the photosphere:

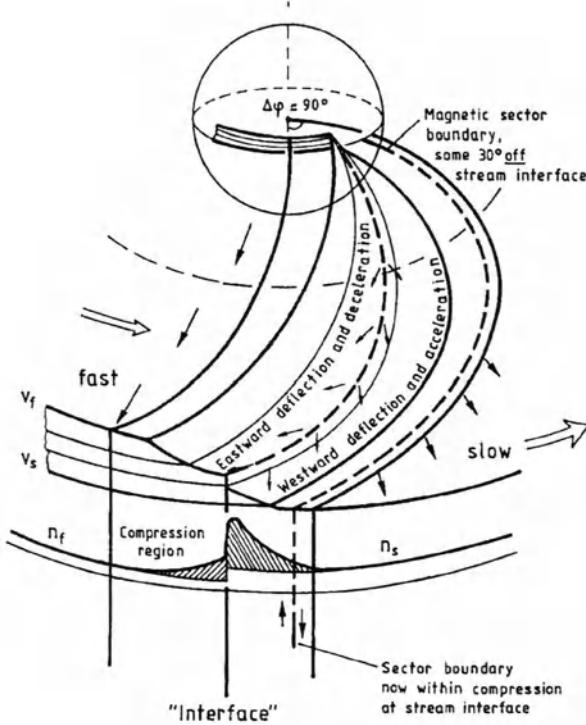


Fig. 8.12. Idealized view of a corotating interaction region (CIR) in the inner heliosphere. Reprinted from R. Schwenn [8.76], in *Physics of the inner heliosphere, vol. I* (eds. R. Schwenn and E. Marsch), Copyright 1990, Springer Verlag

the boundaries of the coronal holes are not related to the neutral line of the photospheric field. In particular, the polar coronal hole can extend into the opposite hemisphere, thereby crossing the solar equator as well as the current sheet.

With increasing distance from the Sun, the characteristic propagation speeds, which are the sound and Alfvén speeds, decrease. At some distance between 2 and 3 AU, the density gradient on both sides of the compression region becomes too large and a shock pair develops, propagating away from the interface. The shock propagating into the slow wind is called the forward shock, the one propagating into the fast wind is the reverse shock.

Corotating interaction regions tend to distort or even destroy all small-scale fluctuations and disturbances propagating outward from the Sun. In the outer heliosphere, the magnetic field and therefore also the shock fronts are more azimuthally aligned, sometimes extending around the entire Sun. Thus finally the spoke-like structure of different solar wind streams close to the Sun is converted into a ring-like shell of concentric shock waves propagating outward like waves from a stone thrown into water. When CIRs or CIRs and traveling interplanetary shocks interact, merged interaction regions result which play a crucial role in the modulation of the galactic cosmic radiation.

8.4 Plasma Waves in Interplanetary Space

The interplanetary magnetic field is highly variable on different temporal and spatial scales. For instance, fast and slow solar wind streams form interaction regions. Coronal mass ejections, sometimes driving an interplanetary shock, are transient disturbances, and on a smaller scale, waves and turbulence are superimposed on the average field.

Figure 8.13 illustrates the variability of the magnetic field on time-scales of minutes to hours. It shows the magnetic field azimuth (angle between magnetic field line and radial direction), the elevation (inclination of the field with respect to the plane of ecliptic), and the flux density for a time period of 7 hours during unusually quiet (left) and turbulent (right) interplanetary conditions. Magnetic field fluctuations are more pronounced in direction than in flux density and are quite irregular in amplitude and frequency.

8.4.1 Power-Density Spectrum

The magnetic field fluctuations can be described by a power-density spectrum [8.8, 8.9]:

$$f(k_{\parallel}) = C \cdot k_{\parallel}^{-q}. \quad (8.18)$$

Here k_{\parallel} is the wave number parallel to the field, q the slope, and C a constant describing the level of the turbulence.

The power-density spectrum in Fig. 8.14 shows magnetic field fluctuations on different scales. Its slope is distinct in different parts of the spectrum, indicating different sources and modes of turbulence. Basically, four regimes can be distinguished:

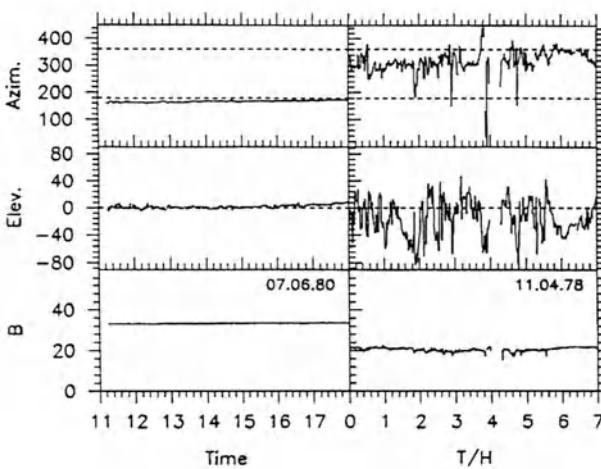


Fig. 8.13. Fluctuations in the magnetic field azimuth, the elevation, and the flux density for extremely quiet conditions (*left*) and during a turbulent phase (*right*). The different azimuth angles indicate an outward direction of the field in the left panel and an inward direction in the right panel. Data from the Helios magnetometer, University of Braunschweig

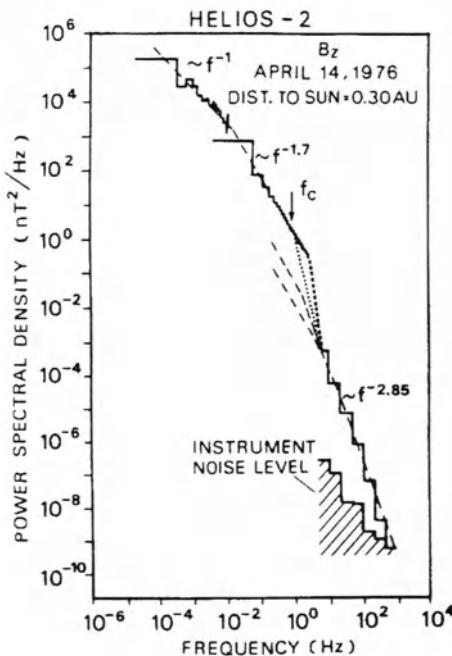


Fig. 8.14. Magnetic field power-density spectrum. Reprinted from K.U. Denskat et al. [8.28], *J. Geophys. Res.* **87**, Copyright 1983, American Geophysical Union

- Large-scale structures lasting a few days up to a solar rotation are related to the stream structure of the solar wind and to solar wind expansion. Both processes are the sources of turbulence on smaller scales; the frequencies of the large-scale turbulence are below 5×10^{-6} Hz.
- Meso-scale fluctuations are associated with the flux-tube structure of the interplanetary medium which originates in the photospheric supergranulation. Frequencies range between 5×10^{-6} and about 10^{-4} Hz.
- In the inertial range, mainly Alfvén waves with periods between some 20 min and more than 15 h are found, corresponding to frequencies between 10^{-4} and 1 Hz. The slope q varies between -1.5 and -1.9 . Magnetic field fluctuations in the inertial range seem to be responsible for the scattering of protons in interplanetary space, cf. Sect. 9.2.
- The smallest scales are in the dissipation range above 1 Hz. Here the spectrum is steeper with a slope close to -3 . The observed fluctuations can be attributed to ion cyclotron waves, ion acoustic waves, and Whistlers.

8.4.2 Waves or Turbulence?

So far we have described the fluctuations in terms of waves. But a single observer in interplanetary space cannot decide whether the fluctuations carried across him by the solar wind are waves or turbulence because he is not able to distinguish between spatial and temporal variations. Thus the question

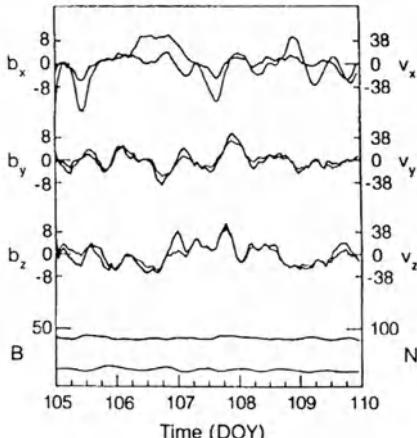


Fig. 8.15. Hourly averages of magnetic field and plasma speed components. The good correlation between the plasma and field fluctuations characteristic of alfvénic fluctuations is evident. Reprinted from R. Bruno et al. [8.15], *J. Geophys. Res.* **90**, Copyright 1985, American Geophysical Union

of whether the magnetic field fluctuations should be interpreted in terms of waves or turbulence has led to a long and fruitless controversy. Only the modern concepts of MHD turbulence, e.g. [8.61], allowed a kind of unification of both approaches: MHD turbulence is not the simple superimposition of different waves, but rather consists of wave-packets which can interact with each other, or can decay and excite new waves. Nevertheless, the interpretation of the interplanetary magnetic field fluctuations is still under debate, see, for example, [8.55, 8.91].

We will not go into the details of this debate, but only introduce the concept of alfvénic turbulence or alfvénicity because it offers a helpful tool in the description of magnetic field turbulence. Alfvén waves are transverse waves propagating along the magnetic field line with the Alfvén speed $v_A = B_0/\sqrt{4\pi\rho}$. Fluctuations are alfvénic if the fluctuations $\delta\mathbf{u}_{\text{sowi}}$ in flow speed and $\delta\mathbf{B}$ in flux density obey

$$\delta\mathbf{u}_{\text{sowi}} = \pm \frac{\delta\mathbf{B}}{\sqrt{4\pi\rho}} . \quad (8.19)$$

In the plasma and field data shown in Fig. 8.15, the good correlation between these fluctuations is evident. Fluctuations are classified as alfvénic if the correlation coefficient is larger than 0.6. Obviously, this is true for Alfvén waves. But there is also a large number of other fluctuations which fulfill (8.19). In particular, structures with variable $|\mathbf{B}|$ can also fulfil (8.19), as can many of the static structures in the solar wind. Alfvén waves contribute only a small amount to the alfvénic fluctuations.

The alfvénicity of fluctuations is useful in the description of the evolution of turbulence from an orderly state (high alfvénicity) to an entirely stochastic one. For instance, the alfvénicity is larger close to the Sun than at the Earth's orbit, indicating that in the inner heliosphere most of the fluctuations are of coronal origin. As these fluctuations decay, the alfvénicity decreases and

the slope of the power-density spectrum evolves towards $-5/3$, which is the Kolmogoroff spectrum of random, uncorrelated turbulence. The alfvénicity is larger in fast solar wind streams than in slower ones, and thus an orderly state is preserved over larger spatial scales. If fast and slow streams interact, the alfvénicity decreases and the spectrum takes the slope of the Kolmogoroff spectrum. The alfvénicity can also be different on both sides of the heliospheric current sheet.

The solar wind and its magnetic field therefore have to be understood as a dynamically evolving, inhomogeneous, anisotropic, turbulent magneto-fluid. With increasing distance, the fluctuations embedded in this fluid evolve from alfvénic turbulence close to the Sun towards a Kolmogoroff spectrum.

8.5 The Three-Dimensional Heliosphere

Until the early 1990s our knowledge of the heliosphere had been limited to the plane of ecliptic. The main goal of the Ulysses mission, launched in October 1990, is the study of the heliosphere's third dimension, i.e. plasmas, particles, and fields in the polar regions of the Sun.

The Ulysses trajectory for the prime mission is shown in Fig. 8.16. A swing-by at Jupiter allowed Ulysses to escape out of the plane of ecliptic into an elliptical orbit around the Sun. This orbit is inclined by 80° relative to the solar equator; the orbital period is 6.3 years. Ulysses flew below the Sun's south pole in autumn 1994 and above her north pole in summer 1995, both polar passes were made during solar minimum. The mission will continue for

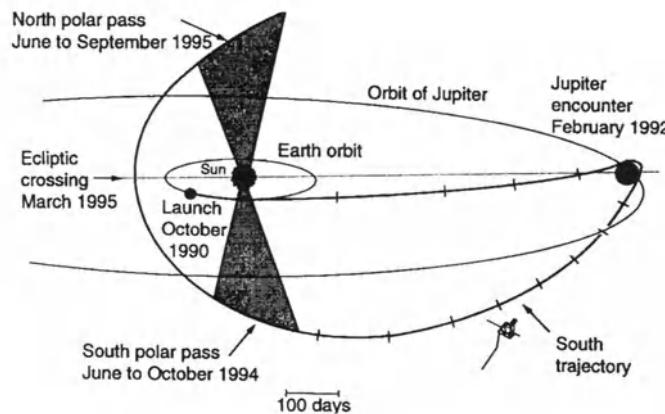


Fig. 8.16. Flight path of Ulysses from launch to the end of the prime mission. After a swing-by at Jupiter in February 1992, the spacecraft left the plane of ecliptic. The mission will continue for another orbit around the Sun. Reprinted from E.J. Smith et al. [8.81], *Science* **268**, Copyright 1995, with kind permission from the American Association for the Advancement of Science

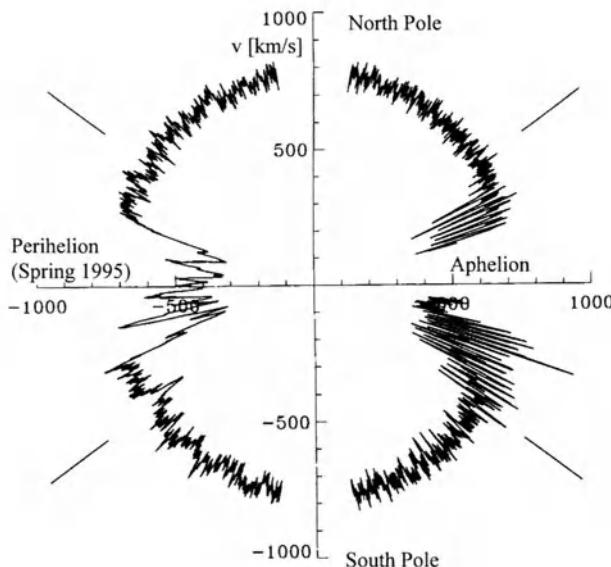


Fig. 8.17. Polar plot of solar wind speed as a function of heliolatitude for the out-of ecliptic phase of the Ulysses mission from February 1992 through January 1997. Reprinted from D.J. McComas et al. [8.58], *J. Geophys. Res.* **103**, Copyright 1998, American Geophysical Union

another orbit, allowing for observations over the poles at solar maximum, too. The most important results of the first polar pass are summarized in a series of papers in *Science* **268**.

Plasma and field observations offered some surprises for the scientists. For instance, the radial component of the magnetic field, which is most easily related to the global solar magnetic field, failed to show any latitudinal gradient [8.5], although the photospheric magnetic field clearly reveals a dipole-like pattern. Thus the magnetic flux is removed from the poles towards the equatorial regions, as had been suggested in the sketch in Fig. 8.11. This observation has consequences for solar wind acceleration models, where the resulting stresses of the magnetic field have not thus far been considered, and for models describing the modulation of galactic cosmic rays.

The plasma measurements showed a pronounced latitudinal variation of the solar wind speed. As shown in Fig. 8.17, it increases from about 450 km/s in the equatorial plane to about 750 km/s above the poles. Up to a latitude of about 30°S, which corresponds to the tilt angle at that time, there is a strong variation between fast and slow streams with a period of about 26 days, resulting in a recurrent CIR. At latitudes higher than 50°S, only the fast solar wind streaming out of the coronal hole is observed [8.68], in agreement with our expectations. The composition of the solar wind, on the other hand, offered some surprises and also a big challenge for theory: the compositions of the fast and slow streams were markedly different, but the abundances were not what would be expected for ions accelerated in the hot corona. Instead, they were more representative of ions formed in the lower temperatures of the chromosphere [8.35].

A comparison between plasma and magnetic field fluctuations inside the coronal hole revealed the existence of large-amplitude, long-period Alfvén waves propagating outward from the Sun [8.5]. Thus it appears that the fluctuations, which originate close to the Sun in the acceleration region of the solar wind, are less likely to decay in the uniform fast solar wind flow than they are in the complex and interacting flows in the plane of ecliptic.

8.6 The Active Sun

The variability of the Sun is most obvious in the number and spatial distribution of sunspots. But other properties, such as the electromagnetic radiation, the solar wind, and the solar and interplanetary magnetic fields, change, too. And, of course, during times of high solar activity there are the phenomena of violent releases of energy and matter, i.e. flares and coronal mass ejections.

8.6.1 The Solar Cycle

The first records of sunspot, the prime indicators for solar activity, date back to the fourth century BC when the Greek astronomers noted dark spots on the solar surface. Ancient Chinese astronomers knew about sunspots, as did the pre-Spanish Peruvians. In Europe, isolated records can be traced back to the ninth century, but systematic observations started only with the development of the telescope early in the seventeenth century. The variability of shape, location, and number of sunspots was recognized early; however, owing to the Maunder minimum, a period of very low solar activity in the seventeenth century, the solar cycle was recognized only in 1843 by H. Schwabe.

Figure 8.18 shows two different representations of the solar cycle. In the top panel, a butterfly diagram gives the latitudinal distribution of sunspots, and in the lower panel, the sunspot number is shown. Alternatively, a sunspot relative number can be used which considers the sunspot size and its relation to other spots or an active region. At the solar minimum, the Sun is almost spotless. Then spots start to appear at latitudes around 30° . These spots are relatively stable and often can be observed over a couple of solar rotations. They move towards the solar equator while at higher latitudes new spots appear. The number of sunspots increases until the solar maximum. Afterwards, only a few new sunspots emerge while the sunspots at low latitudes dissolve. The total number of sunspots decreases until just after the solar minimum when new sunspots begin to emerge at higher latitudes. The average duration of such a cycle is 11 years with variations between 7 and 15 years. The magnetic cycle is twice as long: within the 11 years of a sunspot cycle, the solar field reverses its polarity once, cf. contours in the butterfly diagram in Fig. 8.18. Thus it takes 22 years for the original polarity pattern to be restored. This 22-year cycle also is called the Hale cycle or the solar magnetic cycle.

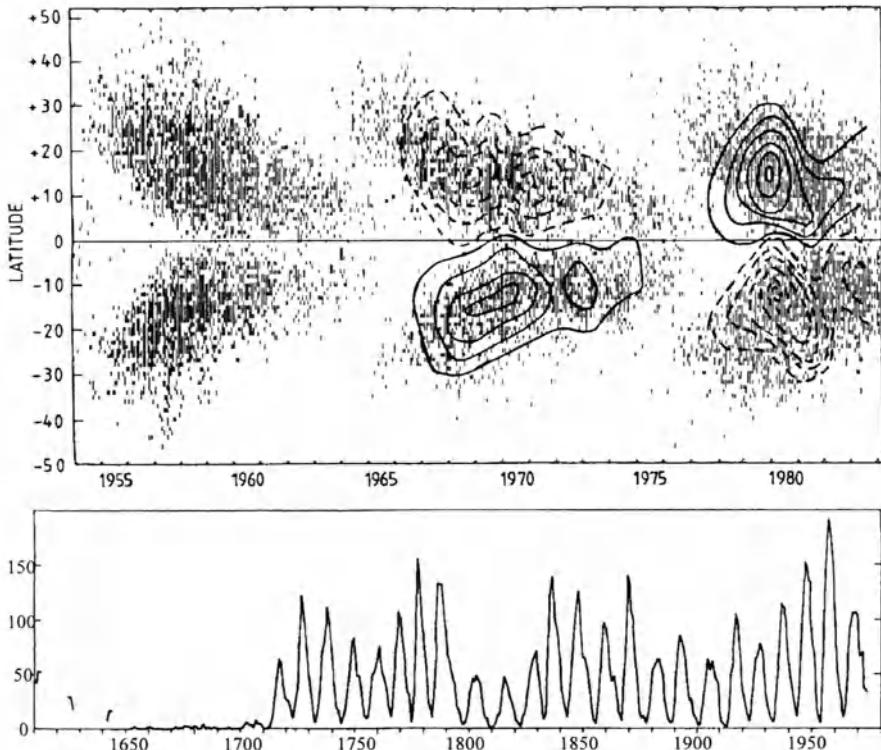


Fig. 8.18. Different representations of the solar cycle. (*Top*) Butterly diagram of the latitudinal distribution of sunspots with each bar marking a sunspot, combined with contours for the radial magnetic field; the contours are for $\pm 27\mu\text{T}$, $\pm 54\mu\text{T}$..., solid positive, dashed negative. From M. Stix [8.84], in *Solar and stellar physics* (eds. E.H. Schröter and M. Schüssler), Copyright 1987, Springer Verlag Berlin. (*Bottom*) Sunspot number versus time. The small numbers in the 1650–1710 period mark the Maunder Minimum. Reprinted from J.A. Eddy [8.31], in *Physics of solar planetary environments* (ed. D.J. Williams), Copyright 1976, American Geophysical Union

From the lower panel in Fig. 8.18 it is evident that solar activity is highly variable in different solar cycles. For instance, the sunspot area in the 1958 solar maximum was about twice as large as the one in the 1885 maximum. Over longer periods, variations up to a factor of 4 have been observed, at some times, e.g. during the Maunder Minimum 1650–1710 solar activity and sunspot numbers can be even smaller.

8.6.2 A Simple Model of the Solar Cycle

Evidently, the clue to understanding the solar cycle lies in the magnetic field and its reversal. Helpful information, in particular for a proof of models, can be found in the details of sunspots:

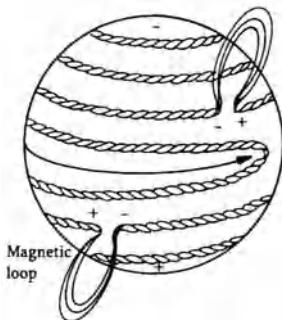


Fig. 8.19. Increased magnetic buoyancy drives flux tubes through the photosphere, creating two sunspots of opposite polarity. Reprinted from K. Lang [8.53], *Sun, earth, and sky*, Copyright 1995, Springer Verlag

1. Sunspots always start at relatively high latitudes and move towards the equator (Spörer's law). During the solar cycle the latitude of emergence also moves towards the equator.
2. Sunspots are observed in bipolar groups with the leading spot (in the direction of apparent motion and closest to the equator) having the same polarity as the hemisphere it appeared in while the following spot has the opposite polarity (Hale's polarity law). The bipolar groups in opposite hemispheres have opposite magnetic orientation and this orientation reverses in each new solar cycle.

The motion of the sunspots reveals another important property of the Sun, namely its differential rotation: the Sun rotates faster at its equator and slower at the poles with a sidereal rotation time of $26.8^\circ/\text{day}$ at the equator and 31.8° at 75° latitude, cf. Table 8.1.

The source of solar activity is a MHD dynamo, as first proposed by Babcock [8.3]. While the details of this process are not completely understood, the underlying principle seems to be valid. The dynamo process works within or at the bottom of the convection zone. During the solar minimum the Sun's magnetic field is poloidal. Differential rotation winds it up to a toroidal field, as already described in the general discussion of the MHD dynamo in Sect. 3.4. The magnetic field is then concentrated in flux tubes with radii of a few hundred kilometers and magnetic field strengths between a few hundred and about 2000 G. The bulk motion in the convection zone twists the field lines, locally increasing the magnetic field strength to up to some thousands of gauss. This high flux density leads to magnetic buoyancy driving up magnetic flux ropes through the photosphere, cf. Fig. 8.19. Where the flux tubes intersect the photosphere, bipolar sunspots emerge with the polarity pattern required by Hale's law. The latitude of the first appearance of sunspots is determined by the interplay of differential rotation and magnetic field strength. With the first sunspots appearing at relatively high latitudes, the magnetic pressure is reduced there and the process continues at lower latitudes, leading to the motion of sunspots towards the equator during the solar cycle. Meridional flows in the convection zone combined with magnetic field diffusion and

dispersion drive the leading spot towards the equator while the following spot stays behind. The leading spots of the opposite hemispheres converge at the equator and dissolve by reconnection. The following spots undergo reconnection with the polar fields. Since the polar field is slightly smaller than the field accumulated in the following spots, the polarity of these spots eventually takes over and the magnetic field is reversed.

This last step, i.e. the pole reversal, is understood least. An alternative explanation for the polarity reversal is the α -effect discussed in Sect. 3.4 that creates a toroidal current which in turn gives rise to a poloidal magnetic field of opposite polarity.

8.6.3 The Heliosphere During the Solar Cycle

Sunspots strongly modify the photospheric magnetic field, subsequently changing the field on the source surface and the tilt angle of the neutral line. These modifications are transported outwards even to the borders of the heliopause and manifest themselves in spatial and temporal variations in solar wind and magnetic field parameters.

The most dramatic variation of the heliospheric structure during the solar cycle is related to the neutral line of the coronal magnetic field and its interplanetary continuation as the heliospheric current sheet. The waviness of the current sheet, as described by the tilt angle, increases towards the solar maximum. Thus the current sheet is rather flat during the solar minimum, as shown in the left panel in Fig. 8.20, while it extends to much higher latitudes during solar maximum. During solar minimum the CIRs are confined to the vicinity of the equatorial plane while during solar maximum conditions stream interactions also can be observed at higher latitudes.

In the solar wind parameters, the solar cycle variations are less pronounced [8.34, 8.76]. In general, any long-term variations apparent in the data are small compared with short-term variations. In addition, each solar cycle seems to be slightly different, and thus a parameter might be related to solar activity quite well in one cycle but not in another. Despite the uncertainties involved, there are some correlations which can be understood easily, for instance: (a) The average solar wind speed is higher during solar minimum than during solar maximum because high-speed solar wind streams are observed more frequently and for longer times during solar minimum. (b) The average solar wind densities are roughly constant except for individual time periods when exclusively slow solar wind streams were observed. (c) The momentum flux is modulated by $\pm 28\%$ with a well-defined minimum at solar maximum, which just is a combination of (a) and (b). For the same reason, the kinetic energy flux is modulated by $\pm 40\%$, again with its minimum around solar maximum. (d) The relative amount of helium has a minimum of 2.8% around solar minimum and a maximum of about 4% around solar maximum. Again, this reflects the fact that during solar minimum conditions the fast solar wind is observed more frequently.

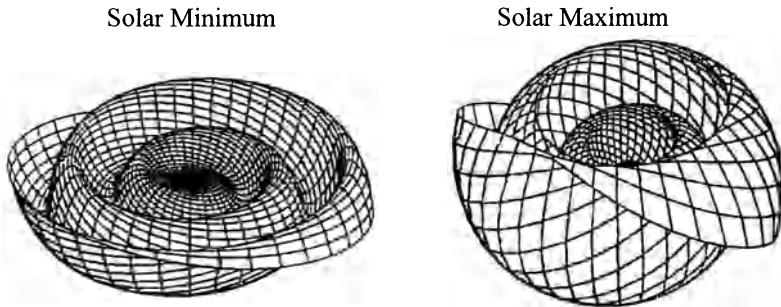


Fig. 8.20. Waviness of the heliospheric current sheet during solar minimum and maximum conditions, based on a sketch by R. Jokipii, University of Arizona

8.7 Flares and Coronal Mass Ejections

Flares and coronal mass ejections (CMEs) are violent manifestations of solar activity. Both are related to the solar magnetic field, sunspots, and filaments. The energy released in these processes had been stored in the field.

The first record of a solar flare dates back to Carrington in 1859, who observed a sudden brightening of a sunspot in white light. After this “explosion”, the sunspot structure had changed and about a day later a violent geomagnetic storm with pronounced auroral activity was observed. Although Carrington himself noted “a swallow does not make a summer”, this was the first direct link between solar activity and its influence on Earth. Coronal mass ejections only have been observed since the early 1970 with the advance of space-borne coronographs. Today, one of the most controversial topics is the relation between flares and CMEs.

8.7.1 Electromagnetic Radiation

A flare is the result of a sudden violent outburst in energy, with energies of up to 10^{25} J being released over a time period of some minutes. White-light flares, like the one observed by Carrington, are rare because even for the largest flares the brightness is less than 1% of the total luminosity of the photosphere. In certain frequency ranges, e.g. at the wings of the black body, the intensity of the electromagnetic radiation can increase by orders of magnitude during a flare. While the flare is defined as an outburst in electromagnetic radiation, often it also is associated with the emission of energetic particles (cf. Sect. 9.2) and huge plasma clouds, the CMEs.

The electromagnetic radiation released in a flare in different frequency ranges shows typical time profiles, cf. Fig. 8.21. These profiles can be used to define the phases of a flare during which distinct physical processes occur.

In a large flare, the electromagnetic emission can be divided into three parts. In the preflare phase, also called the precursor, the flare site weakly

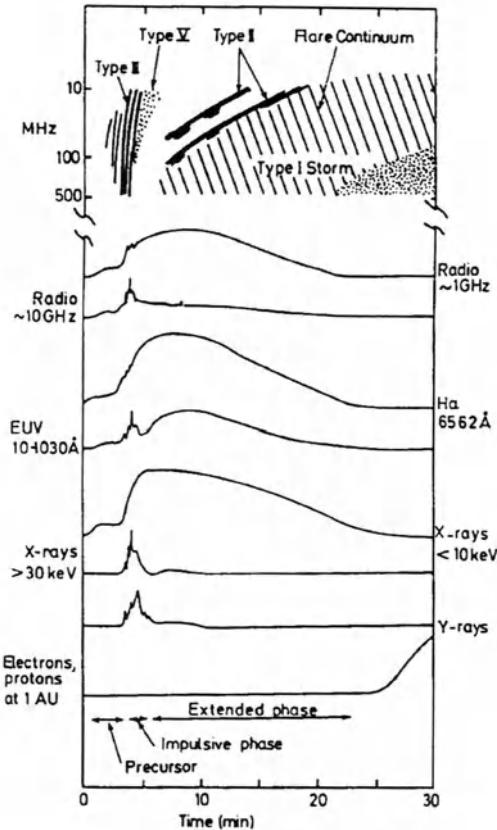


Fig. 8.21. Solar electromagnetic radiation during the different phases of a flare and frequency spectrum of radio bursts. Reprinted from S.R. Kane [8.48], in *Coronal Disturbances* (ed. G. Newkirk), Copyright 1974, with kind permission from Kluwer Academic Publishers

brightens in soft X-rays and $H\alpha$. This phase lasts for some minutes; it is observed in very large flares only, and indicates a heating of the flare site. During the impulsive or flash phase, most of the flare energy is released and the harder parts of the electromagnetic spectrum, such as hard X-rays and γ -rays, are most abundant. This phase lasts for a few minutes and can be followed by a gradual or extended phase during which emission mainly occurs in $H\alpha$ and soft X-rays but microwave and radio emission also can continue. This latter phase can last for some tens of minutes, and occasionally even for a few hours. It is present in the larger events only. Most of the flares are rather small, consisting of an impulsive phase only.

The generation of electromagnetic radiation is a consequence of a rather impulsive energy release, probably reconnection, inside a closed magnetic field loop in the corona, heating the coronal plasma and accelerating particles. The heated plasma emits soft X-rays. Accelerated electrons generate microwaves and hard X-rays at the top of the loop or hard X-rays, γ -rays, UV emission, and part of the $H\alpha$ emission at its footpoints, while accelerated ions generate γ -ray line emission. Streaming electrons also generate radio emis-

sion. Thus the electromagnetic radiation also provides information about the acceleration and propagation of particles accelerated in a flare.

Soft X-Rays and H α . In a solar flare most of the electromagnetic radiation is emitted as soft X-rays with wavelength between 0.1 and 10 nm. Soft X-rays originate as thermal emission in hot plasmas with temperatures of about 10^7 K. Most of the radiation is continuum emission; lines of highly ionized O, Ca, and Fe are also present. In a large flare, soft X-rays are emitted during all three phases; the strong increase at the beginning of the impulsive phase is related to an abrupt increase in the temperature at the flare site up to about 5×10^7 K. The H α emission also is thermal emission, its time profile closely follows the soft X-ray profile.

Hard X-Rays. Hard X-rays are photons with energies between a few tens of kiloelectronvolts and a few hundred kiloelectronvolts, generated as bremsstrahlung of electrons with slightly higher energies. Only a very small amount of the total electron energy (about 1 out of 10^5) is converted into hard X-rays.

Microwaves. Solar microwave emission is generated by the same electron population which also generates the hard X-rays, as can be deduced from the similarities in the time profiles. As in hard X-rays, during the impulsive phase the emission often consists of individual spikes, the ‘elementary flare bursts’. Microwave emission is gyro-synchrotron radiation of electrons with energies between some 10 keV and some 100 keV. If these mildly relativistic particles gyrate in the coronal magnetic field (about 20–100 G), they emit radiation with frequencies between 10 and 100 times their gyro-frequency. The dependence on the magnetic field strength has a remarkable consequence: as the magnetic field decreases with height, microwave emission at a certain frequency, say 17 GHz, is generated by electrons with energies above 200 keV in low lying flares or small loops, while in a loop extending high into the corona, an electron energy of more than 1 MeV is required. While in the impulsive phase the elementary bursts give evidence for individual, isolated energy releases, in the gradual phase the microwave emission most probably is thermal emission.

γ -Rays. γ -ray emission indicates the presence of energetic particles. The spectrum can be divided into three parts: (a) Bremsstrahlung of relativistic electrons and, to a lesser extent, the doppler broadening of closely neighbored γ -ray lines leads to a γ -ray continuum. (b) Nuclear radiation of excited CNO-nuclei leads to a γ -ray line spectrum in the range 4 to 7 MeV. (c) Decaying pions lead to γ -ray continuum emission above 25 MeV.

The most important γ -ray lines are at 2.23 MeV and 4.43 MeV. The 2.23 MeV line is due to neutron capture in the photosphere: ^4He -nuclei decay in p/α or α/p interactions in the corona, emitting neutrons. These reactions require particle energies of at least 30 MeV/nucl. The neutrons can propagate independently of the coronal magnetic field. Elastic collisions decelerate neutrons penetrating into the denser regions of the chromosphere or photosphere. Eventually, the neutron is slowed down to thermal energies and can

be captured by ^1H or ^3He . Neutron capture by ^1H leads to the emission of a γ -quant. The 4.43 MeV line results from the transition of a ^{12}C nucleus from an excited into a lower state, with the excitation being either due to nuclear decay or inelastic collisions with energetic particles.

Radio Emission. Electrons streaming through the coronal plasma excite Langmuir oscillations. Solar radio bursts are metric bursts, their wavelengths are in the meter range. In interplanetary space, radio bursts are kilometric bursts. The bursts are classified depending on their frequency drift. The type I radio burst is a continuous radio emission from the Sun, basically the normal solar radio noise but enhanced during the late phase of the flare. The other four types of bursts can be divided in fast and slow drifting bursts or continua, cf. the upper panel in Fig. 8.21.

The type III radio burst starts early in the impulsive phase and shows a fast drift towards lower frequencies. Since the frequency of a Langmuir oscillation depends on the density of the plasma (cf. (4.44)), the radial speed of the radio source can be determined from this frequency drift using a density model of the corona. The speed of the type III burst is about $c/3$, it is interpreted as a stream of electrons propagating along open field lines into interplanetary space. Impulsive peaks in the hard X-ray emission can be related to individual type III bursts, indicating individual energy releases. Occasionally, the drift of the type III burst is suddenly reversed, indicating electrons captured in a closed magnetic field loop: as the electrons propagate upward, the burst shows the normal frequency drift which is reversed as the electrons propagate downward on the other leg of the loop.

In the metric type II burst, the frequency drift is much slower, indicating a radial propagation speed of its source of about 1000 km/s. It is interpreted as evidence of a shock propagating through the corona. Nonetheless, it is not the shock itself that generates the type II burst but the electrons accelerated at it. As these electrons stream away from the shock, they generate small, type III-like structures, giving the burst the appearance of a herringbone (herringbone burst) in the frequency time diagram with the type II as the backbone and the type III structures as fish-bones. The type II burst is split into two parallel frequency bands, interpreted as forward and reverse shocks.

The metric type IV and V bursts are continuous emission directly following the type II and type III bursts, respectively. The type IV burst is generated by gyro-synchrotron emission of electrons with energies of about 100 keV. It consists of two components: a non-drifting part generated by electrons captured in closed magnetic field loops low in the corona, and a propagating type IV burst generated by electrons moving in the higher corona. The type V burst is a similar burst following the type III burst. But in contrast to the type IV burst it is stationary, showing no frequency drift. Most likely, it is radiation of the plasma itself.

Kilometric radio bursts in interplanetary space are interpreted in the same way: type III bursts show a fast frequency drift, indicating electrons stream-

Table 8.2. Classes of solar flares

	Impulsive	Gradual
Duration of soft X-rays	< 1 h	>1 h
Decay constant of soft X-rays	< 10 min	>10 min
Height in corona	$\leq 10^4$ km	$\sim 5 \cdot 10^4$ km
Volume	$10^{26} - 10^{27}$ cm ³	$10^{28} - 10^{29}$ cm ³
Energy density	high	low
Size in H α	small	large
Duration of hard X-rays	<10 min	>10 min
Duration of microwaves	<5 min	>5 min
Metric type II burst	75%	always
Metric type III burst	always	50%
Metric type IV burst	rare	always
Coronal mass ejection	rare	always

ing along a magnetic field line. If the location of the radio source can be identified, this kilometric type III burst can be used to trace the shape of the interplanetary magnetic field line [8.49]. The kilometric type II burst gives evidence for a shock propagating through interplanetary space.

8.7.2 Classes of Flares

Not all the phases indicated in Fig. 8.21 can be observed in all flares. Instead, flares differ in their electromagnetic radiation, in the acceleration of energetic particles, or the association with a coronal mass ejection.

A useful, although frequently modified, classification scheme for solar flares goes back to Pallavicini et al. [8.65] who used Skylab soft X-ray images of the Sun, combined with the intensity-time profiles. If a flare is observed on the solar limb, the height profile of the electromagnetic emission can be inferred. These limb flares can be divided into three distinct groups: (a) point-like flares, (b) flares in small and compact loop structures, and (c) flares in large systems of more diffuse loops. Flares of classes (a) and (b) are associated with a short duration of the soft X-ray emission, less than one hour, while in flares of class (c) the soft X-ray emission can last for some hours. Therefore, the compact and point-like flares are called impulsive, and the flares in the large diffuse loop are called gradual flares.

The classification scheme, originally introduced for the soft X-rays only, over the years has been extended to other ranges of electromagnetic radiation, as summarized in Table 8.2. These schemes do not always agree. On the basis of the times scales, a flare might appear gradual in soft X-rays but impulsive in hard X-rays or vice versa. These phenomenological criteria provide no sharp separation into two classes but rather a continuous transition from more impulsive to more gradual flares. A better criterion, also pointing to the physical difference, is the occurrence of a coronal mass ejection, leading

to an unambiguous classification of flares into confined (corresponding to impulsive) and eruptive (corresponding to gradual).

We have to be careful not to confuse the classes of flares, i.e. impulsive and gradual, with the phases of flares bearing the same name. An impulsive flare appears to be rather simple in so far as it always has an impulsive phase. However, in some small events, which are observed in interplanetary space as so-called ^3He -rich events, even the impulsive phase is rather small. Some large impulsive events can also show a small gradual phase. In the larger flares (longer duration, larger volumina) the situation is even more complicated. While the gradual phase is well developed, an impulsive phase is not always present, only the largest gradual flares show all three phases. The gradual flares with a gradual phase only in general are small in electromagnetic emission, their main characteristic is the coronal mass ejection. These solar events are often called disappearing filaments because the expelled matter is their main signature while classic flare emission is weak or even absent.

8.7.3 Coronal Mass Ejections

With the aid of a coronograph, the corona can be observed continuously. Basically, a coronograph is a telescope with an occulter screening off the direct photospheric emission. Ground-based coronographs have been in use since the 1930s, observing only selected coronal emission lines. Space-based coronographs, on the other hand, observe the light scattered by the corona. The first space-based coronograph was used on Skylab in 1973/74; the most advanced coronograph is on SOHO [8.14, 8.33]. While the older coronographs had a field of view from a height of about $1.5 r_\odot$ out to 5 or $10 r_\odot$, the combination of different telescopes in the LASCO coronograph on SOHO has a field of view from $1.1 r_\odot$ out to about $30 r_\odot$. In addition, its resolution is much better, thus smaller and fainter mass ejections will be detected.

The most striking feature visible in a record of coronograph images is the coronal mass ejection (CME). A coronal mass ejection is a bright structure propagating outward through the corona, as shown in Fig. 8.22. Large data bases on CMEs exist from the Solwind-Coronograph on P78 and the HAO-coronograph on Solar-Maximum Mission (SMM). The basic features of CMEs can be summarized as follows [8.16, 8.42, 8.44, 8.64]:

Solar Cycle Dependence. During solar maximum, about two CMEs can be observed daily, whereas during solar minimum one CME can be observed in a week [8.93]. This is not too surprising because CMEs in general are observed in connection with flares and filaments which both vary with solar cycle.

Latitude Distribution. Coronal mass ejections are distributed evenly on both hemispheres, the average latitude is 1.5°N . Their distribution is flat within $\pm 30^\circ$ and decreases fast towards higher latitudes. The maximum in

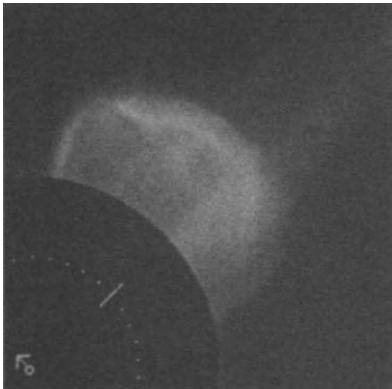


Fig. 8.22. Image of the 4 May 1986 CME, taken by the HAO coronograph. The dark disk in the lower left corner is the occulting disk inside the coronograph, and the dashed circle gives the photosphere. The arrow in the center of the Sun points towards the north. Note that this CME is observed during solar minimum conditions. Reprinted from S.W. Kahler and A.J. Hundhausen [8.47], *J. Geophys. Res.* **97**, Copyright 1992, American Geophysical Union

the $\pm 30^\circ$ region reflects the latitudinal distribution of sunspots and flares. During solar minimum, the CMEs cluster within $\pm 10^\circ$ around the equator.

Width. The projected widths of CMEs show a distribution with an average at 46° and a median at 42° . CMEs smaller than 20° or larger than 60° are rare; however, the largest angular extent is more than 120° . The width of CMEs seems to be independent of the solar cycle.

Speeds. CME speeds range from less than 10 km/s up to greater than 200 km/s with an average of 350 km/s and a median at 285 km/s. The speeds of CMEs do not depend on the solar cycle. Fast and slow CMEs seem to reveal different patterns of energy release: a slow CME can accelerate in the coronograph's field of view, indicating a continuous energy release. A fast CME, on the other hand, does not show evidence of acceleration. Here the energy release appears to be more explosive.

Kinetics. During a CME between 2×10^{14} g and 4×10^{16} g coronal material is ejected; the kinetic energy contained in a CME is between 10^{22} J and 6×10^{24} J. This is at least comparable with the energy released as electromagnetic radiation in a flare.

Structure. As a typical coronal mass ejection normally a loop-like structure is shown, cf. Fig. 8.22. While these are the most impressive and also the most energetic CMEs, other morphologies exist, too, with such picturesque names as spikes, multiple-spikes, clouds, fans, or streamer blow-outs. The basic difference is a smaller extent and a structure distinct from a closed loop. One example will be discussed in connection with Fig. 8.25.

Note that all the geometrical quantities discussed above are apparent quantities only: while the coronal mass ejection is a three-dimensional structure, its image is only a two-dimensional projection into a plane perpendicular to the Sun-Earth axis. Thus sizes and speeds might be underestimated. In particular, if the CME propagates directly towards the observer, its extent and speed cannot be determined. It is even more difficult to detect because it only becomes visible as a halo around the Sun after it has spread far enough.

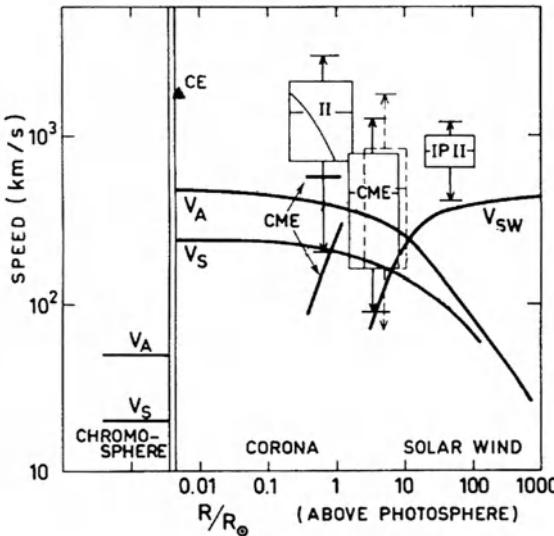


Fig. 8.23. Alfvén and sound speed in the corona and in interplanetary space compared with the solar wind speed v_{sw} and speeds of coronal mass ejections (CMEs), coronal type II bursts, and interplanetary type II bursts. Reprinted from J.-L. Bougeret [8.13], in *Collisionless shocks in the heliosphere: reviews of current research* (eds. B.T. Tsurutani and R.G. Stone), Copyright 1985, American Geophysical Union

8.7.4 Coronal Mass Ejections, Flares, and Coronal Shocks



The speeds of coronal mass ejections are highly variable. In Fig. 8.23 the range of CME speeds is compared with the Alfvén and the sound speed. In the corona, most CMEs are too slow to drive a fast MHD shock wave. Nonetheless, because many of the CMEs are still faster than the sound speed, they might drive slow or intermediate MHD shocks. In at least one CME the curvature of the loop suggests a slow shock [8.45].

In Sect. 8.7.1 we learned about the metric type II burst, interpreted as a shock wave propagating through the corona. The relationship between type II bursts and CMEs is ambiguous. From a statistical study of CMEs and metric type II bursts [8.79], it is evident that about two-thirds of the metric type II bursts are accompanied by a fast CME. However, there are also metric type II bursts without CMEs (one-third) as well as CMEs without metric type II burst (three-fifth). In particular, half of the CMEs without type II bursts are fast CMEs, with speeds above 450 km/s. The situation becomes even more confused if CMEs and type II radio bursts are compared in individual events. In some events, the CME's radial speed is markedly lower than the speed of the radio burst; in other events the source of the metric type II emission is located behind the CME, occasionally overtaking it at later times. Thus a CME is neither a necessary nor a sufficient condition for a coronal shock. Nonetheless, the more energetic CMEs most likely drive a coronal shock.

The relationship between flares and coronal mass ejections also is a hotly debated topic. Traditionally, the CME has been viewed as a phenomenon accompanying large flares. Today, it is suggested that the CME is the primary energy release while the flare is just a secondary process [8.37]. Both models

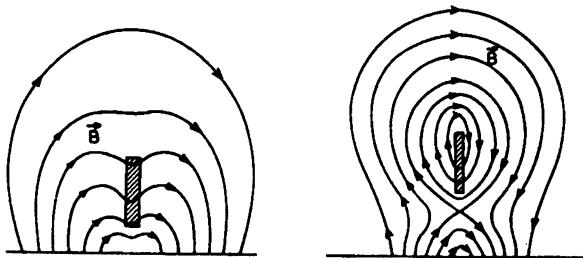


Fig. 8.24. Magnetic field configurations in solar filaments: (*left*) normal configuration, also called K-S configuration, (*right*) inverse configuration, also called R-K configuration. Reprinted from S. Koutchmy et al. [8.51], in *Solar interior and atmosphere* (eds. A.N. Cox, W.C. Livingston, and M.S. Matthews), Copyright 1991, with kind permission of The University of Arizona Press

have their pros and cons, currently we probably should follow the suggestion in [8.32] that the flare neither is the cause nor the consequence of the CME but that both are triggered by a common mechanism, probably an instability. The observations leading to this statement are as follows.

Flares and CMEs can occur together; however, both also can occur separately: in about 90% of the flares no CME is observed, while about 60% of the CMEs go without a flare. The combined flare and CME events are the most energetic events in both groups. In these events the flare, which is small compared with the angular extent of the CME, is not necessarily centered under the CME but more likely is shifted towards one of its legs.

The prime argument for the CME being the cause and the flare the consequence is based on energetics: the energy released in the CME is larger than the one released in the flare. But the mechanism of the energy release is different, too: if a CME is accompanied by a flare, it has a high and constant speed, indicative of an explosive energy release. A CME without flare, on the contrary, often accelerates, indicating that energy is released continuously.

Timing is another crucial factor in this discussion: in about 65% of the combined CME/flare events, the CME leads, while in the other 35% the flare starts before the CME.

8.7.5 Models of Coronal Mass Ejections (CMEs)

Many CMEs originate in filaments, and the magnetic field pattern of the filament can even be recognized if the CME is detected in interplanetary space, see, for example, [8.12].

So far we have learned that magnetic pressure prevents the filament from ‘falling down’ to the photosphere. But how and why can it suddenly be blown out so violently to form a CME? A detailed analysis of the magnetic field pattern of the filament and the photosphere reveals two different configurations. In the normal configuration, which goes back to Kippenhahn and Schlüter

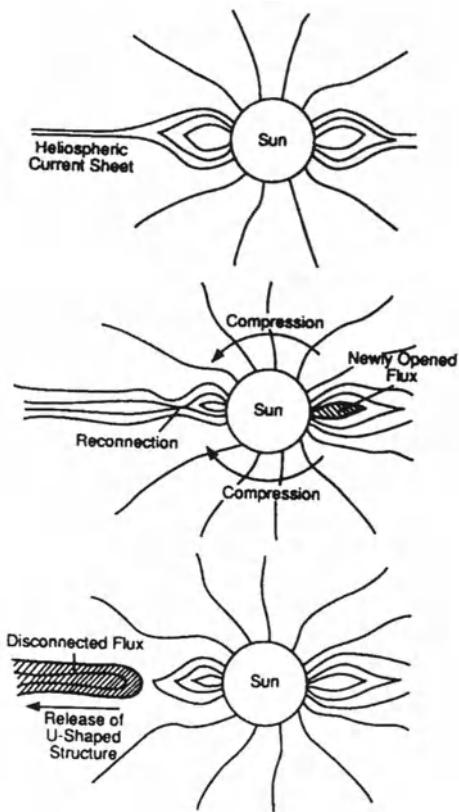


Fig. 8.25. Disconnection event: a CME with an open magnetic field structure is expelled along a helmet streamer. Reprinted from D.J. McCormac [8.59], *Geophys. Res. Lett.* **18**, Copyright 1991, American Geophysical Union

[8.50] and therefore is also called the K-S configuration, the magnetic field inside the filament has the same direction as the photospheric field below it (left panel in Fig. 8.24). We have already used this model in Sect. 3.2.2. The inverse configuration, which is shown in the right panel, is more complex: here the magnetic field inside the filament is directed opposite to the one in the photosphere. This configuration goes back to Raadu and Kuperus [8.71, 8.72] and is sometimes called the R-K configuration. In particular, in the large, high rising filaments, which tend to give rise to CMEs, the R-K configuration seems to be dominant.

The crucial feature is obvious in the R-K configuration: below the filament there is an X-point configuration where magnetic fields of opposite polarity can be found in close proximity. Such a location is favorable for reconnection. The configuration might have been stable for a long time; however, the motion of the magnetic field lines anchoring the filament or a slow rise of the filament due to increased buoyancy can lead to the sudden onset of reconnection at the X-point. Then the magnetic field energy is converted into thermal energy and flow energy, leading to a further rise of the filament. At neighboring X-points,

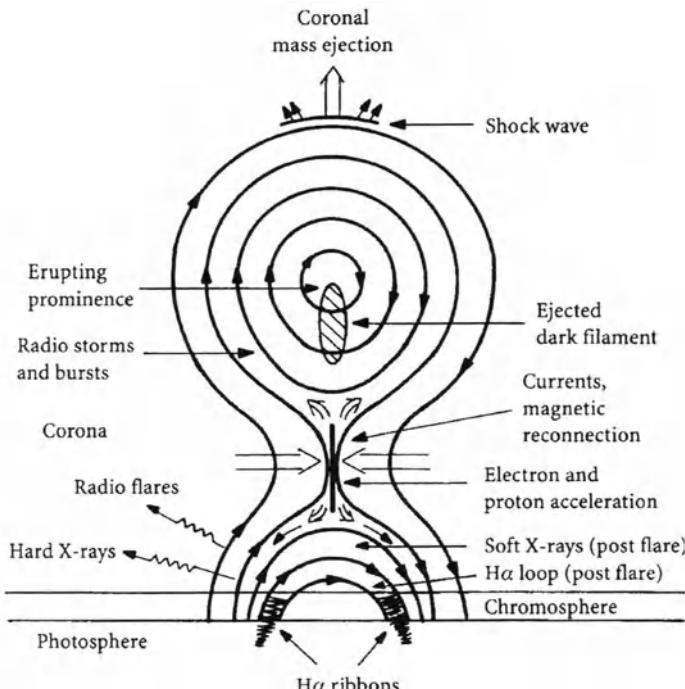


Fig. 8.26. Simplified model of a large eruptive flare. The energy release occurs in the reconnection region below the filament, leading to heating, particle acceleration, and a CME. The heated plasma and the energetic particles give rise to electromagnetic emission in various frequency ranges. From K. Lang [8.53], *Sun, earth and sky*, Copyright 1995, Springer Verlag

reconnection sets in, too, ripping off the filament from its anchoring structure and blowing it out as a CME. A filament of the K-S type can be expelled by the same mechanism, only the forces acting on the anchoring fieldlines or the rise of the filament have to be larger to create a X-point configuration below the filament.

The expulsion of a filament is suitable for explaining the loop-shaped CMEs. Other types of CMEs, however, require different geometries. One example is shown in Fig. 8.25: the required configuration is a helmet streamer. Its outer portion consists of opposing magnetic fields. As somewhere on the Sun new magnetic flux emerges, the coronal magnetic field is deformed and at the current sheet reconnection sets in, blowing out an open magnetic field structure along the streamer. In the coronagraph image, such a CME is seen as a spike- or jet-like structure.

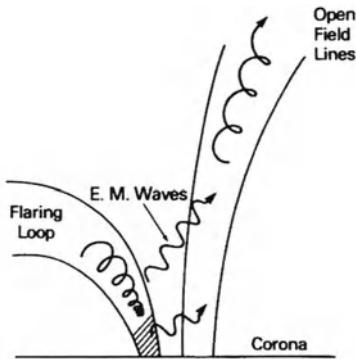


Fig. 8.27. Model of an impulsive ${}^3\text{He}$ -rich flare. Particles accelerated in a closed loop excite waves which in turn accelerate particles on open field lines. These latter can escape into interplanetary space. Reprinted from D.V. Reames [8.74], *Astrophys. J. Suppl.* **73**, Copyright 1990, American Astronomical Society

8.7.6 Models of Flares

We can expand the above model of a loop-like CME to accommodate a flare. By definition, this would be an eruptive or gradual flare. Figure 8.26 shows again a filament in the inverse configuration. As reconnection sets in at the X-point, three different phenomena occur: (a) The plasma is heated, leading to thermal emission in the soft X-ray and visible ranges. (b) Particles are accelerated, either streaming upward along field lines extending into interplanetary space (solar energetic particles, cf. Sect. 9.2) or downward producing the hard electromagnetic radiation. (c) The filament breaks loose and is ejected as a CME. If the CME is fast enough, it drives a shock wave. Particles accelerated at this shock escape into interplanetary space along open field lines. Because the reconnection occurs on all anchoring field lines, the flare occupies a large volume which extends rather high into the corona, thus fulfilling the criteria of a gradual flare as summarized in Table 8.2.

Large confined flares might be explained by the same model, the only difference would be in point (c): although reconnection sets in at the X-point of one or a few anchoring fieldlines, the energy release is not large enough to break all connections between filament and photosphere. Thus the filament is not ejected and the flare is confined in a smaller volume.

In the point-like flares, which also are confined or impulsive flares, the observations of energetic particles in interplanetary space (cf. Sect. 9.2) suggest a different scenario, as sketched in Fig. 8.27. Particles are accelerated inside a closed loop, giving rise to electromagnetic emission. The loop is very stable, preventing particles from escaping into interplanetary space. As the particles bounce back and forth in the closed loop, they excite electromagnetic waves which can propagate in all directions, interacting with the ambient plasma, even accelerating particles. If these 'secondary' particles are accelerated on open field lines, they can escape into interplanetary space. Since the acceleration requires particles and waves to be in resonance, different particles are accelerated by different types of waves. If a particle species is common in the corona, such as H and ${}^4\text{He}$, the waves are absorbed more or less immediately;

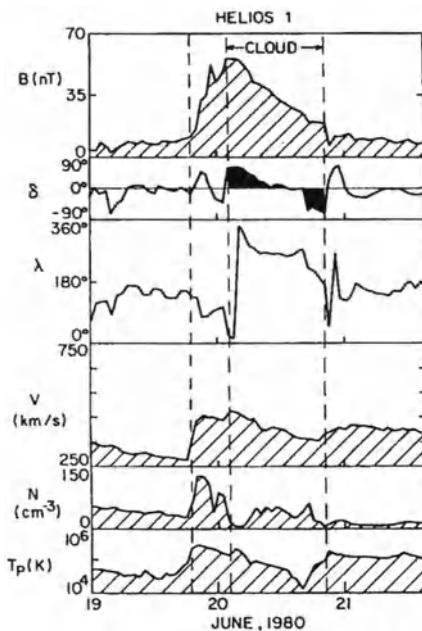


Fig. 8.28. Field and plasma data for an interplanetary shock and the magnetic cloud driving the shock. From top to bottom: magnetic field flux density, elevation, azimuth, solar wind speed, plasma density, and proton plasma temperature. The vertical lines indicate the shock and the boundaries of the magnetic cloud. From L.F. Burlaga [8.17], in *Physics of the inner heliosphere*, vol. II (eds. R. Schwenn and E. Marsch), Copyright 1991, Springer Verlag, Berlin

thus these particles predominately are accelerated inside the closed loop and therefore do not escape into interplanetary space. Other waves, however, can travel larger distances before being absorbed by the minor constituents, such as ${}^3\text{He}$ and the heavy ions, and are thus more likely to accelerate these species on open field lines. Since the escaping particle component is selectively enriched in these minor species, this acceleration process is also called selective heating.

8.7.7 Magnetic Clouds: CMEs in Interplanetary Space

Coronal mass ejections in interplanetary space still carry the magnetic field patterns from their parent filament. These closed magnetic field structures, also called magnetic clouds, have features different from the ambient medium.

Figure 8.28 shows the magnetic field and plasma data for a typical magnetic cloud. Vertical lines indicate the boundaries of the cloud and the shock wave driven by it. The typical signatures of a magnetic cloud can be summarized as follows: (a) a decrease in magnetic field strength inside the cloud, (b) a rotation of the magnetic field vector, in particular its elevation, (c) decreases in plasma density, plasma speed, plasma temperature, and therefore plasma- β , and (d) a bi-directional streaming of suprathermal electrons back and forth along the length of the cloud (not shown in the figure).

The magnetic field configuration of such clouds can be inferred from the variation in magnetic field elevation: at the beginning of the cloud the mag-

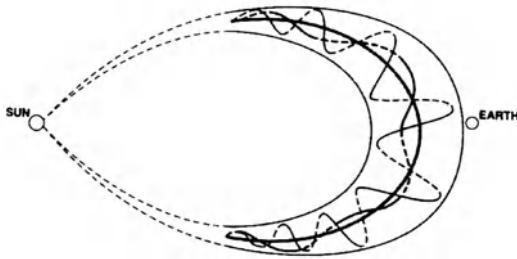


Fig. 8.29. Proposed topology of a magnetic cloud in interplanetary space. Reprinted from L.F. Burlaga [8.17], in *Physics of the inner heliosphere, vol. II* (eds. R. Schwenn and E. Marsch), Copyright 1991, Springer Verlag

netic field is almost perpendicular to the plane of ecliptic. Inside the cloud the elevation decreases until at the end the magnetic field vector is almost opposite to the one at the beginning. This is indicative of a magnetic field wrapped around the ejecta as sketched in Fig. 8.29. In this picture the magnetic cloud is sketched as a bundle of twisted magnetic field lines. The direction of field rotation varies from cloud to cloud, reflecting the field configuration of the parent filament. Just how long the magnetic cloud stays connected to the Sun is still an open question, so therefore the dashed lines in Fig. 8.29 indicate the possibility but not the necessity of such a continued connection.

Magnetic clouds are the main cause of geomagnetic disturbances, their geomagnetic effectiveness depends on whether the field at the leading edge has a strong northward or southward component.

8.7.8 Interplanetary Shocks

In the event in Fig. 8.28 a shock has been observed in front of the magnetic cloud. But only about one-third of the CMEs in space drive an interplanetary shock [8.36], while apparently all traveling interplanetary shocks are driven by CMEs, although the magnetic cloud is not necessarily detected if the observer is located at the flank of the shock.

Interplanetary shocks are identified by characteristic changes in the plasma and field parameters, in particular a sudden increase in plasma density, speed, and temperature, and a jump in the magnetic field strength, cf. also Fig. 7.5. As the driving CMEs, most shocks are observed around the solar maximum.

The properties of interplanetary shocks are highly variable. Between 0.3 AU and 1 AU, the basic characteristics are as follows:

- The compression ratio varies between 1 and 8 with an average close to 2.
- The magnetic compression (ratio between the upstream and downstream magnetic field strengths) varies between 1 and 7 with an average at 1.9.
- Shock speeds in the laboratory frame vary between 300 km/s and 700 km/s with an average of about 600 km/s. Occasionally, shock speeds above 2000 km/s can be observed. Obviously, shocks with speeds only slightly above 300 km/s can be observed in very slow solar wind streams only.

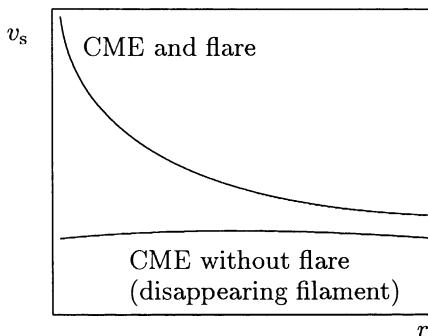


Fig. 8.30. Radial variation of the shock speed from the Sun to 1 AU. If the CME had been very fast close to the Sun, then the shock and CME slow down during propagation. This is generally the case when the CME is accompanied by a flare. If the CME is slow close to the Sun, as in case of a disappearing filament, the propagation speed is roughly constant

Since the shock speed is lower towards the flanks, the tongue-like shape of the shock front closely resembles the shape of the leading edge of the magnetic cloud, as shown in Fig. 8.29.

- The angular extent of the shock varies between a few tens of degrees and up to 180° ; the shock is always wider than the driving CME.
- The Alfvén Mach number is between 1 and 13 with an average at 1.7.

The shock parameters, of course, are related to the properties of the CME, such as speed, angular extent, and total energy released.

An interplanetary shock is a disturbance propagating into the expanding solar wind. The shock should develop absolutely because it expands, too, and also relative to the ambient medium as the latter ones expands differently. In particular, the expansion of the shock leads to a decrease in the plasma and magnetic flux densities. Thus the energy density also decreases. But the latter decreases not only because of the shock's expansion: turbulence created in the wake of the shock and particles accelerated at the shock front (cf. Sect. 9.3) also reduce the shock's energy.

Changes in shock parameters can be quite different from one shock to another. As an example, in Fig. 8.30 the radial variation of the shock speed is shown. Two extreme cases can be distinguished. If a shock is very fast close to the Sun (with CME speeds above 1000 km/s), it is likely to decelerate in interplanetary space. On the other hand, shocks that are rather slow on the Sun do not decelerate but propagate at roughly constant speed. Possible interpretations can be found in the energy release mechanism: it is more explosive in fast shocks and CMEs, which in general are also accompanied by a flare, compared with rather continuous in the slower ones. In addition, the faster shocks in general tend to be more efficient particle accelerators, and thus part of the shock's kinetic energy is converted into kinetic energy of particles. In some sense this relates to the different speed characteristics of the CMEs and to the conversion of shock kinetic energy into particle energy.

8.8 Summary

The heliosphere is structured by the solar wind and the frozen-in magnetic field, which is wound up into Archimedean spirals due to the Sun's rotation. The fast solar wind originates in coronal holes at the poles while the slow wind originates from the streamer belt close to the solar equator where sunspots, filaments, and active regions are located. Where fast and slow solar wind streams meet, corotating interaction regions form. Fluctuations on different scales are superimposed onto the average field. They are related to waves originating in the corona, interactions between different solar wind streams, and transient disturbances such as magnetic clouds and traveling interplanetary shocks. The latter stem from coronal mass ejections, i.e. violent expulsions of huge plasma clouds.

Exercises and Problems

- 8.1** Consider a 10 MeV proton in interplanetary space. Determine its gyro-radius, gyration period, and the wave numbers of Alfvén waves in resonance with the proton (assume different pitch angles: 10°, 30°, and 90°). Compare with the same values for a 1 MeV electron.
- 8.2** For an observer on Earth, calculate the length of the magnetic field line to the Sun and its longitude of origin (connection longitude). Do the same for an observer at 5 AU. (Assume a plane geometry with the field line confined to the plane of ecliptic.)
- 8.3** Imagine a slow solar wind speed starting on the Sun. 30° east of this stream, a fast stream with twice the speed of the slow stream originates. Where would they meet? (Simple assumption of an Archimedian magnetic field spiral.)
- 8.4** An electron beam with $c/3$ propagates through the interplanetary plasma. Assume a decrease in plasma density $\sim 1/r^2$. Calculate the frequency drift under the assumption that the electron beam propagates radially. How do these results change if the curvature of the Archimedian field line, along which the electrons propagate, is considered? How would the curvature of the field line influence the frequency drift of a type II burst in front of an interplanetary shock?
- 8.5** A magnetic loop on the Sun has a parabolic shape with $B = B_o(1 + s^2/H^2)$ with $H = 30\,000$ m being the height of the loop and s the distance from the top of the loop. Calculate the bounce period of particles with a speed of $2c/3$. As the particles interact with the atmosphere at the mirror points, they create hard X-rays. What is the time interval between two subsequent elementary bursts?

9. Energetic Particles in the Heliosphere

The space between Heaven and Earth – is it not like a bellow?

It is empty and yet not depleted;
Move it and more always comes out.

Lao-Tzu, *Tao Te Ching*

Particles in interplanetary space come from sources as diverse as the Sun, the planets, and the vastness of the galaxy. The properties of the different particle populations provide information about the acceleration mechanism(s) and the propagation between source and observer. Thus energetic particles can also be used as probes for the properties of the interplanetary medium. To describe acceleration and propagation, we use concepts such as reconnection, acceleration at shock waves, and wave-particle interactions.

This chapter starts with an overview of the different particle populations and subsequently discusses some of them in detail, in particular solar energetic particles and their propagation, particles accelerated at traveling interplanetary shocks and planetary bow shocks, and galactic cosmic radiation and its modulation.

9.1 Particle Populations in the Heliosphere

Energetic particles in interplanetary space are observed with energies ranging from the supra-thermal up to 10^{20} eV. The main constituents are protons, α -particles, and electrons; heavier particles up to iron can be found in substantially smaller numbers. The particle populations originate in different sources, all having their typical energy spectrum, temporal development, and spatial extent. Figure 9.1 summarizes the different particle populations, and Table 9.1 summarizes their properties.

Galactic Cosmic Rays (GCR) are the high-energy background population with energies extending up to 10^{20} eV. They are incident upon the heliosphere uniformly and isotropically. In the inner heliosphere, the galactic cosmic radiation is modulated by solar activity: the intensity of GCRs is highest during

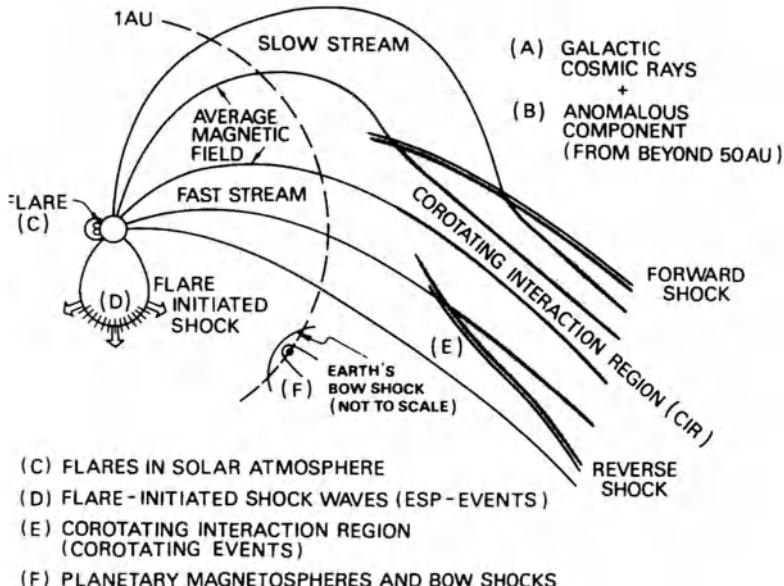


Fig. 9.1. Populations of energetic charged particles in the inner heliosphere. Reprinted from H. Kunow et al. [9.28], in *Physics of the inner heliosphere, vol. II* (eds. R. Schwenn and E. Marsch), Copyright 1991, Springer Verlag

solar minimum and reduced under solar maximum conditions. For reviews see, for example, [9.15, 9.36, 9.45, 9.63].

(B) The **Anomalous Cosmic Rays (ACR)**, also called the anomalous component, energetically connects to the lower end of the GCRs but differs from them with respect to composition, charge states, spectrum, and variation with the solar cycle [9.13, 9.22]. As neutral particles of the interstellar medium travel through interplanetary space, they become ionized at a distance of a few solar radii from the Sun. These charged particles then are convected outward with the solar wind and are accelerated at the termination shock. The accelerated particles then propagate towards the inner heliosphere where they are detected as anomalous component.

(C) **Solar Energetic Particles (SEPs)** are accelerated in solar flares, the injection of these particles into the heliosphere therefore is point-like in space and time. SEP energies extend up to some tens or a few hundred megaelectronvolts, occasionally even into the gigaelectronvolt range. The latter can be observed with neutron monitors on the ground, and the event is then called a ground-level event (GLE). Owing to interplanetary scattering, particle events in interplanetary space last between some hours and a few days, depending on the scattering conditions and the observer's distance from the Sun. SEPs not only provide information about the acceleration processes on the Sun but

Table 9.1. Characteristics of particle populations in interplanetary space. The letter in the first column is the same as in Fig. 9.1, the subsequent columns give the temporal and spatial scales as well as the typical energy range

	Temporal scales	Spatial scales	Energy range	Acceleration mechanisms
A	continuous	global	GeV-> TeV	diffusive shock
B	continuous	global	10–100 MeV	shock?
C	δ	δ	keV–100 MeV	reconnection, stochastic, selective heating, shock
D	days	extended	keV–10 MeV	diffusive shock, shock-drift, stochastic
E	27 days	large-scale	keV–10 MeV	diffusive shock
F	continuous	local	keV–MeV	diffusive shock, shock drift

also can be used as probes for the magnetic structures of the interplanetary medium. Solar energetic particle events exhibit different properties, depending on whether the parent flare is gradual or impulsive [9.5, 9.25]. In gradual flares, the solar energetic particles mix with particles accelerated at the accompanying interplanetary shock. Although SEPs originate in flares, only a small fraction of all flares leads to enhancements in energetic particles above the background level.

(D) **Energetic Storm Particles (ESPs)** are particles accelerated at interplanetary shocks. Originally, ESPs were thought to be particle enhancements related to the passage of an interplanetary shock. The name was chosen to reflect their association with the magnetic storm observed as the shock hits the Earth's magnetosphere. Today, we understand the acceleration of particles at the shock, their escape, and the subsequent propagation through interplanetary space as a continuous process, lasting for days to weeks until the shock finally stops accelerating particles. The properties of particles accelerated at interplanetary shocks are summarized for the tens and hundreds of kiloelectronvolt range in [9.49, 9.54, 9.55, 9.60, 9.64] and for the megaelectronvolt range in [9.4, 9.6, 9.23]. At very strong shocks, protons can be accelerated up to about 100 MeV [9.47].

(E) **Corotating interaction regions** also lead to intensity increases. Protons with energies up to about 10 MeV are accelerated at the shocks [9.1, 9.14, 9.31, 9.39, 9.59]. The energetic particles can even be observed remote from the corotating shocks at distances where the shocks have not yet been formed [9.28, 9.65] or at higher solar latitudes when a spacecraft is above the streamer belt where the CIRs form [9.12, 9.29, 9.58].

(F) Particles accelerated at **planetary bow shocks** are a local particle component with energies extending up to some 10 keV [9.10]. An exception is the jovian magnetosphere where electrons are accelerated up to about 10 MeV. With a suitable magnetic connection between Earth and Jupiter, these jovian electrons can be observed even at Earth's orbit [9.8, 9.9].

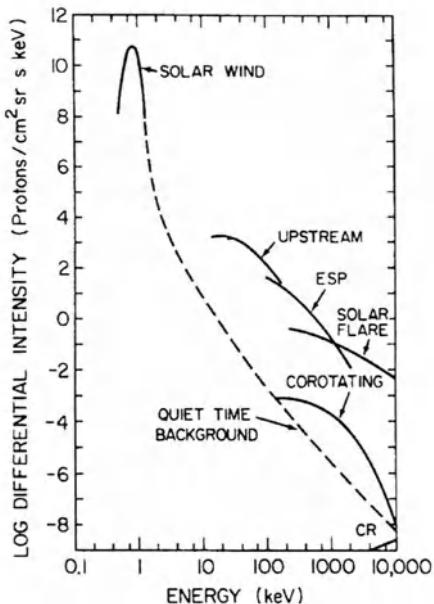


Fig. 9.2. Energy spectra of different ion populations in the heliosphere above a postulated quiescent background. Reprinted from *Adv. Space Research* 4, G. Gloeckler [9.17], Copyright 1984, with kind permission from Elsevier Science Ltd

Figure 9.2 gives the energy ranges and relative intensities for different ion populations in the heliosphere. The figure is limited to energies below 10 MeV; above that energy the higher energetic end of the SEPs can be found as well as the galactic and anomalous cosmic radiation. The spectrum of the GCR is shown in Fig. 9.11.

In the study of energetic particles, we are interested in their source, in particular the acceleration mechanisms, and their propagation. Occasionally, the acceleration mechanism can be studied *in situ*, as in the case of traveling and corotating shocks or planetary bow shocks. For most populations, however, the observer is remote from the source: there is no access to the acceleration regions of solar energetic particles or galactic cosmic rays. And even particles accelerated at shocks in interplanetary space can be observed remotely from the shock. In all these cases interplanetary propagation has modified the particle populations. Therefore the properties of a particle population observed in interplanetary space in general reflect the combined effects of acceleration and propagation.

9.2 Solar Energetic Particles

Particle increases at the Earth's orbit as a consequence of solar flares, first were detected in 1942 in neutron monitor records [9.16]. From the long time scales of these events compared with flare duration, in the 1950s diffusive propagation was proposed [9.37]. Because of the small spatial and temporal

Table 9.2. Properties of energetic particle events following impulsive and gradual flares. The numbers given here refer to particles with energies of a few MeV/nucl

	³ He-rich	gradual
Particles	electron-rich	proton-rich
³ He/ ⁴ He	~ 1 (enrichment 2000 times)	~ 0.0005
Fe/O	~ 1.234 (enrichment 8 times)	~ 0.155
H/He	10	100
Q _{Fe}	~ +20	~ +14
Duration	hours	days
Longitudinal cone	< 30°	≤ 180°
Metric radio bursts	III, V	II, III, IV, V
Coronograph	–	CME
Solar wind	–	ipl. shock
Event rate/a	~ 1000	~ 10

extent of the particle source, SEPs not only provide information about particle acceleration but also can be used to study interplanetary propagation.

9.2.1 Classes of Particle Events

Solar energetic particles are mainly protons, electrons, and α -particles with small admixtures of ³He-nuclei and heavier ions up to Fe. Protons and ions can be accelerated up to some tens or hundreds of MeV/nucl, occasionally even energies in the gigaelectronvolt range can be acquired. The electron acceleration is limited to energies of some megaelectronvolts. The particle increase above background can vary between hardly detectable and some orders of magnitude. While the flare lasts only for a few minutes, maybe even an hour, a solar energetic particle event at the Earth's orbit typically lasts for a day or two, depending on the number of particles injected into space and scattering conditions.

Solar energetic particle events observed in interplanetary space exhibit different features, depending on whether the parent flare was impulsive or gradual [9.5, 9.25], some of them are summarized in Table 9.2.

What do the differences in particle populations tell us? First of all, we can infer that different acceleration mechanisms must be at work. The high charge states of Fe (20) and Si (14) in impulsive events are indicative of acceleration out of a very hot environment with temperatures of about ten million kelvins. The coronal temperature of about two million kelvins would imply lower charge states of about 14 for Fe and 10 for Si, such as observed in gradual events. Thus in the impulsive flares, the plasma must be heated while in gradual flares particles can be accelerated out of the corona without heating the plasma. Particle events following impulsive flares are also enriched in some of the heavier ions, in particular an increase in ³He/⁴He, Fe/C, and Fe/O can be observed. Particle events following gradual flares, on the other hand, show coronal abundances.

These differences can be understood by the two flare models presented in Sect. 8.7.6. Impulsive flares then can be understood in terms of selective heating, while in gradual flares particles are accelerated out of the cool coronal material by the shock in front of the CME. Since the shock accelerates ions, in particular H, more efficiently than electrons, impulsive flares are electron-rich compared with gradual flares and also show a lower H/He ratio.

The shock in front of the CME also has another consequence, namely the different longitudinal cones in which energetic particles can be observed. In the model of selective heating, particles are accelerated on a few open field lines adjacent to the closed loop in which the flare occurred. Thus out of the field lines originating on the Sun, only a small bundle is filled with particles. But these particles do not spread in all directions. Since SEPs are charged particles, they propagate along the magnetic field lines but not perpendicular to it. Thus the particles stay on their field lines and an observer in interplanetary space only can detect particles if he is on one of these field lines, i.e. if he is magnetically connected to the source region. Owing to the curvature of the interplanetary magnetic field line, under normal solar wind conditions an observer on Earth is magnetically connected to a position at 60°W. Thus in impulsive events, only particle events originating in the western hemisphere of the Sun are detected on a spacecraft in an orbit around Earth. In a gradual flare, the situation is different. The shock in front of the CME has a large angular extent. Thus the particles are not accelerated locally but over a wide region. Therefore particles are injected into a much larger cone of field lines open to interplanetary space and the particle events therefore can be detected at larger angular distances from the site of the energy release. Thus close to the Earth, particle events from gradual flares can also be detected if they originated in the eastern hemisphere of the Sun.

So far for a gradual flare we have only considered the particles accelerated at the shock. But in the model of a large flare in Fig. 8.26, the CME and the coronal shock are only one aspect of the event. The processes occurring below the filament, on the other hand, still closely resemble those in a flare without CME. The characteristics of the particles interacting on the Sun can be inferred from the γ -ray line spectrum, and their temporal evolution also from the hard X-ray emission. The observations indicate that the primary energy release must be quite similar in impulsive and gradual events. In particular, although the electromagnetic emission in gradual flares lasts for a longer time, the elementary energy release in both classes of flares occurs in bursts that are similar. In addition, the compositions of the ions interacting on the Sun are similar and both are indicative of some process similar to selective heating. If the observer is magnetically connected close to the flare site in a gradual event, these flare-accelerated particles can also be detected in interplanetary space. The upper panel of Fig. 9.3 shows the temporal evolution of the intensities of protons, O, and Fe in a gradual flare. In the lower panel the Fe/O ratio is shown for different energies. Early in the event, the

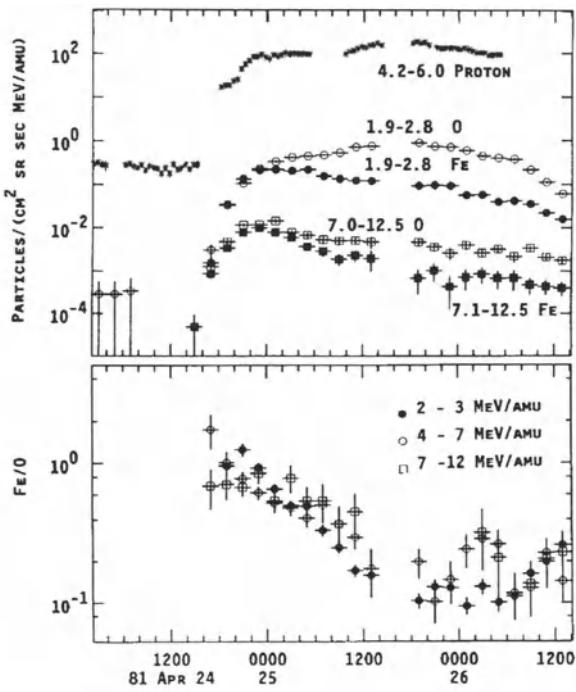


Fig. 9.3. Particle intensities and the development of particle composition in a gradual flare. Reprinted from D.V. Reames et al. [9.48], *Astrophys. J.* **357**, Copyright 1990, American Astronomical Society

Fe/O ratio is high, as expected for an impulsive event. With time, the Fe/O ratio decreases to a value typical of gradual events or the solar wind plasma. Thus early in the event, flare-accelerated particles are present while with increasing time the particles accelerated at the shock become more and more abundant. Note also the differences in the time profiles: the high-energetic Fe channel rises rather fast, and the intensity decreases after its maximum, i.e. a few hours after the flare. This closely resembles a diffusive profile as expected for a short injection on the Sun. In H and O, on the other hand, the intensity stays roughly constant or even increases towards the shock because particles are continuously accelerated and injected from the shock as it propagates towards the observer.

We should be aware that in most cases a classification scheme is used to order a complex data set rather than to reflect physical processes. Nonetheless, the two classes described in Table 9.2 are distinct kinds of events. But whether there really are only these two classes, or whether the two classes represent extreme cases of a more continuous change from impulsive to gradual, is still under debate. In particular, there are large impulsive flares which are not accompanied by a CME but show characteristics which are in between the ones given in Table 9.2, while on the other hand there are disappearing filament events where the particles exhibit the features typical of events following gradual flares, although no flare is observed. In addition, even if a



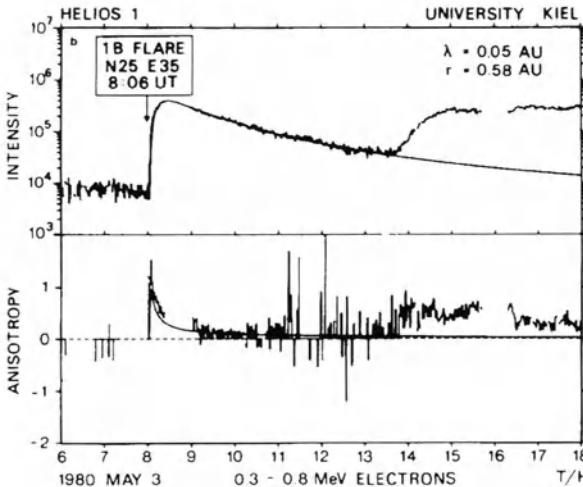


Fig. 9.4. Typical diffusive profile of an electron event in interplanetary space. The solid line gives a fit with the focused transport model; the second increase starting at about 14 UT is due to a second flare. Reprinted from M.-B. Kallenrode et al. [9.26], *Astrophys. J.* **391**, Copyright 1992, American Astronomical Society

flare is accompanied by a coronal mass ejection, the latter does not necessarily drive a shock. This latter point might become more interesting in the future as the first observations with the more sensitive SOHO coronograph suggest that small and slow CMEs are much more common than assumed so far. The particle events following such flares would be some kind of hybrid event, combining the properties of events following impulsive flares with properties of events following gradual ones.

9.2.2 Interplanetary Propagation

The time-scales of particle events in interplanetary space are much longer than the time-scales of the parent flare, which is indicative of diffusive particle propagation. However, diffusion due to scattering at magnetic field irregularities is one aspect of particle propagation only. As we saw in Sect. 8.3, the interplanetary magnetic field is divergent. Since the magnetic moment of a charged particle is constant, the particle's pitch angle decreases as it propagates outward: a particle starting with a pitch angle close to 90° on the Sun will have a pitch angle of only 0.7° at the Earth's orbit. This effect is called focusing.

In the inner heliosphere, i.e. within about 1 AU, particle propagation is influenced mainly by pitch-angle scattering and focusing. The resulting transport equation then is the focused transport equation [9.50]:

$$\frac{\partial f}{\partial t} + mv \frac{\partial f}{\partial s} + \frac{1 - \mu^2}{2\zeta} \cdot v \frac{\partial f}{\partial \mu} - \frac{\partial}{\partial \mu} \left(\kappa(\mu) \frac{\partial f}{\partial \mu} \right) = Q(r, v, t). \quad (9.1)$$

Here s is the length along the magnetic field spiral as described by (8.10), and $\zeta = -B(s)/(\partial B/\partial s)$ is the focusing length. The terms from left to right describe the field parallel propagation, focusing in the diverging magnetic

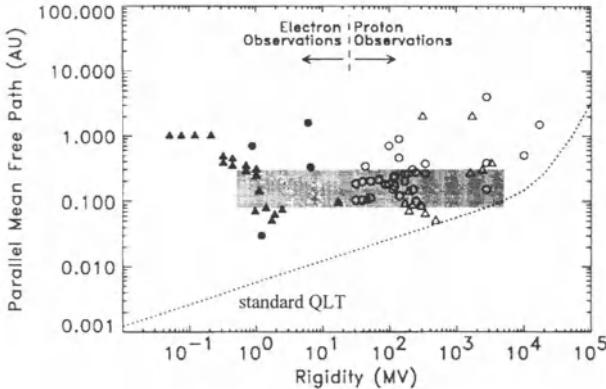


Fig. 9.5. Particle mean free paths in different solar energetic particle events plotted versus the particle's magnetic rigidity. Reprinted from J.W. Bieber et al. [9.2], *Astrophys. J.* **420**, Copyright 1994, American Astronomical Society

field, and pitch-angle scattering. The term on the right-hand side describes the particle source. We have already encountered part of this equation in the discussion of pitch-angle scattering in Sect. 6.1.2. Since convection is ignored, (9.1) should only be applied to particles markedly faster than the solar wind. For low energies the diffusion–convection equation (6.22) should be used instead.

From solutions of the transport equation we can determine the phase space density as a function of time, location, and pitch angle. Thus for a fixed location not only can we determine the intensity–time profile but also the temporal evolution of the pitch-angle distribution. The latter can be described as an anisotropy: if the anisotropy vanishes, particles are coming isotropically from all directions, while for a large anisotropy particles predominately come from one direction.

As an example, Fig. 9.4 shows the intensity–time and anisotropy–time profiles of 0.5 MeV electrons. In the upper panel, the intensity time profile shows the fast rise and slow decay typical of a diffusive profile (cf. Fig. 6.5). The anisotropy (lower panel) is high early in the event because the first particles always come from the solar direction. It decays as particles are scattered back towards the Sun. The solid lines give a fit of (9.1) on the intensity and anisotropy profile with a mean free path λ of 0.05 AU.

Particle mean free paths are different from event to event and also depend on particle rigidity. Figure 9.5 summarizes the results from fits on a large sample of particle events with the mean free path plotted versus magnetic rigidity. Filled symbols refer to electron observations, open ones to protons. The shaded area gives the Palmer consensus range [9.42], where most of the events cluster. Here λ is between 0.08 AU and 0.3 AU and, at least in this statistical study, is independent of rigidity. In individual events, however, a small increase in λ with increasing rigidity can be observed.

The dotted line in Fig. 9.5 shows the expected rigidity dependence determined in an entirely different approach. Since the elementary mechanism is pitch-angle scattering by resonant wave–particle interactions, the magnetic

field turbulence determines the amount of interplanetary scattering. Thus an analysis of the magnetic field fluctuations also gives a particle mean free path, cf. Sect. 6.1.2. The fluctuations are assumed to be small ($\delta B \ll B$), and so quasi-linear theory can be applied. In addition, the wave vectors \mathbf{k} of the fluctuating field are assumed to be parallel to the average magnetic field \mathbf{B}_o (slab model). This approach is also called standard quasi-linear theory or standard QLT. As can be seen from Fig. 9.5, the mean free path determined with standard QLT is always below the one obtained by fitting. In addition, standard QLT predicts an increase of λ with rigidity while the observations suggest λ to be independent of rigidity. This discrepancy has been known since the early 1970s. It is still one of the fundamental problems in space physics, although some promising suggestions for its solution have been made.

The main reason for this discrepancy problem appears to be the interpretation of the magnetic field fluctuations. If their scattering power is overrated, the particle mean free path obtained by standard QLT would be too small. Misinterpretation might be easy: observations with one spacecraft give a scalar power density spectrum of magnetic field fluctuations as these are carried across the observer by the solar wind. However, different three-dimensional fluctuations can lead to the same scalar spectrum. In standard QLT it is assumed that all fluctuations arise from waves with $\mathbf{k} \parallel \mathbf{B}_o$. Thus all power contained in the fluctuations also scatters the particles. To reduce the discrepancy, we can interpret the power-density spectrum so as to remove power from the waves parallel to the magnetic field. For instance, if we assume the fluctuations to be Alfvén waves propagating radially away from the Sun, the field-parallel power available for particle scattering would be reduced. This effect would be particularly strong at larger distances where the angle between the interplanetary magnetic field line and the radial direction is large. Since some observations suggest an increase of the discrepancy with increasing distance, this ansatz is attractive, although it is problematic because Alfvén waves with large angles with respect to the field are easily damped.

A different approach concerns the magnetic field fluctuations not as waves but as two-dimensional dynamical turbulence [9.2]. Observations of the field fluctuations suggest that the correlation function is strongest in the directions perpendicular and parallel to the field [9.35]. Thus the fluctuations seem to consist of two components: slab-like Alfvén waves parallel to the magnetic field, and a two-dimensional turbulent component perpendicular to the field. The relative amount of the slab-component is between 12% and 20%, and the two-dimensional turbulence contributes between 80% and 88% to the power-density spectrum. Since only the slab-component scatters the particles, the mean free paths determined with this modified QLT are much closer to the observed ones [9.2].



Suppl.
→9.3

Nonetheless, even with this ansatz it might be too early to declare the discrepancy problem solved and interplanetary transport understood. In our description of particle propagation we always assume that scattering occurs by fluctuations uniformly distributed in space and time. There are no isolated scattering centers. On the other hand, we know that the level of turbulence in interplanetary space can be quite different. We have already seen from Fig. 8.13 that quiet and turbulent periods exist. On shorter times-scales of about an hour the fluctuation level can also be highly variable. Again, with one spacecraft it is difficult to interpret this: is the level of fluctuations uniform along a magnetic field line but different on neighboring ones and do the changes in fluctuation level result from the motion of the observer relative to the field lines, or do the magnetic field fluctuations show a completely irregular pattern with variations along the field line as well as between neighboring field lines? The latter case would require an entirely new approach to our understanding of propagation in interplanetary space, in particular because we would have to consider a stochastic distribution of scattering centers instead of the continuous wave fields assumed today.

9.3 Particles and Traveling Interplanetary Shocks

In Fig. 9.3 we saw one example of a particle event consisting of both a flare-accelerated component and a particle component accelerated at the shock in front of the CME. Let us now look at the particle distribution right at the shock.

Figure 9.6 shows the energy spectrum between 200 eV and 1.6 MeV for ions in the spacecraft rest frame just downstream of an interplanetary shock. The pronounced peak around 1 keV is the solar wind plasma; the power-law spectrum to higher energies gives the energetic storm particles. The rather smooth transition between these two spectra is in agreement with the assumption that the particles are accelerated out of the solar wind plasma. While the solar wind spectrum is described by a Maxwellian, the combined spectrum can be approximated by a kappa-distribution (5.14). Note the break in the power-law spectrum at energies of a few hundred kiloelectronvolts: here the spectrum is steeper, indicating a less efficient acceleration at higher energies.

It appears reasonable to discuss energetic particles accelerated at interplanetary shocks in two separate energy bands: (a) a low-energy component with ion energies up to a few hundred keV/nucl and electron energies up to some tens of kiloelectronvolts, and (b) a high-energy component with ion energies in the megaelectronvolts and tens of MeV/nucl range and electron energies in the hundreds of kiloelectronvolt to megaelectronvolt range. One reason for such a separation is the break in the ion spectrum, as indicated in Fig. 9.6. Another argument stems from the different particle speeds relative to the shock: while protons in the tens of kiloelectronvolt range on average are only slightly faster than the shock and therefore stick close to the shock

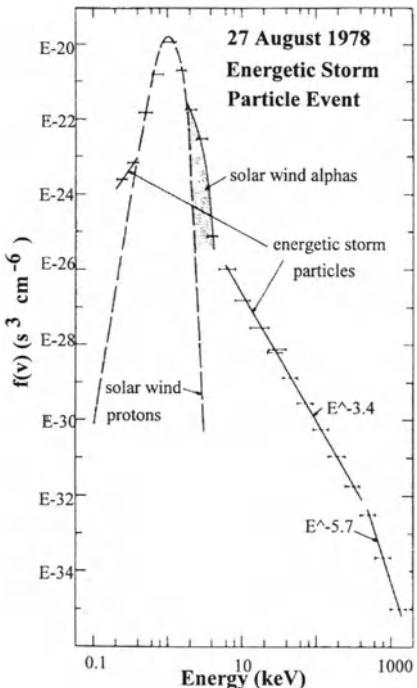


Fig. 9.6. Spectrum of energetic ions from 200 eV to 1.6 MeV at an interplanetary shock in the spacecraft's reference frame. Reprinted from J. Gosling et al. [9.18], *J. Geophys. Res.* **86**, Copyright 1981, American Geophysical Union

front, protons in the megaelectronvolt range are much faster and can escape easily from the shock. Thus an observer in interplanetary space at lower energies detects the shock-accelerated particles mainly around the time of shock passage, while at higher energies particles accelerated at the shock can be detected when the shock is still remote from the observer.

9.3.1 Low Energies (Tens of Kiloelectronvolts)

At low energies, three types of particle events can be distinguished, each related to another acceleration mechanism, cf. Fig. 9.7. Nonetheless, as we saw in Sect. 7.4, combinations of acceleration mechanisms are possible and so too are combinations of different types of events.

In the left panel of Fig. 9.7 an energetic storm particle event is sketched. This is the classical type of particle event at interplanetary shocks. In the particle data, it starts some hours prior to shock arrival at the decaying flank of the preceding solar energetic particle event. The intensity maximum is close to the time of shock passage. Often there is no dip between the first maximum and the maximum at the time of shock passage but a more plateau-like profile or continuous increase as in the protons in Fig. 9.3. In many cases, a few hours after the arrival of the shock, a sudden decrease in intensity is observed as the spacecraft enters the ejecta, the former coronal mass ejection, driv-

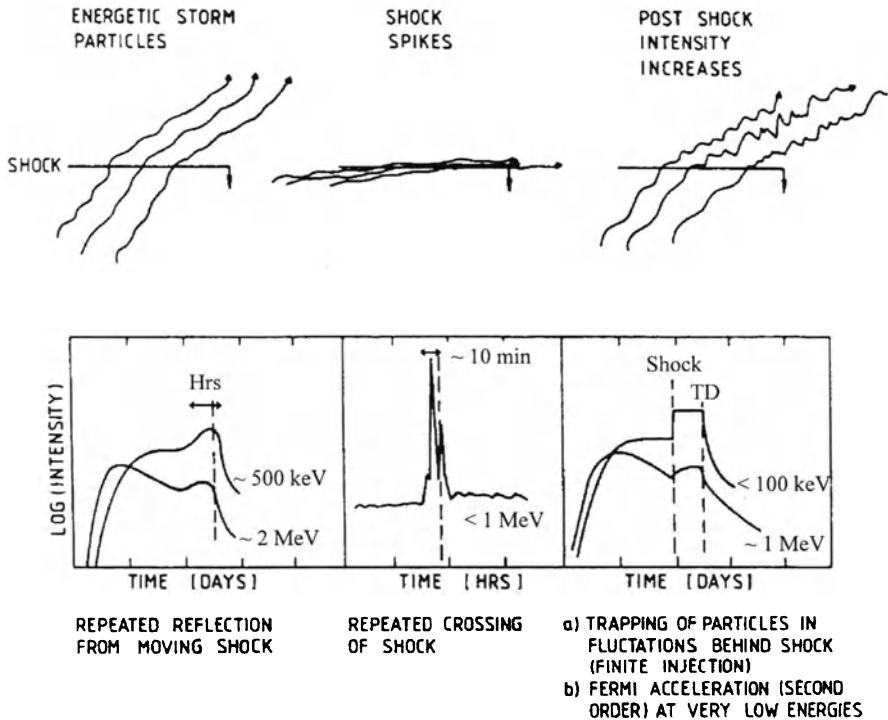


Fig. 9.7. Different types of energetic particle events in the tens to hundreds of kiloelectronvolt range at interplanetary shocks and the acceleration mechanisms. Reprinted from M. Scholer and G. Morfill [9.56], in *Study of traveling interplanetary phenomena*, 1977, Air Force Geophysics Laboratory

ing the shock. With increasing energy, the intensity increase related to the shock becomes smaller, reflecting the steep energy spectrum. These events are connected with quasi-parallel shocks, and thus the particles are accelerated by diffusive shock acceleration. The features of the particle event, for instance the quasi-exponential intensity increase upstream of the shock and the energy spectrum, are in agreement with the predictions of this acceleration mechanism. In addition, the enhanced upstream turbulence is indicative for self-generated turbulence.

At quasi-perpendicular shocks, a different type of particle event can be observed, namely the shock spike (middle panel in Fig. 9.7). Here the intensity increase is limited to a few minutes around the time of shock passage. Despite its short duration, the intensity increase can be quite large. Shock spikes are explained by shock-drift acceleration. However, to acquire the observed energies, the particles must cross the shock more than once, and thus here shock-drift acceleration works in combination with scattering and not as scatter-free shock acceleration. The details of the particle event (single-spike, multiple-spikes, etc.) are more variable than in the classical ESP events. This

structuring, however, is not necessarily related to the acceleration process but might also reflect the differences between neighboring field lines or the variability of the interplanetary medium on short temporal and spatial scales.

The third type of event, the post-shock increase, can occur in isolation, on the decaying flank of a SEP event or in the wake of an ESP event. In the particle data, it can be identified as an abrupt increase in intensity as the shock passes the observer. It often ends abruptly as the observer encounters the ejecta. Post-shock intensity increases are generally associated with strong turbulence in the downstream medium. This turbulence leads to an efficient stochastic acceleration as well as a storage of accelerated particles behind the shock front. In the low-energy electrons, the post-shock increases are the most common type of event; however, in ions and higher energetic electrons they are less frequent.

While the local geometry θ_{Bn} determines the shape of the shock-related particle increase, its size is determined by the compression ratio, the magnetic compression, and the shock speed $v_{SB} = v_s \sec \theta_{Bn}$ parallel to the magnetic field. This latter parameter even seems to distinguish between shocks with little or no particle acceleration ($v_{SB} < 250$ km/s) and shocks that lead to significant particle acceleration [9.60]. Although v_{SB} is less expressive than the shock speed, its physical significance is more important: as the particles are bound to travel along the magnetic field line, the shock speed parallel to the field determines the ability of a particle to escape from the shock. Even at rather slow shocks, a large number of particles might be kept close to the shock for further acceleration as long as θ_{Bn} is large, i.e. the shock is quasi-perpendicular. The physical significance of the magnetic compression also might be simple: the change in B across the shock front creates a magnetic mirror for particles moving from the upstream medium towards the shock. As these particles are reflected at this mirror, they gain energy. This process is faster and more efficient than the reflection in the turbulence downstream of the shock as assumed in standard diffusive shock acceleration. But for efficient particle acceleration multiple encounters with the shock are required, and thus again the upstream turbulence is the crucial factor.

9.3.2 High Energies (Megaelectronvolts)

In megaelectronvolt protons and ions different intensity profiles can be distinguished, too. However, they do not reflect the local acceleration mechanism as in the low energies but the location of the observer relative to the shock. This is a consequence of the higher speeds of the megaelectronvolt particles which allows them to escape from the shock front. An observer in interplanetary space thus samples all particles that the shock has accelerated on the observer's magnetic field line while it propagates outward. The intensity profile therefore is the superimposition of particle injections with different sizes and at different positions.

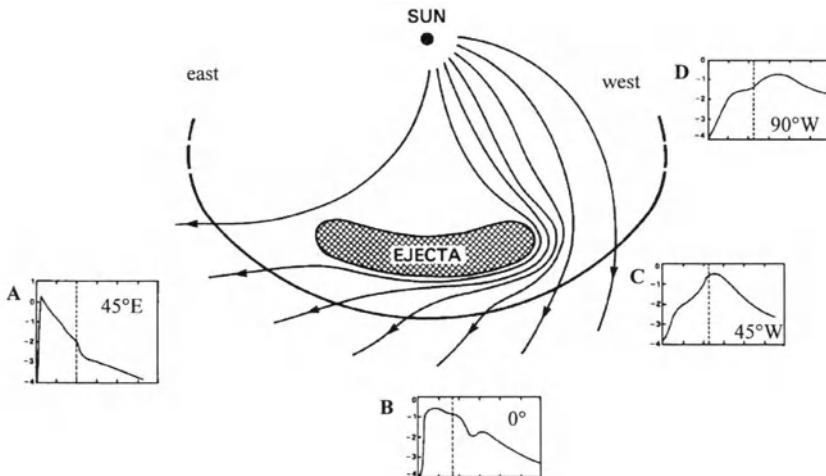


Fig. 9.8. Representative profiles of about 20 MeV proton events for different positions of the observer with respect to the shock. The draping of the field lines around the ejecta is only a suggestion. Reprinted from H.V. Cane et al. [9.6], *J. Geophys. Res.* **93**, Copyright 1988, American Geophysical Union

Figure 9.8 shows sketches of typical intensity profiles of 20 MeV protons for different locations at an interplanetary shock, the latter propagating downward. To allow for easy comparison of the different shapes, all the profiles in Fig. 9.8 are drawn to fit the frames. If we were to draw them to scale to consider the real sizes, we would find that the intensity of the first strong intensity increase is largest for an observer on the eastern flank. Going towards the central meridian or the western flank of the shock, the intensity of the solar component decreases. The intensity of the shock-accelerated component, on the other hand, is largest at the nose of the shock at the central meridian while it decreases towards the flanks.

Now we can easily explain the different profiles. The observer in panel D is located at the eastern flank of the shock and is magnetically connected to the flare site. Thus he sees a strong intensity increase early in the event due to the solar component. Often the flare-accelerated particles might be identified from their charge states and composition, cf. Fig. 9.3. The local acceleration at the shock, however, is rather meager because the observer is located only on its flank. Going to the east, observer A is located at the nose of the shock but is magnetically connected to a position about 60° west of the flare site. The solar component therefore is smaller; however, the particle component accelerated at the shock is maximal. Whether the first increase is larger than the one close to the shock, as depicted in the figure, or vice versa, depends on the properties of the flare, the shock's acceleration efficiency, and the scattering conditions in interplanetary space. Moving farther to the

west, the solar component decreases or becomes undetectable because of an insufficient magnetic connection. The component accelerated at the shock is also smaller than the one at the central meridian; however, it is always larger than the solar component.

Note that Fig. 9.8 is a schematic based on a statistical study. The pattern of changes in profiles along the shock front have also been confirmed by multi-spacecraft observations at individual shocks. Nonetheless, we should be aware that all shocks and flares exhibit their own features and therefore the relative intensities between the first intensity increase and the hump at the time of shock passage can vary by orders of magnitude, even at a fixed location relative to the shock. Thus the typical profiles sketched in Fig. 9.8 can also be observed some ten degrees farther to the west or to the east.

As at low energies, the intensity at the time of shock passage is correlated with the magnetic compression and the shock speed. However, these correlations are rather weak, indicating that other parameters might influence the acceleration efficiency, too. In addition, we should expect a weak correlation only because the shock parameters are measured locally at the position of the observer while the particle event is a superimposition of all injections along the observer's magnetic field line where the shock parameters most likely would have been different.

The observations clearly show that megaelectronvolt protons are accelerated at interplanetary shocks. But how? Of the acceleration mechanisms discussed in Chap. 7, only diffusive shock acceleration might to be able to explain the observations. Stochastic acceleration in the turbulence behind the shock appears to be too slow. In addition, it would keep the particles confined in the downstream medium, leading to a post-shock increase. The latter can be observed in megaelectronvolt protons, although not frequently. Shock-drift acceleration also is not very likely because its characteristic feature, the shock spike, is not observed. In addition, it is debatable whether under interplanetary conditions SDA will be able to accelerate protons to megaelectronvolt energies.

But the explanation of the observations by diffusive shock acceleration also is inconsistent. For instance, if we assume typical scattering conditions, the characteristic acceleration time would be much larger than the travel time of the shock from the Sun to the observer. If, on the other hand, we use the upstream intensity increase to determine λ according to (7.28), we find mean free paths much lower than those typically observed in interplanetary space, even lower than those inferred from the magnetic field fluctuations using standard QLT (which always are a lower limit only). The characteristic acceleration times inferred with these λ s, on the other hand, are still larger than the shock's travel time. Thus the predictions of diffusive shock acceleration cannot be used to derive a consistent picture. Nonetheless, this does not imply that the particles are not accelerated by this mechanism. First of all, all the tests and predictions assume a steady state. Obviously, this is not



Suppl.
→ 9.3.3

acquired because it would require acceleration times larger than the shock's travel time. Instead, a steady state is acquired up to a few hundred kiloelectronvolts only, as is also evident in the break in the power-law spectrum in Fig. 9.6. But even if a steady state is acquired only up to a lower energy, some particles have already been accelerated to much higher energies. In addition, the higher energies might require the consideration of additional effects, in particular particle escape from the shock front or the curvature of the shock front. Thus the acceleration of megaelectronvolt protons at interplanetary shocks is still an open question. One attempt to gain insight into the problem is the simulation of the observed particle profiles with a model that takes into account the propagation of the shock and a continuous, though variable, injection of particles, combined with the subsequent interplanetary propagation of the particles [9.20, 9.27]. Currently the shock is assumed to be a black box, although the inclusion of a real acceleration mechanisms, for instance time-dependent diffusive shock acceleration, should be the target. Nonetheless, such a model still yields valuable results, in particular the variation of the particle injection from the shock as the latter propagates outward can be determined, allowing us to determine the change in shock acceleration efficiency with time.

9.3.3 Forbush Decreases

Interplanetary shocks not only accelerate energetic particles, they also block part of the galactic cosmic radiation in the hundreds of megaelectronvolt to gigaelectronvolt range. This is called a Forbush decrease. Forbush decreases occur in two steps: first, there is an intensity decrease as the shock passes by, followed by a more pronounced decrease as the observer enters the ejecta driving the shock [9.7]. Typical values at 500 MeV are about 2% for the shock decrease and about 5% for the decrease related to the arrival of the ejecta. While both decreases are rather abrupt (within minutes), the recovery phase is long and can last for days.

Decreases in the galactic cosmic radiation are not only found at traveling interplanetary bow shocks but also at the shocks at corotating interaction regions or when a fast solar wind stream is suddenly swept across the observer. The first step of the Forbush decrease is most likely due to the different scattering conditions upstream and downstream of the shock or, in case of a CIR, in the fast and slow solar wind streams. The second step associated with the arrival of the ejecta seems to be influenced by different scattering conditions, too, which regulate the filling of the ejecta originally void of energetic particles.

9.4 Particles at Planetary Bow Shocks

Particles and waves can also be observed upstream of planetary bow shocks. Basically, the bow shock is a parabolic shape symmetric around the Sun–Earth line. At the Earth’s position, the interplanetary magnetic field spiral has an angle of 45° with respect to the Sun–Earth axis. Thus along the bow shock, the local geometry θ_{Bn} is highly variable, cf. Fig. 9.9, ranging from quasi-perpendicular close to the nose of the bow shock to quasi-parallel at the dawn side of the magnetosphere.

Upstream of the bow shock, a foreshock region develops which is characterized by energetic particles streaming away from the shock front and by waves excited by these particles. Depending on the local geometry θ_{Bn} , different particle distributions and different types of waves can be observed. Close to the nose of the shock, the subsolar point, the geometry is quasi-perpendicular and shock-drift acceleration leads to a particle distribution in the form of a rather narrow beam of reflected ions, cf. the left panel of Fig. 9.10. The spiky peak in the middle of the distributions is the solar wind while the broader peaks are reflected and accelerated ions. The wave field consists of low-amplitude waves with frequencies of about 1 Hz.

At the dawn side the geometry is more quasi-parallel and particles are accelerated by diffusive shock acceleration. The resulting particle distribution is a diffusive one, forming a ring around the solar wind peak, as is evident in the right panel in Fig. 9.10. The waves excited by these ions again are low-

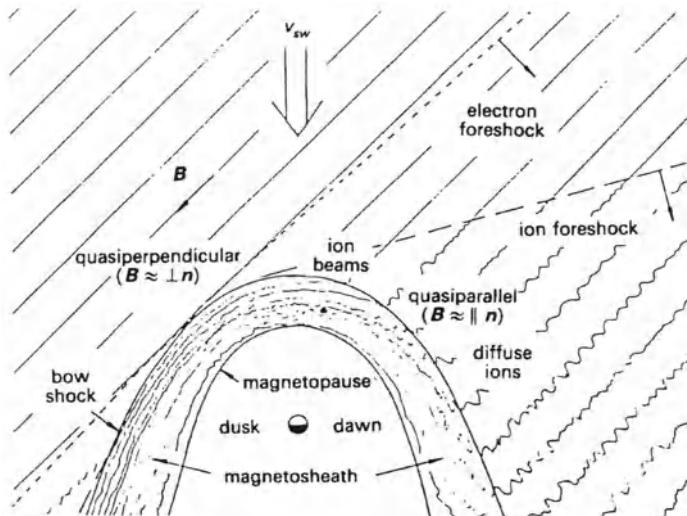


Fig. 9.9. Geometry of the bow shock and waves in the upstream and downstream media. Three different ion distributions are indicated. Reprinted from R. Bingham [9.3], in *Plasma physics* (ed. R. Dendy), Copyright 1993, with kind permission from Cambridge University Press

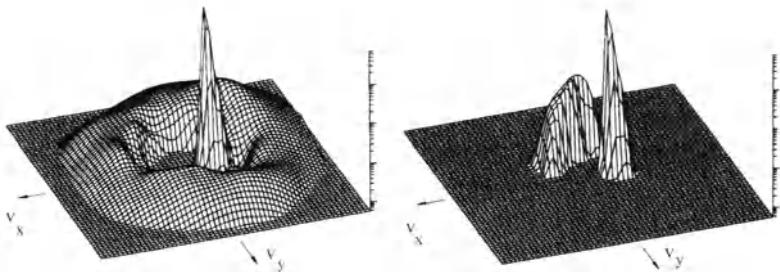


Fig. 9.10. Particle distributions upstream of the Earth's bow shock. In the beam distribution (*left*) particles propagate from the shock front towards the solar wind only while the diffusive population (*right*) covers all pitch angles. In both panels, the sharp peak in the middle is the solar wind incident on the bow shock. Reprinted from G. Paschmann et al. [9.43], in *J. Geophys. Res.* **86**, Copyright 1981, American Geophysical Union

frequency waves with larger amplitudes, occasionally containing shocklets in the sense of discrete wave packets which often are associated with discrete beams of particles. The relationship between particles and waves can be described in the model of coupled MHD wave excitation and ion acceleration [9.30], cf. Sect. 7.4.3.

In between, the local geometry is oblique and both mechanisms contribute to the particle acceleration. The resulting particle distribution is an intermediate distribution in which both the reflected beam and the more diffusive population can be found. The accompanying wave field basically consists of transverse low-frequency waves.

The electrons form a foreshock region, too. It starts close to the nose of the magnetosphere, cf. Fig. 9.9: at the nose of the bow shock, the geometry is quasi-perpendicular and particles are accelerated by shock-drift acceleration. Since the gyro-radii of the electrons are much smaller than those of the protons, their drift path along the shock front is shorter, leading to an earlier escape and therefore a more extended foreshock region.

Compared with the particle populations observed at interplanetary shocks, the bow shock particles, although significantly more energetic than the solar wind, are still low-energy particles. Electron acceleration is observed only up to a few kiloelectronvolts, ion acceleration up to some tens of kiloelectronvolts. Waves and particles upstream of bow shocks cannot only be found on Earth but also on other planets [9.41, 9.52, 9.53]

9.5 Galactic Cosmic Rays

Galactic cosmic radiation is incident on the solar system isotropically and constantly. It basically consists of hydrogen, helium, and electrons, but also heavier ions such as C, O, and Fe. Energetically, the galactic cosmic radiation

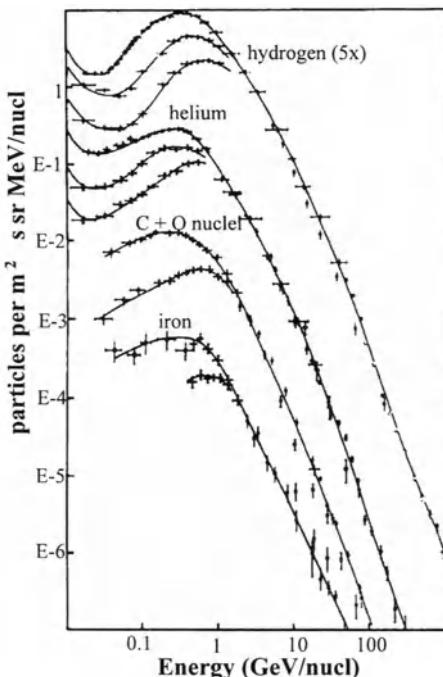


Fig. 9.11. Spectrum of galactic cosmic radiation. At energies below some GeV the three curves for each particle species indicate the spectra for solar minimum (upper curve), average (middle curve), and solar maximum conditions (lower curve). Reprinted from *Physics Today* **27**, P. Meyer et al. [9.38], Copyright 1974, with kind permission from the American Institute of Physics

starts where the spectrum in Fig. 9.2 ended, at some tens of MeV/nucl. The energy spectrum has a positive slope, i.e. the intensity increases with increasing energy, up to some hundreds of megaelectronvolts, cf. Fig. 9.11. At higher energies, the spectrum has a slope of -2.5 .

9.5.1 Modulation

At energies below a few gigaelectronvolts, the intensities of the galactic cosmic radiation show a strong dependence on solar activity with a maximum during the solar minimum. This effect is called modulation. With increasing solar activity, the maximum of the energy spectrum shifts towards higher energies. At proton energies of about 100 MeV the modulation is maximal, while at energies of about 4 GeV the modulation only is 15%–20%. Up to about 10 GeV galactic electrons show a spectrum similar to the protons. They also show a modulation with solar activity in the energy range 0.1 to 1 GeV.

The modulation of the galactic cosmic radiation at some gigaelectronvolts is shown in Fig. 9.12 (lower panel) together with the sunspot number as an indicator of solar activity (upper panel). In both panels, monthly averages are used. The anticorrelation is clearly visible. But how can we understand this behavior? The galactic cosmic rays have to propagate from the heliopause towards the inner heliosphere and it appears that during solar maximum fewer particles manage this task. What is hindering them and which forces do

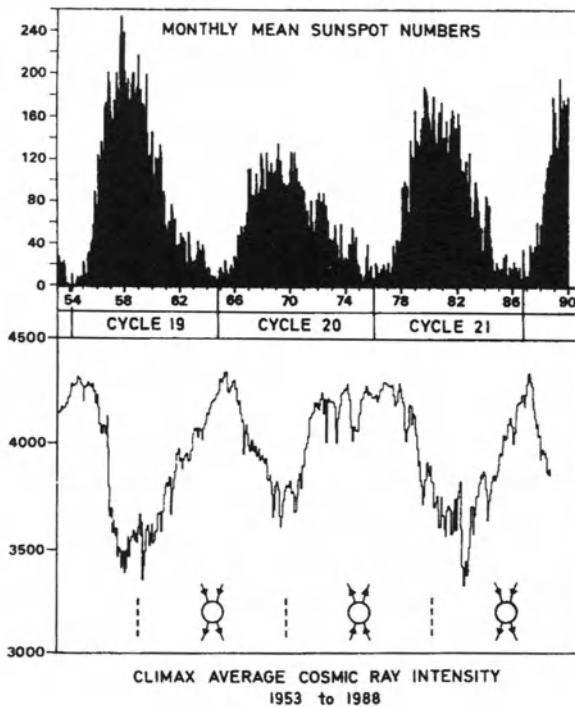


Fig. 9.12. Modulation of the galactic cosmic radiation (*lower panel*) with solar activity (*upper panel*). The small inserts indicate the polarity of the solar magnetic field. Courtesy G. Wibberenz, University of Kiel

they experience? Like solar particles, galactic cosmic rays travel field parallel and are scattered at the magnetic field irregularities. With increasing solar activity the structure of the interplanetary field changes: first, the tilt angle increases, leading to a wavier heliospheric current sheet, cf. Fig. 8.20; second, shocks and the ejecta driving them disturb the interplanetary medium.

Owing to the increased tilt angle and the resulting wavier heliospheric current sheet, a field line from the Sun to the heliopause is much longer during solar maximum than under minimum conditions. Thus the path of the cosmic ray particles towards the inner heliosphere is longer, allowing either fewer particles to intrude or leading to different travel times during solar maximum and minimum which, combined with the constant flux of galactic cosmic rays outside the heliosphere, would also lead to an intensity variation, with higher intensities during solar minimum. Even if we allow for a short-cut due to drift in the heliospheric current sheet, allowing the particles to travel perpendicular to the tightly wound up outer heliospheric magnetic field instead of following the field line around the heliosphere many times, the particle's path is longer during solar maximum because of the wavy current sheet.

The Forbush decreases show that shocks and the driving ejecta locally reduce the intensity in cosmic rays, affecting their propagation towards the inner heliosphere. During solar maximum conditions, large numbers of shocks

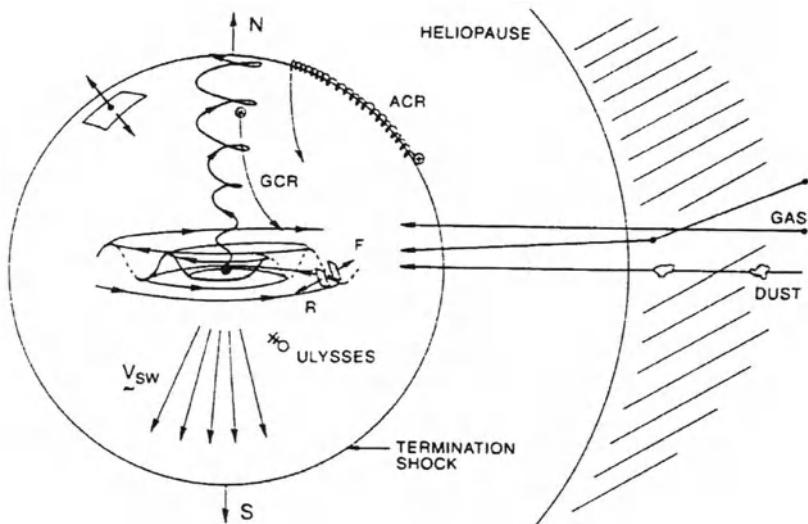


Fig. 9.13. Simplified sketch of the heliosphere and the energetic particles. In reality, the heliosphere has a shape similar to the magnetosphere only that the deformation is due to the interstellar wind. Note the different shapes of magnetic field lines originating from the equatorial and the polar regions. Reprinted from *Space Sci. Rev.* **87**, M.A. Lee [9.32], Ulysses race to the pole: symposium summary, p. 485, Copyright 1995, with kind permission from Kluwer Academic Publisher

and ejecta are flung out from the Sun. During their propagation into the outer heliosphere, shocks interact with the corotating interaction regions as well as with each other forming merged interaction regions. Since these regions occasionally can form a closed torus around the inner heliosphere, galactic cosmic radiation can be blocked efficiently and globally.

9.5.2 Over the Poles



Suppl.
→ 9.6

The above description of the modulation is a very simple one; however, it is graphical. To understand modulation quantitatively, we need a model. Such models exist; however, as in the case of the transport equation, a source has to be specified within these models. This source is the interstellar spectrum, which is not known so far. Since the heliopause is expected at a distance of at least 100 AU, even for the Voyager and Pioneer spacecraft it will still take some time before they enter interstellar space. These spacecraft travel close to the plane of ecliptic. Ulysses, on the other hand, has been built to travel over the Sun's poles where the intensity of the galactic cosmic radiation is expected to be much higher. Figure 9.13 explains this expectation. The phenomena blocking the galactic cosmic rays, i.e. the corotating interaction regions and the traveling shocks with their drivers, are confined to latitudes below about $\pm 40^\circ$. Thus over the poles they would not be able to influence

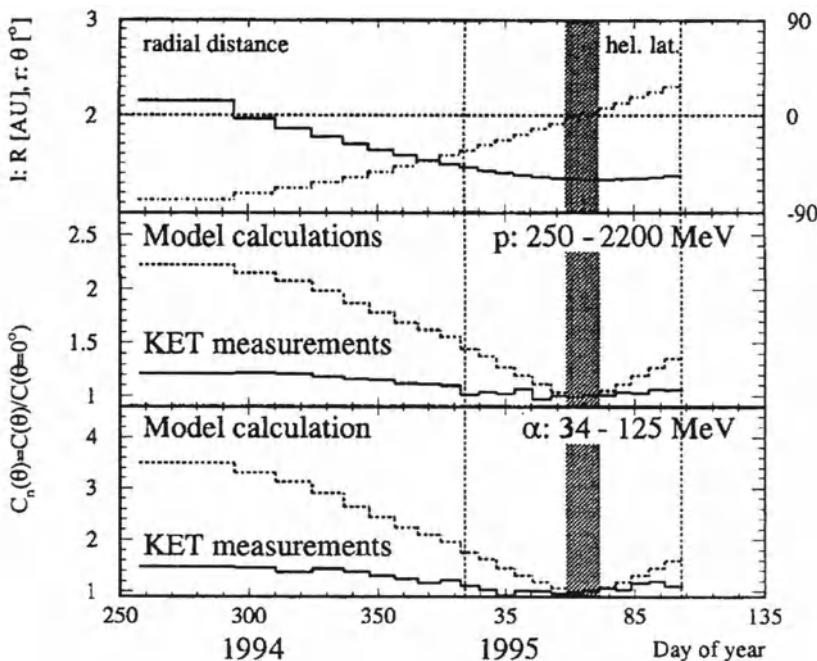


Fig. 9.14. Galactic cosmic radiation over the solar pole. In the upper panel, radial distance and heliographic latitude are shown, the lower panels show the measured (*solid line*) and expected (*dashed line*) intensities in two different energy channels. The dark vertical bar gives the time when Ulysses crosses the equatorial plane. Courtesy B. Heber, University of Kiel

the galactic cosmic radiation. In addition, the Archimedean spiral is wound up tightly in the equatorial plane. Although the field lines over the poles are wound up too, they still go more radial, leading a shorter travel path for the GCRs over the poles than in the plane of the ecliptic. Thus over the poles, scientists expected to observe more or less the interstellar spectrum of the galactic cosmic radiation because there regions provide easy access for the GCRs.

Observations from Ulysses did not meet these expectations, as is shown in Fig. 9.14. The upper panel shows the radial distance and the latitude of Ulysses. In the lower panels, the measured (*solid*) and expected (*dashed*) intensities in two different channels are given. These intensities are relative intensities because they are normalized to the intensity in the equatorial plane. The vertical bar indicates the time when Ulysses crosses the plane of the ecliptic. Although the observations show an increase of galactic cosmic rays over the poles, it is much smaller than predicted.

Two alternative interpretations are discussed. Since the magnetic field turbulence above the poles is unexpectedly high, the particles might experience

strong scattering, making it more difficult to intrude into the solar system. Or, the shape of the heliosphere is different from the expected one. Originally, the heliosphere was assumed to have a structure similar to a magnetosphere, i.e. an almost circular form in the direction towards the interstellar wind flow with a long tail swept away by the interstellar wind. This elongated shape should lie within the plane of the ecliptic. Since the solar wind is fast over the poles, cf. Fig. 8.17, the heliosphere might be much more extended there, giving it a peanut-like shape. Therefore, field lines extending from the poles towards the heliopause would be longer than expected. Alone or combined with the increased turbulence this might explain the unexpectedly low intensity of the galactic cosmic radiation.

Thus Ulysses has not only provided information to fill in some of the white spots in our knowledge but has also challenged some of our older concepts and understanding.

9.6 Summary

Energetic particles in interplanetary space stem from different sources, such as planetary bow shocks, traveling and corotating shocks, solar flares, or from outside the solar system. These populations have characteristic spectra, compositions, and time profiles, providing information about the acceleration and propagation mechanism. For almost all particle populations, shock acceleration is important; in solar flares reconnection and selective heating are also at work. The particles, in particular solar energetic particles, can also be used as probes for the magnetic structure of the interplanetary medium and the superimposed turbulence.

Exercises and Problems

9.1 Determine the Larmor radius, gyro-period, and speed of galactic cosmic rays with an energy of 10 GeV. Compare with the same values for a solar proton with an energy of 10 MeV. Determine the travel time between the Sun and the heliopause at 100 AU for a straight path and a path following an Archimedean magnetic field line.

9.2 Shock acceleration is important for many of the particle populations discussed in this chapter. Describe them and find arguments for the differences, in particular the maximum energy gained by the different populations.

9.3 In Fig. 9.3 the composition slowly evolves from one characteristic of flare acceleration to another one characteristic of shock acceleration. Explain this slow evolution in terms of a δ -like solar acceleration, a continuous acceleration of particles at the shock, and interplanetary propagation.

10. The Terrestrial Magnetosphere

The poet's eye, in a fine frenzy rolling,
Doth glance from Heaven to Earth, from Earth to Heaven,
and, as imagination bodies forth
the *forms of things unknown*, the poet's pen
turns them to shapes, and gives airy nothing
a local habitation and a name.
Shakespeare, *Much Ado About Nothing*

The magnetosphere is shaped by the interaction between a planetary magnetic field and the solar wind. The magnetopause is a discontinuity separating both fields, forming a cavity in the solar wind. Since the solar wind is a supersonic flow, a standing shock wave, the bow shock, develops in front of the magnetopause. In the anti-sunward direction, the magnetosphere is stretched by the solar wind, forming the magnetotail. Inside the magnetosphere, different plasma regimes exist, dominated by ionospheric plasma in the plasmasphere, a highly variable mixture of ionospheric and heliospheric plasma in the geosphere, and by the solar wind plasma in the outer magnetosphere. These different regimes are coupled by fields and currents. Inside the plasmasphere energetic particles are trapped in the radiation belts.

Particles and energy are fed into the magnetosphere from different sources: (a) the solar wind can penetrate into the magnetosphere due to reconnection at the dayside and diffusion into the magnetotail, (b) solar energetic particles can penetrate into the magnetosphere at the polar cusps, (c) galactic cosmic rays traveling along Størmer orbits even can penetrate down to ground level, and (d) plasmas are exchanged between the ionosphere and the magnetosphere.

10.1 The Terrestrial Magnetic Field

Magnetic fields either originate in currents or from magnetized bodies. Earth is not a magnetized body, as can be seen from the variations in the terrestrial field, in particular the pole reversals. Instead, the terrestrial field originates in a dynamo process. Close to the Earth's surface it can be approximated

as a dipole; at higher altitudes or under magnetically disturbed conditions it deviates from the dipole due to currents, the solar wind pressure, and plasma and field exchange with the interplanetary medium.

10.1.1 The Dipole Field

To first order, the Earth can be described as a sphere magnetized uniformly along its dipole axis. This axis intersects the surface in two points, the austral (southern) pole at 78.3°S 111°E close to the Vostok station in Antarctica and the boreal (northern) pole at 78.3°N 69°W close to Thule (Greenland). Both positions are about 800 km from the geographic poles and the magnetic dipole axis is inclined by 11.3° with respect to the axis of rotation. The dipole moment M_E of the Earth is 8×10^{25} G cm³ or 8×10^{22} A m².

Coordinates in a Dipole Field. In a spherical coordinate system orientated along the dipole axis, the position of a point P is described by its distance r from the center of the dipole and the angle θ from the dipole axis to the radius vector of P. With the magnetic moment M_E , the components of the magnetic field can be written as

$$B_r = \frac{M_E}{r^3} \cos \theta \quad \text{and} \quad B_\theta = -\frac{M_E}{r^3} \sin \theta. \quad (10.1)$$

The magnetic flux density then is

$$B = \sqrt{B_r^2 + B_\theta^2} = \frac{M_E}{r^3} \sqrt{1 + 3 \cos^2 \theta}. \quad (10.2)$$

The field strength therefore falls off with distance as $1/r^3$.

Geomagnetic Coordinates. The geomagnetic coordinate system is orientated along the magnetic dipole axis. A plane perpendicular to the dipole axis intersecting the center of the Earth defines the equatorial plane. The intersection of this plane with the Earth's surface gives the geomagnetic equator. The geomagnetic longitude Λ and latitude Φ then are defined analogously to the geographic longitude λ and latitude φ . With $\varphi_o = 78.3^\circ\text{N}$ and $\lambda_o = 291^\circ\text{E}$ as the latitude and longitude of the boreal magnetic pole, the magnetic and geographic coordinates are related by the transformations

$$\sin \Phi = \sin \varphi + \sin \varphi_o \cos \varphi_o \cos(\lambda - \lambda_o) \quad (10.3)$$

and

$$\sin \Lambda = \frac{\cos \varphi \sin(\lambda - \lambda_o)}{\cos \Phi}. \quad (10.4)$$

The magnetic potential at a position \mathbf{r} from the Earth's center is

$$V = \frac{\mathbf{M}_E \cdot \mathbf{r}}{r^3} = -\frac{M_E \sin \varphi}{r^2}. \quad (10.5)$$

From this, the magnetic field strength can be derived as $\mathbf{B} = -\nabla V$.

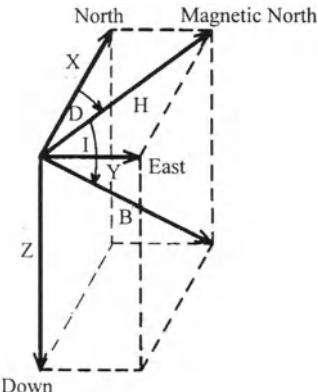


Fig. 10.1. Components of the geomagnetic field at the Earth's surface

Components of the Field. The geomagnetic field can be described in different systems. In a rectangular cartesian system, the triple (X, Y, Z) gives the northward, eastward, and vertical components. In a cylindrical system, the triple (D, H, Z) is used with Z as the vertical intensity, H as the horizontal intensity, and D as the declination of the field. In a spherical system with Z and X as the axes of reference, the field can be described by the triple (B, I, D) with total intensity B , inclination I , and declination D . Lines with constant D are called isogones. Lines with constant I are isoclines. The line with $I = 0^\circ$ is the dip-equator or geomagnetic equator. Figure 10.1 shows the relation between the different systems.

From (10.5) we can derive the field components. The radial (vertical) component then is

$$Z = B_r = \frac{\partial V}{\partial r} = \frac{2M_E \sin \Phi}{r^3}, \quad (10.6)$$

and the horizontal (tangential) component is

$$H = B_\varphi = \frac{1}{r} \frac{\partial V}{\partial \Phi} = -\frac{M_E \cos \varphi}{r^3}. \quad (10.7)$$

At the Earth's surface the magnetic field components in G can be approximated as $H = 0.31 \cos \Phi$ and $Z = 0.62 \sin \Phi$. At the pole B equals Z while at the equator B equals H . Thus the magnetic field strength at the pole is twice that at the equator. This ratio does not change with increasing distance.

The magnetic inclination is $\tan I = Z/H = -2 \tan \Phi$. Thus close to the dip-equator the inclination increases twice as fast as the geomagnetic latitude.

The magnetic field components discussed so far have been intrinsic because their reference direction is the magnetic field itself. The other components are relative ones, their reference direction is the geographic north. The declination D is defined as the angle between the magnetic field direction and the geographic north: $D = Y/X$. The northward and eastward components of the field then are given as $X = H \cos D$ and $Y = H \sin D$.

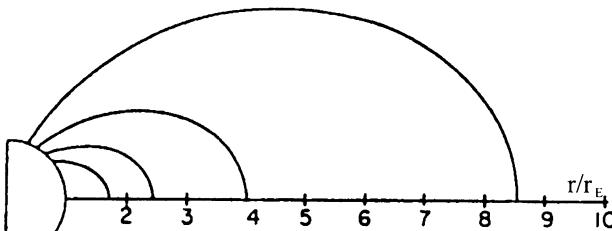


Fig. 10.2. Shape of magnetic dipole field lines and the definition of the L-shell parameter as the intersection between the field line and the equatorial plane

The equation of a field line $r = r(\Phi)$ can be inferred from (10.6) and (10.7). The magnetic field vector always is tangential to the line of force. The angle α between the radius vector and the magnetic field line is given as

$$\frac{B_\Phi}{B_r} = \frac{H}{Z} = \frac{rd\Phi}{dr} = -\frac{1}{2} \tan \Phi . \quad (10.8)$$

Thus we get

$$\frac{dr}{r} = 2d\Phi \tan \varphi = \frac{2 \sin \Phi}{\cos \Phi} d\Phi = 2 \frac{d(\cos \Phi)}{\cos \Phi} . \quad (10.9)$$

Integration yields $\ln r = 2 \ln(\cos \Phi) + \text{const}$, which can be written as

$$r = r_{\text{eq}} \cos^2 \Phi , \quad (10.10)$$

with r_{eq} distance of the field line from the Earth's center above the equator. This is also the largest distance of a field line; it is used to define the L-shell parameter $L_o = r_{\text{eq}}/R_E$, cf. Fig. 10.2. The magnetic field in G then is given as

$$B(L_o, \Phi) = \frac{0.31}{L_o^3} \sqrt{1 + 3 \sin^2 \Phi} \quad (10.11)$$

and the equation of the field line can be written as $L = L_o \cos^2 \Phi$. The field line intersects the Earth's surface at a latitude $\cos \Phi_E = \sqrt{L_o}$, where L equals 1. Physically, the L-shell is the surface traced out by the guiding center of a trapped particle as it drifts around the Earth while oscillating between its northern and southern mirror points. Note that on a given L-shell the particle's physical distance from the Earth's surface might change while its magnetic distance stays constant.

Expansion of the Field. A more correct expression for the geomagnetic field not too high above the Earth's surface is a multipole expansion. In spherical coordinates r , θ , and φ , the potential V can be written as

$$V(r) = \sum_{n=1}^{\infty} \frac{r_E}{r^{n+1}} \sum_{m=0}^n \{g_n^m \cos(m\varphi) + h_n^m \sin(m\varphi)\} P_n^m(\cos \theta) . \quad (10.12)$$

The g_n^m and h_n^m are normalization coefficients and the P_n^m are the Legendre coefficients. For $m > n$, P_n^m equals 0. The quantity n gives the order of the multipole: with $n = 1$, (10.12) describes a dipole field, with $n = 2$ a

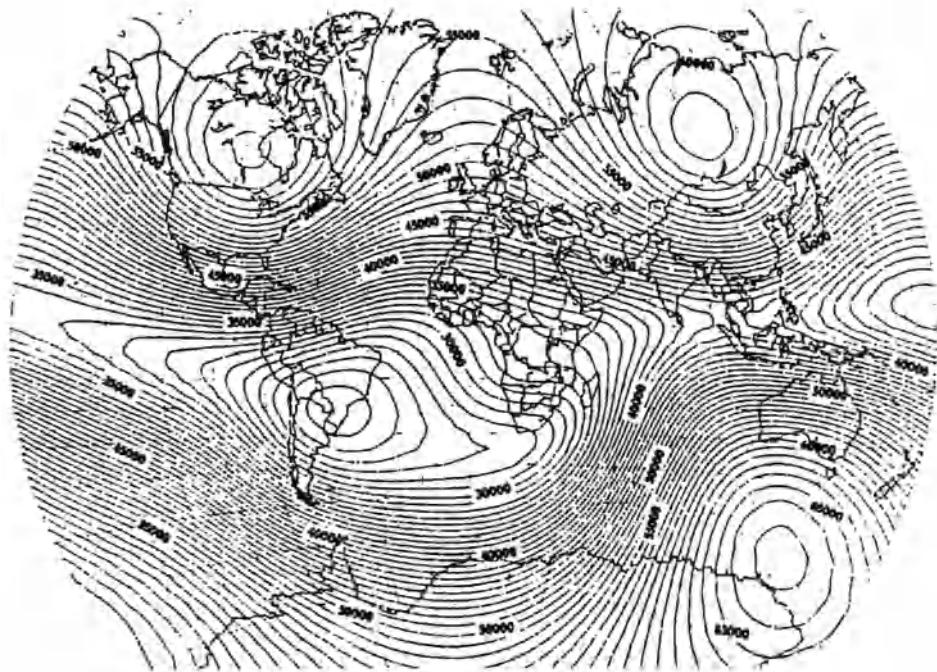


Fig. 10.3. International geomagnetic reference field. Contour intervals are 1000 nT, and a van der Grinten projection is used. In a pure centered dipole field the isointensity lines would be horizontal. Reprinted from IAGA [10.20], *EOS* **67**, Copyright 1985, American Geophysical Union

quadrupole. The magnetic monopole ($n = 0$) is not contained in (10.12). The potential of a dipole field can be inferred from (10.12) as $V = (r_E/r^2)g_1^0 \cos \theta$.

The coefficients in (10.12) are determined from fits to the measured magnetic field. Because the field changes with time, the coefficients have to be adjusted, too. The field determined with the multipole expansion is accurate to 0.5% close to the Earth's surface. With increasing altitude, the higher moments are less important and the field can be approximated by a dipole field. Above a height of 3 to 4 Earth radii, however, (10.12) is no longer useful because the magnetic field is significantly distorted by the solar wind.

The Surface Field: Shift of the Dipole Axis. To describe the field measured close to the Earth's surface, not only the higher moments of the multipole expansion must be considered but also the offset of the magnetic dipole relative to the center of the Earth by 436 km in the direction of the westerly Pacific.

The offset of the magnetic dipole leads to a region of unusually small magnetic flux density in the south Atlantic, just off the coast of Brazil, namely the South Atlantic Anomaly (SAA), cf. Fig. 10.3. Since the radiation belts are roughly symmetric around the dipole axis, they come closest to the Earth's

surface, too, making the SAA suitable for the study of radiation belts by rockets, but also making it a radiation risk for manned space-flight.

Deviations from the Multipole: The Outer Field. With increasing height, the shifted multipole approximation becomes less efficient. The importance of the higher moments of the multipole decrease, but the field does not become dipole-like. Instead, it is distorted by the influence of electric currents in the ionosphere and magnetosphere and by the direct action of the solar wind. While the solar wind strongly modifies the structure of the magnetic field, only the influence of the ring current can be observed at the surface: the ring current results from the opposite drifts of electrons and protons in the radiation belts and gives rise to a magnetic field opposite to the Earth's field, thereby reducing the magnetic flux density.

10.1.2 Variability

With the beginning of systematic and accurate measurements of the terrestrial magnetic field in the middle of the nineteenth century, the variability of the terrestrial magnetic field became evident: a systematic daily variation (Sq variation) was occasionally superimposed by much stronger variations, the geomagnetic disturbances. Later it was discovered that the terrestrial magnetic field also varies with the solar cycle. On geological time scales, variations become more pronounced, including the reversal of the field. The origins of these variations are quite different: daily variations result from the asymmetric shape of the magnetosphere. Magnetic storms and variations with the solar cycle reflect the variability of the solar wind, and the variations on even longer time-scales are related to the MHD dynamo inside the Earth.

In particular, the pole reversals give strong evidence for a dynamo process inside the Earth. Although we have not witnessed a reversal of the Earth's magnetic field, the different sheath of lava at the trenches associated with plate tectonics and sea-floor spreading have preserved a record of magnetic field polarity reversals, cf. Fig. 10.4. It appears that the typical cycle for field reversal is about 500 000 years; however, shorter polarity reversals, also called magnetic events, can be observed on time-scales between a few thousand years and about 200 000 years. The last polarity reversal occurred about 30 000 years ago, a time when the early humans already had spread across the Earth and the Neanderthals still where alive.

At the time of polarity reversal, fossil records often indicate the extinction of different plant species [10.21]. Nonetheless, today it is not understood whether this points to a causal relationship. Four models are discussed. (a) The weak or absent magnetic field at the time of polarity reversal allows the cosmic radiation to penetrate down to the biosphere, causing increased radiation damage in certain plant species, leading to the extinction of at least some of them. Although tempting, this interpretation probably will not hold as most of the cosmic rays will be absorbed by the atmosphere long before

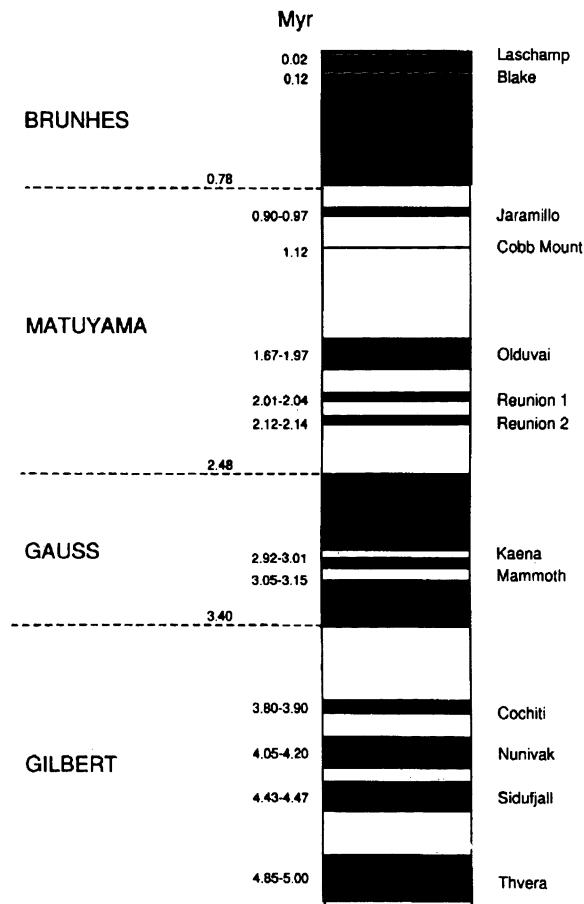


Fig. 10.4. Polarity reversals of the geomagnetic field during the last 5 million years as derived from lava records. Four magnetic field epochs are indicated on the left, and reversals on shorter timescales, magnetic events, are indicated on the right. Reprinted from D. Gubbins [10.12], *Rev. Geophys.* **32**, Copyright 1994, American Geophysical Union

reaching the biosphere. The field reversal can be seen as an increase in the records of cosmogenic nucleids produced by the interaction between cosmic rays and the atmosphere, such as ^{14}C or ^{10}Be ; however, its amplitude is too small to be considered a biological hazard. (b) Certain species in the microfauna are sensitive to the magnetic field. Thus field reversals might either have changed the biochemical processes within these micro-organisms, leading to their extinction, or they might have pushed them from their original habitat into a life-threatening one. Since they are the start of the food chain, the extinction of micro-organisms also can lead to the extinction of other species. (c) Within the framework of catastrophe theory, there is discussion as to whether an external event might have caused both the polarity reversal and the extinction of species. This, however, is highly speculative, because it is difficult to identify what kind of catastrophe this could have been. (d) Magnetic field polarity reversals are often associated with climate changes [10.56]. Although the mechanisms are not yet understood, it appears likely that they

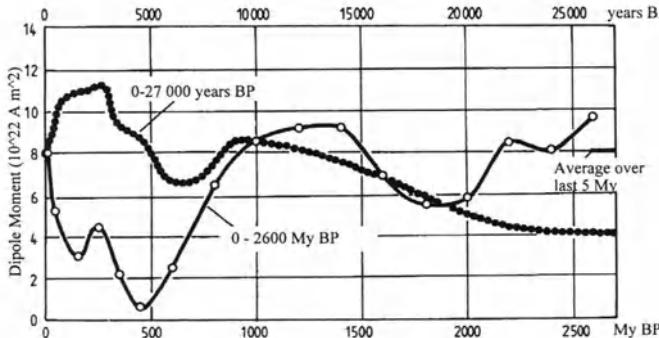


Fig. 10.5. Dipole moment of the terrestrial magnetic field for the last 27 000 years (dotted line) and the last 2600 million years (solid line). Reprinted from K. Strohbach [10.49], *Unser Planet Erde*, Copyright 1991, with kind permission from Gebrüder Borntraeger Verlag

invoke an interaction between galactic cosmic rays and the atmosphere, in particular changes in the global circulation patterns in the stratosphere or upper troposphere, in the ozone column, or in cloud cover.

But even if the magnetic field has a certain polarity, it is not necessarily constant. Instead, variations in the dipole moment can be found, cf. Fig. 10.5. For instance, about 2000 years BP, the magnetic dipole was almost 50% stronger, as can be seen from the dotted curve. About 20 000 years ago, the magnetic field had only half of its present value.

Thus not only the times of polarity reversal but also the strength of the geomagnetic dipole are distributed stochastically. For instance, in the period between 118 million and 83 million years BP, no magnetic field reversal occurred. For the last 5 million years, the statistical distribution of field reversals can be described by an asymmetric random walk distribution [10.31]. Such distributions are found if the observed process, here the polarity reversal, results from a large number of small individual events following a bell shape distribution. In the case of the terrestrial magnetic field, these small events probably are the patterns of the convection cells in the outer core which play an important role in the dynamo process.

10.1.3 The Terrestrial Dynamo

The principle of a MHD dynamo has already been discussed in Sect. 3.4. The special topology of the terrestrial dynamo is shown in Fig. 10.6. Both panels show the Earth's core only, the combined processes give a complete description of the terrestrial dynamo.

The motion of matter is depicted in the left panel: the liquid inside the outer core rotates but the angular speed decreases with increasing distance from the Earth's center. This is the consequence of the vertical transport of

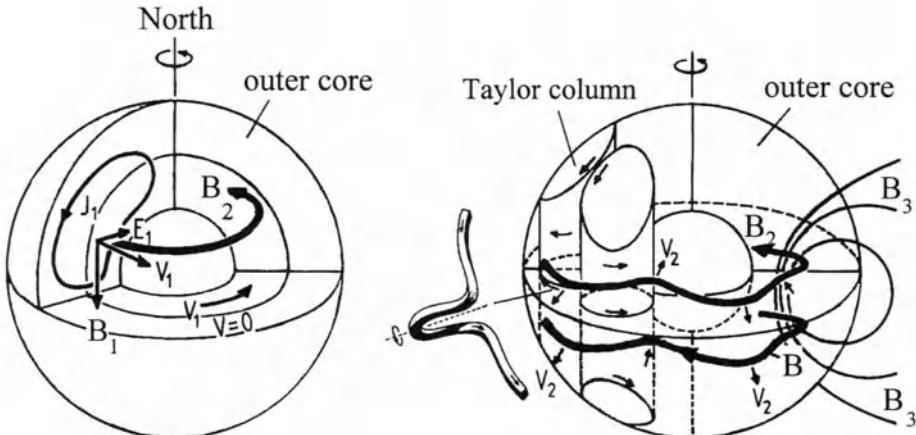


Fig. 10.6. Origin of the terrestrial magnetic field in a dynamo process. Only the core of the Earth is shown. The processes sketched in both panels have to be viewed together. Reprinted from K. Strohbach [10.49], *Unser Planet Erde*, Copyright 1991, with kind permission from Gebrüder Borntraeger Verlag

angular momentum: convection transports matter upwards from the deeper layers of the outer core, where the linear speed is small, while matter from the higher layers, where linear speeds are high, is transported downwards. As a result, the inner part of the outer core rotates faster than its outer one. In Fig. 10.6, this difference is indicated as the speed v_1 of the deeper layers relative to the outer layers. Let us now assume a magnetic seed field B_1 which is at rest relative to the outer layers and parallel to the axis of rotation. In the inner layers, the relative motion of the fluid with respect to the magnetic field causes an electric induction field E_1 directed towards the axis of rotation. Since the matter inside the core is highly conductive, a ring current j_1 results which is directed counter-clockwise in the northern hemisphere and gives rise to a toroidal magnetic field B_2 . In the southern hemisphere, the current j_1 and the magnetic field B_2 have opposite directions.

In addition, in the outer core convection takes place in the Taylor columns. The toroidal magnetic field B_2 is pushed upward by the convective motion and simultaneously twisted by the Coriolis force, as indicated in the small loop in the lower left of the right panel in Fig. 10.6. The resulting magnetic field B_3 is poloidal and adds to the initial field: the small seed field is amplified.

This dynamo process can explain the basic features of the terrestrial magnetic field. Many details, in particular the details of the polarity reversal, are not yet understood. One important ingredient for the terrestrial dynamo is the core's differential rotation, leading to the poloidal field. If the differential rotation were absent, no amplification of the seed field would be possible. This appears to be the case in the extremely slowly rotating Venus and the deep-frozen Mars.

10.2 Topology of the Magnetosphere

In this section we will limit ourselves to a discussion of the magnetic field topology; plasmas and currents will be discussed later. While close to the surface the terrestrial magnetic field can be described as a dipole field, at altitudes above a few Earth radii the influence of the solar wind is no longer negligible: the dipole field is compressed and part of it is violently torn back behind the Earth.

The structure of the magnetosphere is best described in the frame of reference with a fixed Sun–Earth axis. The magnetosphere then stays fixed in space while the Earth rotates inside it. This system divides the magnetosphere into two parts, a dayside directed towards the Sun and a nightside facing the magnetotail. The corresponding directions in the equatorial plane are noon and night, and the direction perpendicular to it dusk and dawn, always referring to local time.

The most important features of the magnetosphere are shown in the noon–midnight cross-section in Fig. 10.7. Typical extensions are about $10 r_E$ in the solar direction and more than hundred r_E tailwards.

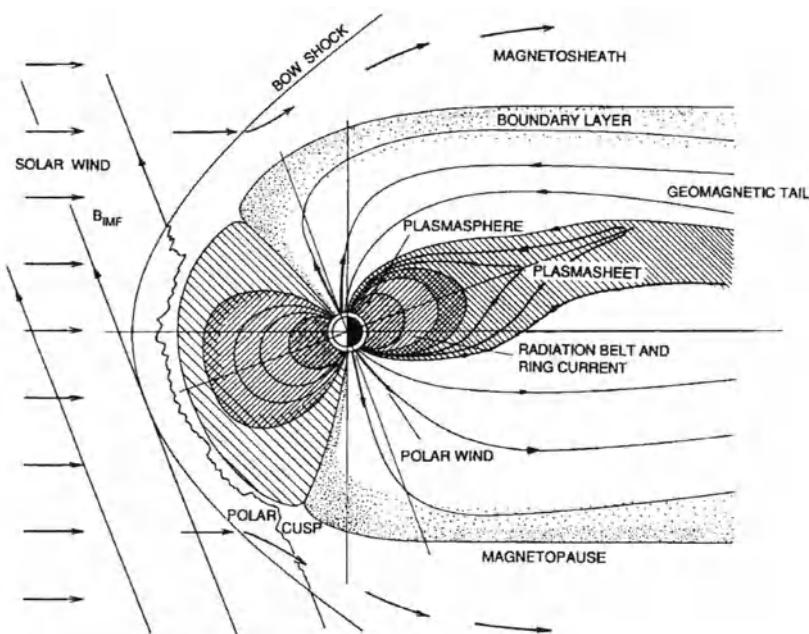


Fig. 10.7. Noon–midnight cross-section through the magnetosphere. Reprinted from G.K. Parks [10.38], *Physics of Space Plasmas*, Copyright 1991, with kind permission from Addison-Wesley Publishing Company

10.2.1 The Magnetopause

The magnetospheric cavity is separated from the interplanetary magnetic field by the magnetopause. It is defined by an equilibrium between the solar wind kinetic pressure and the pressure of the terrestrial magnetic field. The interplanetary magnetic field and the thermal pressure of the solar wind do not contribute to this balance; their combined pressure is less than 1% of the plasma kinetic pressure. For a similar reason, the gas-dynamic pressure inside the magnetosphere does not enter into the balance: it is too small compared with the magnetic pressure.

In the pressure balance we have to consider the geometry of the field and the plasma flow. Since the magnetic pressure is anisotropic, only the tangential magnetic field B_t contributes to the magnetic pressure:

$$p_{\text{mag}} = \frac{B_t^2}{8\pi}. \quad (10.13)$$

If we assume the magnetopause to be a perfect boundary between the solar wind and the terrestrial field, B_n equals zero at the magnetopause. If we describe the solar wind as an electron and ion plasma flowing at an angle ψ with respect to the normal direction on the magnetopause, within a second each surface element of the magnetopause is hit by $n_{\text{sowi}} u_{\text{sowi}} \cos \psi$ particles, with n_{sowi} being the number density and u_{sowi} the bulk speed of the solar wind. These particles transfer a momentum $2mn_{\text{sowi}} u_{\text{sowi}} \cos \varphi = 2\varrho u_{\text{sowi}} \cos \psi$. The pressure balance at the magnetopause therefore can be written as

$$2\varrho u_{\text{sowi}}^2 \cos^2 \psi = \frac{B_t}{8\pi}. \quad (10.14)$$

Note that this is a simplification because we have not considered that the solar wind is slowed down when it passes through the bow shock and part of the flow energy is converted into thermal energy, cf. Sect. 10.2.4. Nonetheless, (10.14) is still valid as long as the factor 2 in the kinetic pressure is substituted by a factor $K < 2$, which at the terrestrial bow shock is about 0.88.

Let us assume the magnetosphere to be axisymmetric. We then can use cylinder coordinates with r the radial distance from the axis of symmetry and s the distance along the magnetopause, measured from the subsolar point. With $dr/ds = \cos \psi$ we can determine the cosine in (10.14). If B_t were known, the position of the magnetopause could be determined. But B_t is known only as the solution of a potential problem with the boundary conditions defined by (10.13). Thus (10.14) has to be solved iteratively: we make an assumption for B_t which is a solution of (10.13). Then we can determine the electric currents inside the magnetopause, which in turn give a new B_t , which again can be used as input.

As a crude measure of the size of the magnetosphere, the position of its subsolar point on the Sun–Earth line is used. Here the plasma flow is perpendicular to the magnetopause and $\cos \psi$ equals 1. B_t can be approximated from B_o , the magnetic field at the Earth's surface. If we assume a mirror dipole

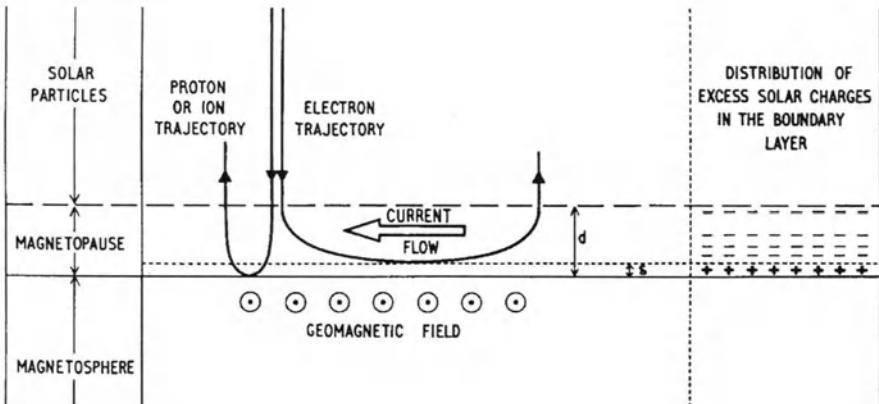


Fig. 10.8. The deflection of solar wind electrons and protons is responsible for both the finite thickness of the magnetopause and the Chapman-Ferraro current inside it. Reprinted from D.M. Willis [10.55], *Rev. Geophys. Space Phys.* 9, Copyright 1971, American Geophysical Union

at a distance $2d$, the normal component of the magnetic field would vanish if the tangential field were doubled in d : $B_t = 2B_o/d^3$. Then the distance of the subsolar point can be determined from (10.14):

$$d_{so} = \sqrt[6]{\frac{4B_o^2}{8\pi K \varrho u_{sowi}^2}}. \quad (10.15)$$

For typical solar wind conditions, the subsolar point, or stand-off distance, is at $10r_E$.

As can be seen from (10.15), the stand-off distance depends on the solar wind speed, and to a lesser extent, also on its density. The variability of the solar wind conditions thus leads to continuous changes in the size and therefore also in the shape of the magnetopause. Depending on the solar wind speed, the subsolar point is between 4.5 and $20 r_E$. Since the solar wind speed changes can be quite abrupt, e.g. across a traveling interplanetary shock or when a fast stream is suddenly swept across Earth, the magnetopause has to adjust to this changed environment rather fast. Satellite observations indicate speeds of the magnetopause between a few kilometers per second in response to solar wind fluctuations and up to about 600 km/s in response to discontinuities. The average speed of the magnetopause is about 40 km/s.

Although the magnetopause is defined as a three-dimensional boundary at which an equilibrium between the solar wind and the planetary magnetic field is established, it is not infinitely thin. Instead, it is an extended sheath with a thickness between a few hundred up to thousand kilometers: the solar wind is reflected at the magnetopause only after it has penetrated into the magnetic field and has been turned around by the Lorentz force, performing half a gyro-orbit during the reflection process, cf. Fig. 10.8. Since

the Lorentz force depends on the charge of the particle, electrons and ions are deflected in opposite directions, forming the Chapman–Ferraro current inside the magnetopause. This charge separation leads to a pile-up of charges at the flanks of the low-latitude magnetosphere (low-latitude boundary layer (LLBL)) with an excess of positive charges on the dawn side and negative charges on the dusk side. This can also be interpreted as a dawn-to-dusk electric field. Field lines intersecting the LLBL map back the potential pattern towards the high-latitude ionosphere.

This deflection can also be described as an extreme case of the grad B drift, with the magnetic field vanishing on one side of the boundary. The drift speed, as given by (2.43), only depends on the particle's Larmor radius which also defines the depth at which the particles penetrate into the magnetopause. Since the Larmor radius depends on $m/|q|$, ions penetrate deeper into the magnetopause than electrons, leading to an excess of negative charges in the outer magnetopause and an excess of ions in the inner one. The resulting electrical polarization field accelerates (decelerates) an electron (ion) on entering the magnetopause and decelerates (accelerates) the particle on leaving it. Thus the particles do not travel along semicircles, but rather their orbits are more elliptical, as sketched in Fig. 10.8.

10.2.2 Polar Cusps

The polar cusps are two singularities in the dayside magnetosphere: here the magnetic field vanishes and particles and plasma can penetrate freely into the magnetosphere. The polar cusps separate closed field lines on the dayside magnetosphere from open field lines swept to its nightside. The cusps are not located at the dipole axis but at lower geomagnetic latitudes because the higher latitude field, which in the dipole field would still be on the dayside, are convected with the solar wind to the nightside magnetosphere. The magnetic field lines connect the cusps back to geomagnetic latitudes of about 78° ; they are the only ones connecting the surface of the Earth to the magnetopause. Thus all field lines of the magnetopause converge at the cusps. The cusps themselves are filled with plasma from the magnetosheath but not from the magnetosphere. Thus at the cusps, plasma of solar wind origin can penetrate deep into the Earth's atmosphere, as can energetic particles.

10.2.3 The Tail

Magnetic field lines extending from the cusps to the nightside border the magnetotail. Close to the Earth, in addition to these open field lines also closed field lines can be found inside the tail, preserving, at least partly, the dipole character of the inner magnetosphere. As on the dayside magnetosphere, the closed field lines originate in geomagnetic latitudes below 78° . At higher latitudes, all field lines are open. This region is called the polar cap.

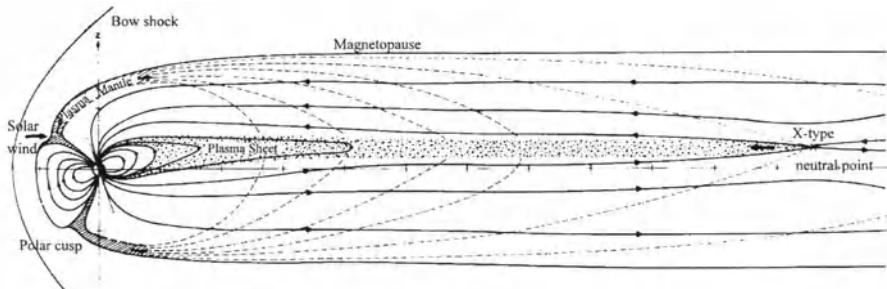


Fig. 10.9. Sketch of the magnetosphere and the magnetotail, drawn to scale. The solid lines are magnetic field lines, and the dashed lines give the trajectories of plasma parcels, filling the plasma sheet from the mantle. X-points inside the plasma sheet are favorable positions for reconnection. Reprinted from W.G. Pilip and G. Morfill [10.40], *J. Geophys. Res.* **83**, Copyright 1978, American Geophysical Union

Figure 10.9 shows a sketch of the magnetosphere, drawn to scale to visualize the extent of the magnetotail. The solid lines give the field lines, and the dashed lines are the trajectories of plasma particles. Two important features are obvious: first, plasma is convected from the plasma mantle at the polar cusps towards the plasma sheet in the magnetotail. Thus a continuous flow of hot solar wind plasma fills part of the magnetosphere. Second, in the equatorial plane there is a plasma sheet separating the oppositely directed magnetic fields of the north and south lobes (fields directed toward and away from the Earth). Inside this plasma sheet, neutral point configurations suitable for reconnection can form. One example is indicated as the X-point. Thus compared with the dayside magnetosphere, the magnetotail obviously has dynamic aspects: it is filled with plasma from the outside and occasionally reconnection in the neutral sheet leads to an ejection of this plasma partly towards the Earth, where it can create beautiful displays of aurora, and partly along the plasma sheet outwards back into the solar wind.

This configuration of the magnetotail requires currents: the Chapman–Ferraro current in the magnetopause, and the tail current inside the plasma sheet separating the north and south lobes. Both currents are perpendicular to the magnetic field lines and form the closed current system sketched in Fig. 10.10.

The radius of the magnetotail can be estimated by a simple approximation. All field lines connect back to the polar caps of the corresponding hemisphere. The magnetic flux leaving the polar caps is defined by the vertical component of the magnetic field integrated over the polar caps:

$$\Phi_{PC} = 2\pi(r_B \cos \theta_{PC})^2 B_o . \quad (10.16)$$

Here θ_{PC} is the latitude of the boundary of the polar cap and B_o is the equatorial magnetic field strength, which is about half the flux density at the polar cap. This flux is convected outwards into the magnetotail. If we assume

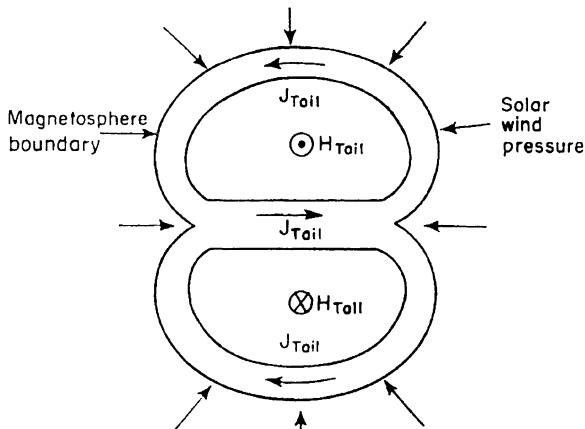


Fig. 10.10. Cross-section of the magnetotail with the currents inside the magnetopause and in the neutral sheet forming a closed loop. From J.W. Dungey [10.9], in *Solar-Terrestrial Physics*, Copyright 1967, with kind permission from Academic Press

the lobe to be semicircular with radius r_{Tail} and magnetic field strength B_{Tail} , the flux inside one lobe is $\Phi_{\text{Tail}} = \pi r_{\text{Tail}} B_{\text{Tail}}/2$. With (10.16) we get

$$\frac{r_{\text{Tail}}}{r_E} = \sqrt{\frac{4B_o}{B_{\text{Tail}}}} \cos \theta_{\text{PC}} . \quad (10.17)$$

With $\theta_{\text{PC}} = 75^\circ$ and $B_o = 31\,000$ nT, the radius of the magnetotail is $20\,r_E$ for $B_{\text{Tail}} = 20$ nT or $29\,r_E$ for $B_{\text{Tail}} = 10$ nT. The latter value is typical for the outer magnetosphere.

The current density inside the magnetopause can be determined from the momentum balance at the magnetopause: $\nabla p_{\text{sowi}} = \mathbf{j} \times \mathbf{B}/c$. On the other hand, the cross-tail current density can also be determined from Ampere's law to be $\delta B = 2B_{\text{Tail}} = 4\pi j/c$, with δB the jump in magnetic field strength across the current sheet. With $B_{\text{Tail}} = 20$ nT we find $j = 30$ mA/m for the cross-tail current and half this value for the magnetopause current.

10.2.4 Magnetosheath and Bow Shock

A prominent feature in front of the magnetopause is the bow shock where the supersonic solar wind is slowed down to subsonic speed. The bow shock is about 2 to 3 r_E ahead of the magnetopause, and its upstream medium is characterized by turbulence and energetic particles, cf. Sect. 9.4.

In the magnetosheath, the region between the bow shock and the magnetopause, the solar wind plasma is deflected and slowed down. Kinetic flow energy is converted into thermal energy, heating the plasma to about 5 to 10 times the solar wind temperature. Spreiter et al. [10.47] used a gas-dynamic model to describe the plasma flow inside the magnetosheath. The magnetic field is considered only in so far as it is convected by the solar wind; however, it does not modify the dynamics of the process. Two sample solutions for the geometry symmetric around the Sun-Earth line are shown in Fig. 10.11. In the left panel, the stream lines give the deflection of the plasma flow around

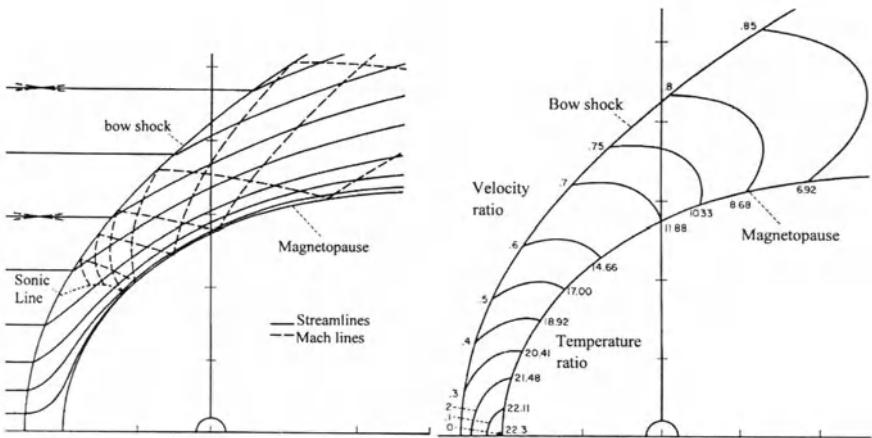


Fig. 10.11. Plasma flow inside the magnetosheath. (*Left*) Streamlines are shown which also are magnetic field lines. The dashed curve marks the transition from supersonic to subsonic flow. (*Right*) Velocity and temperature ratios. Reprinted from J.R. Spreiter et al. [10.47], *Planet. Space Sci.* **14**, Copyright 1966, with kind permission from Elsevier Science Ltd

the magnetopause and, as the magnetic field is convected with the plasma, also the magnetic field configuration in the magnetosheath. The dashed line marks the transition from supersonic to subsonic flow. The right panel shows the velocity and temperature ratios. Their contours are identical, as can be seen from integration of the energy equation (cf. [10.47]) which gives

$$\frac{T}{T_\infty} = 1 + \frac{(\gamma_{\text{ad}} - 1) M_\infty^2}{2} \left(1 - \frac{u^2}{u_\infty^2} \right). \quad (10.18)$$

The increase in plasma temperature can be quite substantial: on the dayside, the solar wind temperature can increase by up to a factor of 20. Since it is the sum of both electron temperature and ion temperature, with the electron temperature often about twice as high as the ion temperature, and since the electron temperature does not change significantly, the increase in ion temperature can be much larger than indicated in Fig. 10.11.

The solar wind flow passes through the bow shock but does not penetrate the magnetopause. Thus the position of the bow shock must be adjusted so as to allow the solar wind to flow around the obstacle magnetopause. At the subsolar point, observational evidence suggests that the ratio between the position of the bow shock and the stand-off distance of the magnetopause is $1.1n$, with n being the density jump at the bow shock. If we assume a gas-dynamic shock, the density jump depends on the Mach number M and on γ_{ad} as $n = [(\gamma_{\text{ad}} - 1) M^2 + 2]/(\gamma_{\text{ad}} + 1) M$. With $M = 8$ and $\gamma_{\text{ad}} = 5/3$, the bow shock is 29% farther out than the stand-off distance of the magnetopause.

In particular, close to the subsolar point the observations are in quite good agreement with these theoretical predictions, as can be seen in Fig. 10.12.

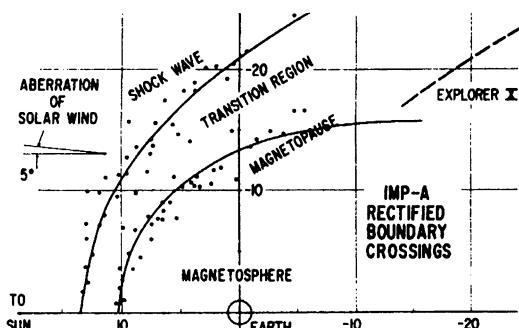


Fig. 10.12. Observed and calculated positions of the bow shock and the magnetopause in the equatorial plane. Observations and calculations are in good agreement; the scatter in the symbols indicates the variability of the magnetopause and bow shock due to solar wind variations. Reprinted from N.F. Ness et al. [10.36], *J. Geophys. Res.* 69, Copyright 1964, American Geophysical Union

Here the positions of the bow shock and magnetopause in the equatorial plane are shown. The solid lines give the calculated magnetopause and bow shock positions for average solar wind conditions, and the symbols give observed distances. The scatter in the symbols reflects the variability in solar wind conditions.

10.2.5 Open or Closed?

We have defined the magnetopause as a boundary separating the interplanetary and the planetary magnetic fields. Thus we have assumed the magnetosphere to be a closed system. For the dayside magnetosphere, this is a reasonable approach, although the flux-transfer events show that plasma can be exchanged between interplanetary space and the magnetosphere. In the nightside magnetosphere, the situation is more complicated. Here we do not even know the magnetic field structure: do the tail field lines close far away from the Earth or do they connect, at least partly, to the interplanetary magnetic field? Figure 10.13 sketches such an open magnetosphere: close to the Earth, three types of magnetic field lines can be observed: (i) an interplanetary magnetic field line of solar origin passing by, (ii) closed dipole field lines of planetary origin, and (iii) a merged planetary and interplanetary magnetic field line which connects the surface of the Earth magnetically to the interplanetary medium. The dashed lines give neutral sheets where the interplanetary magnetic field vanishes.

Whether a magnetosphere is open or closed is determined by the north-south component of the interplanetary magnetic field. In open and closed magnetospheres, different processes dominate the transport of mass, momentum, and energy across the magnetopause. As a consequence, the internal dynamics and the origin of the plasma populations are different.

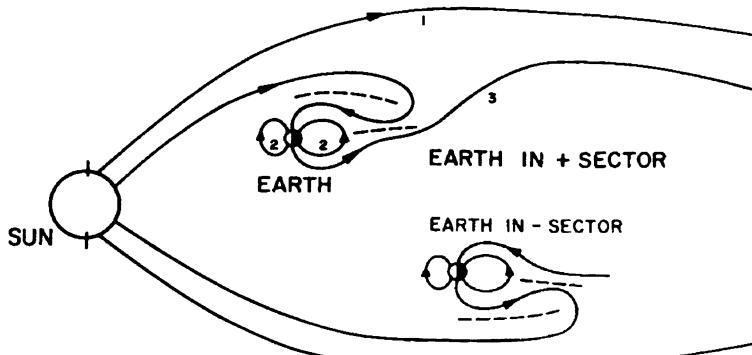


Fig. 10.13. Sketch of an open magnetosphere for two different polarities of the interplanetary magnetic field. Based on K.A. Anderson and R.B. Lin [10.1], *J. Geophys. Res.* **74**, Copyright 1969, American Geophysical Union

10.2.6 Magnetospheric Modeling

So far we have discussed some features of the magnetosphere in rather simple terms which allow the application of concepts introduced in earlier sections. A complete description of the magnetospheric fields and plasmas, however, requires a self-consistent approach. Let us illustrate some neglected effects in the example of the magnetopause. For instance, we have described it as a discontinuity in the magnetic field without adjusting the current systems accordingly. In addition, we have ignored the polarization field inside the current sheet, resulting from the different Larmor radii of solar wind electrons and ions but also influencing the motion of the very same particles, such as is sketched in Fig. 10.8. A self-consistent approach, however, would have to consider quasi-neutrality and the screening of charges in the field of other charges. The characteristic length of the particle motion in the magnetopause then would be the inertial length c/ω_{ce} of an electron instead of the gyro-radius. In the calculation of the magnetopause we have also ignored the magnetosheath, where the field and plasma properties are quite different from the ones in the solar wind, although both field and plasma are of solar wind origin. Putting all these effects together, the thickness of the magnetopause would be some hundred to about 1000 km, much more than the ion gyro-radius of about 100 km but much closer to the observations.

All of these effects can be described analytically; the resulting set of equations, however, can be solved by means of MHD simulations only. In particular, the dynamic aspects such as reconnection of planetary and interplanetary fields and the accompanying plasma exchange can be incorporated into such models. Overviews of magnetospheric modeling can be found in [10.23, 10.52] and some articles in [10.37]. The modeling of the magnetospheric plasma is discussed in [10.35], and models of the radiation belts can be found in [10.28].

10.3 Plasmas and Currents in the Magnetosphere

The magnetospheric plasma has distinct properties in three different regions: the lowest layer is the plasmasphere, filled with a plasma of ionospheric origin. In contrast, the outer heliosphere is filled with solar wind plasma. In between, the geosphere is filled with a highly variable mixture of ionospheric and solar wind plasma. These different regions are coupled by currents.

10.3.1 The Ionosphere

The ionosphere starts in a height of about 70 km as a charged-particle component inside the atmosphere. The ionosphere often is described as the base of the magnetosphere. Because of its high density and the existence of a large neutral component it does not obey the definition of a magnetosphere, namely that particle motion is determined by the magnetic field only. Nonetheless, it is vital for the understanding of the magnetosphere because it provides a highly conducting bottom layer and ionosphere–magnetosphere coupling is important for the energetics of the magnetosphere.

The ionosphere is formed by the ionization of atmospheric constituents by hard electromagnetic radiation in the UV and EUV range. The ionosphere is conveniently described as consisting of different layers. To derive the height profile of such a layer, the Chapman profile, we can use a simplified model considering only one atomic species and monochromatic electromagnetic radiation. The variation of density n with height z is described by the barometric height formula $n(z) = n_0 \exp\{-z/H\}$, with H being the scale height. Thus the intensity decreases exponentially with increasing height. The intensity I of the ionizing electromagnetic radiation, on the other hand, increases with increasing height: it is maximal at the top of the atmosphere and then is absorbed according to Bougert–Lambert–Beer’s law $dI/dz = -I_\infty \sigma_a n$, with σ_a being the absorption cross-section of the particle species and frequency

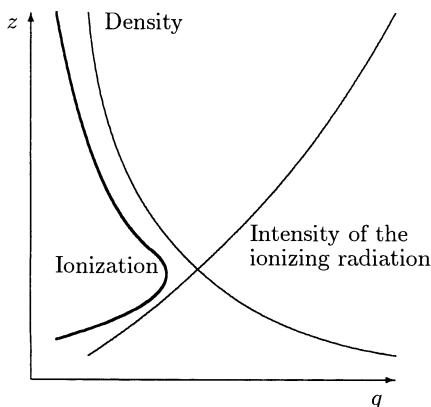


Fig. 10.14. The Chapman profile of an ionospheric layer results from the superimposition of the height dependences of the particle density and the flux of the ionizing electromagnetic radiation

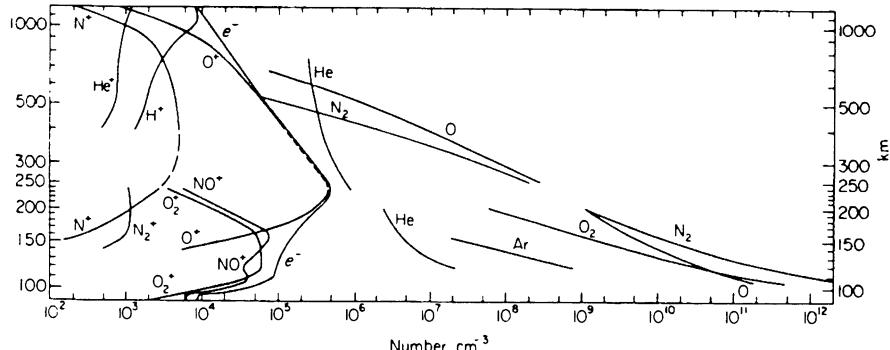


Fig. 10.15. Height dependence of different constituents of the ionosphere and atmosphere for quiet solar conditions. Reprinted from C.Y. Johnson [10.22], in *Annals of the international years of the quiet Sun*, Copyright 1969, with kind permission from MIT Press

range under consideration. The electromagnetic radiation therefore leads to a height-dependent ionization rate $q(z) = n\sigma_i I(z)$, with σ_i being the ionization cross-section. The combination of these two profiles gives the Chapman profile, cf. Fig. 10.14: at a certain height, the ionization, and therefore also the charge density, is highest. Below, it decreases as the intensity of the ionizing radiation decreases. At higher altitudes, although the intensity of the ionizing radiation is higher, the charge density decreases, too, because the density of particles available for ionization is lower. Formally, the charge density in the Chapman layer can be written as

$$q(z) = \sigma_i n_0 I_\infty \exp \left\{ - \int_z^\infty \sigma_a n(z) dz / \cos \theta - \frac{z}{H} \right\}, \quad (10.19)$$

with θ as the Sun's altitude. If the Sun is in the zenith, the charge density is largest and the maximum of the Chapman layer is at lower altitudes. With decreasing solar altitude, the Chapman layer shrinks and its maximum shifts to higher altitudes.

Since the atmosphere consists of different particle species and the incoming radiation covers a broad spectrum, for each particle species such a layer forms in the ionosphere at its typical height. Figure 10.15 summarizes these layers: on the right, the densities of the neutrals are shown, and on the left, the ionospheric layers are depicted for the different ion species as well as for the electrons. The electron distribution is the sum of the different ion layers. Note that with increasing height the relative importance of the ionized component increases: while at a height of about 100 km only 10^{-7} of the atoms and molecules are ionized, above a height of about 800 km only ionized particles exist.

The profiles of the different Chapman layers in Fig. 10.15 are shown for quiet solar conditions. Thus the hard electromagnetic radiation is at a rather low level and no additional radiation is emitted in flares. During solar maximum, the intensity of the hard electromagnetic radiation is higher, leading to a stronger ionization. In addition, during solar flares the hard electromagnetic radiation can be enhanced even further. Then the charge density in the ionosphere can become so large that electromagnetic waves are no longer reflected but are absorbed, leading to a break-down in long-wave communication. Such an event is called a sudden ionospheric disturbance (SID).

Currents and plasma flows couple the ionosphere and the magnetosphere. The ionosphere differs from the magnetosphere in so far as collisions of charged particles with the neutrals of the atmosphere occur frequently; the ionosphere therefore is characterized by a collision-dominated plasma. Thus the conductivity is finite and the frozen-in flux approximation is no longer valid. But the conductivity is not only finite but also highly anisotropic. Three typical conductivities can be defined. The field-aligned conductivity parallel to \mathbf{B} depends on the masses m_e and m_i of the electrons and ions and on their collision frequencies ν_e and ν_i :

$$\sigma_{\parallel} = \left(\frac{1}{m_i \nu_i} + \frac{1}{m_e \nu_e} \right) n e^2. \quad (10.20)$$

The Pederson conductivity is concerned with currents parallel to the electric field. With ω_i as the gyro-frequencies we have

$$\sigma_{\text{Ped}} = \left[\frac{\nu_i}{m_i(\nu_i^2 + \omega_i^2)} + \frac{\nu_e}{m_e(\nu_e^2 + \omega_e^2)} \right] n e^2. \quad (10.21)$$

Pederson currents dissipate energy since $\mathbf{E} \cdot \mathbf{j} > 0$. The Hall conductivity is concerned with currents perpendicular to both the electric and the magnetic fields. It is free of dissipation and can be written as

$$\sigma_{\text{Hall}} = \left[\frac{\omega_i}{m_i(\nu_i^2 + \omega_i^2)} + \frac{\omega_e}{m_e(\nu_e^2 + \omega_e^2)} \right] n e^2. \quad (10.22)$$

The total current in the ionosphere therefore can be written as

$$\mathbf{j} = \sigma_{\parallel} \mathbf{E}_{\parallel} + \sigma_{\text{Ped}} \mathbf{E} + \sigma_{\text{Hall}} \frac{\mathbf{B} \times \mathbf{E}}{B}. \quad (10.23)$$

Of these conductivities, the field-aligned one generally is the largest. To maintain current continuity, the electric field component \mathbf{E}_{\parallel} parallel to the magnetic field has to be very small. In particular, at high latitudes the magnetic field is almost perpendicular to an ionospheric layer and therefore allows for an efficient current between the lower ionosphere and higher altitudes.

10.3.2 The Plasmasphere

The plasmasphere is dominated by a dense and cold plasma of ionospheric origin, as is evident from the high O^+/H^+ ratio and the existence of other ion

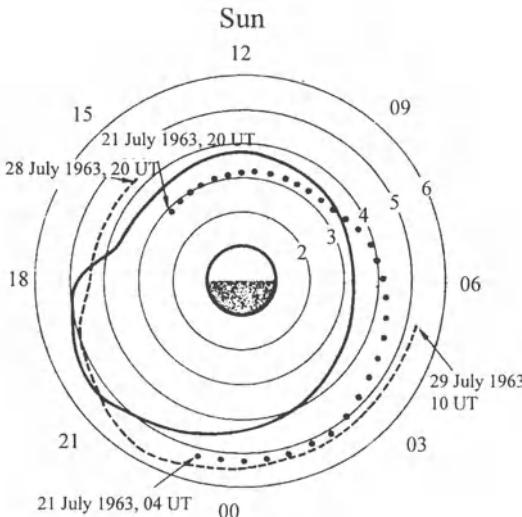


Fig. 10.16. Cross-section through the plasmasphere in the equatorial plane. The location of the plasmapause, as determined from Whistler observations, is indicated for different times corresponding to different levels of geomagnetic activity. Reprinted from D.L. Carpenter [10.4], *J. Geophys. Res.* **71**, Copyright 1966, American Geophysical Union

species such as He^+ , O^{2+} , N^+ , and N^{2+} which cannot be found in the solar wind. Spatially, it coexists with the radiation belts, extending up to heights of about 3 to 5 Earth radii. The particles have energies close to 1 eV, and the density varies between 10^4 cm^{-3} at about 1000 km and $10\text{--}100 \text{ cm}^{-3}$ at the outer boundary of the plasmasphere. The plasmasphere is filled from the ionosphere by the polar wind.

The plasmapause as the relatively sharp outer boundary of the plasmasphere was first proposed from the properties of Whistler waves in the magnetosphere. Figure 10.16 shows the location of the plasmapause in the equatorial plane for three different times corresponding to different levels of geomagnetic activity. The bulge at the duskside is a persistent feature, although it can rotate somewhat in local time, depending on magnetospheric conditions. The plasmasphere – and with it the plasmapause – basically is a field-aligned structure. It can be traced from the equatorial plane down to the ionosphere.

Figure 10.17 shows the plasma density plotted versus the L-shell parameter for the nightside plasmasphere for different levels of geomagnetic activity. With increasing geomagnetic activity, the plasmasphere shrinks and its outer boundary becomes more pronounced.

The plasmasphere corotates with the Earth, leading to an electric induction field $\mathbf{E}_{\text{corot}} = -(\omega \times \mathbf{r}) \times \mathbf{B}(r)$, with ω being the rotation frequency and $\mathbf{B}(r)$ the magnetic flux density at position \mathbf{r} . This field is smaller than the electric field driving the current system in the magnetotail; the plasmapause as the outer boundary of the plasmasphere separates these two current systems.

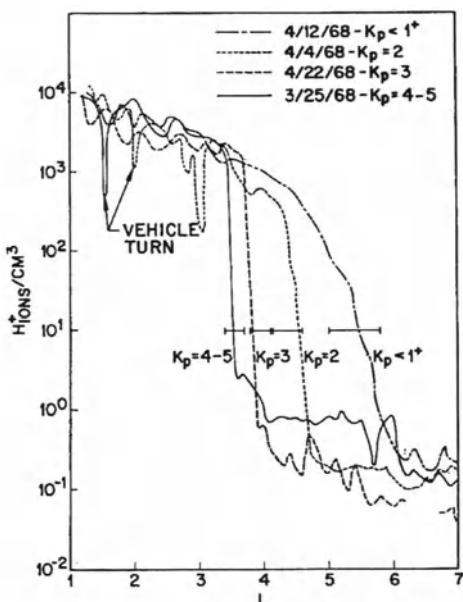


Fig. 10.17. Variation of the nightside plasmapause with geomagnetic activity. Reprinted from C.R. Chappel et al. [10.5], *J. Geophys. Res.* **75**, Copyright 1970, American Geophysical Union

10.3.3 The Geosphere

The plasmasphere is embedded in the geosphere, a highly variable region filled with a hot plasma of low density. The plasma inside the geosphere has two sources, the ionosphere and the solar wind, as is evident from the composition. Figure 10.18 shows the size and location of the geosphere together with the three main current systems. In the upper panel, the familiar meridional cross-section is given. The middle panel corresponds to a view from high above the North Pole towards the equatorial plane, and in the lower panel two cross-sections with views from the subsolar point tailwards are shown. The shading indicates the plasma density, the arrows indicate the Chapman-Ferraro current in the magnetopause, the tail current across the tail in the equatorial plane, and the ring current around the Earth associated with the radiation belt particles.

The plasma density in the geosphere is much lower than in the plasmasphere; it is also lower than in the outer magnetosphere. Evidence for a contribution of the ionospheric plasma to the geosphere again comes from the presence of O^+ and other heavier ions. The dominant species, H^+ and He^{2+} , in number are roughly independent of solar and geomagnetic activity, while their energy as well as the relative amount of ionospheric ions dramatically increases with increasing geomagnetic activity. Thus the solar wind ions H^+ and He^{2+} are fed into the geosphere roughly independent of activity while geomagnetic activity strongly enhances the energy imparted to these particles as well as the density of the ionospheric component. In addition, the

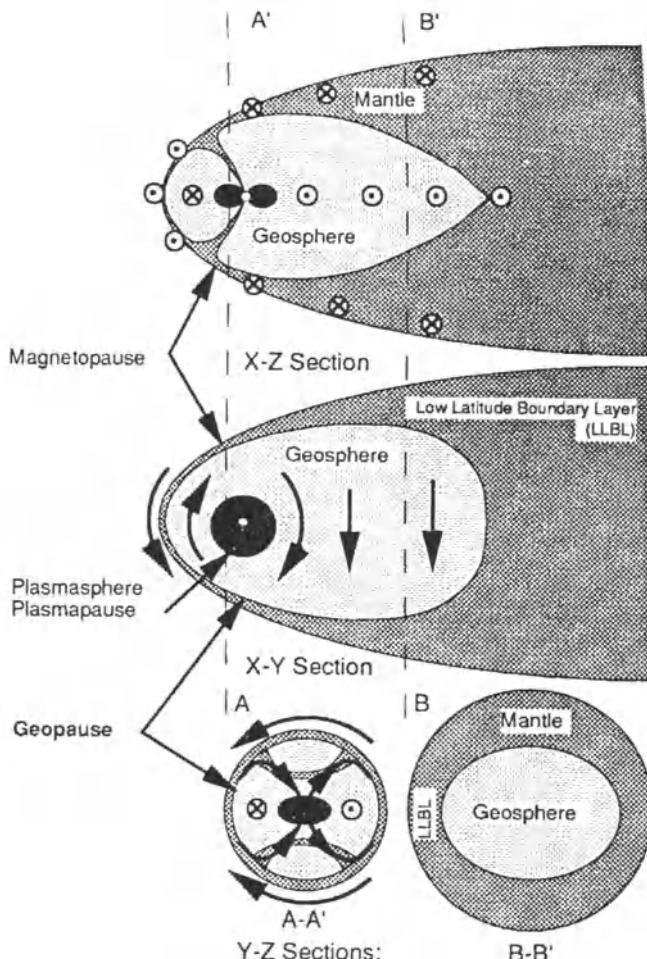


Fig. 10.18. Plasma regimes in the magnetosphere with the darkness of the shading indicating the plasma density. Views from the duskside to the dawn (*upper panel*), from pole to equator (*middle panel*), and from the subsolar point tailwards (*lower panel*). The arrows give currents. Reprinted from T.E. Moore and D.C. Delcourt [10.34], *Rev. Geophys.* **33**, Copyright 1995, American Geophysical Union

ionospheric outflow into the geosphere tends to be larger by up to a factor of 4 if the interplanetary magnetic field has a northward component.

10.3.4 The Outer Magnetosphere

The outer magnetosphere is dominated by solar wind plasma. The basic mechanism for feeding plasma into the magnetosphere is reconnection. The magnetopause therefore is not a simple boundary separating two completely

decoupled systems. Instead, plasma transfer takes place almost everywhere along the magnetopause. Since the field-lines of the magnetopause converge at the polar cusps, the high-latitude ionosphere is directly influenced by the solar wind plasma penetrating through the magnetopause.

Even the simplest models of the magnetosphere had predicted an direct access of solar wind at the polar cusps. But the plasma in some properties, in particular in thermal energy, is different from the solar wind plasma: because it comes from the magnetosheath, the temperature is higher than in the solar wind because the latter had been slowed down at the bow shock. This downward plasma flow can be detected at latitudes of about 78° for about ± 3 h around local noon, i.e. exactly below the cusps, as an increase in electron density and temperature. While the electrons form a narrow beam penetrating downwards through the cusps, the ions are spread back towards the tail. This is caused by a $\mathbf{E} \times \mathbf{B}$ drift in the dawn-to-dusk electric field. Since the electrons have a very short travel time, their displacement by this drift is negligible. The ions, on the other hand, are much slower and therefore are affected by the drift. Their displacement increases with increasing travel time, i.e. with decreasing energy.

Plasma transfer across the magnetopause due to reconnection requires an X-point configuration or neutral line where fields of opposite polarity meet. This is most likely to occur at the dayside magnetopause if the interplanetary magnetic field has a southward component. Then the magnetopause is a rotational discontinuity while at times of a northward interplanetary magnetic field it is a tangential discontinuity completely separating the interplanetary and the planetary plasmas. The corresponding configurations are an open and a closed magnetosphere.

Figure 10.19 sketches the consequences of such a process: the interplanetary magnetic field line $1'$ reconnects with the planetary field line 1 . Thus two mixed planetary/interplanetary field lines result (2 and $2'$) which are convected tailwards with the solar wind, eventually becoming field lines of the geomagnetic tail (5 and $5'$). During this process, plasma from the magnetosheath is convected into the tail, first filling the mantle and later also moving down to lower latitudes, as indicated by the dashed lines in Fig. 10.9. If this process were to continue indefinitely, the entire geomagnetic field would soon be connected with the interplanetary field. Thus the magnetic flux must be returned to the closed magnetic field. This is archived on the nightside magnetosphere: as the mixed field lines are pushed further towards the equatorial plane, a new X-point results where lines 6 and $6'$ meet. Here reconnection sets in, forming a closed geomagnetic field line and a purely interplanetary magnetic field line. Magnetic tension will relax the geomagnetic field lines ($7, 8$), which then will return to the dayside magnetosphere at lower altitudes. Magnetic tension also allows the interplanetary field line $7'$ to shorten and therefore to be pulled outwards through the geomagnetic tail.

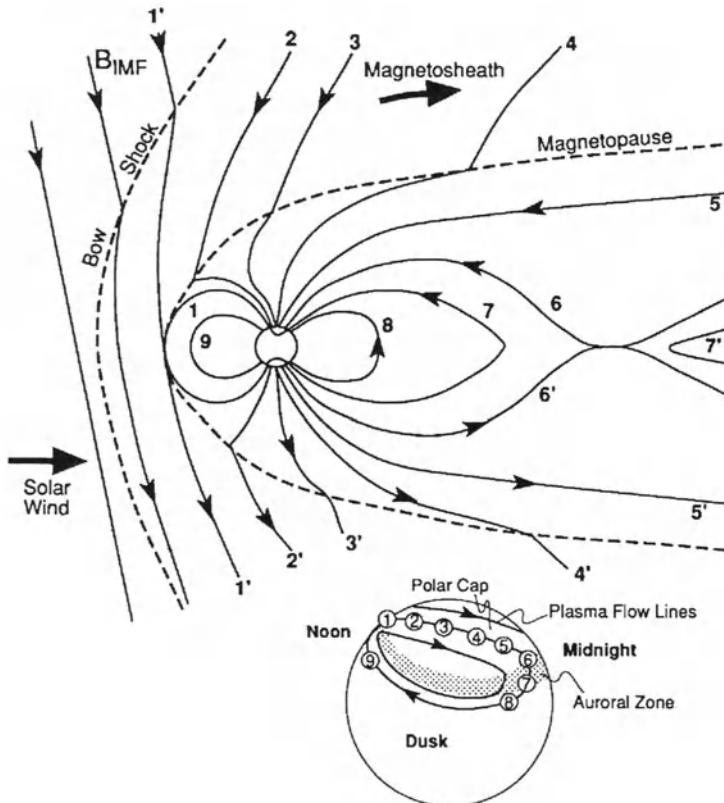


Fig. 10.19. The convection of plasma into the magnetosphere. The numbered field lines show a succession of configurations of an interplanetary magnetic field line 1' at the front of the magnetosphere. Field lines 6 and 6' reconnect in the tail and return to the dayside at lower latitudes. The small figure shows the resulting convection pattern of the plasma in the northern high-latitude ionosphere: an anti-sunward flow in the polar cap and a return flow at lower latitudes. The shaded area is the auroral oval. Reprinted from W.J. Hughes [10.15], in *Introduction to Space Physics*, Copyright 1995, with kind permission from Cambridge University Press

This picture is grossly simplified since in reality the entire process will be essentially non-steady. Thus although the time-averaged reconnection rates at the dayside magnetopause and in the tail must be equal, at any given time they can be quite different. Nonetheless, there is observational evidence for such a process to occur. In the late 1950s, it was realized that the plasma flow in the polar and auroral ionospheres must map outward and be related to a magnetospheric flow pattern. Magnetometer measurements indicated a plasma flow over the polar regions from noon to midnight (points 1–6 in Fig. 10.19) and a flow back towards the dayside at lower latitudes (points 7–9). This pattern is roughly stationary in local time, i.e. in a frame of reference

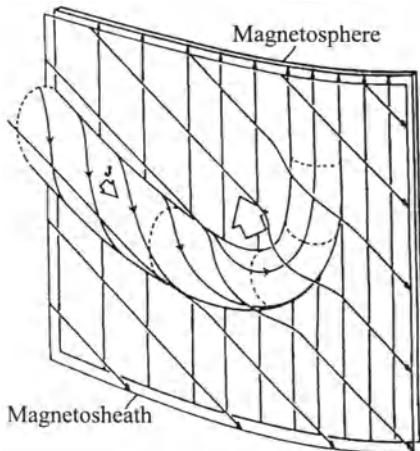


Fig. 10.20. Sketch of a magnetic flux tube through the magnetopause. The motion of such a flux tube leads to the signature of a flux-transfer event in the magnetic field data. Reprinted from C.T. Russell and R.C. Elphic [10.43], *Geophys. Res. Lett.* **6**, Copyright 1979, American Geophysical Union

with a fixed Sun–Earth line. An observer on Earth rotates underneath this flow pattern, seeing it as a diurnal magnetic field variation. Since the flow pattern resembles a thermally driven flow cell, it has been termed a convection pattern, although it is not thermally driven. The lower panel in Fig. 10.19 shows this flow pattern on the duskside. Since the magnetic field is frozen into the plasma, we can map back the plasma flow to a pattern of motion of a magnetospheric field line which is swept across the polar cap and then returns to the dayside magnetosphere at lower latitudes, which is exactly the motion of the field line shown in the upper panel of Fig. 10.19. Historically, the convective flow pattern is not used to support the idea of a continuous reconnection between the planetary and the interplanetary magnetic fields, but it led Dungey in 1961 [10.8] to propose such an open magnetosphere with a continuous flux and field exchange to explain it.

More recent evidence for reconnection between the planetary and the interplanetary magnetic fields stems from *in situ* observations at the dayside magnetopause. These observations are not related to the convection pattern sketched above, but they confirm what has been the primary concern of this section, namely the exchange of plasma and field between the magnetosphere and the interplanetary medium. As a satellite passes through the magnetopause, short events in the magnetic field data, plasma data, and particle data provide evidence for such flux exchange, and the events therefore are termed flux-transfer events. The earliest observations were based on magnetic field data only. A flux-transfer event can be identified as two consecutive short excursions of the magnetic field component normal to the magnetopause: first to positive values, later to negative ones, and afterwards returning to the undisturbed value around zero if the satellite is in the northern hemisphere or in the opposite sequence for an observer in the southern hemisphere. These signatures are interpreted as an isolated magnetic flux

Solar Wind–Magnetosphere Coupling

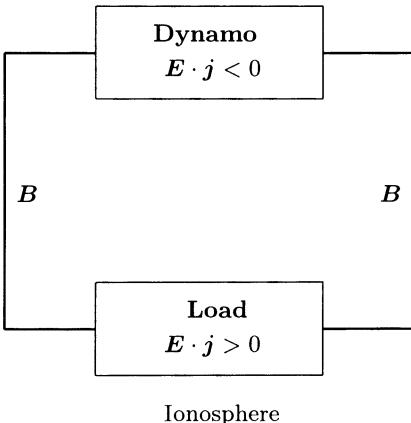


Fig. 10.21. Magnetosphere–ionosphere coupling as a closed circuit

tube connected through the magnetopause to the geomagnetic field and convected northwards across the satellite by the solar wind, cf. Fig. 10.20. The composition of the plasma as well as the energetic particle inside the flux tube were both indicative of a magnetospheric origin. Since reconnection at the dayside magnetosphere causes the magnetic field configuration visible in the flux-transfer events, these events mainly are observed at times when the interplanetary magnetic field has a southward component.

10.3.5 Magnetosphere–Ionosphere Coupling

The motion of particles, plasmas, and magnetic fields gives rise to currents. Currents in the magnetosphere associated with its large-scale structure are the Chapman–Ferraro current inside the magnetopause and the tail current separating the southern and northern lobes. In addition, the drift of charged particles trapped in the radiation belts gives rise to a ring current. These currents are perpendicular to the magnetic field. Another kind of current flows parallel to the magnetic field and therefore can provide coupling between the ionosphere and the magnetosphere. Since the magnetic field is perpendicular to the ionospheric layer only at high latitudes, these field-parallel currents are a phenomenon typical of the polar ionosphere and magnetosphere.

The entire configuration then can be interpreted as a circuit with the solar wind–magnetosphere interaction working as a dynamo ($E \cdot j > 0$) and the ionosphere being a load with dissipative losses ($E \cdot j < 0$). The circuit then is closed by currents parallel to B , cf. Fig. 10.21. The driving force, and thus also the source of energy, is the solar wind. Its interaction with the terrestrial magnetic field gives rise to the Chapman–Ferraro current inside the magnetopause, the dawn-to-dusk electric field, and the tail current. These currents do not affect the terrestrial magnetic field, and so their fluctuations are not recorded by ground-based magnetometers.

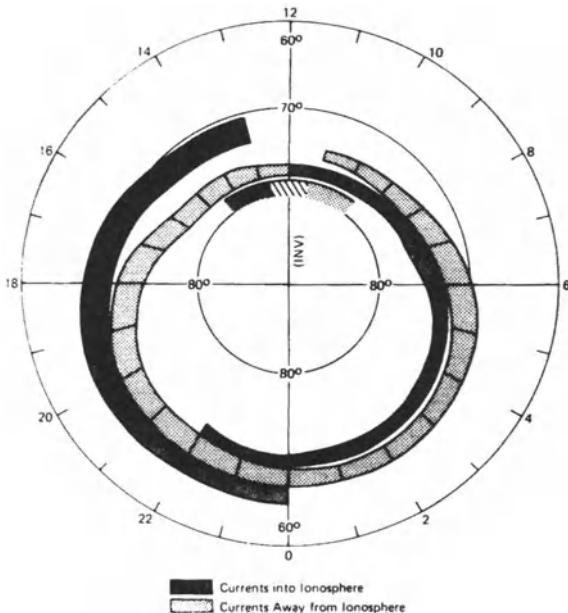


Fig. 10.22. Distribution of the field-parallel Birkeland currents in the polar ionosphere. Note that the numbers refer to the geomagnetic latitude. Reprinted from T. Iijima and T.A. Potemra [10.17], *J. Geophys. Res.* **81**, Copyright 1976, American Geophysical Union

The field-parallel currents can be observed in the polar ionosphere where the magnetic field lines are almost perpendicular. The average patterns of these Birkeland currents are shown in Fig. 10.22. The currents are plotted versus geomagnetic longitude with the dark areas indicating currents into the ionosphere and the lighter shaded areas indicating currents out of it. At high altitudes, currents are flowing out of the ionosphere in the evening side and into it at the morning side. They are called region 1 currents. At somewhat lower latitudes, the region 2 currents show the opposite pattern: they flow into the ionosphere in the evening side and out of it in the morning side. At very high latitudes around noon, i.e. below the polar cusps, the pattern of the field-parallel current is highly variable and strongly depends on the northwards or southwards component of the interplanetary magnetic field. Around midnight, the currents in regions 1 and 2 overlap without a clear separation.

Most of the field-parallel currents are carried by the electrons. Thus an inward current implies an outward motion of electrons and vice versa. The pattern of Birkeland currents also reflects the spatial distribution of aurorae at geomagnetic quiet periods: aurorae are observed where electrons stream down to the ionosphere, i.e. where the Birkeland current is directed upwards.

The upward and downward Birkeland currents are closed by ionospheric currents. The auroral zone electric field \mathbf{E}_a in the ionosphere is directed northwards in the dusk sector and southwards in the dawn sector, i.e. from dusk to dawn. Since the conductivity of the ionosphere is finite, the Birkeland

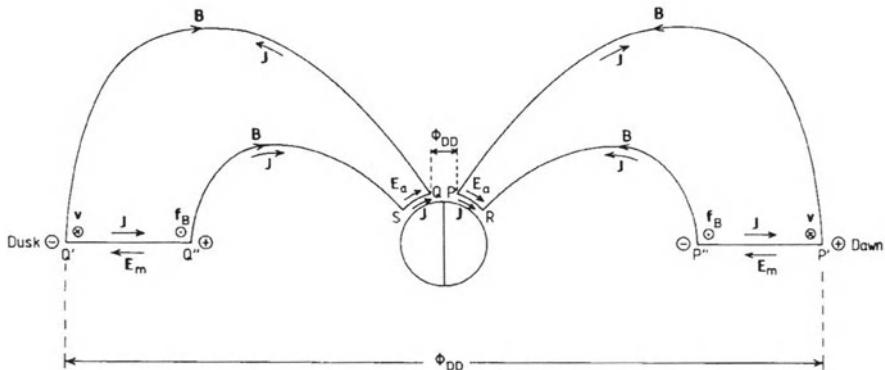


Fig. 10.23. Possible scenario for the closure of the ionosphere-magnetosphere-current system in a cross-section in the dusk-to-dawn plane viewed from the tail towards the Sun. Reprinted from A. Brekke [10.2], *Physics of the upper polar atmosphere*, Copyright 1997, with kind permission from Wiley & Sons, Inc.

currents will be closed in the ionosphere by field-parallel currents (Pederson currents) northwards in the dusk sector and southwards in the dawn sector. A possible closure of the circuit is sketched in Fig. 10.23. Here a dusk-to-dawn cross-section of the magnetosphere is shown viewed from the tail towards the Sun. The closure of the ionosphere-magnetosphere circuit in the equatorial plane is suggested to be radial in the dusk-to-dawn direction. Note that the dusk-to-dawn auroral zone electric field is reversed if mapped outward, in agreement with the dawn-to-dusk electric field driven by the solar wind and also with the cross-tail electric field. Thus in the magnetosphere, the

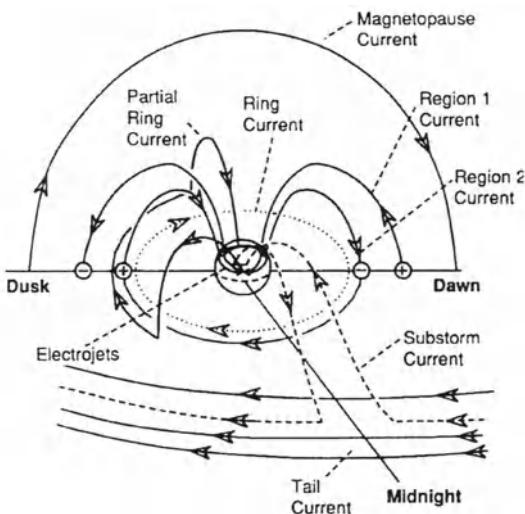


Fig. 10.24. Schematic representation of the various current systems in the magnetosphere and ionosphere. Reprinted from R.L. McPheron [10.30], in *Introduction to space Physics* (eds. M.G. Kivelson and C.T. Russell), Copyright 1995, with kind permission from Cambridge University Press

electric field and currents are antiparallel, forming the generator proposed in Fig. 10.21, while they are parallel in the ionospheric load.

Part of the ring current is also involved in the magnetosphere–ionosphere coupling. It flows near dusk in the equatorial magnetosphere and is closed through the ionosphere by field-parallel currents. And part of the tail current is diverted into the ionosphere by field-aligned currents, forming the substorm current. The substorm current, as its name suggests, plays an important role in geomagnetic and auroral activity. These currents, together with the current systems discussed so far, are summarized in Fig. 10.24.

10.4 Particles in the Magnetosphere

Particle populations in the magnetosphere have different sources and properties. A simple distinction is based on rigidity. Particles with high rigidity are able to traverse the magnetosphere, and thus no long-lived trapped particle components with high rigidities exists. High-rigidity particles coming from the outside, such as galactic cosmic rays, depending on their direction of incidence, penetrate deep into the magnetosphere and interact with the upper atmosphere or are deflected back into space. For low-rigidity particles, three cases can be distinguished, cf. the right-hand side in Fig. 10.25. Particles hitting the low-latitude magnetosphere from the outside perform half a gyro-orbit inside the magnetosphere and then are reflected back into space. Only at the polar cusps can these particles penetrate into the magnetosphere, and on interaction with the atmosphere produce polar cap absorption (PCA) events. The third low-rigidity particle component is different from all particle populations discussed above in so far as it is not a transient but a long-lived component: particles are trapped inside the radiation belts. Their motion is regulated by the adiabatic invariants; nonetheless, radiation belts are not a static phenomenon but a dynamic one with sources and losses depending on the other particle populations and on geomagnetic activity.

10.4.1 The Radiation Belts

The discovery of the radiation belts was the first significant scientific result of space research with satellites. The first observations were made with a Geiger counter on board Explorer 1, launched on 31 January 1958. The discovery of the radiation belt was accidental. The Geiger counter had been designed by a group around J. van Allen from the University of Iowa to measure cosmic rays in the high atmosphere and had been adjusted to accommodate the expected fluxes. During short parts of the satellite orbit, the observations met the expectations; for long times, however, the observed fluxes either were much too large or way too small, raising doubts about the performance of the instrument. But from the pattern of expected and unexpected counting rates,

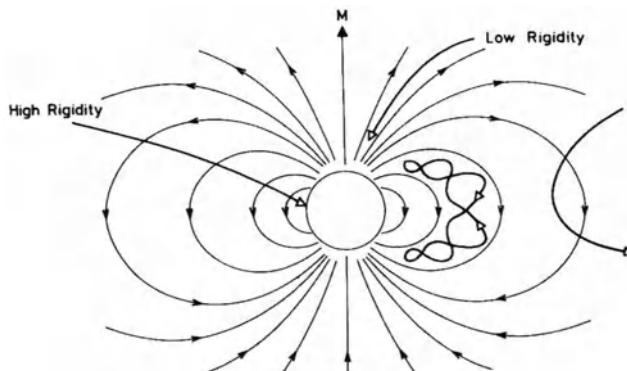


Fig. 10.25. Orbits of particles with low (right) and high (left) magnetic rigidity. Reprinted from D.A. Bryant [10.3], in *Plasma physics* (ed. R. Dendy), Copyright 1993, with kind permission from Cambridge University Press

it became evident that the unexpected signals always indicated particle fluxes much higher than expected, with the low counting rates being a saturation effect. Subsequent measurements with Explorer 3 and Sputnik 3, both in 1958, confirmed the existence of the radiation belts.

First Observations. The first measurements did not identify the particle species or energy. Electrons with energies between 50 keV and 150 keV were expected to be responsible for most of the counts. A strong dependence of counting rate on height was observed. Observations with Pioneer 3 in a highly elliptical orbit with an apogee at 107 400 km suggested a double structure with an inner radiation belt starting at about 400 km with a maximum at about $1.5 r_E$. The counting rates then decreased, but a second radiation belt was found between $3 r_E$ and $4 r_E$ with a maximum at about $3.5 r_E$. The depleted region between these two radiation belts is called the slot.

Subsequent measurements changed this rather simple picture. First, it was discovered that the trapped particles not only are electrons in the tens and hundreds of kiloelectronvolt range but also protons with energies up to more than 30 MeV. And second, the two distinct radiation belts are fictitious.

Properties and Orbits of Radiation Belt Particles. Figure 10.26 gives a more detailed description of the radiation belts. The upper panel is concerned with protons, the lower ones with electrons. In the upper panel, solid lines give the distribution of protons with energies greater than 30 MeV, and the dotted lines represent protons with energies between 0.1 MeV and 5 MeV. The high energetic protons dominate the inner part of the radiation belt: its lower edge is at about $1.15 r_E$. At lower altitudes the losses of particles due to interaction with the atmosphere are too large to allow for a stable trapped population. The maximum of the high energetic protons is at about $1.5 r_E$; with increasing height the density decreases. The picture is different for the

low energetic proton component. Here the fluxes are much higher and a broad maximum can be found around $3.5 r_E$.

If we consider the electrons as shown in the lower panel in Fig. 10.26, the pattern is slightly different. Low-energy electrons (dotted line, 0.04–1 MeV) and high-energy electrons (solid lines, energy above 1.6 MeV) have their maxima at similar positions, i.e. about $3.5 r_E$. Nonetheless, the high-energy component is confined to a smaller spatial region.

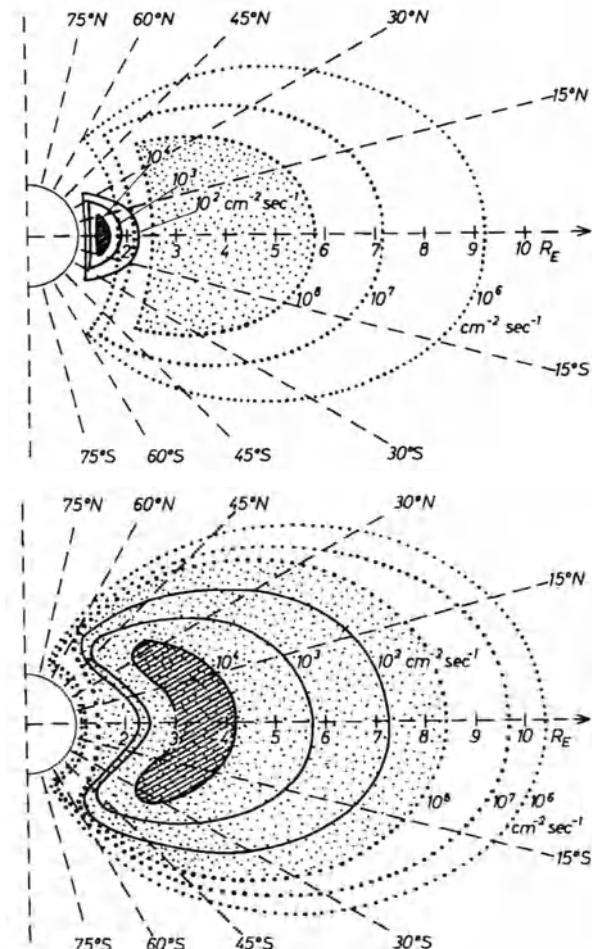


Fig. 10.26. Distribution of particles inside the radiation belts. In the upper panel, the solid lines give the distribution of protons with energies above 30 MeV, and the dotted lines represent protons with energies between 0.1 MeV and 5 MeV. In the lower panel, the solid lines give electrons with energies above 1.6 MeV, while the dotted lines correspond to energies between 0.04 MeV and 1 MeV. Reprinted from W. Kertz [10.24], *Einführung in die Geophysik*, Copyright 1971, with kind permission from Spektrum Akademischer Verlag

Note that in Fig. 10.26 the radiation belts are given in geomagnetic coordinates under the assumption of axisymmetry around the dipole axis. Since the dipole axis is offset with respect to the center of Earth, in spatial coordinates the radiation belts are asymmetric. In particular at the SAA off the coast of Brazil, the radiation belts can be found in rather low altitudes.

Although the upper panel of Fig. 10.26 still suggests a description in terms of two separate radiation belts, consideration of the intermediate energies not shown in the figure suggests a different picture. Today, we understand the radiation belt as a zone of trapped particles with the properties of the particles changing continuously with increasing distance from the Earth. The higher energies can be found predominately close to the Earth, thus in the inner part of the radiation belt the energy spectrum is rather hard while it steepens with increasing distance.

Nonetheless, the radiation belt often is divided into two zones, an inner one with $L < 2$ and an outer one with $L > 2$. This distinction is not so much concerned with the properties of the particles as with stability: in the inner zone, the particle populations are very stable and long-lived, while the particle populations in the outer zone vary with solar and geomagnetic activity.

The motion of the radiation belt particles under undisturbed conditions is described by the adiabatic invariants, cf. Sect. 2.4. The typical time-scales can be found in Fig. 10.27: the particles gyrate around their guiding field line with periods in the order of 10^{-5} s. They bounce back and forth along their guiding center field line with periods of about 1 s (north–south oscillation), and they drift around the Earth with a period of about 500 s, forming a ring current. Detailed discussions of particles in the radiation belts can be found in [10.28, 10.42, 10.44, 10.53].

The particles trapped inside the magnetosphere show a typical pitch-angle distribution, the loss-cone distribution: particles with small pitch angles are absent because during the north–south oscillation these particles are not mirrored back at high altitudes but penetrate deep enough into the atmosphere to be lost by interaction with atmospheric constituents.

Gains and Losses. The radiation belts would be stable were it not for two processes: first, magnetic field fluctuations scatter particles during their north–south oscillation into the loss cone, and second, during geomagnetic activity, the entire structure of the magnetosphere is distorted. Thus the adiabatic invariants can be violated and particles are fed into the loss cone. On the other hand, the radiation belts do exist, and thus there are sources replenishing them. Despite their structure and temporal variability, the average properties of the radiation belts are remarkably constant, thus an equilibrium between sources and sinks appears to exist.

The first indications for the lifetime of radiation belt particles came from atmospheric A-bomb tests. In the Starfish experiment in 1962, an artificial electron population had been injected into the radiation belt by an A-bomb explosion. After only about 10 years, this population had vanished into the

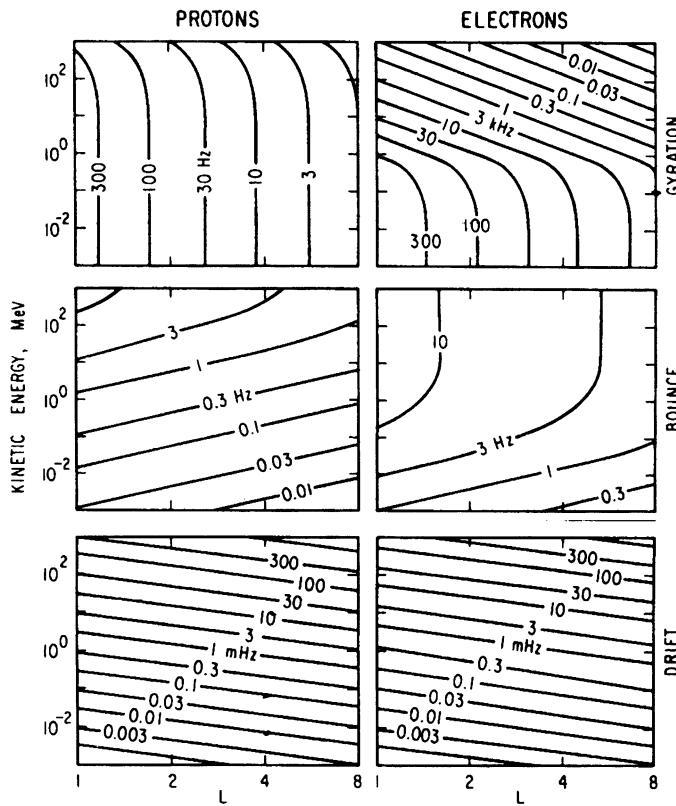


Fig. 10.27. Scales of particle motion in the Earth's magnetosphere. Reprinted from M. Schulz and L.J. Lanzerotti [10.44], *Particle diffusion in radiation belts*, Copyright 1974, Springer Verlag

background. However, the lifetime of trapped particles is not a universal constant, but rather depends on altitude and particle properties. The lifetime increases fast from the lower edge of the radiation belts to some years at an altitude of about 8000 km ($1.25 r_E$) and decreases to minutes at the outer edge of the radiation belts [10.51]. The exact variation depends on the external circumstances, in particular the influence of solar and geomagnetic activity: under undisturbed conditions, the lifetime is larger than during strong solar or geomagnetic activity.

Particle losses are always due to the interaction of radiation belt particles with the atmospheric gas. Thus the losses happen at the mirror points where the particles come closest to the atmosphere. The chance of getting lost therefore is largest for particles with small pitch angles, and overall losses increase when pitch-angle scattering increases. The physical mechanisms for losses are:

- Charge exchange with a particle from the neutral atmosphere: a fast radiation belt proton captures an electron from the atmospheric hydrogen, leaving behind a slow proton and continuing itself as a fast hydrogen atom through the atmosphere. This mechanism is of particular importance in the inner radiation belt for protons with energies below about 100 keV. With increasing energy, the interaction time between radiation belt protons and atmospheric hydrogen decreases, making the interaction less likely.
- Nuclear collisions between protons and atmospheric atoms and molecules: these are important loss mechanism for protons with energies above 75 MeV in the inner radiation belts.
- Scattering into the loss cone is the main loss mechanism for electrons and protons in the outer radiation belt. Scattering can occur either by violation of the second adiabatic invariant due to changes in the field at time-scales shorter than the north–south oscillation time, or due to pitch-angle scattering at electrostatic or electromagnetic waves, cf. Sect. 6.2.4.

Sources. The sources of radiation belt particles can be divided into the ‘creation’ of particles inside the radiation belts or the motion of particles into the radiation belts under violation of the second adiabatic invariant.

The high energetic particles in the inner radiation belts in general are created there. The main mechanism is CRAND (Cosmic Ray Albedo Neutron Decay). Nuclei from the galactic cosmic radiation penetrate deep into the atmosphere and interact with the atmospheric gas. A 5 GeV proton, for instance, typically creates about seven neutrons during these interactions. Some of the neutrons are slowed down, creating the cosmogenic nuclides such as radiocarbon (capture of thermal neutrons by atmospheric nitrogen: $^{14}\text{N}(\text{n},\text{p})^{14}\text{C}$) or ^{10}Be (spallation of nitrogen or oxygen due to the capture of fast protons or neutrons). Other neutrons simply escape without interaction. Since neutrons are neutrals, their motion is not influenced by the geomagnetic field, thus allowing some of them to propagate into the radiation belts. But neutrons are not stable and decay into a proton, an electron, and an antineutrino. If this decay happens inside the radiation belt, electrons and protons are trapped, thus replenishing the radiation belt population. This process is called CRAND because the primaries are Cosmic Rays, creating secondaries which are partly reflected (Albedo = reflectivity) into the radiation belts, where the Neutrons Decay.

Another source, also based on *in situ* creation, is the influx of particles from the outer magnetosphere. In the outer magnetosphere, particles with low energies dominate. Occasionally, these particles can recombine, forming neutrals. As neutrals, the particles are no longer guided by the magnetic field and eventually can propagate into the radiation belts. On the dayside, these neutrals are not stable but immediately become ionized by the hard electromagnetic radiation from the Sun. If this process happens while the neutral is inside the radiation belt, an additional electron and ion are fed into the belt population.

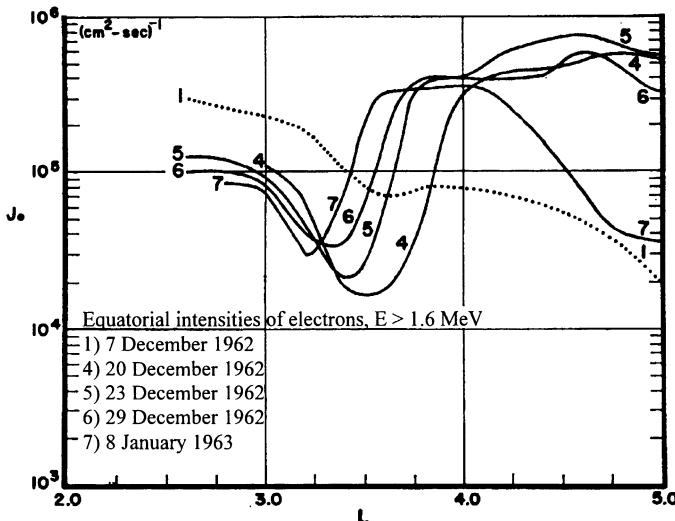


Fig. 10.28. Fluxes of trapped electrons with energies above 1.6 MeV for different times after a geomagnetic storm. The dashed curve (1) gives the undisturbed conditions on 7 December 1962. The other curves are taken after the storm on 20 December (4), 23 December (5), 29 December (6), and 8 January (7). The motion of the electron distribution towards the inner magnetosphere is obvious. Reprinted from L.A. Frank et al. [10.10], *J. Geophys. Res.* **69**, Copyright 1964, American Geophysical Union

In the outer radiation belt ($L > 3$), the sources and sinks become more complex and are much more intimately related to solar and geomagnetic activity. For instance, the fluxes of energetic electrons are increased during higher geomagnetic activity and relax only slowly after a strong geomagnetic storm. While the maximum of the radiation belt is depleted during the storm, at larger distances the flux densities are higher, slowly propagating towards smaller L-shells, cf. Fig. 10.28. Thus the radiation belt is refilled with particles from the outer magnetosphere.

For a fixed distance in space, changes in particle flux can be observed as a consequence of changes in the solar wind speed and the related geomagnetic activity. In Fig. 10.29 the dashed line gives the variation of the solar wind speed, and the solid line the changes in electron fluxes in a geostationary orbit at a height of 36 000 km, well outside the maximum of the radiation belts. The electron flux increases in response to an increase in solar wind speed with an energy-dependent delay between about 1 and 2 days due to the inward propagation of the particles. Thus compression of the magnetosphere is connected with an increase in particle fluxes, although the latter is delayed depending on the energy of the particles and the position of the observer.

L-Shell Diffusion. Particles refilling the radiation belts have entered the magnetosphere at the polar cusps, either from the outside as solar wind

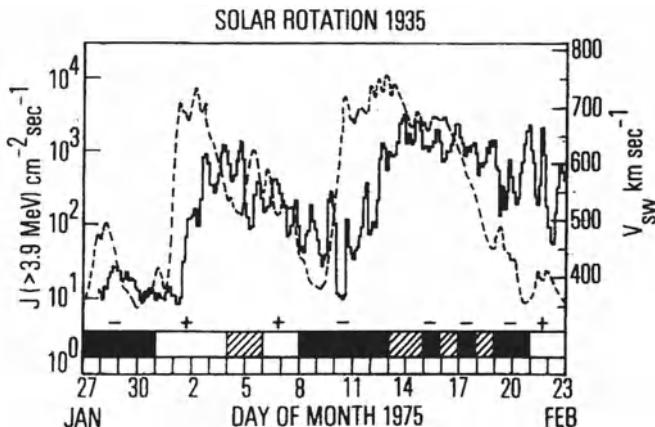


Fig. 10.29. Flux of energetic electrons above 3.9 MeV in a geostationary orbit (*solid line*) compared with the solar wind speed (*dashed line*). The polarity pattern of the interplanetary magnetic field is shown at the bottom. Reprinted from G.A. Paulikas and J.B. Blake [10.39], in *Quantitative modeling of magnetospheric processes* (ed. W.P. Olson), Copyright 1979, American Geophysical Union

plasma or solar energetic particles, or from the ionosphere. Although some of these particles might have rather high energies on entering the magnetosphere, most of them will have energies much closer to the plasma's thermal energy than to the energies typically observed in the radiation belts. Thus not only do we need a mechanism to transport these particles into the radiation belts but we also need to accelerate them.

Both transport as well as particle acceleration is induced by magnetic fluctuations, as is evident from the observations described in connection with Figs. 10.28 and 10.29. The transport mechanism basically is diffusion, although convection and drift also contribute to the transport [10.45]. The acceleration is related to this transport because during the inward transport the magnetic flux density increases. The acceleration mechanisms are the betatron effect and Fermi acceleration.

In the betatron effect a charged particle gyrates in a magnetic field with slowly increasing magnetic flux density. As the flux increases, the gyro-frequency increases, too, cf. (2.27). Since the angular momentum $mvr_L = mv^2/\omega_c$ stays constant, the particle's kinetic energy increases as $E = mv^2/2 \sim \omega_c \sim B$. Thus a particle moving inwards in the magnetosphere gains energy as the magnetic field increases.

Whereas the betatron effect accelerates particles moving perpendicular to the field towards regions of higher flux density, the Fermi effect accelerates the motion of a particle parallel to the field: according to the second adiabatic invariant (2.73), a particle gains energy as its travel path along the magnetic field line shortens. As a particle moves inwards in the magnetosphere, the field line along which it bounces shortens and the particle gains energy.

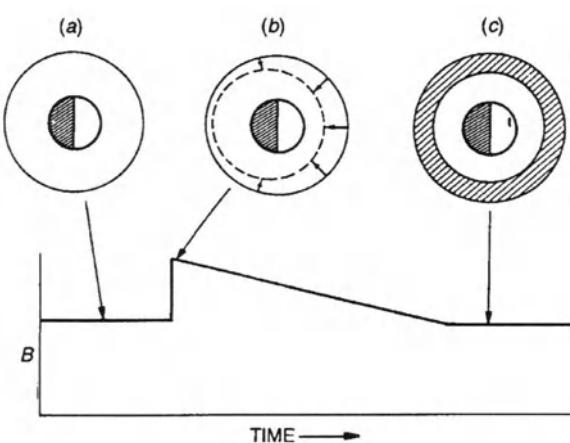


Fig. 10.30. Effect of an asymmetric sudden compression and slow relaxation of the geomagnetic field on a narrow band of equatorially trapped particles. After the recovery period, the particles fill the shaded band. Reprinted from M. Walt [10.53], *Introduction to geomagnetically trapped radiation*, Copyright 1994, with kind permission from Cambridge University Press

The inward motion of the particles is a diffusive process based on the violation of the third adiabatic invariant (flux invariant) when marked variations of the magnetosphere occur on time-scales smaller than the drift period of the particles. Figure 10.30 gives a very simple model: assume a narrow particle distribution in the equatorial plane (a). A sudden variation in the solar wind leads to a fast compression (expansion) of the magnetosphere (b). Since this is a fast change, the magnetic flux inside the particle's drift orbit is suddenly increased (decreased). The field now relaxes slowly (adiabatically) back to its original state. As the flux is now conserved, the drift orbit expands (shrinks) and the particle orbit has moved in radius. Since the magnetosphere is asymmetric, a change in solar wind leads to a stronger variation in the field on the dayside than on the nightside. Thus the effect on the particle not only depends on the properties of the disturbance but also on the particle's position in its drift orbit. Therefore the initial distribution is spread in radius (c). This process is diffusive, and it is called L-shell diffusion or radial diffusion.

The process shown in Fig. 10.30 leads to a spread of the particle distribution but not necessarily to an inward motion. Since particle streaming in diffusion is driven by a gradient, cf. (6.10), for undisturbed magnetospheric conditions, such as in the particle distributions in Fig. 10.26, we would not expect an inward streaming but rather one directed outwards. After a strong geomagnetic storm, however, the particle density increases in the outer magnetosphere due to compression as well as to an injection of plasma from the tail, leading to a density gradient and therefore a particle streaming towards the inner magnetosphere as is evident in the observations shown in Fig. 10.28.

How Many Radiation Belts are There? In connection with Fig. 10.26 we have seen that the trapped particles can be described either as one population that varies with radius and energy, or as two radiation belts, a stable inner van Allen belt dominated by protons and a more variable outer one dominated by electrons. More recent observations, in particular by SAM-

PEX (Solar, Anomalous, and Magnetospheric Particle EXplorer), indicate the existence of a distinct trapped particle component inside the inner radiation belt. The particles of this New Radiation Belt differ from the other radiation belt particles in so far as their composition and charge states closely resemble those of the anomalous component instead of the galactic cosmic rays. The flux densities are about two orders of magnitude larger than those of the anomalous component outside the magnetosphere. The dynamics of the new radiation belt are similar to those of the van Allen belt.

10.4.2 Galactic Cosmic Rays – Størmer Orbits

Galactic cosmic rays have high rigidities, and the resulting gyro-radius is comparable with the size of the magnetosphere. Thus concepts such as the adiabatic invariants cannot be applied to the motion of these particles: instead, their equation of motion has to be integrated along the particle orbit. This was first done by C. Størmer, therefore the particle orbits are called Størmer orbits.

In interplanetary space, a galactic proton has a gyro-radius of the order of 1/100 AU. If such a proton hits the magnetosphere, it senses a strong change in the magnetic field on a spatial scale short compared with its Larmor radius. Thus the simplifying assumption of a uniform magnetic field made for the radiation belt particles cannot be applied here. The fate of the particle then depends on its pitch angle and the direction and location of incidence: particles propagating nearly field-parallel into the polar cusps are less influenced by the geomagnetic field than particles hitting the magnetosphere in equatorial regions perpendicular to the magnetic field. In the latter case, the particle would have to cross a large inhomogeneous bundle of field lines before reaching the ground or the denser atmosphere.

The motion of such a high energetic particle is still determined by the Lorentz force; however, since the magnetic field varies strongly on a scale smaller than the gyro-radius, the equation of motion has to be integrated. An example of the resulting orbit is shown in Fig. 10.31 where the orbit of a 1 GV particle coming from the east is shown in the equatorial plane (a) and the meridional plane (b) of a magnetic dipole field. The particle trajectory is irregular with wide excursions in latitude and longitude until it finally hits the Earth.

The particle's orbit depends on the location and direction of incidence. The most important result of Størmer's work has been the definition of allowed and forbidden regions on the ground which can be reached by a charged particle traveling towards the Earth. For instance, to hit the Earth at a certain magnetic latitude ϕ_c , the particle's rigidity must be at least $P_c = 14.9 \cos^4 \phi_c$, with P_c in units of 10^9 V or 1 GV. This rigidity is also called the cut-off rigidity: particles with a rigidity R_c can reach all latitudes of ϕ_c or above, but particles with smaller rigidities can hit the ground only at higher latitudes.

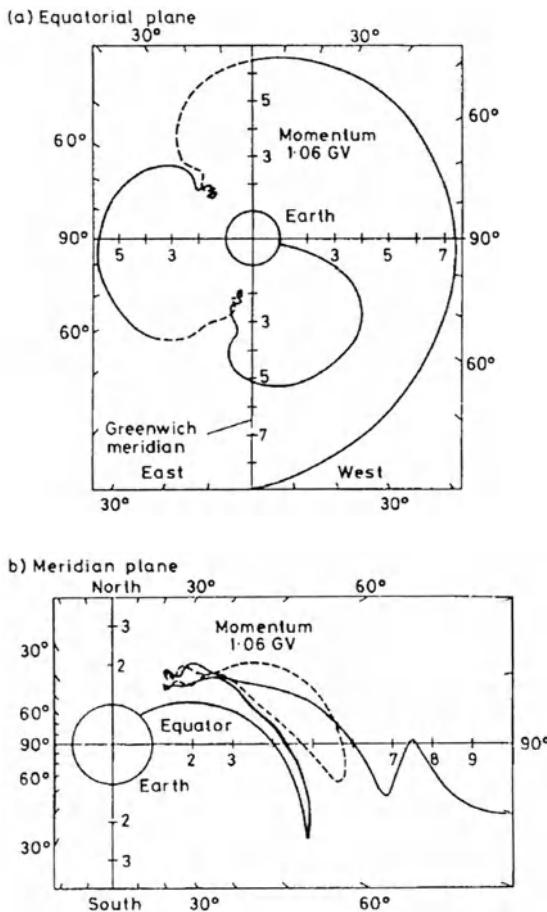


Fig. 10.31. Orbit of a high energetic particle in a magnetic dipole field in (a) the equatorial plane and (b) in a meridional plane. Reprinted from J.K. Hargreaves [10.13] (based on [10.29], *The solar-terrestrial environment*, Copyright 1992, with kind permission from Cambridge University Press)

The extent of the forbidden region can be approximated by the Størmer unit $C_{St} = \sqrt{M_E/(Br_L)} = \sqrt{M_E/P}$. For a solar wind proton with an energy of 1 keV, the forbidden region has an extent of about $100 r_E$, and thus it will not reach the ground. For a solar wind electron, the forbidden region is even wider.

If a proton of the galactic cosmic radiation has been detected on the ground, its origin can be determined by the reverse procedure: we let a negatively charged particle with the same rigidity propagate out through the geomagnetic field towards infinity. Thus the direction of incidence of the galactic cosmic radiation can be determined. On Earth, the galactic cosmic radiation is isotropic, while during ground-level events solar protons with gigaelectron-volt energies can be detected coming predominantly along the interplanetary magnetic field line connecting the Earth to the Sun.

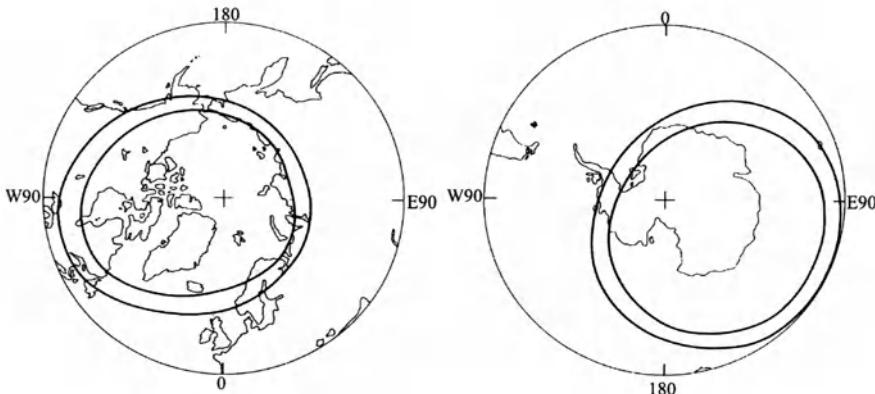


Fig. 10.32. Areas affected by PCA: the regions inside the inner curves give the polar plateaus while the regions outside the outer curves usually are unaffected, except during severe geomagnetic disturbances. Reprinted from G.C. Reid [10.41], in *Physics of the Sun, vol. III* (eds. P.A. Sturrock, T.E. Holzer, D.M. Mihalas, and K. Ulrich), Copyright 1986, with kind permission from Kluwer Academic Publisher

10.4.3 Solar Energetic Particles – Polar Cap Absorption

Except for ground-level events, the energies of solar energetic particles are too small to allow them to penetrate down to the ground. This is true for most of the magnetosphere. At the polar cusps, however, lower energetic solar particles also can reach the denser atmosphere. Their interaction with the atmosphere leads to increased ionization, which in turn leads to the absorption of radio waves. Since this absorption is limited to the polar regions, this phenomenon is called polar cap absorption (PCA). Owing to the propagation time between the Sun and the Earth, a PCA generally starts a few hours after the flare. A similar effect on the dayside ionosphere, i.e. a sudden ionospheric disturbance, on the other hand, starts immediately after flare onset because here the increased ionization is due to the flare's hard electromagnetic radiation.

The protons causing PCAs typically have energies between 1 MeV and 100 MeV and can penetrate down to heights between about 30 km and 90 km. Thus the ion population created during a PCA lies below the ionosphere in an atmospheric layer which normally is neutral, allowing for ion–chemical interactions that normally do not occur at these heights.

Figure 10.32 shows the areas typically affected by PCA. PCAs are limited to a small latitudinal ring around the geomagnetic pole. Outside this ring, PCAs only occur during very strong geomagnetic disturbances when the geomagnetic cut-offs are reduced, allowing particles to reach the denser atmosphere at lower geomagnetic latitudes.

Part of the energy transferred from the solar particles to the atmosphere leads to electromagnetic radiation in the visible range, called the polar glow

aurora (PGA). In contrast to the normal aurora, which is highly variable and often well structured, the polar glow aurora is a diffuse red glow of the entire sky in high latitudes. Most often it is red since the excited atoms mainly are hydrogen, emitting in the visible the H α line and in the UV range the L α line. Even if its luminosity is well above the visibility threshold, it often is difficult to detect because of its total lack of any contrasts. Note that the polar glow aurora is the only visible effect caused by solar energetic particles; the much more impressive aurora on the other hand is caused by plasma from the solar wind and the plasma sheet in the geomagnetic tail.

10.5 Summary

The magnetosphere is shaped by the interaction between the solar wind and the terrestrial magnetic field. Since both are variable the magnetosphere shows a slow evolution as well as fast fluctuations. The basic features in the magnetospheric field topology are the magnetopause, the polar cusps, and the tail. The magnetopause marks the separation between the planetary and the interplanetary magnetic fields. In front of the magnetopause a bow shock develops where the solar wind is slowed down to subsonic speeds. Thus the region between bow shock and magnetopause, the magnetosheath, is filled by a hot but slow solar wind flow.

Although the magnetopause separates the magnetic fields, the solar wind plasma can still penetrate through it, filling part of the outer magnetosphere. In the inner magnetosphere, the cold and dense ionospheric plasma fills the plasma sheet, as is evident from the high O $^+$ content. The region between this inner plasmasphere and the outer magnetosphere contains a rarefied plasma of mixed origin: under undisturbed conditions, the geosphere is dominated by the solar wind plasma while during geomagnetically active periods an outflow of ionospheric plasma is added.

The particle populations in the magnetosphere have different sources and can be divided into stable and transient populations. The only stable component is the radiation belt particles, forming the ring current. Its variations lead to geomagnetic activity, detected on the ground as geomagnetic storms (cf. Chap. 11). The radiation belts are only a quasi-stable population since particles are lost due to interaction with the denser atmosphere and are created due to cosmic rays (CRAND) or the transport and acceleration of lower energetic particles or solar wind plasma from the outer magnetosphere towards the Earth. Galactic cosmic rays have much higher rigidities and are not trapped inside the geomagnetic field. Instead, depending on their direction of incidence, they can either penetrate down to the denser atmosphere or are deflected back into space. Particles with lower rigidities, such as solar energetic particles or the solar wind, can penetrate into the magnetosphere only at the polar cusps and, on interaction with the atmosphere, lead to polar cap absorption events.

Exercises and Problems

10.1 Determine the magnetic flux density and the direction of the field for your home town (assume a simple dipole field). Compare with Fig. 10.3. To which L-shell is your home town connected magnetically? Determine also the cut-off rigidity and the Størmer unit for this rigidity.

10.2 A 10 keV proton with an equatorial pitch angle of 40° moves from $L = 6$ to $L = 1.5$. Assume a dipole field and calculate the energy gained by Fermi acceleration (second adiabatic invariant) during this motion.

10.3 Assume a sinusoidal variation with period $T = 1$ h in the solar wind speed with an amplitude of ± 40 km/s around an average of 400 km/s. Determine the speed of the magnetopause and the maximum and minimum stand-off distances.

10.4 Størmer orbits are calculated for a terrestrial dipole field. Give a qualitative statement about the errors made in neglecting the actual shape of the magnetosphere. Try to consider the influence of the shifted dipole as well as the different topologies in the noon and midnight directions.

10.5 Fig. 10.28 shows the inward L-shell diffusion of energetic electrons. Try to estimate the diffusion coefficient for this process (assume that the energy of the particles does not change during the inward motion).

11. The Dynamic Magnetosphere

Forward, forward let us range
let the great world spin forever down the ranging grooves of change.
A. Tennyson, *Locksley Hall*

The magnetosphere has been recognized as a dynamic phenomenon as the first systematic magnetic field measurements at the ground became available: aside from diurnal variations, strong transient disturbances can be observed. These magnetic storms often are accompanied by other phenomena, such as aurorae, or influence our technical environment, as evident in disruptions in radio communication or power-line breakdowns. All these phenomena are caused by strong fluctuations or discontinuities in the solar wind and can be related to changes in magnetospheric structure, in particular in the magnetospheric current system and in the plasma sheet inside the magnetotail. This chapter provides an introduction to these phenomena and a supplementary section about the aurora and the history of aurora research.

11.1 Geomagnetic Variability

The magnetic field measured on the ground is determined by two parts: (a) the inner field created by the geomagnetic dynamo and (b) an outer field caused by the magnetospheric–ionospheric current system, in particular the ring current and the ionospheric current system. Although the currents inside the magnetopause and the tail are strongly influenced by solar wind variations, their fluctuations cannot be detected on the ground.

11.1.1 Daily Variations

The magnetogram of a geomagnetically quiet day shows systematic variations in all three field components. The most pronounced variation can be observed around local noon. These variations are regular excursions, which are repeated each day. Their direction and magnitude depends on the geomagnetic latitude of the observer. These Sq, or solar quiet, variations are

related to the ionospheric current system which is fixed in space while the Earth rotates below it. Note that this ionospheric current system is different from the current systems that provide the ionosphere–magnetosphere coupling shown in Fig. 10.24. The ionospheric current system consists of two vortices centered close to local noon at latitudes of about 30° . At the equator, the currents of both cells flow from dawn to dusk. These currents are created by winds which in turn are driven by solar heating and, to a lesser extent, also by solar and lunar tides.

Occasionally, the Sq variations are enhanced by a solar flare: since its hard electromagnetic radiation leads to a stronger ionization of the dayside ionosphere, the ionospheric current system is enhanced. In this case the change in magnetic field is called the sfe variation (sfe: solar flare effect).

11.1.2 Geomagnetic Disturbances

On these periodic quiet time variations, irregular disturbances on different time scales are superimposed. Fluctuations with time-scales below about 0.2 s are waves, fluctuations with time-scales between 0.2 s and 600 s are pulsations of the magnetosphere, and variations with time-scales above 10 min are geomagnetic disturbances. These geomagnetic disturbances can be observed worldwide; however, their amplitudes generally are largest at high geomagnetic latitudes and smaller, or even vanishing, at low latitudes.

Geomagnetic disturbances are also called magnetic storms or substorms, depending on their temporal and spatial extent. Magnetospheric substorms are the most frequent type of geomagnetic activity. Its most obvious manifestation is the sudden explosion of a quiet auroral arc to more brilliant colors and moving structures. Over a period of an hour, they develop through an orderly sequence that depends on time and location. Simultaneously, a magnetometer on the ground below the aurora will record intense disturbances caused by the electric currents accompanying the aurora. These auroral electrojets are roughly parallel to a geomagnetic parallel circle, flowing at a height of about 120 km in concentrated channels of high conductivity produced by the same particles that generate the auroral emission. The disturbances in the magnetic field are in the range 200 nT–2000 nT, they typically last between 1 h and 3 h, and are most pronounced at high geomagnetic latitudes.

If the coupling of matter and energy from the solar wind into the magnetosphere is stronger and lasts longer, a magnetic storm develops. Its temporal development can best be seen in the D_{st} index, defined as the worldwide average of the equatorial H disturbance. Figure 11.1 shows the variation of the D_{st} index for a large geomagnetic storm. The storm often begins with a sudden increase in the magnetic field. This sudden commencement may last for many hours. This initial phase is followed by a rapid and sometimes highly disturbed decrease in D_{st} , which defines the storm's main phase. Subsequently, D_{st} recovers, first rather quickly, later more slowly. A storm lasts between 1 and 5 days with the initial phase anything between zero and about

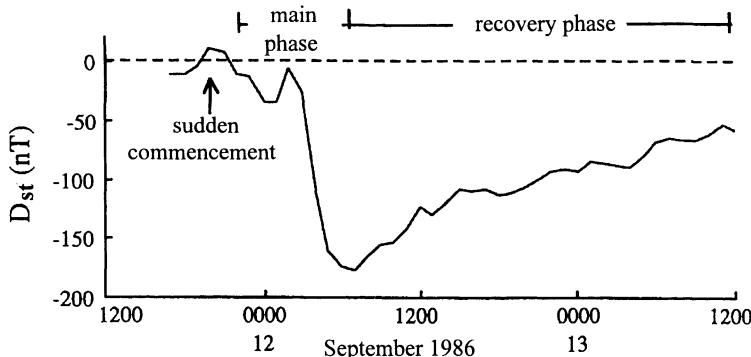


Fig. 11.1. The variation of the D_{st} index for the large magnetic storm on 12 September 1986. From *Solar Geophysical Data*

1 day, the main phase normally about 1 day, and the recovery phase lasting for several days. The distribution of storm magnitudes obeys a power law: storms with D_{st} between 50 nT and 150 nT occur about once in a month. Disturbances with D_{st} between 150 nT and 300 nT occur several times a year, while only a few disturbances with D_{st} above 500 nT can be observed over an entire solar cycle.

The scale of a magnetic storm depends on the kind of disturbance in the solar wind. In general, the storm is more pronounced if the solar wind disturbance is a fast shock and if the interplanetary magnetic field (or the ejecta driving the shock) has a southward component. In this case the magnetosphere has an open configuration, and plasma and energy are fed into it easily.

A simple interpretation of the phases of a geomagnetic disturbance is as follows: the increase in magnetic field strength at the beginning of the disturbance can be attributed to the compression of the magnetosphere as the magnetopause is pushed inward by the increased solar wind speed. The decrease in field strength during the main phase of the storm is due to an increase in the ring current, which creates a magnetic field opposite to the terrestrial one. A typical current density in the undisturbed ring current is about 10^{-8} A m^{-2} , mainly carried by particles with energies between 10 keV and 100 keV at a height between $3 r_E$ and $6 r_E$. During a strong magnetic storm, particles are injected from the plasma sheet in the magnetotail into the radiation belts, enhancing its density by an order of magnitude on a time-scale as short as 10 min. This enhanced ring current then reduces the magnetic field measured on the ground.

11.1.3 Geomagnetic Indices

For a quantitative description of geomagnetic disturbances the K and A indices are used. The basis for these indices are the magnetograms. The K

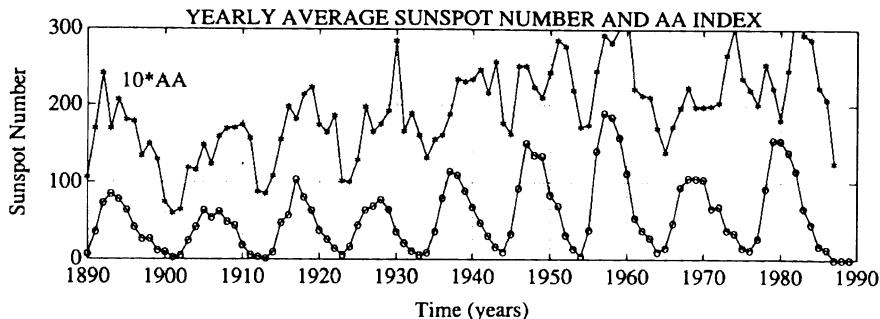


Fig. 11.2. Variation of sunspots and the level of geomagnetic activity with the solar cycle. Reprinted from R.L. McPherron [11.21], in *Introduction to space physics* (eds. M.G. Kivelson and C.T. Russell), Copyright 1995, with kind permission from Cambridge University Press

index is a quasi-logarithmic number between 0 and 9 determined at the end of specified 3-h intervals as the maximum deviation of the observed magnetic field from the expected quiet field. It is determined for each of the three magnetic field components separately. The largest of the maxima is converted to a standardized K index taking into account the geomagnetic properties of the observation site. Thus K indices of different stations, in particular of stations at high and low latitudes, can be compared and can be combined to give a planetary K index.

At individual stations, the eight daily K indices are linearized to give an a index and than averaged arithmetically to give the A index describing the daily averaged magnetic activity. The construction of the K and A indices and the global network of observatories is described in [11.22].

Geomagnetic disturbances generally are described by two indices. Magnetic storms can be quantified by the D_{st} index which gives the excursion of the equatorial H component compared with quiet times. Physically, the D_{st} index is related to the ring current. At high latitudes, an AE index is used, related to the auroral electrojet. It is also determined from the excursion of the H component compared with quiet times. Both indices are determined globally by combining different observatories at comparable geomagnetic latitudes but different longitudes.

11.1.4 Geomagnetic Activity Through the Solar Cycle

Geomagnetic disturbances are not distributed uniformly in time. Instead, characteristic dependences can be observed which directly point to the physical mechanisms responsible for them. In the late 1930s Chapman and Bartels [11.9] showed that the number of geomagnetic disturbances is related directly to the number of sunspots and thus to solar activity. Figure 11.2 shows this close relation for the last hundred years using yearly averages of the sunspot

number (lower curve) and the AA index (upper curve). The latter describes the difference between the observed and expected horizontal component of the magnetic field at mid-latitude and is the activity index with the largest record.

11.1.5 Annual Variations and Recurrent Geomagnetic Activity

Geomagnetic activity shows systematic variations on shorter time-scales, too. In particular, there is an enhanced geomagnetic activity during the equinoxes. Thus there is an annual variation, and there is also a recurrent geomagnetic activity on a time-scale of 27 days.

Figure 11.3 shows the D_{st} index plotted versus time for 16 solar rotations during the 1974 solar minimum. Strong deviations of the D_{st} index to lower values are indicative of geomagnetic activity. For a better identification, these periods are blackened. Both systematic variations can be identified in this figure. Most obvious is the recurrence of the geomagnetic disturbances during each solar rotation. These disturbances are related to fast solar winds, which can be observed best during the solar minimum. But this recurrence is not observed for all 16 rotations: in the spring, the strongest geomagnetic disturbances are observed in the middle of each rotation, while in the autumn they are at the beginning of the rotation. In addition, during summer and winter the disturbances are weak, while in spring and autumn they are more pronounced.

To understand this change in pattern between the two equinoxes let us first look at one of these disturbances, the one beginning in the middle of rotation 1921 in Fig. 11.3. Figure 11.4 shows the density and speed of the solar wind, the flux density and north–south component of the interplanetary magnetic field, and the AE (auroral electrojet) index as a measure of geomagnetic activity related to aurorae and substorms. For solar rotation 1921 two fast streams can be identified, one starting in the data gap on 15 January, the other starting on 25 January. Immediately before the beginning of this stream the increase in plasma and magnetic flux density indicates the compression region in front of the fast stream. The envelope on the AE index traces the solar wind speed, thus changes in the solar wind are related to geomagnetic disturbances, as expected. But even at times of rather constant solar wind speed there can be intense although short geomagnetic disturbances (for instance the one on the evening of 18 January). These disturbances are strongest when the interplanetary magnetic field has a southward component.

With this information we can go back to the interpretation of Fig. 11.3. The fact that geomagnetic disturbances are recurrent is related to the existence of two recurrent fast solar wind streams, one at the beginning and the other in the middle of each solar rotation. The geomagnetic effectiveness, on the other hand, depends on whether the interplanetary magnetic field

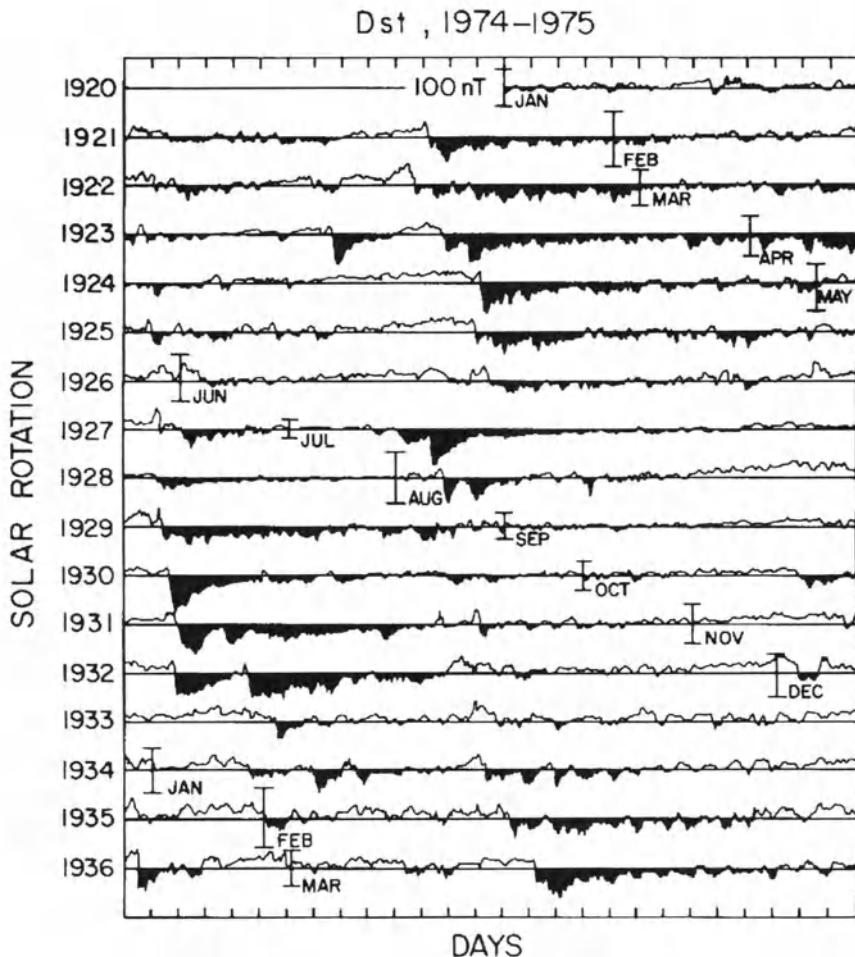


Fig. 11.3. Variation of the D_{st} index versus time for 16 solar rotations during the 1974 solar minimum. Reprinted from N.U. Crooker and G.L. Siscoe [11.10], in *Physics of the Sun, vol. III* (eds. P.A. Sturrock, T.E. Holzer, D.M. Mihalas, and K. Ulrich), Copyright 1986, with kind permission from Kluwer Academic Publishers

has a southward component or not, i.e. whether the magnetosphere is open or closed. As discussed earlier, an open magnetosphere also means an easier exchange of energy and matter between the interplanetary medium and the terrestrial environment, which is essential to drive geomagnetic activity. But the annual variation of geomagnetic effectiveness does not indicate that the fast streams change polarity during the course of the year. Since they are related to the coronal structure and magnetic field, their polarity stays constant. The change occurs in the position of the Earth and the terrestrial magnetic field relative to this structure. The solar wind and thus the inter-

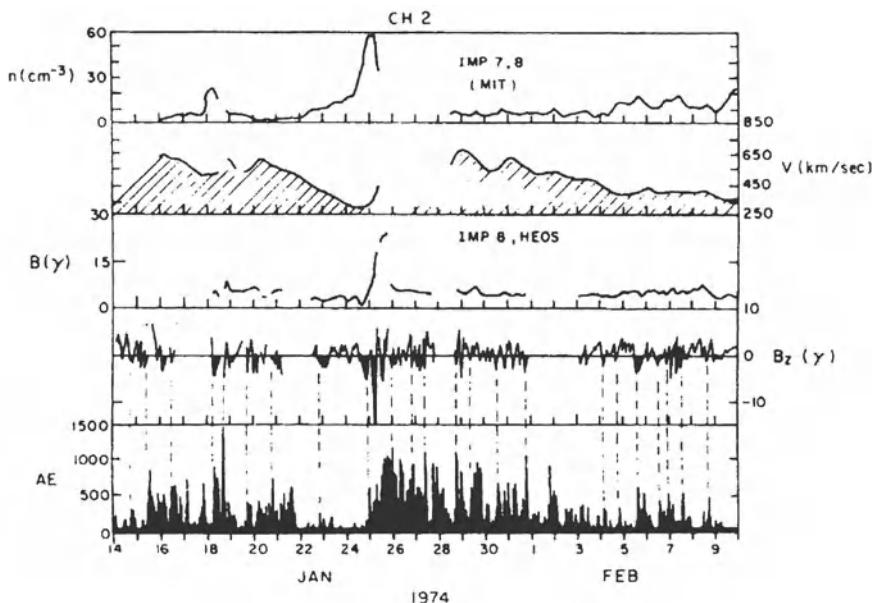


Fig. 11.4. Variation of solar wind density, speed, magnetic flux density, the north-south component of the interplanetary magnetic field with negative values indicating a southward interplanetary magnetic field, and geomagnetic activity over solar rotation 1921. The dashed vertical lines link the times of a southward excursions by the interplanetary magnetic field and enhanced geomagnetic activity. Reprinted from L. Burlaga and R.P. Lepping [11.8], *Planet. Space Sci.* **25**, Copyright 1977, American Geophysical Union

planetary magnetic field is fixed with respect to the Sun's equatorial plane. The plane of ecliptic, which also defines the orbit of the Earth, is inclined by 7.2° with respect to this plane, with both planes intersecting in the equinoxes while the Earth is above (below) the solar equator at the summer (winter) solstices. In addition, the axis of the Earth's rotation is tipped at 23° to the ecliptic plane. The combined effects regulate the southward component of the interplanetary magnetic field with respect to the terrestrial field. If the magnetic field is directed away from the Sun, in spring it has a strong southward component in geomagnetic coordinates while in autumn it has a very weak southward or even a northward component. If the interplanetary magnetic field is directed inwards, the pattern is just the opposite. Thus a fast solar wind stream which has a high geomagnetic effectiveness in spring is less efficient in autumn, while a stream with an opposite direction of the interplanetary magnetic field shows the reverse pattern. At the solstices, in both streams the southward component is diminished, leading to a reduced geomagnetic effectiveness.

11.1.6 Magnetic Storms in Response to Transient Phenomena

Geomagnetic disturbances are not only caused by fast solar wind streams but also by transient disturbances, such as interplanetary shocks and magnetic clouds. The same pattern as for recurrent solar wind disturbances emerges: the geomagnetic activity increases with a change in solar wind flux and is stronger if the interplanetary magnetic field or the field at the leading edge of the magnetic cloud has a southward component [11.16, 11.29]. As a rule of thumb, an intense geomagnetic storm requires a southward component of the interplanetary magnetic field of more than 10 nT for at least 3 h. The reaction of the magnetosphere to such transient disturbances is responsible for the correlation shown in Fig. 11.2.

11.1.7 A Simple Model of Geomagnetic Activity

Figure 11.5 offers a simple illustration of this dependence of geomagnetic activity on solar wind flow and the southward component of the interplanetary magnetic field in terms of the superimposition of two energy fluxes. The solar wind feeds energy continuously into the magnetosphere. The energy partly is stored as magnetic field energy in the tail (small bucket to the right) and partly is converted to geomagnetic activity (outflow to the left). During weak geomagnetic activity as well as during large storms the solar wind energy is fed directly into the magnetosphere and ionosphere. At times of moderate geomagnetic activity, however, the energy is stored in the magnetosphere before it is released. As discussed in connection with Fig. 10.19, solar wind energy is fed more efficiently into the magnetosphere if the latter is open, i.e.

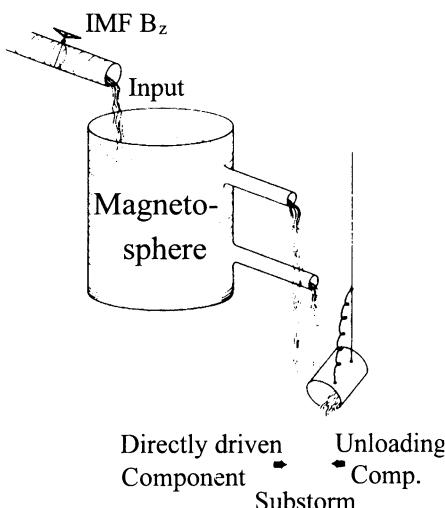


Fig. 11.5. Energy flux from the solar wind into the magnetosphere. Reprinted from S.-I. Akasofu [11.2], *EOS* **70**, Copyright 1989, American Geophysical Union

if the interplanetary magnetic field has a southward component, indicated by the handle at the outflow. Thus more energy is available for release in the form of geomagnetic activity. This release, however, is not continuous but energy is stored over a time period of typically an hour and then liberated abruptly in a substorm.

11.2 Substorms

Magnetic substorms can be seen as magnetic field variations, in particular in the AE index, as well as in the aurora. The primary location of energy storage is the tail; the physical mechanism for the energy release is reconnection in its plasma sheet. The basic process can be likened to a dripping fountain: a drop at the outflow fountain forms by the interplay between gravity pulling the water down and surface tension keeping it up. If enough water has been collected at the outflow, gravity takes over and the drop falls. In a magnetospheric substorm, a plasmoid in the magnetotail takes over the role of the drop: solar wind drag pulls on the magnetosphere, feeding plasma and energy into it. As the plasmoid grows, a magnetic neutral point forms close to the Earth. Here reconnection sets in, expelling the plasmoid tailwards and accelerating a small amount of plasma towards the Earth, causing the aurora at the Earth's high-latitude nightside and the accompanying geomagnetic disturbance.

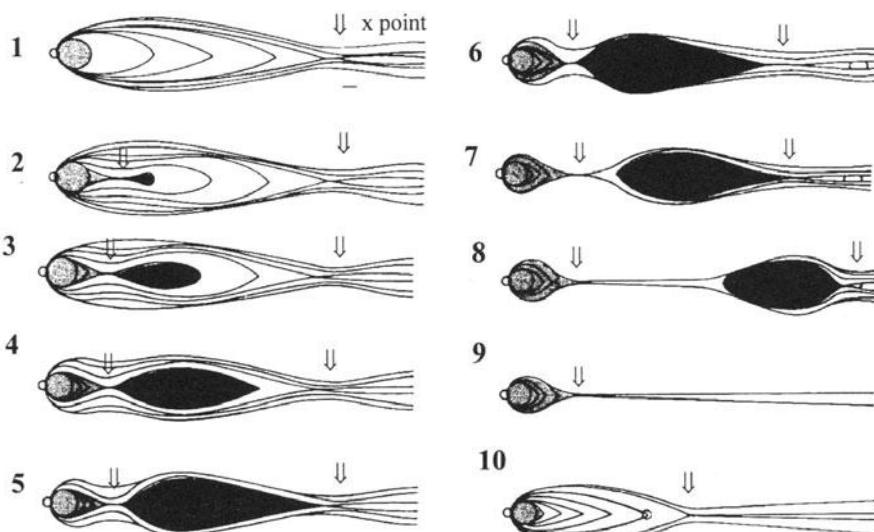


Fig. 11.6. Sequence of events leading to a magnetospheric substorm, cf. text. Based on E.W. Hones [11.17], in *Magnetic reconnection*, Copyright 1984, American Geophysical Union

This scenario is sketched in more physical terms in Fig. 11.6. Panel 1 gives the quiet time configuration of the magnetosphere with an X-point at a radial distance of about $100 r_E$. This field line is the last closed field line: in the direction of Earth, all field lines are closed, at larger distances all field lines are open. As reconnection sets in at this X-point, the energetics of the plasma sheet are changed so that a second X-point forms much closer to the Earth (panel 2). As more energy is fed from the solar wind into the magnetosphere, reconnection at this inner X-point continues and a plasmoid is formed between the inner and the outer X-points (panels 3–6) until finally the plasmoid becomes detached at the inner X-point (panel 7) and is accelerated, leaving the magnetotail (panel 8). The bulk of the energy fed from the solar wind into the magnetosphere is contained inside the plasmoid and thus is fed back into the solar wind. Only a small amount is converted to kinetic energy of plasma moving towards the Earth. As the latter interacts with the high-latitude ionosphere, it causes the aurora and enhances the auroral electrojet. The tail slowly fills with new solar wind plasma (panels 9 and 10), until the initial configuration (1) is restored and the cycle starts anew.

The plasma traveling towards Earth during the discharge of the magnetotail leads to the substorm current, which has already been shown in Fig. 10.24. It is associated with the tail current, which during a substorm, is diverted partly as a field-parallel current towards the ionosphere in the evening sector, continues through the ionosphere as auroral electrojet, and flows back towards the tail as field-parallel current. The excursion of the tail field happens when, during the onset of reconnection, the tail field collapses. This current system is called the substorm current wedge because a projection of the current system into the equatorial plane takes the form of a wedge. The opening angle of the wedge typically is 70° , the current is about $2 \times 10^6 \text{ A}$, and it extends from the Earth into the tail for about $5 r_E$.

During the sequence shown in Fig. 11.6, an observer on Earth sees characteristic changes in the aurora. As the plasmoid in the tail grows, the auroral oval slowly expands equatorwards with the aurora still being a quiet arc. As reconnection sets in at the inner X-point, the initial auroral arc brightens, exhibits more structures and fast changing features, and moves rapidly equatorwards. During the recovery, a rather quiet auroral arc or curtain contracts to the size of the initial auroral oval.

11.3 Aurorae

Aurorae, or polar lights, are a prime example of solar–terrestrial relationships. They also provide an interesting example of the development in geophysical research. And they are simply beautiful and, depending on his habitat, have fascinated or frightened mankind since historical times.

The aurora, also called the northern light, is a typical phenomenon of the high latitudes, where under suitable conditions (clear sky, no full moon), it

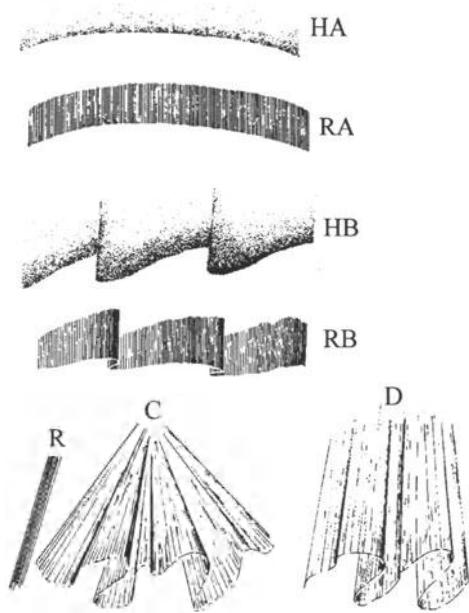


Fig. 11.7. Typical shapes of the aurora: homogeneous arc HA, rayed arc RA, homogeneous bands HB, rayed band RB, rays R, coronae C, and draperies D. Reprinted from A. Brekke and A. Egeland [11.7], *Northern lights*, Copyright 1983, Springer Verlag

can be observed almost constantly. Under normal conditions, the aurora is a colored arc extending roughly from east to west, changing its appearance in a typical pattern during the night. Under geomagnetically disturbed conditions, the aurora brightens, becomes highly structured, moves equatorwards across the sky, and changes its appearance fast. Typical auroral structures are shown in Fig. 11.7, with the arcs and bands more typical of geomagnetically quiet conditions and the draperies, rays, and corona more often observed during geomagnetic activity. The aurora is less bright than the full moon, and thus although even in mid-Europe some aurorae occur each year often they are difficult to detect because of a city's counterglow in the sky.

11.3.1 Historical Excursion

Records of the aurora can be traced back for the last 2500 years. The ancient Chinese described dragons winding in the sky, calling the aurora ‘flying dragons’. More detailed records date back to the Romans. Although made at low latitudes (Mediterranean), these observations describe the many shapes and colors normally only observed in the auroral oval. But the most frequent aurora in low latitudes is a reddish glow at the horizon, often misinterpreted as a burning farmstead or village. For instance, Seneca writes that the emperor Tiberius sent troops to the village of Ostia because an aurora evoked the impression of the village being in flames. Such misinterpretation can be traced throughout history. Even in this century reddish glows have caused





Fig. 11.8. Fantastic illustration of an aurora observed from Bamberg, Germany, in December 1560. The flashing lights in the northern sky are interpreted as sparks from clashing swords in a heavenly battle

false alarms of distant forest fires. The ancient Greeks, too, saw the aurora and named it “chasmata”, which means frightening apparition.

At middle and low latitudes the reddish color, combined with the small number of sightings, has led to interpretations of the aurora in terms of bad omens of war, famine, fire, and pestilence, or clerics interpreted it as a battle in the heavens between good and evil, with rays appearing as swords, spears, or faculae. A typical example of such an interpretation is shown in Fig. 11.8 for an aurora observed in 1560 from the German town of Bamberg.

Since aurorae were observed only occasionally, such a superstitious interpretation in an end-of-the world mood was typical of mid-Europe. Cultures, which from the very same scholars were regarded as primitive (note that at the time of the woodcut shown in Fig. 11.8 the discovery of America was already history), often had a healthier attitude towards these lights in the night sky. One of the reasons obviously is the higher frequency of occurrence for people living at higher geomagnetic latitudes, taking away much of the horror and sometimes even encouraging the search for natural explanations. In the *King's Mirror*, a Norse chronicle of 1259 (300 years before the woodcut in Fig. 11.8) were as many as three alternative explanations offered, which in the framework of the then world picture are quite reasonable:

The men who have thought about and discussed these lights have guessed at three sources, one of which, it seems, ought to be true. Some hold that fire circles about the ocean and all the bodies of water that stream about on the outer side of the globe; and since Greenland lies on the outermost edge of the Earth to the north, they think it is possible that these fires shine forth from the fires that encircle the outer ocean. Others have suggested that during the hours of night, when the Sun's course is beneath the Earth, an occasional gleam of light may shoot up into the sky, for they insist that Greenland lies so far out on the Earth's edge that the curved

surface which shuts out the sunlight must be less prominent there. But there are still others who believe (and it seems to me not unlikely) that the frost and the glaciers have become so powerful there that they are able to radiate forth these flames.

With the image of the Earth as a flat disk and the daily experience of glittering water and gleaming glaciers, all three explanations are consistent and natural.

Nonetheless, although aurorae were part of the daily experience, other cultures close to the northern Arctic Circle have offered less scientific and more mythological interpretations. Many Inuit tribes interpreted aurorae as torches in the hand of Gods leading the souls of their deceased or as the souls of the departed playing ball on the heavenly meadows. Other Eskimo tribes interpreted them as the dances of their Gods. A similar interpretation was offered by the Scotts about 2000 years ago (they called the aurorae ‘merry dancers’) or by the Aboriginals. Their neighbors, the Maori, interpreted the aurorae as fires set ablaze by their ancestors who, in their canoes, had drifted too far to the south and now had lightened fires to keep themselves warm.

But even cultures that rather often saw aurorae were not immune against superstition. The North American native Indians, for instance, also interpreted an aurora as a bad omen or fighting tribes. And even the Laps, who should be acquainted with the aurora rather well, interpreted it as a sign of violent death, as the souls of the victims.

11.3.2 Beginning of the Scientific Analysis

At the beginning of the scientific analysis, the aurora had been interpreted as the counterglow or reflection of a natural light source, for instance the reflection of the Sun, already below the horizon, from clouds or a flat surface of water, ice, or snow. Alternative interpretations included natural terrestrial sources such as volcanos, streams of lava, or bog fires. These explanations are particularly attractive for an aurora observed as a reddish glow just above the horizon, such as observed most often from mid-Europe.

This line of thought survived for rather a long time, although in the middle of the eighteenth century Cavendish, using triangulation, found the heights of the aurora to be at about 100 km. Thus the aurora cannot be explained by reflection from clouds at much lower altitudes. A very systematic analysis of the heights of the lower edges of the aurora was performed by Størmer. Figure 11.9 shows his results: aurorae are observed in a wide band of altitudes ranging from just above 80 km to more than 500 km. Nonetheless, most aurorae had their lower edge at an altitude between 90 km and 150 km.

In the 1860s, a spectroscopic analysis of the aurora by Ångström brought an end to all reflection theories. Since pressure in the high atmosphere is very low, the auroral emission consists of many forbidden lines, at that time unknown from laboratory experiments. But although these lines were identified only in the 1920s, the limitation of auroral emissions to a few lines

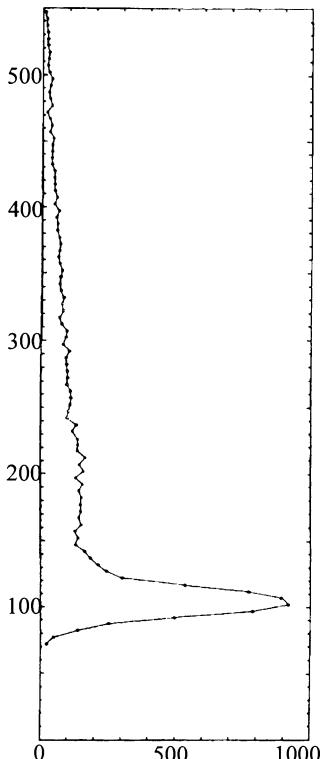


Fig. 11.9. Height distribution of 20 000 aurorae. Reprinted from C. Størmer [11.28], *The polar aurora*, Copyright 1955, with kind permission from Oxford University Press

gave evidence against the reflection of a continuous solar spectrum. Thus the aurora had been identified as an independent phenomenon, originating in the discharge of excited gases.

Hints at its origin had developed only slowly; in particular, the relationship between solar and auroral activity had long been debated. In 1730, Lewis proposed that sunspots were responsible for auroral activity because the number of aurorae and their southward extent, which also leads to higher detectability and therefore higher frequency at lower latitudes, increases with sunspot number. Since it was difficult to establish a causal chain between these phenomena, Lewis's hypothesis was disregarded. More than a century later it was revived by two different observations. In May 1859, Carrington observed a flare in white light, which not only gave rise to the first record of a flare but two days later also to a violent geomagnetic storm and strong auroral activity even at mid-latitudes. Carrington himself speculated on a relationship between the flare and the aurora; however, he also cautioned against hasty conclusions since one swallow does not make a summer. A more systematic analysis was published in 1873 by H. Fritz, cf. Fig. 11.10, confirming the close correlation between sunspot number and aurora proposed about

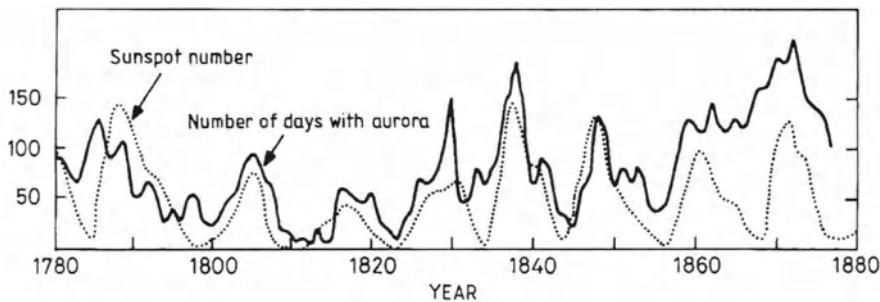


Fig. 11.10. Correlation between sunspot number and frequency of aurora. Reprinted from A. Brekke and A. Egeland [11.7], *Northern lights*, Copyright 1983, Springer Verlag

150 years earlier by Lewis. Fritz described the results of his analysis, which considered data from the past 100 years, as follows:

The aurora is a periodical phenomenon and closely related to the formation of dark spots on the Sun. The times of the richest exhibition of spots on the central body of our planetary system are characterized by rich and magnificent light phenomena around the poles of our Earth, while times of sunspot minima correspond to rareness, weak development, and small spatial extent of aurora. During these times, in the mid-latitudes of our planet the aurora vanishes while during maximum times it extends downward from both poles, occasionally even close to the equator.

Because of this close relationship between aurora and solar activity, the frequency of aurorae at mid-latitudes can be used as a proxy for solar activity, in particular when historical times are concerned where no or only few records of sunspots are available.

Fritz not only studied the frequency of aurorae but also their spatial distribution, as did Muncke, Franklin, and Loomis before him. The spatial distribution of auroral activity can be described in a map of isochasms, cf. Fig. 11.11, with an isochasm being a line of equal frequency of an unusual or frightening apparition. The isochasms were found to be ovals around the geomagnetic pole with the maximum at geomagnetic latitudes around 70° . For the auroral oval around the North Pole, the southernmost extension of about 60° latitude is at about 90° western longitude. Then the isochasm deviates northward from the parallel and goes through Baffin Bay, around the southern tip of Greenland, crosses Iceland and the northern part of Spitzbergen, reaching the highest northern latitude at about 40° eastern longitude. Then it deviates southward from the parallel, turning back to its starting point on a route across the Siberian Ice Sea and just north of the Bering Strait. Farther north or south of this line, the aurora is less frequent. Nonetheless, the southernmost northern light recorded so far was observed on 15 September 1909, from Singapore, just a few degrees north of the equator.

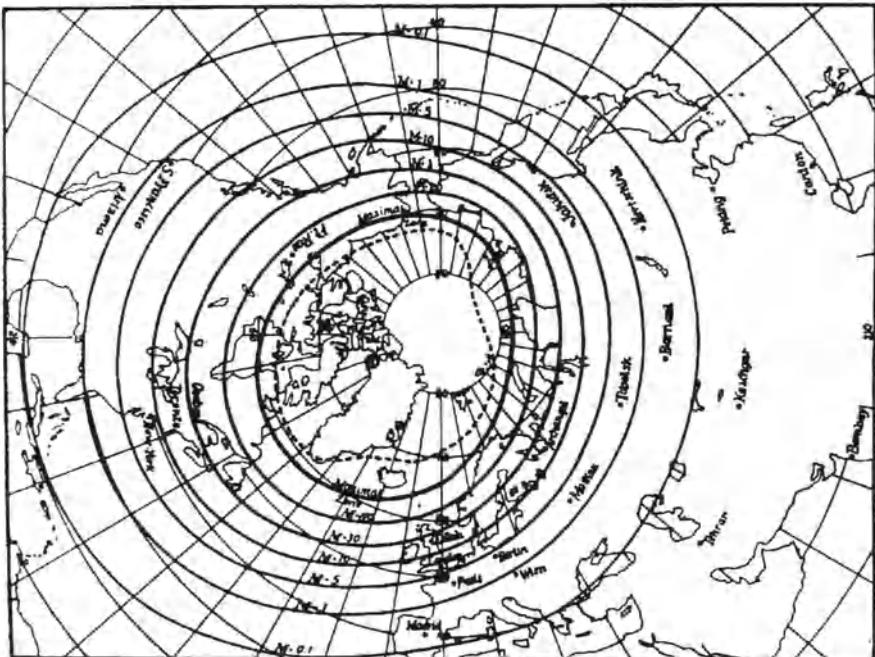


Fig. 11.11. Chart of isochasms, i.e. lines of constant aurora frequency. Reprinted from H. Fritz [11.14], *Das Polarlicht*, Copyright 1881, with kind permission from Spektrum Akademischer Verlag

Thus the spatial distribution of aurorae somehow is governed by the terrestrial magnetic field. A connection between aurora and magnetic field observations was observed even earlier. For instance, in 1749, Celsius and Hiorter published reports on the relationship between magnetic field disturbances and the aurora, wondering ‘who would have thought that there is a relation between the aurora and the compass needle, and that the northern light, if moving southwards across the zenith, can cause a deviation of the magnetic needle by some degrees?’.

11.3.3 Modern Interpretation

A simple, though incomplete, analogy of our understanding of the aurora is the cathode-ray tube in a television set: particles accelerated at the cathode (during a substorm in the plasma sheet in the magnetotail) are deflected by a magnetic field (the terrestrial field) and emit light on hitting a fluorescent screen (the upper atmosphere). Early this century, long before the development of the TV, Birkeland built the terrella, a model that allowed simulation of the auroral distribution in the laboratory. It completely illustrates this analogy: a cathode-ray tube in a vacuum chamber emits electrons towards a

sphere, the terrella, covered with a fluorescent material. If an electromagnet inside the terrella is switched on, the terrella emits light only from two circles around its poles, while with the electromagnet turned off, light is emitted from all over the hemisphere viewing the cathode-ray tube.

Although this analogy is vivid, it is incomplete. The basic difference between the aurora and a TV set lies in the motion of the particles just before the screen. As the particles are accelerated in the plasma sheet, they propagate towards the Earth, are deflected by the magnetic field from equatorial regions towards higher latitudes, and then have to penetrate into the atmosphere to excite the atoms and molecules which in turn emit the light seen as the aurora. The fundamental problem in this chain of events is the propagation of the particles deep into the atmosphere. Moving closer towards the Earth, the particles find themselves in a converging magnetic field. The constancy of the magnetic moment then leads to a reflection back towards the magnetosphere. This reflection occurs at heights of about 1000 km, where densities are too small for recognizable light emission. Particles could penetrate deeper into the atmosphere if they were accelerated during their motion. Although the details of the acceleration are still under debate, the spectra of electrons measured by rockets at heights of some hundreds of kilometers clearly give evidence for it: outside of auroral arcs, the spectrum is roughly a power law up to energies of at least 10 keV. Over an auroral arc, on the other hand, there is a pronounced peak superimposed on this power law, exceeding the power law intensities by up to two orders of magnitude at energies of a few kiloelectronvolts. Thus the additional electron component is nearly monoenergetic which strongly suggests acceleration in an electric field.

Today it is assumed that at heights of some thousands of kilometers a potential structure develops along the field line with a higher positive potential at lower altitudes. Although such a structure would explain the observed electron spectra, an explanation of the structure itself is difficult. Since the plasma is assumed to be collisionless, the conductivity σ_{\parallel} parallel to the magnetic field is infinite according to (10.20) and the magnetic field lines are equipotential lines. All parallel electric fields E_{\parallel} would be canceled immediately.

Different processes for the development of the potential structure are discussed, cf. [11.4, 11.15]. All mechanisms agree that the fundamental driving mechanism is an increase in the magnetospheric convection due to changes in solar wind properties. One of the mechanisms under discussion is the development of double layers. A double layer forms when currents are flowing between plasmas with different properties. In geospace, these are the cold and dense ionospheric plasma and the hot and rarefied magnetospheric plasma. For a double layer to form and to be stable, a field-parallel current must already flow. In the high-latitude ionosphere this would be the Birkeland current, connecting the ionosphere with the magnetosphere. While it is difficult to describe the formation of a double layer, we can at least explain how



it can stay stable. If the double layer has formed, a potential drop exists. The particle populations encountering this drop can be divided into electrons and ions, ionospheric and magnetospheric plasma, and thermal and suprathermal particles. Let us start with the ionospheric population. This is a thermal population, consisting of ions and electrons. Ions moving upward from the ionosphere towards the double layer see a decreasing potential and thus are accelerated, leaving the ionosphere through the double layer. Ionospheric electrons, on the other hand, are reflected back into the ionosphere. For the magnetospheric plasma, the situation is just the opposite: flowing towards the Earth it encounters a positive potential drop, leading to an acceleration of electrons towards the ionosphere while the ions are deflected. In addition, electrons are created inside the double layer by the interaction between the accelerated magnetospheric ions and neutrals. These electrons are accelerated downwards, too. The net effect is an acceleration of the electrons without destroying the potential drop. Note that this works only when the flow speed of the magnetospheric electrons is larger than their thermal speed. Thus a current already flows. This is the upward Birkeland current shown in Fig. 10.24. This explains the earlier statement that an aurora only is observed where the Birkeland current flows from the ionosphere into the magnetosphere, i.e. where electrons move from the magnetosphere to the ionosphere.

A modification of the double layer concept is the electrostatic shock. This is a double layer which is not stationary like the one discussed above but moves with the average ion speed along the magnetic field line.

Observations suggest that a different mechanism also is important in the electron acceleration: the pitch-angle distributions of electrons and ions above auroral arcs are different. Thus both populations are reflected at different positions in the converging magnetic field. Since the particles stay at the reflection point for rather a long time (their velocity parallel to the field vanishes), a charge separation and thus an electric field results which in turn accelerates the lighter species, the electrons.

On hitting the denser atmosphere, the electrons cause electromagnetic emissions due to excitation ($M + e^- \rightarrow M^* + e^-$) or excitation and ionization ($M + e \rightarrow M^{**} + 2e^-$) of the neutrals. Here M denotes an atmospheric constituent, such as N, N₂, O, and O₂. Auroral lines are emitted in the entire range from UV to IR. The most intense lines in the visible are the green oxygen line at 557.7 nm and the red double line of oxygen at 630 nm and 636.4 nm. Since both are forbidden lines, early observers were not able to identify them. Another intense line is emitted by the nitrogen molecule at 427 nm, a weaker one at 470 nm. Both lines result from transitions between different vibrational states and can be observed only at the lower edge of the aurora since nitrogen molecules are rare at heights above 120 km, cf. Fig. 11.9. Table 11.1 summarizes the most important auroral lines and the heights where they are emitted.

Table 11.1. Frequent lines in the auroral emission. The H α line is only observed in proton aurorae

Wavelength (nm)	Emitting species	Altitude (km)	Visual color
391.4	N $^+$	1000	violet-purple
427.8	N $^+$	1000	violet-purple
557.7	O	90–150	green
630.0	O	>150	red
636.4	O	>150	red
656.3	H α	20–60	red
661.1	N $_2$	65–90	red
669.6	N $_2$	65–90	red
676.8	N $_2$	65–90	red
686.1	N $_2$	65–90	red

Not only electrons but also protons can excite the neutral atmosphere. Then the excitation is due to charge exchange: the proton is decelerated and becomes an excited hydrogen atom. This, in turn, emits either the L α line in UV or the H α line in the red. Thus proton aurorae are always red. In addition, they are less structured than electron aurorae, cover larger areas, and are observed only at quite high altitudes, i.e. between 300 km and 500 km. Proton and electron aurorae can occur simultaneously. A special example of a proton aurora is the polar glow that forms when solar protons penetrate into the dayside magnetosphere along the cusps.

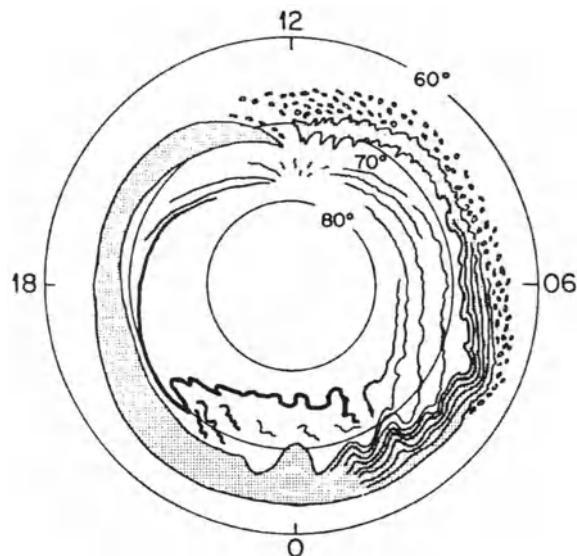


Fig. 11.12. Shape of the aurora in dependence on local time. Reprinted from Schlegel [11.25], in *Plasmaphysik im Sonnensystem* (eds. K.-H. Glassmeier and M. Scholer), Copyright 1992, with kind permission from Spektrum Akademischer Verlag

But the aurora is not only limited to the cusps and the nightside. Satellite observations indicate that often a closed auroral oval can be observed and observations from the ground show that auroral activity is not only limited to a few hours around local midnight but can be observed during the entire night (and during the polar night even at day-time). In this case a variation of the aurora with local time can be observed, as shown in Fig. 11.12. In the shaded area between noon and midnight, diffuse aurora form or quiet and stable arcs. After about 20 LT, these arcs become more wavy, forming complex patterns of bands which particularly during a substorm, move across the sky in a westward traveling surge. During the decay phase of a substorm, these bands resolve into isolated patches traveling eastward. These patches frequently are observed during the morning hours. The small arcs on the dayside at latitudes of about 75° are formed when solar particles or the solar wind penetrate into the polar cusps.

A special case is the theta aurora. It occasionally can be observed from high-flying satellites as a closed auroral oval supplemented by an arc extending across the polar cap from the dayside to the nightside, giving the aurora the shape of the greek letter Θ . It is observed only at times of a northward interplanetary magnetic field, i.e. at times of a closed magnetosphere. The existence of this arc is difficult to understand since the field lines from the polar cap connect back to the lobes where the plasma density is very low. Current interpretations involve boundaries along the Sun-Earth line with Pederson currents converging over the caps at this boundary and then flowing upwards.

11.4 Summary

The dynamic magnetosphere manifests itself in geomagnetic disturbances and aurorae. The energy source for these phenomena is the solar wind; the coupling between the solar wind and the magnetosphere is due to magnetospheric convection. The level of geomagnetic activity is determined by the solar wind fluctuations and the direction of the interplanetary magnetic field. If the latter has a southward component, the magnetosphere has an open configuration, more energy and plasma are fed into it, and its responses to solar wind variations are much stronger. The variations in the solar wind responsible for geomagnetic disturbances are fast solar wind streams, often leading to recurrent geomagnetic disturbances, as well as shocks and the ejecta driving them. While the recurrent disturbances can best be observed during the solar minimum when stable and recurrent fast and slow solar wind streams have developed, geomagnetic disturbances in response to transient phenomena are much more frequent during the solar maximum, since flares, CMEs, and interplanetary shocks are more frequent at these times.

12. Planetary Magnetospheres

Empty space is like a kingdom, and heaven and earth are no more than a single individual person in that kingdom... How unreasonable it would be to suppose that besides the heaven and earth which we can see there are no other heavens and no other earths?

Tang Mu, thirteenth century

A magnetosphere is not a typical terrestrial phenomenon. Instead, magnetospheres can be found around all magnetized bodies surrounded by a plasma flow. Even around unmagnetized bodies (comets, planets without a magnetic field) a cavity is formed by the interplanetary magnetic field frozen into the deflected solar wind flow. In the solar system, all planets except Mars and Venus have a magnetosphere. Although they are different sizes, for most of them the shape and the properties are similar to those in the terrestrial magnetosphere. Special features are the large size and the flat inner structure of the Jovian magnetosphere and the oscillation of Neptune's magnetosphere between a pole-on and an Earth-like.

12.1 Overview

12.1.1 The Planets

Table 12.1 lists the planets with their distance from the Sun, their sizes, and basic orbital parameters. The planets are divided into two groups, the inner, Earth-like planets Mercury, Venus, Earth, and Mars, and the outer, gaseous giants Jupiter, Saturn, Uranus, and Neptune. The Earth-like planets all have comparable radii (Mercury is the smallest, its radius is only about one-third of the Earth's radius). Since they all have a solid crust and a heavy, iron-rich core, their densities are relatively high, ranging between 4 g/cm^3 and 6 g/cm^3 . All Earth-like planets are surrounded by atmospheres, although these atmospheres are vastly different in density, composition, and temperature. For instance, Mercury has a very thin atmosphere which can hardly be recognized as an atmosphere, while the carbon-dioxide atmosphere on Venus

Table 12.1. Properties of the planets in the solar system

	Solar distance (AU)	Sidereal period	Spin period (days)	Average density (gm/cm ³)	Surface gravity (N/kg)
Mercury	0.39	88 d	56.8	5.4	3.6
Venus	0.72	225 d	243	5.1	8.7
Earth	1.00	365 d	1	5.5	9.8
Mars	1.52	1 yr 322 d	1.03	4.0	3.7
Jupiter	5.20	11 yr 315 d	0.41	1.3	26.0
Saturn	9.55	29 yr 167 d	0.44	0.7	11.2
Uranus	19.22	84 yr 7 d	0.72	1.18	9.4
Neptune	30.11	164 yr 280 d	0.67	1.56	15.0

is very thick, leading to a strong greenhouse effect with surface temperatures of about 750 K. The inner planets have few or no moons (Earth has one, Mars has two) and no rings.

The outer planets, Jupiter, Saturn, Uranus, and Neptune, again are similar. They are large, gaseous giants, mainly consisting of hydrogen and helium. Thus their average densities are low with values close to 1 g/cm³. From the outer atmosphere towards the center of the planet, the pressure increases up to more than a million times the surface pressure on Earth. Here the hydrogen first becomes liquid and later metallic. It is still under debate whether the outer planets have solid cores. They have many moons and ring systems affecting the plasma and particle populations in the magnetospheres.

The energetics of the outer planets are surprising: they radiate more energy back into space than they receive from the Sun. Jupiter radiates back twice the solar energy; on Saturn this ratio is about 3.5. The additional energy either is gravitational, released during the contraction of the planet, or a remnant from the creation of the solar system. Thus the large gaseous planets are more similar to the Sun than to the inner planets.

The outermost planet, Pluto, cannot be fitted into this scheme: it appears to be a solid piece of rock but shows no similarities with the gaseous giants. Pluto's orbit is unusual, too. It is highly eccentric and inclined with respect to the plane of the ecliptic. It is likely that Pluto was not formed together with the solar system but is a captured asteroid.

A review of planetary magnetospheres can be found in [12.4], and the magnetospheres of the outer planets are discussed in [12.14].

12.1.2 Structures of Planetary Magnetospheres

Although the physical mechanisms shaping a magnetosphere are the same for all planets, the properties of these magnetospheres can be different, depending on the magnetic moment of the planet, its rotation period, plasma sources and sinks, and the local solar wind properties.

Table 12.2. Parameters of planetary magnetic fields and magnetospheres. The column ‘angle between axes’ gives the angle between the planet’s magnetic field axis and its rotation axis. In the column ‘plasma sources’ the ‘W’ indicates solar wind, ‘A’ indicates planetary atmospheres and ionospheres, and ‘S’ satellites

	Equatorial field (gauss)	Dipole moment (gauss cm ³)	Angle between axes	Typical stand-off distance (r_p)	Calculated stand-off distance	Plasma sources
Mercury	0.002	3×10^{22}	< 10°	1.45	1.74	W
Venus	<0.0003	< 10 ²¹	?	1.1	?	W,A
Earth	0.305	7.9×10^{25}	11.5°	10.7	10.7	W,A
Mars	0.0004?	1.4×10^{22}	?	?	?	?
Jupiter	4.2	1.5×10^{30}	9.5°	47–97	45	W,A,S
Saturn	0.2	4.3×10^{28}	< 1°	17–24	20	W,A,S
Uranus	0.23	3.8×10^{27}	58.6°	18–25	26	A
Neptune	0.06–1.2	2×10^{27}	46.8°	23–26.5	25	W,A,S

In front of all magnetospheres a bow shock develops where the supersonic solar wind is slowed down to subsonic speeds. Behind the bow shock, structures are different. The two neighbors of Earth, Mars and Venus, both have only weak magnetic fields (cf. Table 12.2). Thus no well-structured magnetosphere develops, instead an ionopause/magnetopause separates the interplanetary and planetary plasmas. This does not result from an interaction between the solar wind and the planetary magnetic field but between the solar wind and the planetary ionosphere. The interplanetary magnetic field then is frozen-out of the planet’s ionospheric plasma.

The most common type of magnetosphere is Earth-like. It also can be found at Mercury, Saturn, Uranus, and, at least for part of its rotation, around Neptune. The complexity of these magnetospheres depends on the angles of the planetary dipole and the axis of rotation with respect to each other and to the solar wind flow. The sizes of the Earth-like magnetospheres are different, depending on the planet’s magnetic moment.

The most complex magnetosphere can be found on Jupiter: although the axis of rotation and the magnetic field axis parallel to each other and are nearly perpendicular to the plane of the ecliptic, the magnetosphere is far more complex than the Earth-like ones because Jupiter’s fast rotation causes centrifugal forces, stretching the magnetosphere outward in the equatorial plane.

The parameters of the planetary magnetospheres, such as the magnetic moment, the angle between the magnetic field axis and the axis of rotation, the typical distance of the magnetopause in planetary radii r_p as observed and as calculated from the pressure balance, as well as plasma sources are listed in Table 12.2. In the last column, the plasma sources are ‘W’ for the solar wind, ‘A’ for planetary atmospheres and ionospheres, and ‘S’ for the

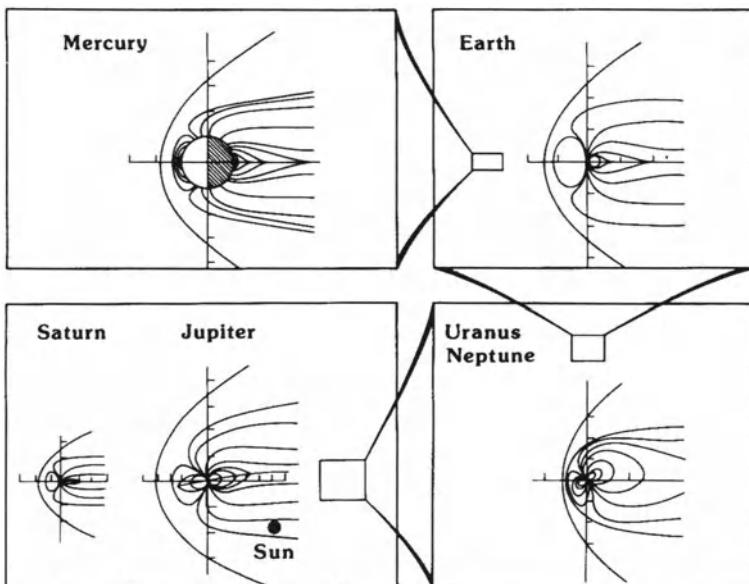


Fig. 12.1. Relative sizes of the planetary magnetospheres. Reprinted from C.T. Russell and R.J. Walker [12.14], *Introduction to Space Physics*, (eds. M.G. Kivelson and C.T. Russell), Copyright 1995, with kind permission of Cambridge University Press

planetary satellites. They will be explained in more detail in the discussion of the planets. Note the relation between the rotation period and the dipole moment: with decreasing rotation period, the dipole moment increases because the MHD dynamo works more efficiently. Mars is an exception to this rule: most likely, the planet has cooled down too much to allow for a liquid core. Thus the dynamo process does not work. Nonetheless, the visit of Mars Global Surveyor to Mars revealed a magnetic field close to the planet about an order of magnitude larger than expected, and in size comparable to Mercury's field. But this field does not result from a planetary dynamo, but rather appears to originate in ferromagnetic rocks in the planet's crust[12.1].

12.1.3 Sizes

Jupiter not only has the most complex magnetosphere but also has the largest one. This is true not only if the subsolar distance of the magnetopause is expressed in units of the planet's radius but, since Jupiter is the largest planet, also in absolute numbers. The Jovian magnetosphere is also the largest structure in the solar system, even dwarfing the Sun in size, cf. Fig. 12.1.

One parameter that defines the size of a magnetosphere is the stand-off distance expressed in units of the planetary radius r_p . This 'typical magnetopause distance' is given twice in Table 12.2: first as the observed value and

second as calculated from the balance between the magnetic pressure of the planetary field and the kinetic pressure of the solar wind. Both values agree quite well, and the variability of the observed magnetopause distances reflects the high variability in solar wind pressure.

The typical magnetopause distances order with the magnetic moment of the planet: the inner planets all have very small magnetospheres, extending less than one planetary radius above the planet's surface. The only exception is Earth with its relatively high magnetic moment. The outer planets have even higher magnetic moments, corresponding to larger stand-off distances. Although at the outer planets the stand-off distances expressed in planetary radii are not much different from the Earth's, these magnetospheres are much larger because the planets themselves are larger, cf. Fig. 12.1.

12.1.4 Plasma Sources

A magnetosphere often is called a magnetic cavity: the solar wind plasma cannot penetrate into this cavity. Nonetheless, as we have already seen in Chap. 10, the cavity is not void of any plasmas or particle populations. Plasma sources in the Earth's magnetosphere are the solar wind and the terrestrial atmospheres and ionospheres, cf. last column in Table 12.2. The same sources are available in Venus's magnetosphere. Mercury, on the other hand, which is too small and too close to the Sun to support its own atmosphere, picks up all the magnetospheric plasma from the solar wind.

Going farther out to the gaseous giants, we find another plasma source: for these planets at least some of the moons and the ring systems are inside the magnetosphere. These plasma sources easily can be identified by their unusual composition: in Chap. 10 we mentioned that the relatively high contribution of O^+ points to a partly atmospheric origin of the terrestrial magnetospheric plasma. Moons and rings can contribute to the magnetospheric plasma populations by even more complex mechanisms, occasionally leading also to more exotic species. (a) On hitting the solid surface of a moon, energetic ions can knock out neutral atoms. These neutrals become ionized by charge-exchange with the surrounding plasma. This process, called sputtering, appears to be relevant in the inner magnetosphere of Saturn. (b) Interactions between the dust of the planetary ring system and the surrounding magnetospheric plasma might create an additional plasma component. This process also is discussed as a likely explanation of the spokes in Saturn's rings. (c) Vulcanism on a planetary moon directly injects particles into the magnetosphere. A prominent example is Io, a small moon in the Jovian system which is kneaded so thoroughly by Jupiter's strong gravitational pull that its core is molten, allowing for vulcanism. Io mainly injects S^+ , S^{2+} , and O^+ into the Jovian magnetosphere. In time, a plasma torus has formed along Io's orbit, consisting of these ions. (d) The atmospheres of moons also directly inject particles into planetary magnetospheres. In the solar system, two moons with atmosphere are known: Titan, with an atmosphere basically consisting of nitrogen

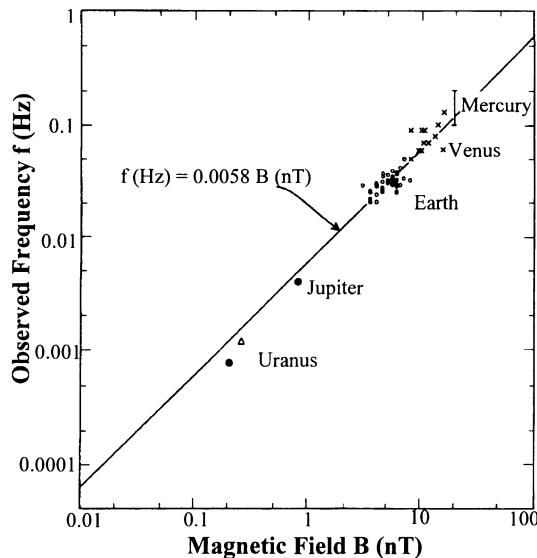


Fig. 12.2. Frequencies of waves upstream of planetary magnetosphere in dependence of the interplanetary magnetic field strength. Reprinted from C.T. Russell et al. [12.13], *J. Geophys. Res.* **95**, Copyright 1990, American Geophysical Union

and methane with methane being available in all three phases, comparable with the water on Earth, orbits around Saturn, and Triton, with a thin atmosphere of nitrogen and methane, revolves around Neptune. Along their orbits, these moons have also created plasma tori; however, the intensities are much smaller than the ones observed along Io's orbit. (e) And finally, moons also provide a loss mechanism for the magnetospheric plasma: on hitting the moon's surface, a particle might be absorbed.

12.1.5 Upstream of the Bow Shock: The Foreshocks

Since all planets in the solar system have a bow shock, the shock enthusiast might wonder whether all these shocks accelerate particles. If they do, we would expect to observe a foreshock with plasma turbulence and energetic particles similar to the one in front of the terrestrial bow shock, cf. Sect. 9.4. And indeed, a foreshock can be found in front of all planets. But like the magnetospheres themselves, these foreshocks are variable. Their properties depend on the angle between the interplanetary magnetic field and the bow shock. Since the curvature of the Archimedean field line increases with increasing radial distance, the angle between the field line and the bow shock changes, too. With increasing radial distances other parameters also change, such as the strength of the interplanetary magnetic field and the temperature and kinetic pressure of the solar wind plasma.

Despite these vastly different local conditions, all planets have foreshocks. Although they differ from the terrestrial foreshock, they nonetheless show the same general features, such as a resonance between waves and particles and, depending on θ_{Bn} , regions with different particle distributions, reflecting

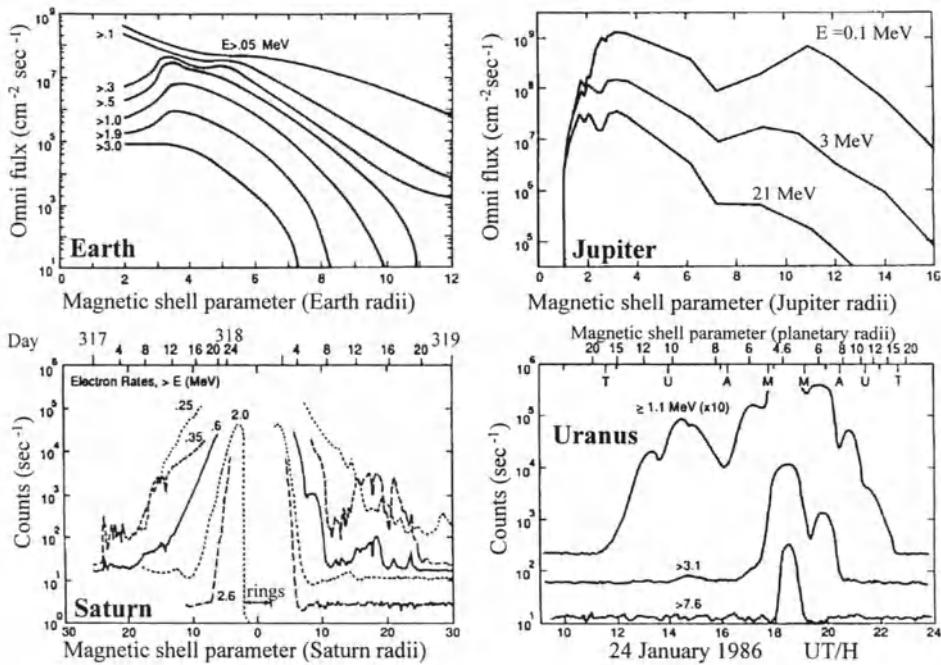


Fig. 12.3. Fluxes of energetic electrons in the magnetospheres of Earth, Jupiter, Saturn, and Uranus. Reprinted from C.T. Russell and R.J. Walker [12.14], in *Introduction to Space Physics* (eds. M.G. Kivelson and C.T. Russell), Copyright 1995, with kind permission from Cambridge University Press

the underlying acceleration mechanisms. Except for the Jovian electrons, the accelerated particles have rather low energies in the kiloelectronvolt to tens of kiloelectronvolt range. Depending on the planet's distance from the Sun and the magnetic field strength at this distance, the frequencies of the waves resonating with the particles change: the lowest frequencies are observed at the outermost planets while the foreshock of the innermost planet shows the highest frequencies, cf. Fig. 12.2.

12.1.6 Radiation Belts

Except for Jupiter, all the outer planets have radiation belts that closely resemble the terrestrial one. Radial diffusion transports particles perpendicular to the magnetic field and pitch-angle scattering scatters particles into the loss cone from where they can penetrate deep into the atmosphere. The Jovian magnetosphere is different: Io lies deep inside the magnetosphere and is a significant source of matter and energy. The resulting radiation belt is humangeous compared with the typical radiation belts of Earth or the other outer planets. It is not only the large spatial extent, but also the high energies

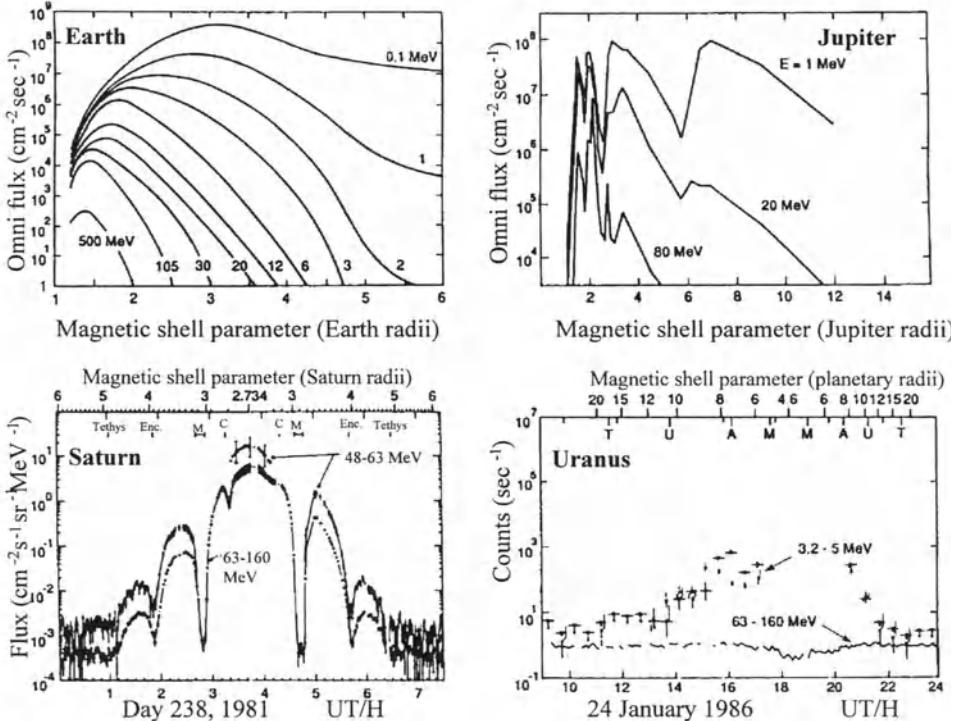


Fig. 12.4. Fluxes of energetic protons in the magnetospheres of Earth, Jupiter, Saturn, and Uranus. Reprinted from C.T. Russell and R.J. Walker [12.14], in *Introduction to space physics* (eds. M.G. Kivelson and C.T. Russell), Copyright 1995, with kind permission from Cambridge University Press

and fluxes of the particles which make the Jovian radiation belts a hazard even for unmanned spacecraft.

Figures 12.3 and 12.4 show cuts through the radiation belts of electrons and protons for Earth, Jupiter, Saturn, and Uranus. Neptune's radiation is not shown; it is very similar to the one of Uranus, although the particle fluxes are significantly lower. The radiation belts of the different planets are quite similar: the highest fluxes always are observed just above the planetary atmosphere. Saturn is an exception to this rule in so far as the maximum of the radiation belts is just outside the ring system. In all magnetospheres particles with high energies are limited to the inner radiation belts while lower energy particles can also be observed farther out. Thus the energy spectrum is hardest close to the planetary atmosphere; with increasing radial distances the spectrum steepens. Particle fluxes and energies, on the other hand, are quite different. For instance, the flux of >3 MeV electrons in the Jovian magnetosphere exceeds the one in the Earth's atmosphere by 3 orders of magnitude, which in turn is still one order of magnitude higher than in the magnetosphere

of Uranus. In addition, the electron energies in the Jovian magnetosphere are much higher than in the other magnetospheres. The differences in energies and fluxes of protons are even more pronounced.

Magnetospheric size cannot be the parameter responsible for the different fluxes and energies because in the rather small terrestrial magnetosphere higher fluxes and more energetic particles can be observed than on Saturn, Uranus, and Neptune. The differences more likely result from the interaction between the magnetosphere and the solar wind as the ultimate source of energy. In particular, differences in the reconnection rates at the magnetopause might feed more energy into the Earth and Jupiter systems, which in turn leads to more efficient acceleration. Loss mechanisms for the accelerated particles, in particular pitch-angle scattering into the loss cone, might modify the particle distributions; however, they are not likely to explain the large differences in energy and fluxes.

12.2 Planets with a Magnetic Field

Aside from the outer gaseous giants, only Earth and Mercury have a magnetic moment strong enough to support a magnetosphere. We will now discuss in more detail the structures of the magnetospheres together with the plasma populations and their relationships to moons and rings.

12.2.1 Mercury

Mercury's magnetosphere is rather simple because both the planet's rotation axis as well as the magnetic field axis are roughly perpendicular to the plane of ecliptic. Since Mercury's magnetosphere is very small with a stand-off distance of only $1.1 r_p$, it neither accommodates radiation belts nor a plasma population.

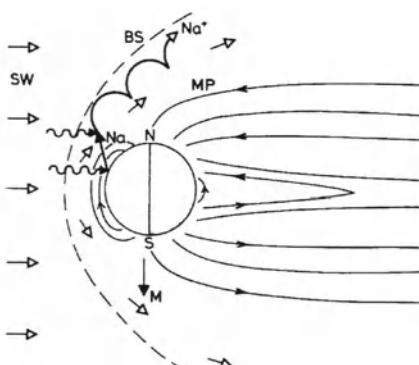


Fig. 12.5. Structure of Mercury's magnetosphere. Reprinted from D.A. Bryant [12.4], in *Plasma physics* (ed. R. Dendy), Copyright 1993, with kind permission from Cambridge University Press

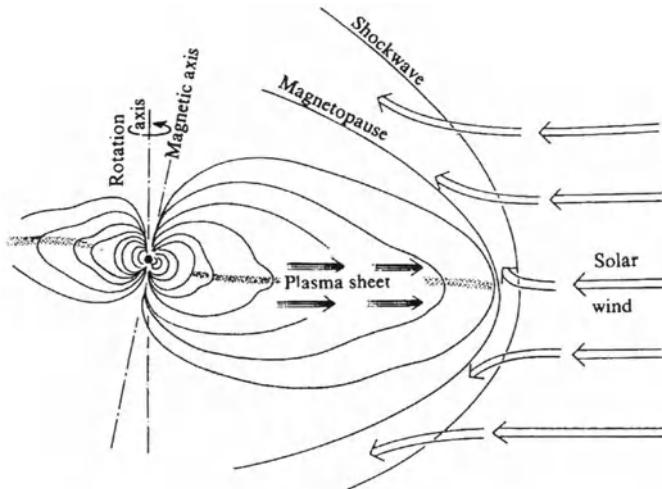


Fig. 12.6. Jovian magnetosphere: the planet's fast rotation leads to a concentration of plasma in the equatorial plane which carries out the magnetic field. Reprinted from J.A. Simpson and B.R. McKibben [12.15], in *Jupiter* (ed. T. Gehrels), Copyright 1976, with kind permission from The University of Arizona Press

Despite this simplicity, Mercury's magnetosphere is very dynamic. Compared with the other planets, Mercury's orbit is highly eccentric ($\varepsilon = 0.206$) with a perihelion¹ at 0.308 AU. In the perihelion, fluctuations in the solar wind can create such a strong kinetic pressure that the typical magnetopause distance is smaller than the planet's radius. Thus a magnetosphere in the sense of a planetary magnetic field enclosing the planet no longer exists, and the magnetic field is swept into the magnetotail. This variability of the magnetosphere makes it impossible to support stable radiation belts.

Nonetheless, Mercury's magnetosphere contains a thermal plasma consisting mainly of H, He, O, Na, and K ions as well as electrons. The lighter species, H and He, are swept up out of the solar wind. The heavier ions are created on the dayside of the planet, as indicated in Fig. 12.5: solar UV radiation, and to a lesser extent also the solar wind, knock Na, K, and O out of the planet's surface. These neutrals immediately become ionized by the UV radiation and then are kept in the planet's magnetic field.

The tail of Mercury's magnetosphere even shows dynamic phenomena which closely resemble the substorms known from the terrestrial magnetosphere. Because Mercury has neither an atmosphere nor an ionosphere, this observation indicates that the ionosphere is not a necessary condition for the occurrence of substorms. For the planetary consequences of a substorm, nonetheless, the existence of an ionosphere makes a difference.

¹ In an elliptical orbit the perihelion is the position closest to the Sun. Its counterpart is the aphelion.

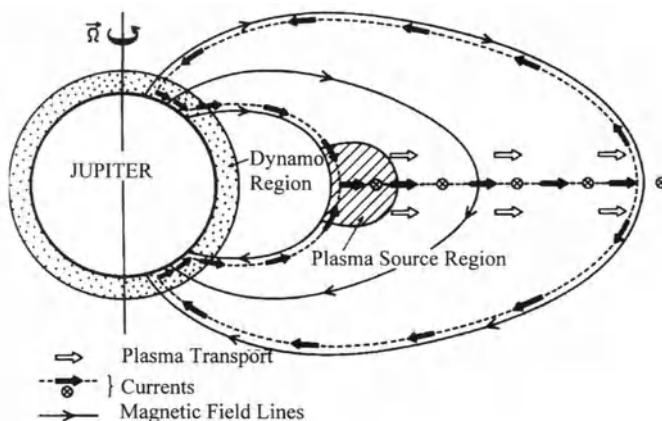


Fig. 12.7. Plasma flows, currents, and magnetic field lines in the inner Jovian magnetosphere. Based on F.M. Neubauer [12.10], in *Plasmaphysik im Sonnensystem* (eds. K.-H. Glassmeier and M. Scholer), Copyright 1991, reprinted with kind permission from Spektrum Akademischer Verlag

12.2.2 Jupiter

While Mercury's magnetosphere is the smallest and simplest inside the solar system, the Jovian one not only is the largest but also the most complex one. A very attentive and detailed description based on the Pioneer and Voyager fly-bys can be found in [12.5]; the first results obtained by Galileo are summarized in *Science* **272**. In addition, the Jovian magnetosphere appears to be highly variable, as can be inferred from a comparison of the magnetic field data obtained by Voyager in 1979, Ulysses in 1992, and Galileo in 1995.

The Jovian planetary magnetic field can be approximated as a dipole with its dipole axis inclined by 11° with respect to the planet's rotation axis. In addition, the magnetic dipole is slightly offset with respect to the axis of rotation. The magnetic moment of Jupiter exceeds that of Earth by four orders of magnitude, and the magnetic field at the planet's surface is about one order of magnitude larger than the terrestrial field. Since also the solar wind intensity has decreased markedly between Earth and Jupiter, we expect a large magnetosphere. The fast rotation of the planet with a period of slightly less than 10 h adds another feature to the Jovian magnetosphere: the centrifugal forces in the equatorial plane are very strong, resulting in an outward plasma flow. The planetary magnetic field frozen into this plasma is carried out, too, leading to the magnetic field configuration shown in Fig. 12.6. Therefore the inner Jovian magnetosphere is much flatter than the dipole-like inner magnetospheres of the other planets.

This flattening of the magnetosphere is associated with a current system, cf. Fig. 12.7: the outward plasma flow is accompanied by an electric current directed radially outward. Without this current the plasma would not

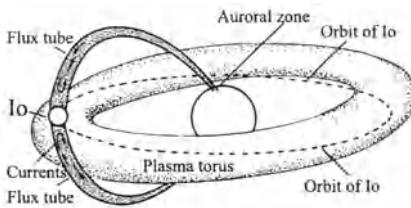


Fig. 12.8. Plasma torus around Io's orbit and flux tubes connecting to the auroral regions. Because the planet's magnetic axis is tilted with respect to the axis of rotation, Io weaves up and down within the plasma torus. Reprinted from K. Lang and C.A. Whitney [12.9], *Wanderers in space*, Copyright 1991, with kind permission from Cambridge University Press

corotate with the planet but stay behind. The current system is closed by field-parallel currents through the dynamo region in Jupiter's ionosphere. An additional current, separating the fields of opposite polarity, is flowing around Jupiter in the equatorial plane. Currents inside the Jovian magnetosphere can exceed 10^9 A .

Within about 15 planetary radii, the dipole field is dominant. In this region, the radiation belts are observed, cf. Figs. 12.3 and 12.4. The particles gyrating inside the radiation belts emit synchrotron radiation with frequencies up to some megahertz, making Jupiter a strong radio source.

Some of Jupiter's moons are inside the magnetosphere. Since the relative speed between the moons and the magnetospheric plasma is smaller than the Alfvén speed, no bow shocks develop in front of the moons. They are unmagnetized bodies, and thus no magnetospheres develop around them. Therefore Jupiter's magnetospheric plasma can interact freely with the moon's surface. Since absorption exceeds the sputtering, the particle intensities in the radiation belts are reduced at the moon's orbits, cf. Figs. 12.3 and 12.4.

One moon, Io, is also a plasma and particle source: vulcanism on Io injects particles, in particular S, cf. Table 12.3. The matter released by Io immediately becomes ionized and is accelerated to the speed of the corotating magnetosphere. Small differences in the initial conditions of each particle and the interaction between charged and neutral particles lead to the formation of a torus along the moon's orbit instead of just a cloud of ionized gas around the moon itself, cf. Fig. 12.8. Charged particles flowing field-parallel out of this torus create Whistler waves and, on interaction with Jupiter's ionosphere, a glowing aurora in the polar regions. Io's plasma torus has a sharp inner boundary, while its outer boundary is less well-defined and extends outwards into the magnetosphere.

12.2.3 Saturn

Saturn's magnetosphere again is a rather simple one. It is axial symmetric because both the magnetic field axis and the axis of rotation are perpendicular to the plane of ecliptic. Since Saturn's magnetic field is much weaker than Jupiter's, the magnetosphere is smaller, more closely resembling the one at Earth. Plasma tori with different dominant species exist, as indicated

Table 12.3. Sources of thermal plasma in the magnetospheres of the outer planets. The contributions of the planetary ionospheres and moons can be identified by their unusual composition, in particular the large amount of ions heavier than He

	Jupiter	Saturn	Uranus	Neptune
Sources	Io ionosphere	ionosphere, icy moons, solar wind, Titan	ionosphere, H-corona	ionosphere, solar wind, Triton
Composition	H, Na, O, K, S, O, H ₂ O, N	H, OH, H ₂ , O,	H	H, N
Source strength	$\sim 3 \times 10^{28}$ i/s	10^{26} i/s	10^{25} i/s	3×10^{25} i/s
Life time	months, years	months	days	weeks

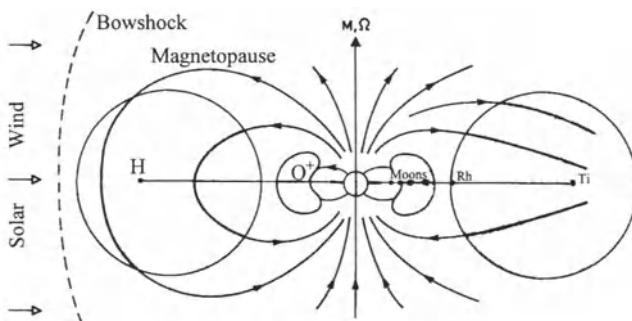


Fig. 12.9. Saturn's magnetosphere with the hydrogen torus created by Titan, the oxygen torus of Thetis, and the positions of the inner moons and the ring system. Reprinted from D.A. Bryant [12.4], in *Plasma Physics* (ed. R. Dendy), Copyright 1993, with kind permission from Cambridge University Press

in Fig. 12.9, the densities, however, are much smaller than in the Jovian magnetosphere.

Saturn's magnetosphere can be divided into four distinct parts. The outer part is filled with a plasma corotating with the planet. The highest plasma density is observed at about 6 planetary radii, which is still deep inside the magnetosphere. The plasma mainly consists of N, O, and OH. Part of this plasma most likely originates in the ring system, and trapped particles in the sense of a radiation belt are not observed.

Moving further inwards, the radiation belts start and an observer encounters the slot region, extending from about $6.5 r_p$ inwards to about $4 r_p$. Here the particle fluxes are reduced because of absorption by the moons Dione, Thetis, and Enceladus, cf. Figs. 12.3 and 12.4. In the region between 4 and $3.1 r_p$, i.e. between the orbit's of Enceladus and Mimas, the fluxes of radiation belt particles increase again. In addition, the energy of the particles increases. Farther in, in the region occupied by the ring system, particle fluxes drop sharply.

Titan, one of Saturn's moons, has a relatively thick atmosphere, mainly consisting of methane and nitrogen. As Titan loses particles out of this atmosphere, it creates a torus around its orbit with a width of about 2 Saturn radii inside the orbital plane and 7 to 23 planetary radii perpendicular to it. The interaction between Titan and Saturn's magnetosphere is not as continuous as the one between the Jovian magnetosphere and Io because part of Titan's orbit lies outside the magnetopause in the magnetosheath or even in interplanetary space.

Saturn's magnetosphere contains another plasma component which we have not encountered in the other magnetospheres: strongly ionized dust particles with sizes between a few microns and the size of a small moon. These dust particles modify the plasma properties. Surrounded by an electron cloud, they behave like the nuclei of large quasi-atoms.

12.2.4 Uranus

Uranus is a special planet in the solar system in so far as its axis of rotation nearly parallels the plane of ecliptic. Thus one pole of the planet always is directed towards the Sun. For the other planets, the axis of rotation is more or less perpendicular to the ecliptical plane, corresponding to our concepts about the formation of the solar system. It is speculated that a collision between Uranus and a relatively large and massive body might have tipped over the planet's axis of rotation.

On the basis of our observations at Mercury, Earth, Jupiter, and Saturn, as well as our understanding of the magnetohydrodynamic dynamo, we would expect the magnetic field axis to be approximately parallel to the axis of rotation. If this were true, a pole-on magnetosphere should result: the magnetic field axis lies in the plane of ecliptic and the solar wind blows directly in the direction of the polar cusp. As a consequence, the magnetotail would be quite different: instead of one current sheet separating the northern and southern lobes, within each lobe a current sheet would result. A sketch of such a pole-on magnetosphere is shown in the right panel of Fig. 12.12 below.

The Voyager encounter with Uranus, however, revealed a magnetosphere similar to the one at Earth, cf. Fig. 12.1. The magnetometer measurements on Voyager indicated an inclination of the dipole axis of 58.6° with respect to the rotation axis. In addition, the magnetic dipole is offset with respect to the center of the planet, cf. Fig. 12.10. Thus the magnetosphere is not axial-symmetric as expected for a pole-on magnetosphere but rotates around the planet's rotation axis.

Currently, two interpretations for the large angle between the dipole axis and the axis of rotation are suggested. First, we might have observed Uranus just at the time of a polarity reversal. Second, the tilt between the two axes might originate in an unusual inner structure of Uranus, in particular in locations where the planetary dynamo operates. This second interpretation is plausible in so far as on Neptune, which has a similar inner structure, a

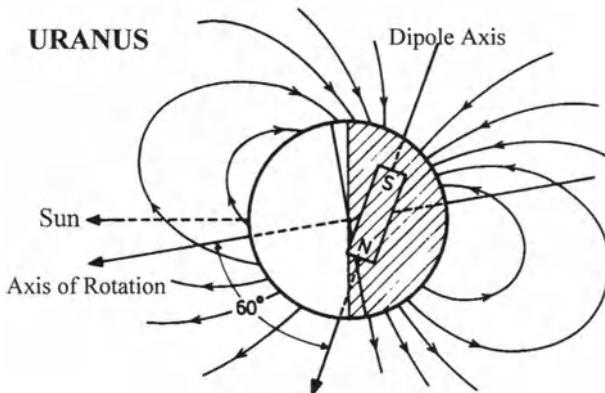


Fig. 12.10. Magnetic dipole of Uranus. Based on F.M. Neubauer [12.10], in *Plasmaphysik im Sonnensystem* (eds. K.-H. Glassmeier and M. Scholer), Copyright 1991, reprinted with kind permission from Spektrum Akademischer Verlag

similar tilt between the dipole and the rotation axes is observed, although Neptune's axis of rotation is inclined with respect to the plane of ecliptic by only 28.8° .

Inside the magnetosphere, radiation belts exist; however, the particle fluxes and energies are rather small, cf. Figs. 12.3 and 12.4. In particular, in the electron fluxes the influence of the moon on the radiation belt populations is obvious: along the moon's orbit, the particle fluxes are reduced.

Compared with all other magnetospheres in the solar system, the magnetosphere of Uranus seems to be the only closed one: because the magnetospheric plasma does not contain any α -particles, the solar wind seems to be completely frozen-out. Plasma sources most likely are the ionosphere, ionization of the vast hydrogen corona of the planet, and, to a lesser extent, sputtering from the moon's surfaces.

12.2.5 Neptune

Neptune's magnetic field and the resulting magnetosphere proved to be even more complex. Neptune's dipole axis is inclined by 47° with respect to the axis of rotation, which in turn is inclined by 28.2° with respect to the plane of ecliptic, cf. Fig. 12.11. The dipole not only is eccentric but also shifted along its axis towards its south pole. During one rotation of the planet, the inclination of the dipole axis with respect to the plane of ecliptic varies between $90^\circ - 28.8^\circ - 47^\circ = 14.2^\circ$ and $90^\circ + 28.8^\circ - 47^\circ = 71.8^\circ$. Thus within 8 h, Neptune's magnetosphere oscillates between a nearly pole-on and an Earth-like magnetosphere, cf. Fig. 12.12. These oscillations were observed during the Voyager fly-by.

Within the orbit of Triton, a radiation belt exists in Neptune's magnetosphere with fluxes much lower than those observed at Uranus. Plasma sources

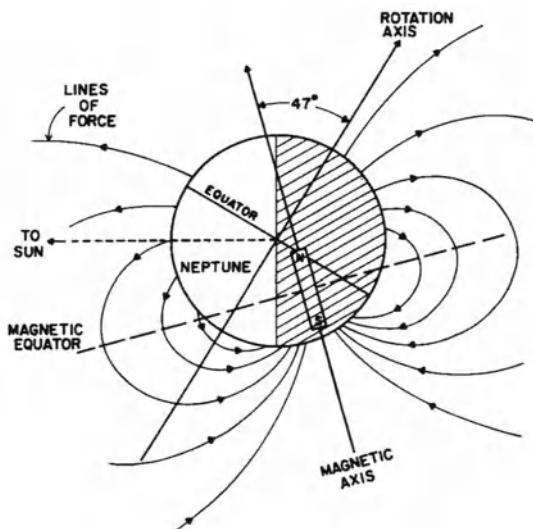


Fig. 12.11. The magnetic dipole of Neptune is inclined by 47° with respect to the rotation axis of the planet and also shifted towards one of the poles. Reprinted from N.F. Ness et al. [12.11], *Science* **246**, Copyright 1989, with kind permission from the American Academy for the Advancement of Science

are the solar wind, the atmosphere, sputtering from the moons, and heavier ions from Triton. The plasma density in the outer magnetosphere is too small to distort the field significantly. Neptune's magnetosphere appears to be the least active in the solar system, being only weakly influenced by the variable solar wind.

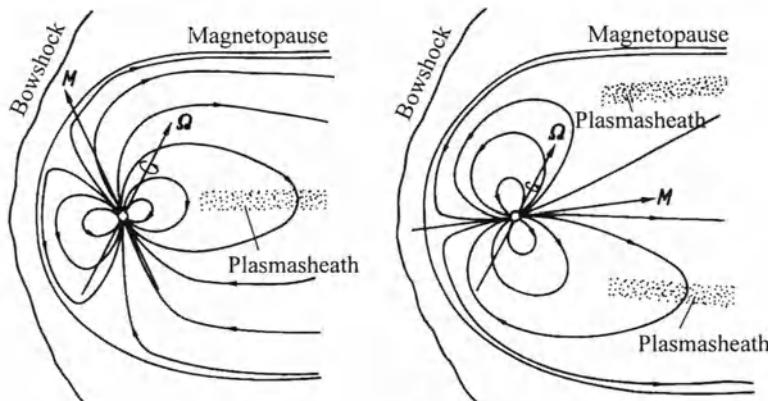


Fig. 12.12. Neptune's magnetosphere can be similar to the terrestrial one (left panel) while half a rotation later it is a pole-on magnetosphere. Reprinted from F. Bagenal [12.2], *Annual review of earth and planetary sciences* **20**, Copyright 1992, with kind permission from Annual Reviews Inc.

12.3 Planets Without a Magnetic Field

The interaction of the solar wind with unmagnetized planets is studied best for Venus our knowledge about the Martian magnetosphere at best is poor. Venus interacts with the solar wind in the same way comets do.

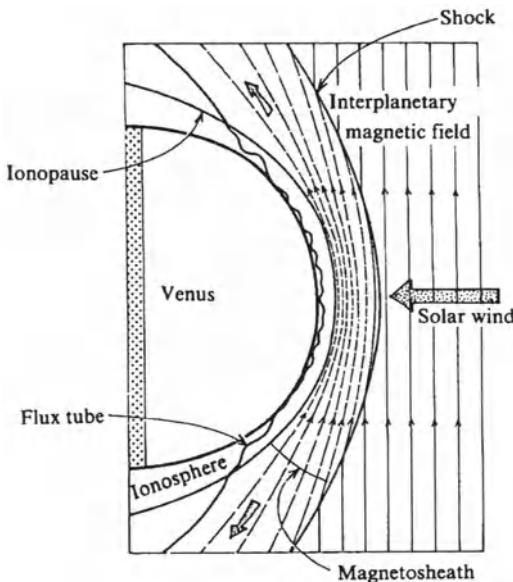


Fig. 12.13. The interaction between the solar wind and the dayside of Venus leads to an ionopause and a bow shock. Reprinted from T. Encrenaz and J.J. Bibring [12.6], *The solar system*, Copyright 1990, Springer Verlag

The dayside of Venus's 'magnetosphere' is shown in Fig. 12.13. As the solar wind is slowed down to subsonic speed, a bow shock develops. In the magnetosheath behind the bow shock, the plasma and the interplanetary field are compressed. The magnetosheath is separated from Venus's ionosphere by the ionopause. Analogous to the magnetopause, the ionopause is defined as an equilibrium between the kinetic plasma pressure and the magnetic field pressure, only this time it is the planetary plasma and the combined pressure of the solar wind and the interplanetary magnetic field. The latter is frozen out of the ionospheric plasma. Although Venus has a magnetic field, it is too weak to support a magnetosphere in the classical sense: the subsolar magnetopause would lie inside the planet.

The solar wind deflected around the planet forms a cavity which is nearly void of any plasma. Since the magnetic field lines forming the magnetotail are of opposite polarity, a current system similar to the one observed in the geomagnetic tail evolves. Depending on the solar wind conditions, strong magnetic fields can be observed at a height of a few hundred kilometers above the planet's surface. These fields, however, are not of planetary origin but stem from the compressed and deflected interplanetary field.

12.4 Summary

Almost all planets in the solar system, except Mars and Venus, have a magnetosphere. The structures, sizes, and dynamics of these magnetospheres can be quite different. The most important parameters determining magnetospheric properties can be summarized as follows:

- The planet's magnetic moment combined with the local solar wind pressure determine the typical size of the magnetopause. If this is smaller than the planet's radius, no classical magnetosphere is observed.
- The structure of the magnetosphere is determined by the inclination of the dipole axis with respect to the axis of rotation and the plane of ecliptic, leading to Earth-like magnetospheres, pole-on magnetospheres, or oscillations between these two states, as observed in Neptune's magnetosphere.
- Jupiter's magnetosphere is the largest object in the solar system. Owing to strong centrifugal forces, the Jovian magnetosphere is rather flat with a plasma sheet in the equatorial plane.
- Sources of magnetospheric plasma include the solar wind (except for Uranus), the atmospheres and ionospheres (except for Mercury), and at the outer planets also the moons and rings.
- Radiation belts can be observed inside the Earth's magnetosphere and the magnetospheres of the outer planets. At the orbits of the moons particle fluxes are significantly lower because the moons absorb particles. Some moons create plasma tori along their orbits, such as Io in the Jupiter system and Titan orbiting Saturn.

Exercises and Problems

12.1 Determine the energies of electrons and protons in resonance with the waves in the foreshock of the different planets. Use the numbers given in Fig. 12.2.

12.2 Confirm the calculated stand-off distances given in Table 12.2 (assume that the solar wind density decreases as r^{-2}).

12.3 Discuss the relationship between the spin period and the magnetic moment of the planet in terms of the magnetohydrodynamic dynamo. Would this relationship be in agreement with the assumption of a similar process working in all planets?

12.4 Discuss the possibility of aurorae and their shapes and detectability on other planets. Use your knowledge about the aurorae on Earth.

12.5 Describe and discuss the plasma sources in the different magnetospheres.

13. Solar–Terrestrial Relationships

It is a good morning exercise for a research scientist to discard a pet hypothesis every day before breakfast. It keeps him young.

K. Lorenz, *The So-Called Evil*¹

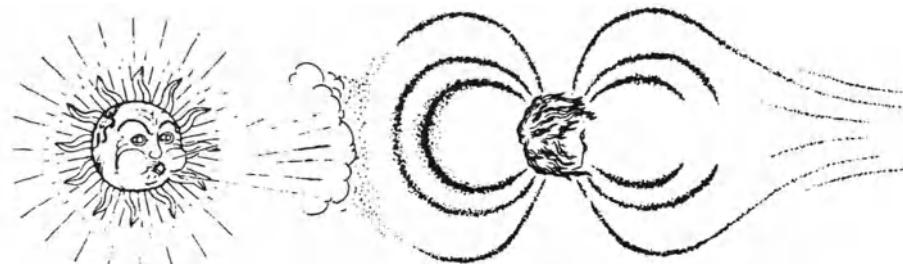
Solar–terrestrial relationships deal with the influence of the Sun and solar activity on our terrestrial environment. The driving force is the input of energy and matter into geospace. From the viewpoint of space plasma physics, the most important (and also the scientifically soundest) consequences have been discussed in Chap. 11. But there are also side-effects to these phenomena, such as the influence of auroral particles on the chemistry of the atmosphere, or the uproar caused in our technical environment due to severe geomagnetic disturbances. Other connections exist, too, relating solar cycle variations to weather and climate. These latter bear the seed of controversial discussions; nonetheless, some of the ideas will be reported here.

13.1 Solar–Terrestrial Relationships: Overview

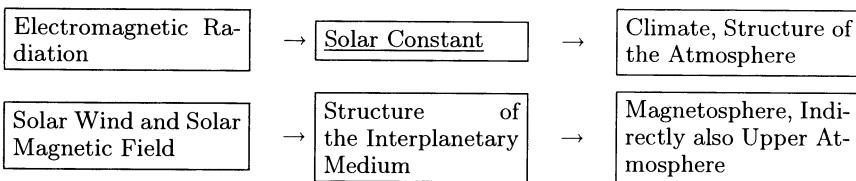
The Sun is a source of electromagnetic radiation, plasmas, fields, and energetic particles. These emissions have a constant component and a variable, the solar-cycle-dependent one, cf. Fig. 13.1. Ordered by decreasing energy, they shape our terrestrial environment as follows:

(a) The continuous solar electromagnetic radiation, described by the solar constant, determines the structure of the atmosphere and the climate. Since the solar constant varies by less than 0.1% during the solar cycle, its variation cannot have a recognizable influence on climate. Although the solar constant is roughly constant, the hard electromagnetic emission is enhanced by up to an order of magnitude during solar maximum, causing solar cycle variations in the atmospheric height, the thermospheric temperature, and the ionospheric layers.

¹ Reprinted from K. Lorenz *On Aggression* (transl. M. Latzken), Routledge, London, Copyright 1983, with kind permission from Deutscher Taschenbuch Verlag



Continuous Emission:



Solar-Cycle-Dependent Emission:

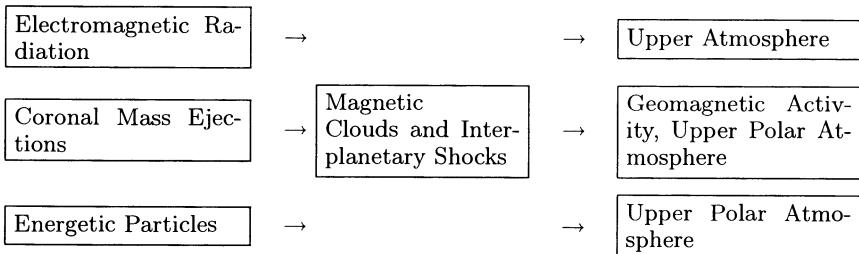


Fig. 13.1. Simplified overview of solar-terrestrial relationships. The top panel is from a poster for a 1978 workshop on solar-terrestrial influences on weather and climate at the Ohio State University

(b) The solar wind and the frozen-in magnetic field shape interplanetary space and, on interaction with the terrestrial magnetic field, form the Earth's magnetosphere. The energy density in the solar wind is six orders of magnitude smaller than in solar electromagnetic radiation. Nonetheless, disturbances in the solar wind cause much stronger modifications in the terrestrial environment than does the flare's increased electromagnetic radiation. In addition, the structure of the heliosphere leads to modulation of the galactic cosmic radiation. These particles influence the chemistry and ionization of the upper atmosphere.

(c) Solar flares emit electromagnetic radiation from the gamma to the radio range. As a consequence, the temperature, circulation, and chemistry

of the upper atmosphere, and even the upper stratosphere, can be modified, leading to an increased drag on satellites in low orbits or to sudden ionospheric disturbances. Since the travel time of light from the Sun to the Earth is only 8 min, these effects start immediately after the flare.

(d) Coronal mass ejections and the shocks driven by them cause geomagnetic disturbances. Their influence is delayed by 1 to 3 days with respect to the flare. Shock and mass ejection transfer energy to the magnetosphere which, during auroral activity, partly is transferred to the upper polar atmosphere. In addition, energetic particles accelerated at the shock can penetrate into the atmosphere at the polar cusps, causing additional ionization. This in turn influences the temperature, chemistry, and circulation of the thermosphere.

(e) Solar energetic particles arrive at Earth a few hours after the flare. Like the particles accelerated at the shock, they can interact with the upper atmosphere in the cusp regions, feeding energy into it and modifying its chemistry.

Although the electromagnetic radiation has the strongest influence on the terrestrial system, shaping our climate and weather patterns and making our world a habitable one, here we will limit our discussion to the topic of this book, namely particles and plasmas. The variability of the Sun's radiative output is discussed in [13.15], her role in climate change in [13.11] and in the extremely controversial popular account [13.2].

In solar-terrestrial relationships, three “levels of knowledge” can be distinguished: (i) accepted relationships where the observations are unambiguous and plausible explanations exist (for instance, the variations in the upper atmosphere and the disruptions to our technical environment), (ii) relationships where the observations appear to be sound but an explanation is lacking or insufficient (such as the long-term relationship between solar activity and climate, as evidenced, for instance, in the Little Ice Age), and (iii) relationships for which there is observational evidence but also a remaining doubt, in particular since the explanations are lacking (such as the influence of the Sun on the drought cycle in the western US). In the latter cases, a relationship often appears unreasonable since the energy fed into the terrestrial system is small compared with the energy contained in the phenomenon.

13.2 The Atmosphere

The solar electromagnetic radiation determines the structure and dynamics of the atmosphere. Depending on its wavelength, it interacts with the atmosphere at different altitudes. Since interaction always is associated with heating, a characteristic temperature profile develops which can be used to define atmospheric layers, cf. Fig. 13.2.

The bottom layer of the atmosphere, the troposphere, has a thickness between about 16 km at the equator and less than 10 km close to the poles.

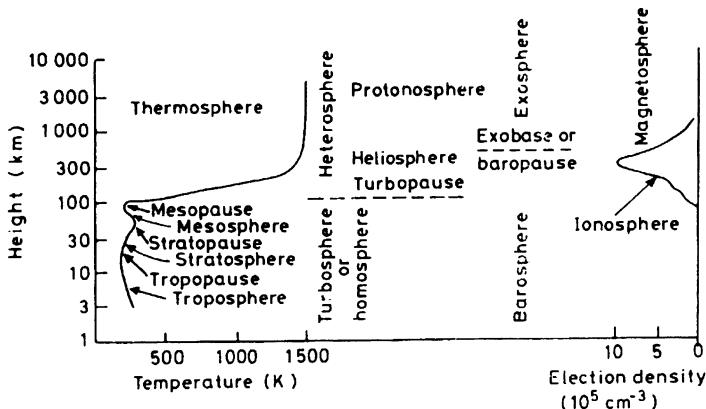


Fig. 13.2. Horizontal structure of the terrestrial atmosphere with different layers defined by extrema in the temperature profile, mixing of the components, the possibility of particle escape, and the ionization. Reprinted from J.K. Hargreaves [13.8], *The solar-terrestrial environment*, Copyright 1992, with kind permission from Cambridge University Press

This layer contains more than three-quarters of the atmospheric matter and therefore energywise is the most important layer. Since the presence of water vapor allows the formation of clouds and precipitation, it is the weather layer of our planet. The combined effects of radiation, convection, and the transport of latent heat cause a negative temperature gradient of about 6.5 K/km . Out of the incident solar radiation, only the visible and infra-red penetrate down to the troposphere, the shorter wavelengths are absorbed at higher altitudes. The top of the troposphere is the tropopause, the local minimum in the temperature profile.

The next layer, the stratosphere, is characterized by a positive temperature gradient. This is caused by the absorption of UV in the ozone layer. Its shape can be described by the same combination of effects which is responsible for the formation of the ionospheric Chapman layers: the incoming electromagnetic radiation increases with height while the density decreases. The maximum of the ozone layer is at a height of about 25 km. The stratosphere is a dry place, and its water vapor content is almost negligible, although water vapor plays an important role in the ozone chemistry (see below). Since the tropopause is a temperature inversion, ideally there would be no exchange of matter across it. Exchange, however, happens, as is evident for instance from the influence of the anthropogenic CFCs on the ozone layer. Two effects allow for such an exchange: violent processes, such as nuclear explosions or erupting volcanoes, and a seasonal slow exchange at the jet streams where the tropopause is 'leaky'. The consequences of this slow exchange are twofold: luckily, it is rather difficult for man-made gases to reach the stratosphere, but unfortunately once such substances or the aerosols from volcanic eruptions

have reached the stratosphere, they will be removed only extremely slowly. The stratosphere extends up to altitudes of about 40–50 km.

In the mesosphere, the temperature gradient again is negative. The mesosphere extends up to a height of about 80 km, it is characterized by photo-dissociation, ionization (the D-layer of the ionosphere is inside the mesosphere), and a variety of chemical processes.

Above the mesosphere, the temperature increases again because the hard electromagnetic radiation is absorbed, leading to the formation of the main ionospheric layers. Thus the lower thermosphere is characterized by a positive temperature gradient. Above an altitude of 150–200 km the thermosphere is isothermal with temperatures between about 1300 K (nightside at solar minimum) and 2000 K (dayside at solar maximum). It is characterized by extended circulation systems which vary seasonally and with the solar cycle. It also is the atmospheric layer that is connected directly to the magnetospheric processes: the currents providing the ionosphere–magnetosphere coupling run through the thermosphere and the particles causing aurora penetrate down to its bottom.

13.3 Responses of the Upper Atmosphere to Solar Variability

Solar cycle variations in the upper atmosphere as well as its responses to solar activity are supported by many observations and can be understood quite well. The solar-cycle-dependent input into the upper atmosphere comprises hard electromagnetic radiation (its variation with the solar cycle as well as enhancement during flares), auroral particles, solar energetic particles, and galactic cosmic radiation. The consequences of these inputs are heating and ionization. The energy input follows different spatial patterns: changes in the solar electromagnetic radiation directly affect the entire dayside atmosphere or, since the exchange between dayside and nightside is fast and it rotates with the planet, often the entire atmosphere. The galactic cosmic radiation basically influences the entire atmosphere, too. The energy input of auroral and solar energetic particles, on the other hand, is limited to the high-latitude atmosphere.

This section will be concerned with the consequences of variations in the particle input into the atmosphere only. The particle energies range from a few kiloelectronvolts (auroral particles) to megaelectronvolts (solar energetic particles) and even into the gigaelectronvolt range. Let us consider the direct energy transfer only, i.e. we do not follow the path of the particles but assume the particles to be incident isotropically from the upward-looking hemisphere into the upper atmosphere. The primary energy transfer mechanism is ionization. Figure 13.3 shows the ion production by incoming protons for different energies. Protons with energies in the kiloelectronvolt range cause ionization

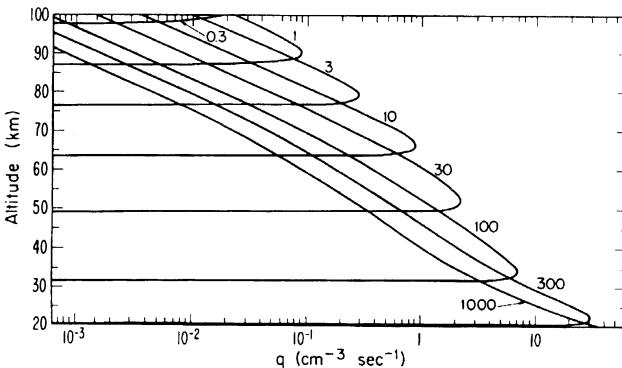


Fig. 13.3. Ion production in the atmosphere for different monoenergetic proton distributions incident over the upward-looking hemisphere. The flux of the incoming particles is 1 proton $\text{cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1}$, and the particle energy is given in megaelectronvolts. Reprinted from G.C. Reid [13.17], in *Physics of the Sun, vol. III* (eds. P.A. Sturrock, T.E. Holzer, D.M. Mihalas, and R.K. Ulrich), Copyright 1986, with kind permission from Kluwer Academic Publishers

at altitudes where ionized particles already exist. Protons with energies in the 1 MeV to 100 MeV range, which is the bulk of the solar energetic protons, ionize at altitudes between 30 km and 90 km where normally the atmosphere is neutral. Thus the changes compared with the quiet atmosphere are more pronounced, particularly since they allow for interactions between atmospheric species which, under normal conditions, would not occur at these altitudes.

13.3.1 Polar Cap Absorptions and Ozone

The primary energy loss mechanism for the incoming protons is ionization. The consequences for the atmosphere are heating, a modified atmospheric electricity, and changes in atmospheric chemistry. The bulk of ions produced in the middle atmosphere, i.e. at altitudes from a few tens up to about 90 km, are N_2^+ and O_2^+ ; smaller amounts of Ar^+ , CO_2^+ , O_3^+ , O^+ , and N^+ are produced, too. The electrons released in these interactions are either lost by recombination with a positive ion or form negative ions with a previously neutral atmospheric constituent.

The most important chemical reactions in the middle atmosphere involve the ozone layer. Some of the ions created by the incident protons are ozone-destroying radicals. The reduction of the ozone layer locally can be as strong as a factor of 2 to 4, as has been observed in the strong polar cap absorption (PCA) event following a flare in November 1969 [13.22]. Both O_2^+ as well as N_2^+ contribute to ozone destruction.

The ozone destruction by O_2^+ is based on a chain of events. First, the ionized oxygen molecule has to attach itself to a water molecule. The latter one reacts with another water molecule:

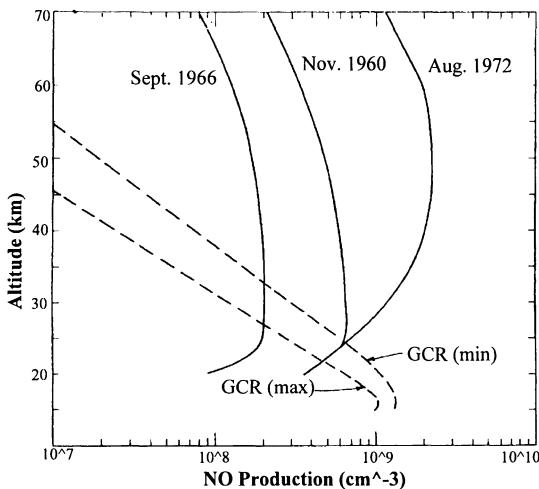
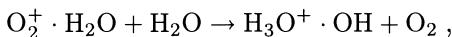
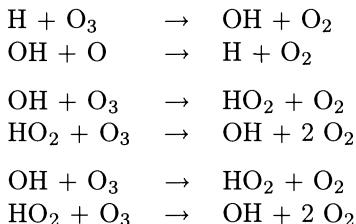


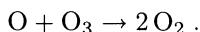
Fig. 13.4. Calculated NO production in the atmosphere for three solar energetic particle events and for the galactic cosmic radiation under solar minimum and maximum conditions. Reprinted from G.C. Reid [13.17], in *Physics of the Sun, vol. III* (eds. P.A. Sturrock, T.E. Holzer, D.M. Mihalas, and R.K. Ulrich), Copyright 1986, with kind permission from Kluwer Academic Publishers



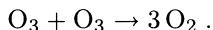
forming a hydroxyl-radical OH and an H⁺. The following pairs of reactions can result:



In the upper chain, the net effect is the recombination of an ozon molecule and an oxygen atom yielding two oxygen molecules



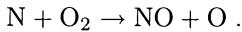
In the lower two chains, the net effect is a recombination of two ozone molecules yielding three oxygen molecules:



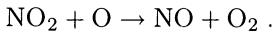
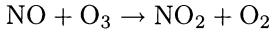
In these reactions, H and OH act as catalysts. They only are lost from the reaction chain when two of them meet, forming either H₂ or H₂O. These reactions have time constants of a few hours, thus the reaction chain cannot go on indefinitely and the ozone will recover after some days. In addition, these reactions reduce the water content in the middle atmosphere since during recombination more H₂ is formed than H₂O. Thus further ozone reduction, for instance during the next PCA, will be reduced

Since they are long-lived, stable ions function as catalysts only; the main hazard for the ozone layer are the nitric oxides, in particular NO. The latter

is produced by a two-step reaction from N_2^+ . First, N_2^+ undergoes dissociative recombination, giving two neutral nitrogen atoms. If these interact with oxygen, then NO is formed:



The ozone destruction then is given by the reactions



One or two large PCA events, lasting for 2 or 3 days, produce more nitric oxides than the galactic cosmic radiation during an entire year, cf. Fig. 13.4. Thus at times of high solar activity, the NO production in the mesosphere basically is determined by solar activity. The nitric oxides, however, are produced at higher altitudes than the ones due to ionization by galactic cosmic rays since the solar energetic particles have lower energies. But the removal of the ozone-destroying NO is slower at higher altitudes: the loss mechanism for NO above about 40 km is photodissociation, leading to the formation of N_2 . NO produced by the galactic cosmic radiation close to the tropopause, on the other hand, can be washed out rather easily, forming nitric acids.

Although at times of high solar activity the nitric oxides generated by solar energetic particles have a recognizable influence on the ozone layer, we should keep in mind that this is limited to high latitudes. Figure 13.5 shows the ozone concentration above the 4 mbar level for a period of about 2 months in late summer 1972. In the upper panel, a latitudinal band of $\pm 5^\circ$ around the equator is considered, the middle panel gives a band of the same width centered around 60°N , and the lower panel considers latitudes between 75° and 80°N . Up to day 217, the ozone concentration in all three panels is roughly constant, but following the strong flare and PCA event on day 217 there is a marked decrease in ozone at high latitudes lasting for more than a month. The variations in mid-latitude starting around the same time probably are caused by latitudinal exchange processes, in particular the transport of ozone-depleted air from higher to lower latitudes. In addition, a disturbance of the magnetospheric structure during the geomagnetic storm following that flare might have caused a far southward extension of the auroral oval, allowing particles to penetrate the atmosphere at mid-latitudes. The variation in ozone close to the equator appears to be related to transport processes, in particular an exchange between low latitudes and mid-latitudes since both panels at least partly show opposing trends.

13.3.2 Thermospheric Circulation

The latitudinal temperature gradient inside the thermosphere drives a meridional circulation system. This gradient is caused by the latitudinal variation of the incident solar radiation and energetic particles. As in the troposphere,

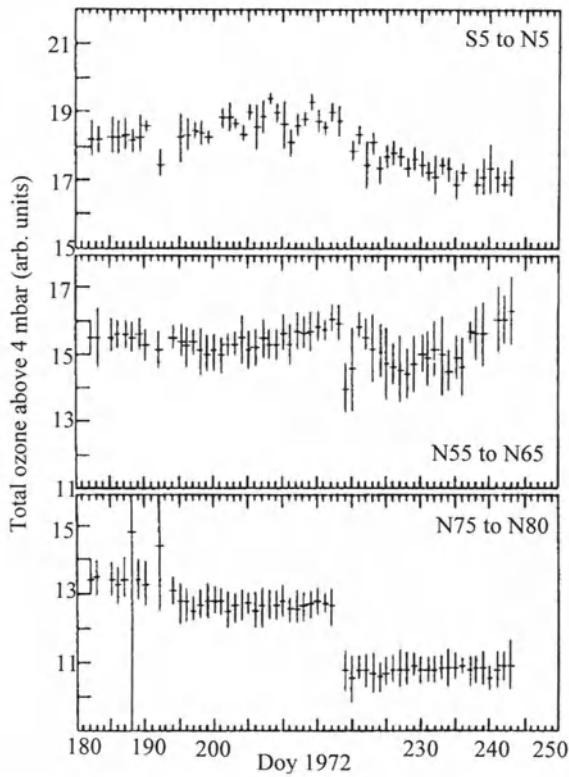


Fig. 13.5. Observed variation in the ozone column above the 4 mbar level in different latitude bands. The sharp decrease at high latitudes follows the PCA of the August 1972 flare. Reprinted from D.F. Heath et al. [13.9], *Science* **197**, Copyright 1977, with kind permission from the American Association for the Advancement of Science

more radiative energy is absorbed close to the equator than at high latitudes. In the thermosphere the energetic particles incident at high latitudes provide an additional heat source. Thus the thermospheric circulation system shows a seasonal dependence (latitude dependence of the electromagnetic radiation) as well as a superimposed dependence on the solar cycle.

Figure 13.6 shows model calculations for the thermospheric circulation for the equinoxes (left) and solstices (right) for different levels of solar activity: (a) an extremely quiet Sun, (b) average solar activity, and (c) a highly active Sun. The influence of the latitudinal dependence of the incident electromagnetic radiation can be seen best for very quiet solar conditions: the thermospheric circulation then consists of a thermally driven Hadley cell. During the equinoxes, heating is strongest close to the equator, leading to an updwelling of hot air which, at higher altitudes, is transported polewards where it cools, sinks, and moves as cold air equatorwards through the stratosphere (that would be at the bottom of the figure), closing the circulation cell. During solstices, only one Hadley cell evolves, spanning the entire globe. Heating is strongest at high latitudes in the summer hemisphere, thus air dwells up there and is transported across the equator towards the high latitudes of the winter hemisphere where it sinks.

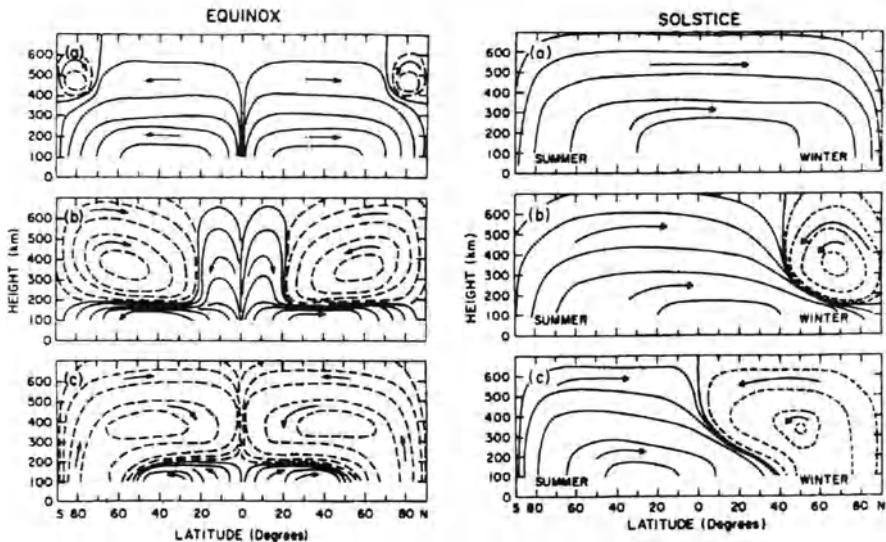


Fig. 13.6. Thermospheric circulation at the equinoxes and the solstices for extremely quiet solar conditions (a), average solar activity (b), and a very active Sun (c). Reprinted from R.G. Roble [13.19], *The upper atmosphere and magnetosphere*, Copyright 1977, with kind permission from National Academy Press

With increasing solar activity, heat is supplied to the high-latitude upper thermosphere by energetic particles. Thus close to the poles, secondary circulation cells evolve providing heat transport towards lower latitudes. During equinoxes, two such cells are observed, one over each pole. With increasing solar activity, these cells expand, nearly suppressing the thermally driven circulation. During the solstices only one such cell develops in the winter hemisphere. Wind speeds in the thermospheric circulation can be up to 2500 km/h.

Figure 13.6 is a grossly simplified meridional cross-section. Since the Earth rotates, the real thermospheric circulation consists of a thermally driven meridional component and a zonal deflection due to the Coriolis force. Whereas in the troposphere the Coriolis force breaks up the Headley cell into three cells in each hemisphere, the thermospheric circulation pattern basically consists of the one cell shown in Fig. 13.6. Only around the geomagnetic poles can closed zonal circulation patterns exist, depending on solar activity. In the southern hemisphere, closed vortices around the pole appear to be a dominant feature which is observed in the oceans, in the stratosphere (where it plays a prominent role in the seasonal appearance of the ozone hole), as well as in the thermosphere.

13.4 The Technical Environment and Solar Activity

Solar-cycle-related disturbances in our technical environment stem from magnetic storms caused by the variable solar wind. The changing magnetic flux at the Earth's surface causes induction currents in long leads, in particular in power lines or pipelines. Since these currents can amount to 50–100 A, they might trigger the overload protectors or might cause damage to transformers, leading to a disruption of the electrical power supply. In recent years, this problem has become increasingly important because power supplies are connected over larger and larger distances. Since a large contribution to the disturbed geomagnetic field is due to the enhanced auroral electrojet, high latitudes, such as Scandinavia, Canada, or the northern US, are more frequently affected by these effects. For instance, the large geomagnetic storm on 13 March 1989, one of the largest ever recorded, caused a disruption of the entire 9000 MW power supply network of the province of Quebec for more than 9 hours. Computer networks also are affected or damaged by induction currents.

Pipelines for natural gas or oil also are vulnerable to large geomagnetic storms because the induction currents cause corrosion. This effect has been observed best at the Alaska oil pipeline, which stretches for about 1300 km from north to south through the auroral oval. Where the pipeline is connected to the ground, the induction currents lead to strong corrosion.

The effects of geomagnetic disturbances are not only limited to the Earth's surface. Satellites also can be damaged by induction currents in their electronics or on sensitive surfaces, leading to the failure of the satellite or at least of some instruments. Occasionally, a satellite might lose its spatial orientation in a large geomagnetic storm: some satellites are oriented by reference to the geomagnetic field. If the latter is strongly disturbed, in particular if fluctuations are fast or if the magnetopause has moved so far inwards that the satellite suddenly finds itself in interplanetary space where the orientation of the magnetic field is entirely different, communication with the satellite might be lost because the satellite is not able to keep the proper orientation of its antenna. Although important at the time of the geomagnetic storm, the disorientation does not cause permanent damage since the satellite is able to orientate itself properly as soon as average magnetic conditions are restored.

Higher magnetospheric electron fluxes during magnetic storms also might cause satellite failure if the spacecraft becomes electrically charged. This, by itself, does not cause harm, but becomes harmful if discharge occurs between different components. In addition, satellites are also vulnerable to the direct action of energetic particles: high energetic particles, although only few in number, might cause radiation damage to the electronics. Obviously, increased fluxes of energetic particles also are a risk to manned spaceflights.

The disturbance of communication systems is not related to geomagnetic activity but is a direct consequence of the increased ionization in the ionosphere due to the hard electromagnetic radiation (and to a much lesser and

spatially limited extent also to PCA). Ground-ground communication with short and long waves is disturbed because these waves are no longer reflected by the ionosphere but are absorbed. Satellite-ground communication also can be affected due to enhanced absorption. Even if the signals pass through the ionosphere they might be distorted or their travel time might be elongated, making navaids such as the GPS less reliable at times of high solar activity.

The increased hard electromagnetic radiation during solar maximum conditions as well as in flares not only enhances ionization but also causes the atmosphere to expand outwards. Or, viewed from a fixed altitude, the density and temperature increase with increasing solar activity. A satellite flying at low altitudes, typically a few hundred kilometers, an altitude used for instance by astrophysical and solar research satellites, then experiences an increased drag, reducing its speed and therefore also its altitude. In time, the satellite will move closer to the denser atmosphere and finally plunge down. Prominent victims of increased solar activity have been Spacelab and the Solar Maximum Mission, SolarMax.

And the Biosphere? While the influence of solar activity on our technical environment is undoubtedly, the influences of solar activity on the biosphere are still under discussion. If there are any links, most likely the magnetosphere and geomagnetic activity should provide the connection because solar-activity-related enhancements in hard electromagnetic radiation are absorbed high in the atmosphere. Some influences are obvious or can be tested easily in the laboratory. Migrating animals, such as pigeons and whales, use the geomagnetic field for orientation and navigation. They become disturbed during geomagnetic active periods. For instance, carrier pigeons launched in a magnetic storm need much longer to find their home (days compared with some hours). Some micro-organisms seem to have a magnetic field sensor, too, using the field for their orientation. Laboratory experiments show that changes in the magnetic field can affect the vital functions of these organisms.

The influence of geomagnetic disturbances on humans is still under debate. Although we still might have a magnetic field sensor, as is suggested by some tests, our means of orientation and navigation rely much more on visual information. Thus most people probably would not even be able to deliberately access the information provided by such a field sensor. Nonetheless, since currents are the carriers of information in the human body, the influences of geomagnetic activity should not be ruled out completely, see, for example, [13.12]. An often cited example is the higher incidence of cardiovascular diseases, in particular cardiac infarction, during geomagnetic storms, which is suggested by different statistical studies at different places over the last 30 years. These correlations often have been doubted since the changes in the surface magnetic field are small compared with the magnetic fields created by our technical environment; however, recent research suggests that the link is probably not the magnitude but the frequency on the fluctuations. And it appears that geomagnetic disturbances have frequencies in the range

of slightly less than and up to a few hertz which in principle can lock with the cardiac rhythm. If such locking occurs during the vulnerable phase, cardiac arrest may result.

13.5 The Solar Cycle, Sector Boundaries, Droughts, and Thunderstorms

Another controversial topic is the relationship between solar activity, the weather, and the climate. Recently, the discussion has become more aggressive since some authors claim that the temperature increase observed over the last hundred years can be explained solely by variations in the solar output while others claim that it is solely anthropogenic. Since the atmosphere is a complex system, strongly coupled to the oceans, it is not likely that there is only one cause; rather climate variability most likely is influenced by many factors which include internal oscillations of the climate system, extraterrestrial causes, as well as anthropogenic influences. To delve deeply into this debate would be the topic of another book; nonetheless, some observations relating weather and climate to solar activity will be presented here.

Attempts at correlations between weather or climate and the solar cycle or the geomagnetic activity are many; a comprehensive, though somewhat older overview can be found in [13.16]. Some of these correlations, such as the droughts in the western US, can be traced back for centuries, while others cover only shorter time spans of 3 to 10 solar cycles because older records are not available.

13.5.1 Solar Activity, Climate, and Culture

Although evidence is only indirect, a widely accepted relationship between solar activity and terrestrial climate exists for the last 7000 years. This correlation is closely tied to the name of Eddy [13.5, 13.6]; however, attempts to correlate the Sun and climate can be traced back for at least 200 years.

Figure 13.7 correlates solar activity and climate parameters. In the upper panel, the deviation of the ^{14}C content from its long-term trend is shown. Since ^{14}C is a cosmogenic nuclide, it provides a measure of the intensity of the galactic cosmic radiation, which in turn is modulated by solar activity: high solar activity corresponds to a reduction in galactic cosmic radiation and therefore also in ^{14}C , while during the solar minimum both are enhanced. Thus the ^{14}C record can be used as a long-term record of solar activity, which is shown in the middle panel. The panel does not show individual cycles because it is not possible to extract individual solar cycles from the ^{14}C record since the isotope is produced in the stratosphere but stored in organic matter. The combined transport and assimilation processes have time-scales of some years, smearing out the individual solar cycles. In the bottom panel,

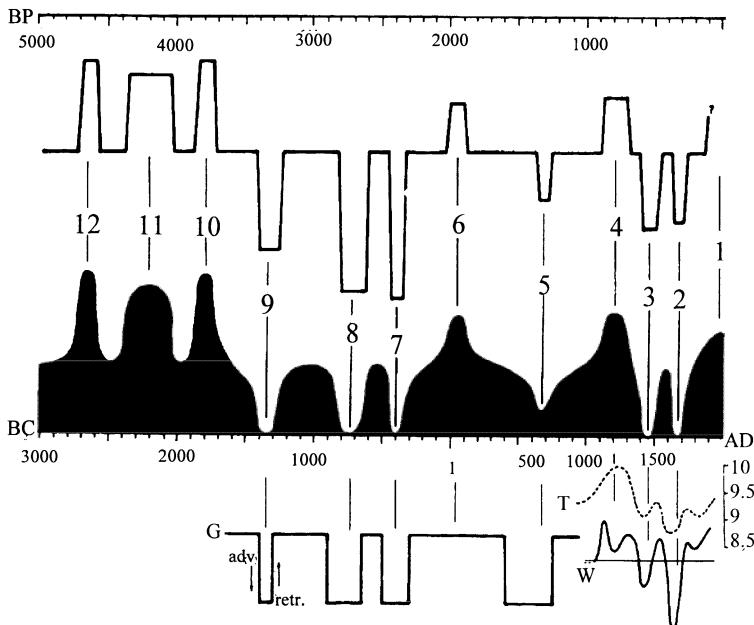


Fig. 13.7. Changes in solar activity as inferred from cosmogenic nuclides and related changes in climate. Reprinted from J.A. Eddy [13.6], in *Physics of solar planetary environments* (ed. D.J. William), Copyright 1976, American Geophysical Union

climate parameters are shown with T being the temperature, and W the winter index as an indirect measure of the strength of the winter, inferred from the time of freezing and thawing of certain lakes or rivers. The curve G gives the advance and retreat of glaciers in Europe.

The figure suggests that times of high solar activity are connected with a retreat in glaciers while lower solar activity corresponds to their advance. On shorter time scales of some hundreds to about a thousand years, the crude climate measure 'glacier' is supplemented by historical records and in the last 400 years also by direct measurements. Times of high solar activity not only are connected with higher temperatures but also with cultural growth. Examples are indicated by the vertical lines: (12) marks the Sumerian maximum, the first advanced civilization, between about 2700 BC and 2600 BC, (11) the Pyramid Maximum (2300 BC to 2050 BC), (10) the Stonehenge Maximum (1850 BC to 1750 BC), (6) the Roman Maximum, and (4) the Medieval Maximum, when the climate was warm enough to allow the Vikings to cross the Atlantic Ocean and to discover Greenland as a green land. Times of low solar activity, such as the Spörer Minimum (2) and the Maunder Minimum (3), both marking the Little Ice Age, or (5) the Medieval Minimum often are related to cultural depression and retreat or to migration of nations in the hope of finding a more suitable environment. Evidence for

higher solar activity during the Roman Maximum is also given by the rather frequently observed and reported aurora which today is rather unusual in the Mediterranean.

Although the observations are accepted, the explanation of the correlation between weak solar activity and a colder climate is still not completely understood. When the first measurements of the solar constant over an entire solar cycle became available in the late 1980s, they showed only a small variation of 0.08% in the solar constant with a maximum during solar maximum. Thus despite the larger number of dark spots, the Sun emits more radiation at maximum conditions. This additional radiation comes from a bright chromospheric network related to the magnetic field structure. Nonetheless, for the climate system a variation in the solar constant by less than 0.1% is much too small to produce any recognizable signal, even if this reduction lasts for some tens of years.

Thus if we assume that the minima in solar activity indicated in Fig. 13.7 are nothing more than a prolonged ordinary solar minimum, changes in solar irradiance would not explain the variations in climate. But long-lasting minima might be different from the ordinary sunspot minima known today. A careful analysis of observations of the Sun during the Maunder Minimum in the late seventeenth century suggested a larger solar diameter and a slower but stronger differential rotation than observed during the more recent solar cycles. In addition, although no sunspots were observed, the cosmogenic nuclide ^{10}Be suggested the continuation of a solar cycle with a reduced period of about 9 years.² Thus the solar dynamo was still at work although at a faster pace and with weaker photospheric fields, as is evident from the lack of sunspots. From these changes in solar properties, reductions of the solar constant between 0.14% and 0.5% have been suggested. Presently it is still debated which of these numbers is more reliable and whether it alone would be sufficient to explain the Little Ice Age.

13.5.2 Sun and Weather

Changes in the solar constant only are one aspect of reduced solar activity. Other aspects include an enhanced galactic cosmic radiation and reduced geomagnetic activity. The possibility of an influence of these aspects on weather and climate often is ruled out by the argument that the energy fed into the terrestrial system by these fluctuations is too small to cause any changes. Nonetheless, correlations suggesting such connections have been reported.

Droughts in the Western US. A prominent example of the correlation of weather phenomena and the solar cycle are droughts in the western US, in the dust bowl. Droughts severely affect plant growth, and thus a drought

² ^{10}Be records are stored in the Arctic and Antarctic ice shields. In contrast to ^{14}C , the Be isotope is immediately washed out of the atmosphere by precipitation, allowing the detection of variations in time-scales of about a few years.

index can be derived from tree rings: a wide tree ring indicates plenty of water available while a smaller one suggests a reduction in the water supply. To reduce the influence of temperature on tree ring size, plant species with different dependences on water supply and temperature can be compared. With these data in hand, a drought index for the western US can be traced back to about 1600, the time when the first systematic records of sunspots and solar activity started. There is a correlation between droughts and solar activity at the 99% confidence level; however, this correlation is not at the 11-year sunspot cycle but at the 22-year magnetic cycle. The largest extension of the area affected by drought is about 2 to 3 years after the minimum at which the Hale cycle starts. The severity of the droughts is also modulated with a 90-year cycle, which also modulates the sunspot cycle.

The dependence on the 22-year cycle is amazing in so far as it suggests a relationship between climate and the magnetic properties of the Sun rather than with its radiation. Although the underlying processes are not completely understood, it appears likely that the energy that triggers the atmospheric change, i.e. the drought, is fed into the terrestrial system by geomagnetic activity. During times of low solar activity, recurrent geomagnetic disturbances can be observed, cf. Fig. 11.3. A comparison of field variations in subsequent solar rotations allows the definition of a geomagnetic recurrence index. The latter one is large only if the geomagnetic disturbance can be observed again during the next rotation. Thus the recurrence index is anti-correlated with the solar cycle since at times of high solar activity transient disturbances modify the structure of interplanetary space and recurrent fast solar wind streams are observed less frequently. The recurrence index is larger and stays high for a longer time period in the minimum between an even and an odd solar cycle, i.e. at the beginning of a new magnetic cycle when the droughts in the western US are also observed.

Sector Boundaries and Weather Forecast. Evidence of the involvement of the interplanetary magnetic field also stems from a correlation between the vorticity area index (VAI) and geomagnetic activity. The VAI describes the size of the vortex field on the 500 mb surface and can be used as a measure of the strength of the high- and low-pressure regions. Thus the VAI is directly related to the large-scale atmospheric phenomena that determine our weather. On the basis of a data set of some hundred sector boundary crossings collected over 20 years, a pattern is established with a reduction of the VAI by up to 10% on the day after the sector boundary crossing. The reduction is strongest if the interplanetary magnetic field has a southward component; it is weaker if the interplanetary field is directed northwards. The strongest reductions in VAI are observed if the interplanetary magnetic field has a southward component and the sector boundary is accompanied by a stream of fast protons. Thus somehow energy fed from the interplanetary medium through the magnetosphere into the upper atmosphere modifies the flow pattern in the troposphere.

Modifications to the VAI directly influence the weather. This can be shown indirectly: the reliability of the 12 h and 24 h weather forecast in the middle latitudes of the northern hemisphere is reduced from 85% to 65% on the day following the sector boundary crossing. In addition, the thunderstorm area then is largest, too.

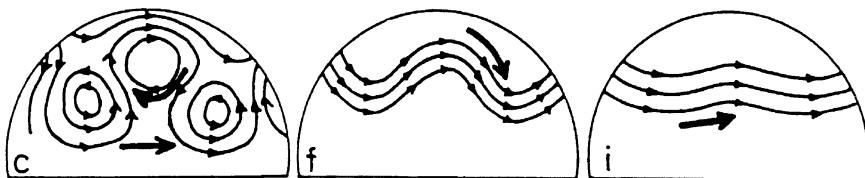


Fig. 13.8. Upper tropospheric circulation pattern for very low (*left*), average (*middle*), and high geomagnetic activity (*right*). Reprinted from V. Bucha [13.1], *Ann. Geophys.* 6, Copyright 1980, Springer Verlag

Both correlations give some clues to the possible mechanism. Since the tropospheric response is strongest for a southward interplanetary magnetic field accompanied by a stream of energetic protons, particles fed into the upper polar atmosphere appear to trigger the tropospheric changes, probably by modifying the stratospheric heat balance and circulation which then modifies the tropospheric circulation pattern. The larger thunderstorm area suggests that the additional ionization provided by the energetic particles directly influences atmospheric electricity, which in turn might lead to more efficient cloud formation.

Solar Cycle Signals in Tropospheric Winds. There are also suggestions for changes in the tropospheric wind patterns with the solar cycle. In the northern hemisphere, winter storm tracks are 2.5° farther south during solar maximum than during solar minimum. This is related to an equatorwards motion of the jet stream, a fast zonal wind stream in the upper troposphere which guides the motion of the pressure systems. For an observer at northern mid-latitudes the weather depends on whether she is north or south of the jet stream: south of the jet stream, she is met by the low-pressure regions and accompanying storms, giving a mild and rainy winter, while north of the jet stream the arctic high with clear and chilly skies dominates the weather.

Not only the average latitudes of the jet stream and the storm tracks shift but also their northward and southward excursions. Figure 13.8 shows the circulation pattern over the northern Atlantic ocean for extremely low (left), average (middle), and high (right) geomagnetic activity. The times of high geomagnetic activity are associated with strong westerlies over the Atlantic ocean, thus storms move from west to east, leading to mild and rainy weather in mid-Europe. With decreasing solar activity (middle), a blocking weather pattern develops over the Atlantic Ocean, leading to the penetration of Arctic

air into mid-Europe, as indicated by the arrow in the middle panel. If the geomagnetic activity is very low, the blocking pattern becomes even stronger with westerlies only at relatively low latitudes and the intrusion of Arctic air into mid-latitudes.

The changes in the tropospheric wind pattern are most likely a superimposition of two effects: the increased ionization and heating of the upper atmosphere by hard electromagnetic radiation during solar maximum as well as the increased precipitation of particles into the atmosphere at high latitudes. The latter ones obviously change the thermospheric/stratospheric circulation but do not have enough energy to change the tropospheric pattern. Nonetheless, it appears possible that atmospheric waves might provide a sufficient coupling between the different atmospheric layers to trigger such a straightening of the tropospheric flow pattern.

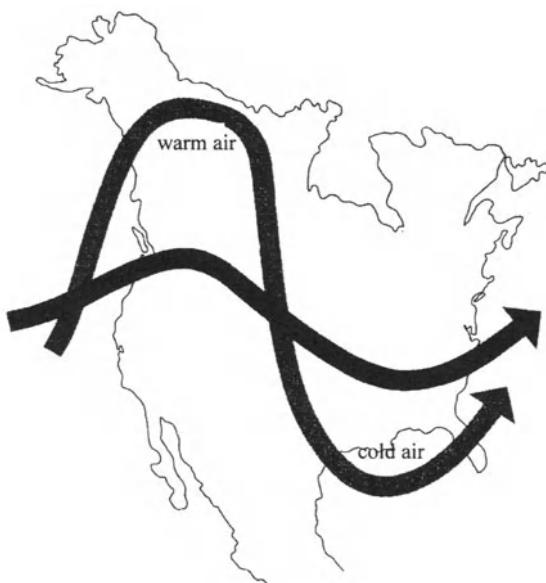


Fig. 13.9. Westerly flow and blocking weather pattern over North America. Reprinted from Roberts and Lansford [13.18], *The climate mandate*, Copyright 1979, with kind permission from Freeman Co.

There is another interesting aspect related to this interpretation. The change in the flow pattern might require only a relatively small amount of energy. In fact, changes between westerly and blocking patterns are frequent and natural, it is only the persistence and higher frequency of the blocking pattern at times of low solar activity that cause recognizable tropospheric effects. The consequences of this changed pattern, however, might evoke the impression of much larger energy inputs. In mid-Europe, a larger number of blocking weather patterns, which were observed, for instance, also during the Little Ice Age, always mean cold and relatively dry winters. Thus the average temperature in mid-Europe would be reduced. But this does not necessarily

imply a global temperature reduction. Instead, it is more a redistribution of heat; thus while in one place the temperatures are lower in another they might be higher. Figure 13.9 illustrates this. It shows the main direction of the upper tropospheric motion for a westerly flow and a blocking weather pattern over North America. The westerly flow brings mild and wet winters all over the US while the blocking pattern leads to a far northward excursion of warm air, leading to unusually high temperatures in Alaska. This air cools over Alaska and moves south towards Florida, leading to unusually cold eastern winters with a lot of snow. As a consequence, during one such blocking weather pattern in winter 1977, on some days in January temperatures were higher in parts of Alaska than in Florida. The western US, on the other hand, experiences a period of drought since the moist air coming from the Pacific Ocean has moved northwards. Thus even the dependence of the western US droughts on the solar cycle can be fitted, at least, qualitatively into this picture.

Caveat. I want to caution the reader that the above discussion represents only a tiny glimpse of a small facet of the complex climate system and its connection with solar activity. The correlations and interpretations given here are examples of investigations and lines of thought, pointing to mechanisms that might contribute to a connection between solar activity and climate. The real phenomenon “climate system”, however, cannot be reduced to one parameter and a sole cause. And the tampering of humankind with the system Earth should not be ignored, too.

14. Instrumentation

Measurement began our might.
W.B. Yeats, *Last Poems*

Observations in the solar-terrestrial environment are made from the ground, such as magnetic field measurements, *in situ* from rockets and satellites, and by means of remote sensing, for instance radio sounding of the ionosphere. Measured parameters cover fields and their fluctuations, plasmas, energetic particles, and electromagnetic radiation. In this chapter we give a brief overview of the basic principles of field, plasma, and particle measurements.

To study complex phenomena, not a single parameter but a set of parameters has to be determined. To study the propagation of energetic particles, for instance, not only the particles themselves have to be measured but also the magnetic field and its fluctuations. Thus not only does each instrument have to be designed cleverly to fulfill the specifications for space-flight (such as small mass, low power consumption) but also the combination of instruments on the spacecraft or rocket has to be chosen to yield the maximum of information for the topic under study. Some special publications contain summaries of mission goals as well as detailed technical descriptions of the instruments, for instance [14.3] for the SOHO mission, [14.11] for Ulysses, [14.12] for Wind, [14.7] for Polar, and [14.2] for the Galileo spacecraft.

14.1 Field Instruments

The measurement of the electromagnetic field can divided into the measurement of the magnetic and the electric fields. Since the spacecraft itself generates electric and magnetic fields, field sensors are always mounted on booms that extend up to some tens of meters from the spacecraft. These booms consist of thin strong wires which, during the launch of the spacecraft, were tightly wound on a reel. They are extended by the centrifugal force caused by the rotation of the spacecraft. Thus the sensors can be at very long beams in a plane perpendicular to the spacecraft's axis of rotation, while parallel to the spin axis only rather short, folded metal rods can be used as booms.

14.1.1 The Magnetic Field

The magnetic field is described by the magnetic field vector, i.e. the field direction and the flux density. The measurements can be done either by observing changes in the magnetic flux, for instance in the pulsation or fluxgate magnetometer, or by utilizing effects at the atomic level, such as proton precession or the Zeeman effect. These latter effects are used for absolute ground-based magnetic field measurements.

Pulsation Magnetometer. A pulsation magnetometer is a tri-axial arrangement of three coils. According to Faraday's induction law (2.6), changes in the magnetic flux through a surface cause an electromotoric force in its circumference. Thus the fluctuating field causes an induction current in the coil which is proportional to the change in the magnetic flux. The pulsation magnetometer is used to measure magnetic field fluctuations in three perpendicular directions with each coil giving one fluctuating magnetic field component. Since it cannot be used to determine the absolute flux density, it is only used in ground-based observatories, preferentially at high geomagnetic latitudes, to give detailed records of magnetic field fluctuations.

Fluxgate Magnetometer. Similar to the pulsation magnetometer, a fluxgate magnetometer is a tri-axial arrangement of three sensors. Fluxgate magnetometers not only measure the magnetic field fluctuations but also the components of a constant field. However, a fluxgate magnetometer is not an absolute instrument but has to be calibrated.

The sensor is a small transformer, wound around a high-permeability core. Its primary winding is excited by a high frequency-current, typically some kilohertz. Current and permeability are chosen so as to drive the core to saturation during each half-cycle of excitation. The secondary winding then detects a time-varying voltage which is related to the input signal by the hysteresis curve of the core material. For a high-permeability core, these curves are very non-linear and the output signal is highly distorted, containing all higher harmonics of the input signal. If there was no external magnetic field along the axis of the transformer, the hysteresis loop would be traversed in a symmetric manner. In this case, only the odd harmonics of the input frequency would show up at the output. In the presence of a magnetic field, however, saturation is acquired earlier in one half-cycle of excitation than in the other. This asymmetry adds even harmonics to the output signal. The amplitudes and phases of the even harmonics are then proportional to the field component parallel to the axis of the transformer.

Since the odd harmonics are much stronger than the even ones, their signals must be eliminated. This can be done with the ring-core or race-track transformer: two cores, each with its own primary coil, are operated in parallel with a common secondary winding. The primary windings are excited by equal currents of opposing direction. Thus if the external magnetic field was zero, the output at the secondary coil would be zero, too, because the odd

harmonics of both primaries cancel. If an external field is present, the odd harmonics still cancel, reducing the output signal to the desired quantities, the even harmonics.

Alkali–Vapor Magnetometers. The alkali-vapour magnetometer utilizes the Zeeman effect, i.e. the splitting of atomic lines in the presence of a magnetic field. The distance between these splitted lines is proportional to the magnetic field strength.

The sensor consists of two cells filled with atomic vapor, for instance Rb, aligned with a photocell. As the first cell is heated, it emits light with frequencies corresponding to the transition between the different sublevels of the two lowest atomic levels. This light passes through a circularly polarizing filter into the second cell. Here it is absorbed, raising the electrons to specific sublevels of the next highest level. Falling back to the lower level, these electrons emit light. Since this light is circularly polarized, only certain transitions between different sublevels are allowed, and after a brief time all electrons will fill the highest sublevel of the lowest level. Since transitions from the highest sublevel are forbidden, no further absorption occurs. The photocell thus detects the following chain of events: as the first cell is heated, light is emitted and detected by the photocell. As absorption sets in inside the second cell, this light is reduced until finally all electrons occupy the highest sublevel and the full stream of light hits the photocell again.

The second cell is surrounded by a coil which allows the transmission of a radio signal into the cell, redistributing the electrons to different sublevels. Thus the pumping process can be repeated. For redistribution to occur, the frequency of the radio signal must correspond to the frequency difference between the higher states. It is generated by using the output of the photocell to control the frequency of an oscillator. Thus the entire system oscillates between absorption during pumping and radiation when all electrons occupy the highest sublevel. The frequency of the radio signal required to maintain the oscillation depends on the frequency of the splitted lines and therefore is a measure for the magnetic field strength.

The alkali–vapor magnetometer is an extremely fast and accurate instrument; however, it only gives the total field but not its components.

Proton-Precession Magnetometers. Proton-precession magnetometers, too, are absolute instruments measuring field strength only. They utilize the fact that the magnetic moment of the proton makes it a small bar magnet. Thus protons can be aligned field-parallel, for instance by putting a container with a liquid rich in protons into a coil through which a strong current passes. The protons than align along this field. If the current is switched off, they try to re-align to the external magnetic field, the one to be measured. Since the electrons rotate, they behave like small gyroscopes, precessing around the field line. This precession frequency is inversely proportional to the gyromagnetic ratio g_p and proportional to the magnetic field. It can be determined from the magnetic effects of the precessing protons on the coil.

14.1.2 Electric Field Measurements

The basic idea in electric field measurement is simple: use two (or more) probes extending on booms from the spacecraft and measure the potential between them. But this information alone is not sufficient to determine the electric field because two effects modify this potential difference. First, the motion of the satellite through a magnetic field, which always is present in space plasmas, causes an electric induction field. In the inner magnetosphere, this might be as large as 0.5 V/m. Since this is a well-defined effect, it can easily be corrected for, using either simultaneously performed magnetic field measurements or the field strength at the satellite's position predicted from a magnetic field model. The second effect is more serious and has to be taken into account in instrument design: the probes (and the satellite) interact with the ambient plasma, and thus the potential measured at the probe does not reflect the electric field alone. Electrons and ions accumulate at the probe, forming a plasma sheet around it. If the probe, or part of it, is lit by sunlight, electrons also can be removed by photo-ionization. The sum of these currents gives the probe's floating potential, cf. Fig. 14.1. This floating potential can be balanced by applying a negative bias current to the probe.

A modern electric field instrument, such as the one on board the Freja satellite, uses a set of six probes extended from booms with lengths between 5 and 15 m; the probes on the Viking satellite are mounted on 40 m long wire booms.

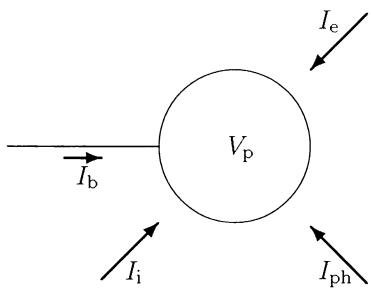


Fig. 14.1. Currents influencing the potential V_p measured by an electric field probe. Currents are due to photo-ionization of the probe (I_{ph}) and flows of ions (I_i) and electrons (I_e) towards the probe. These latter currents create the floating potential, which has to be balanced by a bias current I_b

An entirely different method for the electric field measurement utilizes the drift of electrons in a crossed electric and magnetic field, cf. Sect. 2.3.2. Such an instrument is far more complex; however, its measurements are not influenced by the interaction between the spacecraft and the plasma. Such an electric field sensor consists of an electron gun producing electrons with a well-defined energy, a system of coils to generate a homogeneous magnetic field perpendicular to the electron beam, and a detector with high spatial resolution. The latter is mounted perpendicular to the direction of electron injection close to the position of the electron gun, cf. Fig. 14.2. Thus an

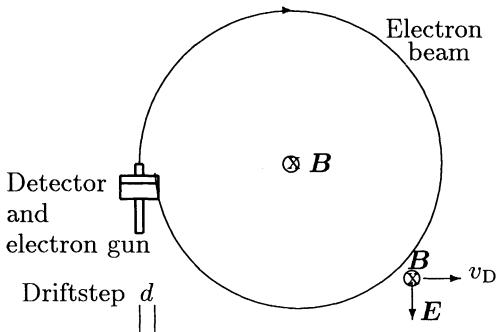


Fig. 14.2. Basic measurement principle to determine the electric field from the drift of electrons

electron is captured just after it has finished one gyro-orbit. Therefore, its drift speed and the direction of drift are known, which, combined with the prescribed electron speed and magnetic field strength, give the electric field.

14.1.3 Wave Measurements

To identify waves in space we have to measure the electric and magnetic fields, as well as variations in the plasma parameters, such as density and temperature. The maximum frequency which can be measured is determined by the temporal resolution of the field and plasma instruments. It should be sufficiently high to allow the identification of all waves of interest in the physical question.

The measurement of waves in space plasmas is a difficult task. In a textbook, a wave is sinusoidal of indefinite length. Space plasmas consist of a mixture of different waves and fluctuations; the waves often can be bursty, localized wave packets, occasionally even solitary waves. In addition, the observer is not fixed with respect to the wave field but moves relative to it, either because of his own motion, which would be the case in the quiet magnetosphere, or because the plasma and fields are swept over him, such as in interplanetary space or in the active magnetosphere. Thus the observed field and plasma variations consist of both temporal and spatial fluctuations.

The detection of waves in space plasmas therefore requires plasma and field measurements with various sensors and high sampling rates. Multipoint measurements are advantageous. For small wavelengths, these can be realized with antennas of various orientation and length. But even if very accurate, all these measurements give the fluctuations only. To identify the waves, a sound knowledge of wave physics and good guesses on the expected waves are required, too. Occasionally, particle measurements can be used to check the interpretation of the field fluctuations in terms of waves. One example was discussed in Sect. 9.2.2.

14.2 Plasma Instruments

Plasma instruments are designed to measure the density, temperature, speed, and composition. They can be divided into two groups: in dense plasmas particle energies are low and the instruments measure the collective behavior of the plasma, while in rarefied plasmas individual particles are collected.

14.2.1 Instruments for Dense Plasmas

The plasma densities inside the magnetosphere are high, and thus plasma properties can be determined from the plasma's collective behavior, such as currents. Rather simple probes and traps can be used to measure densities and temperatures. The measurement principle is based on a sensor extending into the plasma, and drawing electrons or ions from it, depending on the potential of the sensor. A trap utilizes the motion of the spacecraft relative to the plasma, like a snowplow piling up electrons and ions in front of it.

Langmuir Probes. The basic plasma sensor is a Langmuir probe. It can even detect particles with very low energies by the current they carry. Figure 14.3 gives the current–voltage characteristic of a Langmuir probe. The space potential V_s is used as a universal reference potential. Its meaning can be visualized as follows: imagine a large open-wire grid placed inside an ionized medium. Its potential relative to a large but remote body, such as Earth, can be varied. When the flow of electrons and ions through the holes in the grid no longer is modified by its presence, the grid has the space potential.

Let us now substitute the grid by a plate at the same potential. Then electrons and ions are collected at its surface. The electron current is $n_e e v_e$, and the ion current $n_i e v_i$. Assuming quasi-neutrality, we have $n = n_e = n_i$. Thus the net current towards the plate is $n e (v_i - v_e)$. Since the electron speed by far exceeds the proton speed, a negative current towards the plate

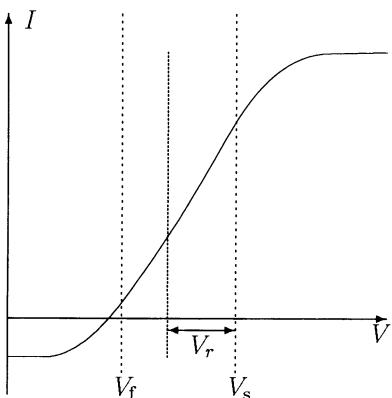


Fig. 14.3. Voltage–current characteristic of a Langmuir probe

results, representing the preferred electron current. If we now disconnect the plate from its original potential, it soon will be charged negatively up to the floating potential V_f . At this point the net flow vanishes and $e(n_i v_i - n_e v_e)$ becomes zero. Note that now the densities of electrons and ions are different. Since $v_e > v_i$ still holds, close to the plate the density of the electrons must be smaller than the one of the ions. Thus at the floating potential a negatively charged plate is surrounded by a cloud of positively charged ions, controlling the direction of the current towards the plate, comparable to the space charge in a diode. The thickness of the sheath is the Debye length.

The physical principle of the Langmuir probe is thus based on the different masses and speeds of electrons and ions and the resulting dependence of the net current on the probe's potential. The characteristic in Fig. 14.3 has three well-defined regions which can be utilized for measurement:

(a) If the probe's potential V exceeds the space potential V_s , electrons are drawn towards the probe while ions are repulsed. Thus the current is determined by the electron density n_e only.

(b) If the probe's potential is between the space potential and the floating potential, the current still is determined by the electrons. But only the faster electrons are collected at the probe since the retarding potential $V_r = V_s - V$ prevents the slower ones from reaching it. Thus the retarding potential determines the minimum kinetic energy an electron must have to be captured by the probe. Variation of V_r allows us to determine the change in electron density with energy, yielding the electron distribution. Since the electron distribution is a Maxwellian, the temperature can be determined.

(c) If the probe's potential V is smaller than the floating potential V_f , electrons are repulsed while ions are collected and the ion density can be determined. Since the speed of the ions in general is smaller than the spacecraft's speed v_{sc} , the theoretical treatment is more difficult than for the electrons. The collection condition then becomes $eV_r \leq m_i v_{sc}^2 / 2$, allowing us to determine ion masses and abundances.

Retarding Potential Analyzer. The retarding potential analyzer basically is a Langmuir probe supplemented by some grids. Since with this configuration a more detailed analysis of the ions becomes possible, it is also called an ion trap. At the upper grid, all ions arrive with approximately the same speed, the speed of the satellite. The main purposes of the grids are to deflect the electrons and to screen the detector from potential changes on other grids. But one grid serves a special purpose. Its potential can be varied between 0 and some tens of volts, deflecting also the lighter ions. Thus the ion composition can be determined. The analysis of the current-voltage characteristics gives the ion temperatures as well.

As well as retarding potential analyzers Langmuir probes can be built in different shapes, either planar, as discussed above, or spherical or cylindrical. The operating principle, however, is the same.

Impedance and Resonance Probes. An entirely different approach is chosen in the impedance and resonance probes. These probes do not measure the plasma itself but the electrical properties of the medium, which then allow us to derive the plasma density from theory.

In the impedance probe, the dependence of the dielectric constant ϵ on the electron density and radio frequency is used. Thus the probe has to measure the dielectric constant in space. In the laboratory, the dielectric constant of a material is determined by placing it inside a capacitor and then measure the change in its capacity. The same principle is used in space plasmas. Here the capacitor is designed to be part of an oscillator. Since its frequency depends on the capacity, the electron density can be inferred from it. Impedance probes often are flown on rockets. Since they are fast instruments, they allow for a very detailed analysis of the height profile of atmospheric ionization.

Resonance probes are based on a tuneable transmitter and a receiver. There are characteristic frequencies at which the medium between the transmitter and the receiver starts to resonate. Plasma parameters than can be determined from these resonances.

14.2.2 Instruments for Rarefied Plasmas

While the application of the plasma instruments discussed so far is limited to the magnetosphere, we will now discuss plasma instruments that also are used in the rarefied plasmas in interplanetary space. Nonetheless, today such instruments also are used in the denser magnetospheric plasmas.

Mass Spectrometer. A standard instrument for the analysis of a gaseous medium is the mass spectrum meter. Mass spectrometers in space obey the same principles as earth-bound ones, but also have two natural advantages: first, in space a spectrometer does not need a vacuum system since above an altitude of about 100 km collisions have become negligible. Second, there is no need to ionize the sample because all matter is ionized.

The basic principle of a mass spectrometer is the deflection of particles in a magnetic field by the Lorentz force. Using an acceleration voltage and a system of shutters, the particle speed can be prescribed. The deflection then depends on the particle's mass-to-charge ratio m/q . In the upper atmosphere, ions normally are singly ionized, and thus their mass can be determined.

Mass spectrometers are only rarely used on spacecraft since the main part of the spectrometer is a large magnet, adding to the weight of the instrument. In addition, the high magnetic fluxes might disturb other instruments.

Time-of-Flight Spectrometer. Time-of-flight spectrometers avoid the problems just outlined. The principle of a time-of-flight spectrometer is shown in Fig. 14.4. To reach the detector, the ions must pass through a series of grids, some of them biased by a steady potential, the central one having an oscillating potential. The potential drop V_1 between grids 1 and 2 would accelerate the ion to a speed $\sqrt{2q_i V_1 / m_i}$ which depends on the mass-to-charge

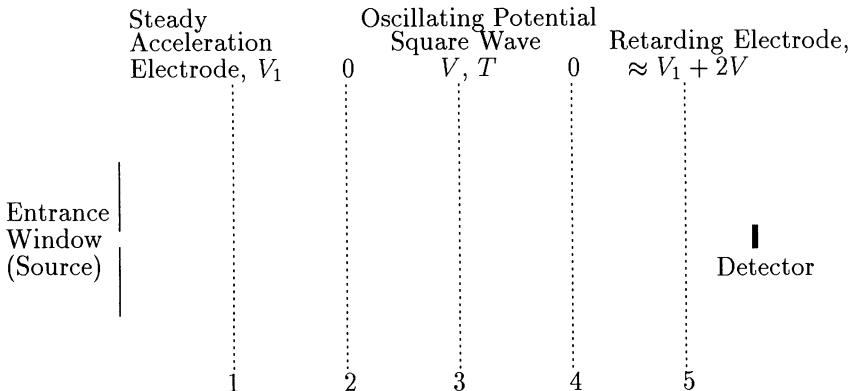


Fig. 14.4. Basic principle of a time-of-flight spectrometer

ratio m/q . However, depending on their initial speed, the ions still have different speeds. The oscillating potential at grid 3 only accelerates ions with speeds close to $2d/T$, with d being the distance between grid 2 and 3 and T being the oscillation period. Grid 5 now is used as a retarding electrode, preventing all ions that have not acquired the maximum possible speed from reaching the detector. Thus the ion beam passing grid 2 is velocity filtered and only ions with a certain mass-to-charge ratio reach the detector.

Time-of-flight spectrometers often are combined with electrostatic deflection systems. The flight path then more closely resembles the one in a mass spectrometer. One example of an extremely complex, but very advanced, combination of different sensors is CELIAS, the Charge Element and Isotope Analysis System on board SOHO [14.3].

Modern time-of-flight spectrometers are used to measure ions with energies up to some MeV/nuc. Thus the instruments not only measure the plasma composition but also energetic particles. However, only the lower end of the energetic particle population can be analyzed in a mass spectrometer since the higher energies would require longer travel paths and therefore larger instruments.

14.2.3 Energetic Particle Instruments

The measurement principle for energetic particles is based on their interaction with matter. Energy losses are due to ionization, bremsstrahlung, Cerenkov radiation, or nuclear interactions, cf. [14.10]. The energy loss inside the detector can be measured, giving us clues to the particle's properties. The most commonly used particle detector is a semiconductor.

If an energetic charged particle traverses matter, electrons will be shifted to higher energy levels (excitation) or will be removed from the atom (ionization). The energy loss of the energetic particle is described by the Bethe-

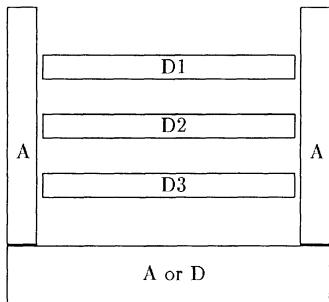


Fig. 14.5. Basic principle of a particle telescope: a stack of detectors D is surrounded by an anticoincidence detector A

Bloch relation and depends on the particle's parameters, i.e. charge and speed, but not on the particle's mass. Thus with only one detector, particles cannot be identified. Instead, a particle telescope is used, consisting of a stack of detectors surrounded by an anticoincidence detector, cf. Fig. 14.5. The anticoincidence detector serves two purposes: together with the upper two detectors, it defines an entrance aperture, i.e. the solid angle out of which particles can be detected. Particles coming from outside this angle might also hit the telescope and lose energy, for instance in detectors D2 and D3. But these particles can be singled out since they also have produced a signal in the anticoincidence detector. Since the anticoincidence detector only needs to detect the particle without determining its energy loss, often scintillation counters are used.

The information obtained with such a particle telescope are the energy losses dE/dx for each detector and the rest energy dE of the detector in which the particle comes to rest. Combined, this gives the particle's energy, which of course could also have been determined with a single sufficiently thick detector. But the use of a stack of detectors also allows the identification of the particles. A first separation between different particle species can be done due to a suitable choice of the first detector D1. The thickness of this detector can be chosen such that even a slow electron loses less energy than a fast proton and that even a fast proton loses less energy than a slow α particle. Thus by defining certain thresholds for the energy loss, electrons, protons, and heavier ions can be distinguished. This is a crude but simple method and can be used in onboard data analysis to assign the particles to energy channels. A more careful analysis is based on a combination of the energy losses in individual detectors. Here even different heavier ions and isotopes such as ${}^3\text{He}$ and ${}^4\text{He}$ can be distinguished.

Since particle telescopes have a well-defined aperture, they can be used to determine the angular distributions of energetic particles. On an axis-stabilized spacecraft, this can be done only by combining instruments looking in different directions. On a spinning spacecraft, the situation is simpler: if mounted perpendicular to the spin axis, the telescope scans all directions

during one rotation. Thus sectorized data can be obtained, which in turn allow us to determine the anisotropy.

The telescope in Fig. 14.5 gives only a very simple sketch. The thickness and the number of detectors determines the maximum particle energy to be measured. For higher energies, either more detectors have to be added or absorbers can be placed between the detectors. If these are passive absorbers, there will be a gap in the energy spectrum obtained by the instrument. Active absorbers, on the other hand, do not necessarily give the energy loss exactly, but at least some information on whether the particle is stuck in the last detector before the absorber or whether it has managed to pass into the absorber and is stuck there. Thus a continuous energy spectrum extending over a larger energy range can be obtained.

14.3 Supplementary Ground-Based Observations

Historically, ground-based observations are the oldest means for observing geospace. Although today rockets and satellites are attractive means to study the terrestrial environment *in situ*, ground-based observations nonetheless still are important, basically for two reasons. First, *in situ* observations are one-point measurements, giving extremely accurate data for one isolated spot without considering its surroundings. Ground-based observatories, on the other hand, often give a more regional overview, or combined even the global picture. For some questions, such as the extent and strength of geomagnetic disturbances, this global information is more useful than the *in situ* observation at one spot only. Thus ground-based observations often can be used to complement *in situ* measurements. In addition, ground-based measurements in many cases can provide a continuous baseline which allows the study of long-term variations. Second, ground-based measurements can provide access to regions which cannot be covered by *in situ* measurements, in particular the lower ionosphere up to a few hundred kilometer altitude where the atmospheric density is still too high to allow for long-living satellites.

Ground-based measurements comprise the magnetic field measurements, which in turn give information about the ionospheric currents, optical observations of the aurora, and measurements of ionospheric parameters by radio sounding and radar. Instruments for magnetic field measurements have already been described at the beginning of this chapter; the principle of radio sounding was discussed in Sect. 4.5. Auroral observations from the ground are performed with different cameras, in particular all-sky cameras showing the evolution of the aurora during a geomagnetic storm.

The most comprehensive information on ionospheric parameters can be obtained by radar. The basic principle in ionospheric radar is the same as used in radar on ships and airports: a radio wave is scattered back by the object of interest. During this scattering process the properties of the wave change, depending on the properties of the scattering object/medium. Radar

can be operated in two modes: in a monostatic radar, the transmitter and the receiver are located at the same position; in bistatic mode, the transmitter and the receiver are located at different positions. In the latter case the two antenna beams define the scattering volume while in the former case the scattering volume is defined by the transmitter modulation in the range direction.

In the first instance, the reflected signal is used to derive quantities describing the spectrum, such as width, power, or Doppler shift. From knowledge of how the properties of the medium influence the spectral shape, the parameters of the medium, such as densities, temperatures, or velocities, can be inferred. With the EISCAT radar system located in northern Scandinavia, the following plasma parameters can be measured: electron density, temperature, and velocity, ion temperature and velocity, ion mass, the relative number of negative ions, the gyro-frequency of ions, and the ion neutral collision frequency. From these parameters, other parameters can be deduced, for instance the electric field, the velocities of the neutral component, the ion composition, the current density, the energy deposition $E \cdot j$, and the particle's energy. Although this list is impressive, we should be aware of the fact that data are noisy and imperfect and thus assumptions have to be made about some of the parameters, allowing us to infer the others. For instance, temperature and collision frequency cannot be fitted independently. Instead, for heights below about 105 km, assumptions about the temperature are made and the collision frequency can be inferred, while at higher altitudes the modeled collision frequencies are used to infer the temperature.

A. Appendix

A.1 List of Symbols

α	pitch angle
β	plasma- β
γ	spectral index
γ_a	adiabatic exponent, c_p/c_v
η	viscosity
φ	electric potential
Φ	magnetic flux
$\kappa(\mu)$	pitch-angle diffusion coefficient
λ	mean free path
λ_c	Coulomb logarithm $\lambda_c = \ln \Lambda$
λ_D	Debye length
λ_{De}	electron Debye length
μ	magnetic moment
ν	kinematic viscosity
ν_c	collision frequency
ν_{ei}	electron-ion collision frequency
ω	angular frequency
Ω	angular speed
Ω	solid angle
ω_c	cyclotron frequency
ω_{ce}	electron cyclotron frequency
ω_{cp}	proton cyclotron frequency
ω_p	plasma frequency
ω_{pe}	electron plasma frequency
ω_{pp}	proton plasma frequency
ψ	azimuthal angle in a circular motion
Ψ	stream function
ρ	mass density
ρ_c	charge density
ρ_e	electron density
σ	conductivity
σ	scattering cross-section

τ	collision time
θ_{Bn}	angle between shock normal and magnetic field direction
A	area
A	vector potential
B	magnetic field
c	speed of light
c_p	specific heat at constant pressure
c_v	specific heat at constant volume
D	spatial diffusion coefficient
D	diffusion tensor
e	elementary charge
E	electric field
e	unit vector
F	force
f	force density
$f(\mathbf{r}, \mathbf{v}, t)$	distribution function (phase space density)
g	gravitation
G	universal constant of gravitation
j	current density
j_D	drift current density
J	differential flux
J_i	action integral
k	wave vector
k	wave number
k_B	Boltzmann's constant
L	characteristic length scale
l	vector along a path
m	mass
m_e	electron mass
m_p	proton mass
n	number density
n_e	number density of electrons
n_p	number density of protons
p	momentum
p_i	generalized momentum
p	pressure
p_M	magnetic pressure
P	stress tensor
P	rigidity
q	charge
q_i	generalized spatial coordinate
r	radius vector
r_c	radius of curvature
r_L	Larmor radius

r_{Le}	electron Larmor radius
R_M	magnetic Reynolds number
\mathbf{S}	vector defining a surface
s	path length (along the Archimedean spiral)
\mathbf{S}_P	Poynting vector
T	temperature
T_e	electron temperature
T_c	gyration time (cyclotron period)
\mathbf{u}	bulk velocity
U	particle flux
v_{sowi}	solar wind speed
\mathbf{v}	velocity of a particle
v_{\perp}	speed perpendicular to the field
v_{\parallel}	speed parallel to the field
v_A	Alfvén speed
v_D	drift velocity
v_e	electron speed
v_g	group speed
\mathbf{v}_{gc}	velocity of the guiding center
v_p	proton speed
v_{ph}	phase speed
v_s	sound speed
v_{th}	thermal speed
$v_{\text{th},e}$	thermal speed of electrons

A.2 Useful Relations

A.2.1 Vector Identities

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A} \tag{A.1}$$

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{B} \cdot \mathbf{A} \tag{A.2}$$

$$\mathbf{A} \times \mathbf{B} = -\mathbf{B} \times \mathbf{A} \tag{A.3}$$

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \mathbf{C} \cdot (\mathbf{A} \times \mathbf{B}) = \mathbf{B} \cdot (\mathbf{C} \times \mathbf{A}) \tag{A.4}$$

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B}) \tag{A.5}$$

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) + \mathbf{B} \times (\mathbf{C} \times \mathbf{A}) + \mathbf{C} \times (\mathbf{A} \times \mathbf{B}) = 0 \tag{A.6}$$

$$(\mathbf{A} \times \mathbf{B}) \cdot (\mathbf{C} \times \mathbf{D}) = (\mathbf{A} \cdot \mathbf{C})(\mathbf{B} \cdot \mathbf{D}) - (\mathbf{A} \cdot \mathbf{D})(\mathbf{B} \cdot \mathbf{C}) \tag{A.7}$$

$$(\mathbf{A} \times \mathbf{B}) \times (\mathbf{C} \times \mathbf{D}) = (\mathbf{A} \times \mathbf{B} \cdot \mathbf{D})\mathbf{C} - (\mathbf{A} \times \mathbf{B} \cdot \mathbf{C})\mathbf{D} \tag{A.8}$$

A.2.2 Vector Calculus

$$\nabla = e_x \frac{\partial}{\partial x} + e_y \frac{\partial}{\partial y} + e_z \frac{\partial}{\partial z} \quad (\text{A.9})$$

$$\nabla \mathbf{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \quad (\text{A.10})$$

$$\nabla \times \mathbf{A} = \begin{pmatrix} e_x & e_y & e_z \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ A_x & A_y & A_z \end{pmatrix} \quad (\text{A.11})$$

$$\nabla \cdot (\nabla \times \mathbf{A}) = 0 \quad (\text{A.12})$$

$$\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} \quad (\text{A.13})$$

$$\nabla \times \nabla \phi = 0 \quad (\text{A.14})$$

$$\nabla \cdot \nabla \phi = \nabla^2 \phi \quad (\text{A.15})$$

$$\nabla(\phi\rho) = \nabla(\rho\phi) = \phi\nabla\rho + \rho\nabla\phi \quad (\text{A.16})$$

$$\nabla \cdot (\phi \mathbf{A}) = \phi \nabla \cdot \mathbf{A} + \mathbf{A} \cdot \nabla \phi \quad (\text{A.17})$$

$$\nabla \times (\phi \mathbf{A}) = \phi \nabla \times \mathbf{A} + (\nabla \phi) \times \mathbf{A} \quad (\text{A.18})$$

$$\nabla \cdot (\mathbf{A} \times \mathbf{B}) = \mathbf{B} \cdot (\nabla \times \mathbf{B}) - \mathbf{A} \cdot (\nabla \times \mathbf{B}) \quad (\text{A.19})$$

$$\nabla \times (\mathbf{A} \times \mathbf{B}) = \mathbf{A}(\nabla \cdot \mathbf{B}) - \mathbf{B}(\nabla \cdot \mathbf{A}) + (\mathbf{B} \cdot \nabla)\mathbf{A} - (\mathbf{A} \cdot \nabla)\mathbf{B} \quad (\text{A.20})$$

$$\nabla(\mathbf{A} \cdot \mathbf{B}) = \mathbf{A} \times (\nabla \times \mathbf{B}) + \mathbf{B} \times (\nabla \times \mathbf{A}) + (\mathbf{A} \cdot \mathbf{B})\mathbf{B} + (\mathbf{B} \cdot \nabla)\mathbf{A} \quad (\text{A.21})$$

V is a volume enclosed by a surface \mathbf{S} . $d\mathbf{S} = e_n dS$ is a vector normal to this surface. Then we have

$$\int_V = \nabla \phi dV = \int_S \phi d\mathbf{S} \quad (\text{A.22})$$

Divergence theorem (Gauss's theorem):

$$\int_V \nabla \cdot \mathbf{A} d^3x = \int_S \mathbf{A} \cdot d\mathbf{S} \quad (\text{A.23})$$

$$\int_V \nabla \cdot \mathbf{T} dV = \int_S \mathbf{T} \cdot d\mathbf{S} \quad (\text{A.24})$$

$$\int_V \nabla \times \mathbf{A} dV = \int_S d\mathbf{S} \times \mathbf{A} \quad (\text{A.25})$$

$$\int_V (\phi \nabla^2 \rho - \rho \nabla^2 \phi) dV = \int_S (\phi \nabla \rho - \rho \nabla \phi) \cdot d\mathbf{S} \quad (\text{A.26})$$

$$\int_V (\mathbf{A} \cdot \nabla \times \nabla \times \mathbf{B} - \mathbf{B} \cdot \nabla \times \nabla \times \mathbf{A}) = \int_S (\mathbf{B} \times \nabla \times \mathbf{A} - \mathbf{A} \times \nabla \times \mathbf{B}) \cdot d\mathbf{S} \quad (\text{A.27})$$

S is an open surface enclosed by a curve C with length element dl :

$$\int_S d\mathbf{S} \times \nabla \phi = \oint_C dl \phi \quad (\text{A.28})$$

Stokes' law:

$$\int_S (\nabla \times \mathbf{A}) \cdot d\mathbf{S} = \oint_C \mathbf{A} \cdot dl \quad (\text{A.29})$$

$$\int_S (d\mathbf{S} \times \nabla) \times \mathbf{A} = \oint_C dl \times \mathbf{A} \quad (\text{A.30})$$

$$\int_S (d\mathbf{S} \cdot (\nabla \phi)) \times \nabla \rho = \oint_C \phi d\rho = - \oint_C \rho d\phi \quad (\text{A.31})$$

A.2.3 Cylindrical Coordinates

$$\nabla \cdot \mathbf{A} = \frac{1}{r} \frac{\partial}{\partial r} (r A_r) + \frac{1}{r} \frac{\partial A_\theta}{\partial \theta} + \frac{\partial A_z}{\partial z} \quad (\text{A.32})$$

$$\nabla \phi = \left(\frac{\partial \phi}{\partial r}, \frac{1}{r} \frac{\partial \phi}{\partial \theta}, \frac{\partial \phi}{\partial z} \right) \quad (\text{A.33})$$

$$\nabla \times \mathbf{A} = \left(\frac{1}{r} \frac{\partial A_z}{\partial \theta} - \frac{\partial A_\theta}{\partial z}, \frac{\partial A_r}{\partial z} - \frac{\partial A_z}{\partial r}, \frac{1}{r} \frac{\partial (r A_\theta)}{\partial r} - \frac{1}{r} \frac{\partial A_r}{\partial \theta} \right) \quad (\text{A.34})$$

$$\nabla^2 \phi = \frac{1}{r} \left(\frac{\partial \phi}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 \phi}{\partial \theta^2} + \frac{\partial^2 \phi}{\partial z^2} \quad (\text{A.35})$$

$$\nabla^2 \mathbf{A} = \left(\nabla^2 A_r - \frac{2}{r^2} \frac{\partial A_\theta}{\partial \theta} - \frac{A_r}{r^2}, \nabla^2 A_\theta + \frac{2}{r^2} \frac{\partial A_\theta}{\partial \theta} - \frac{A_\theta}{\partial r^2}, \nabla^2 A_z \right) \quad (\text{A.36})$$

A.2.4 Spherical Coordinates

$$\nabla \cdot \mathbf{A} = \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 A_r) + \frac{1}{r \sin \theta} \frac{\partial}{\partial \theta} (A_\theta \sin \theta) + \frac{1}{r \sin \theta} \frac{\partial A_\phi}{\partial \phi} \quad (\text{A.37})$$

$$\nabla \rho = \left(\frac{\partial \rho}{\partial r}, \frac{1}{r} \frac{\partial \rho}{\partial \theta}, \frac{1}{r \sin \theta} \frac{\partial \rho}{\partial \phi} \right) \quad (\text{A.38})$$

$$\begin{aligned} \nabla \times \mathbf{A} = & \left(\frac{1}{r \sin \theta} \frac{\partial (A_\phi \sin \theta)}{\partial \theta} - \frac{1}{r \sin \theta} \frac{\partial A_\theta}{\partial \phi}, \right. \\ & \left. \frac{1}{r \sin \theta} \frac{\partial A_r}{\partial \varphi} - \frac{1}{r} \frac{\partial (r A_\phi)}{\partial r}, \frac{1}{r} \frac{\partial (r A_\theta)}{\partial r} - \frac{1}{r} \frac{\partial A_r}{\partial \theta} \right) \quad (\text{A.39}) \end{aligned}$$

$$\nabla^2 \rho = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \rho}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial \rho}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 \rho}{\partial \phi^2} \quad (\text{A.40})$$

A.3 Useful Numbers

A.3.1 Fundamental Constants

c	speed of light in a vacuum	2.998×10^8 m/s
ϵ_0	permittivity in a vacuum	8.854×10^{-12} F/m
μ_0	permeability in a vacuum	$4\pi \times 10^{-7}$ H/m
e	elementary charge	1.602×10^{-19} C
m_e	electron mass	9.109×10^{-31} kg
m_p	proton mass	1.673×10^{-27} kg
m_p/m_e	proton-to-electron mass ratio	1.836×10^3
σ	Stefan–Boltzmann number	5.6708×10^{-8} Jm $^{-2}$ s $^{-1}$ K $^{-4}$
k_B	Boltzmann number	1.381×10^{-23} J/K
eV	electronvolt	1.602×10^{-19} J
eV/k_B	temperature equivalent for 1 eV	1.160×10^4 K
G	gravitational constant	6.673×10^{-11} Nm 2 kg $^{-2}$

A.3.2 Numbers in Plasmas

In the following numerical expressions the plasma parameters have to be inserted in SI units [as long as no other units are indicated in square brackets]: the electron density n_e has to be inserted in m $^{-3}$, B in Tesla, v_\perp in m/s and the electron temperature T_e in K. The fundamental frequencies in plasmas then are given as:

ω_{pe}	electron plasma frequency	$0.564 \times \sqrt{n_e}$ rad/s
ω_{pp}	proton plasma frequency	$1.32 \times \sqrt{n_p}$ rad/s
ω_{ce}	electron cyclotron frequency	$1.76 \times 10^{11} B$ rad/s = $28B[nT]$ Hz
ω_{cp}	proton cyclotron frequency	$9.58 \times 10^7 B$ rad/s = $0.01525B[nT]$ Hz

These are angular frequencies. Frequencies can be obtained by division by 2π .

Important length scales and speeds in a plasma can be calculated according to the following rules:

$v_{Th,e}$	electron thermal speed, $\sqrt{2k_B T_e/m}$	$5.51 \times 10^3 \sqrt{T_e}$ m/s
$v_{Th,i}$	ion thermal speed, $\sqrt{2k_B T_i/M_i}$	$129 \times \sqrt{T_i \cdot m_p/M_i}$ m/s
λ_{De}	electron Debye length, $\sqrt{k_B T_e/m}/\omega_{pe}$	$69 \times \sqrt{T_e/n_e}$ m
r_{Le}	electron Larmor radius, v_\perp/ω_{ce}	$5.68 \times 10^{-12} v_\perp/B$ m
r_{Lp}	proton Larmor radius, v_\perp/ω_{cp}	$1.04 \times 10^{-8} v_\perp/B$ m
v_s	sound speed	$0.117 \sqrt{T_e}$ km/s

References

Chapter 1

- 1.1 Bequerel, H. (1879): *Aurora, their characters and spectra*, Spon, London
- 1.2 Birkeland, K. (1908): *The Norwegian aurora polaris expedition 1902 - 1903, vol. 1: On the cause of magnetic storms and the origin of terrestrial magnetism*, H. Ashebourg, Christiania
- 1.3 Celsius, A. (1847): Bemerkungen über der Magnetnadel stündliche Veränderungen in ihrer Abweichung, *Svenska Vetensk. Handl.* **8**, 296
- 1.4 Chapman, S. and V.c.A. Ferraro (1931): A new theory of magnetic storms, *Terr. Magn.* **36**, 77
- 1.5 Franklin, J. (1975): *A narrative of a journey to the shores of the polar sea in the years 1819, 1820, 1821 and 1822*, Hurtig, Edmonton
- 1.6 Fritz, H. (1881): *Das Polarlicht*, Brockhaus, Leipzig
- 1.7 Gillmor, C.S. and J.R. Spreiter (eds.) (1997): *Discovery of the magnetosphere*, History of Geophysics Vol. 7, American Geophysical Union, Washington DC
- 1.8 Hiorter, O.P. (1847): Von der Magnetnadel verschiedener Bewegungen, *Svenska Vetensk. Handl.* **8**, 27
- 1.9 Humboldt, A. (1808): Die vollständigste aller bisherigen Beobachtungen über den Einfluß des Nordlichts auf die Magnetnadel, *Glob. Ann.* **29**
- 1.10 Loomis, E. (1860): On the geographical distribution of auroras in the northern hemisphere, *Amer. J. Sci. Arts* **30**, 89
- 1.11 Rosenbauer, H., H. Grünwaldt, M.D. Montgomery, G. Paschmann, and N. Sckopke (1975): Heos 2 plasma observations in the distant polar magnetosphere: the plasma mantle, *J. Geophys. Res.* **80**, 2723
- 1.12 Russell, C.T. (1995): A brief history of solar-terrestrial physics, in *Introduction to space physics* (eds. M.G. Kivelson and C.T. Russel), Cambridge University Press, Cambridge
- 1.13 Stern, D.P. (1989): A brief history of magnetospheric physics before the space-flight era, *Rev. Geophys.* **27**, 103
- 1.14 Stern, D.P. (1996): A brief history of magnetospheric physics during the space age, *Rev. Geophys.* **34**, 1

Chapter 2

- 2.1 Alfvén, H. and G. Fälthammar (1963): *Cosmical electrodynamics*, Clarendon Press, Oxford
- 2.2 Chen, F.F. (1984): *Plasma physics and controlled fusion*, vol. 1, Plenum Press, New York

- 2.3 Dendy, R. (ed.) (1993): *Plasma physics*, Cambridge University Press, Cambridge
- 2.4 Fermi, E. (1949): On the origin of cosmic radiation, *Phys. Rev.* **5**, 1169
- 2.5 Fermi, E. (1954): Galactic magnetic fields and the origin of cosmic radiation, *Astrophys. J.* **119**, 1
- 2.6 Kivelson, M.G. (1995): Physics of space plasmas, in *Introduction to space physics* (eds. M.G. Kivelson and C.T. Russell), Cambridge University Press, Cambridge
- 2.7 Landau, L.D. and E.M. Lifschitz (1982): *Mechanics*, Butterworth-Heinemann, Oxford
- 2.8 Northrop, T.G. (1963): *The adiabatic motion of charged particles*, Interscience, New York
- 2.9 Parks, G.K. (1991): *Physics of space plasmas – an introduction*, Addison-Wesley, Redwood City, CA

Chapter 3

- 3.1 Alfvén, H. and G. Fälthammar (1963): *Cosmical electrodynamics*, Clarendon Press, Oxford
- 3.2 Baumjohann, W. and R.A. Treumann (1996): *Basic space plasma physics*, Imperial College Press, London
- 3.3 Chen, F.F. (1984): *Plasma physics and controlled fusion*, vol. 1, Plenum Press, New York
- 3.4 Cowling, T.G. (1957): *Magnetohydrodynamics*, Interscience, New York
- 3.5 Dendy, R.O. (1990): *Plasma dynamics*, Oxford Science Publications, Oxford
- 3.6 Dendy, R. (ed.) (1993): *Plasma physics*, Cambridge University Press, Cambridge
- 3.7 Dorman, L.I. and G.I. Freidman (1959): *Problems of magnetohydrodynamics and plasma dynamics*, Zinātne, Riga
- 3.8 Faber, T.E. (1995): *Fluid dynamics for physicists*, Cambridge University Press, Cambridge
- 3.9 Goldston, R.J. and P.H. Rutherford (1995): *Plasma physics*, IOP, Bristol
- 3.10 Hones, E.W. (ed.) (1984): *Magnetic reconnection*, Geophys. Monogr. 30, American Geophys. Union, Washington, DC
- 3.11 Kippenhahn, R. and C. Möllenhoff (1975): *Elementare Plasmaphysik*, BI-Wissenschaftsverlag, Mannheim
- 3.12 Kippenhahn, R. and A. Schlüter (1957): *Z. Astrophys.* **43**, 36
- 3.13 Kivelson, M.G. and C.T. Russell (eds.) (1995): *Introduction to space physics*, Cambridge University Press, Cambridge
- 3.14 Kivelson, M.G. (1995): Physics of space plasmas, in [3.13], p. 27
- 3.15 Krause, F. and K.-H. Rädler (1980): *Mean-field magnetohydrodynamics and dynamo theory*, Akademie-Verlag, Berlin
- 3.16 Parks, G.K. (1991): *Physics of space plasmas – an introduction*, Addison-Wesley, Redwood City, CA
- 3.17 Petschek, H.E. (1964): Magnetic field annihilation, in *Symp. on Physics of Solar Flares*, 425, NASA SP-50, Goddard Space Flight Center
- 3.18 Piddington, J.H. (1969): *Cosmic electrodynamics*, Wiley, New York
- 3.19 Priest, E.R. (1982): *Solar magnetohydrodynamics*, Dordrecht, Reidel
- 3.20 Sakai, J.-I. and Y. Ohsawa (1987): Particle acceleration by magnetic reconnection and shocks during current loop coalescence in solar flares, *Space Sci. Rev.* **46**, 113

- 3.21 Spitzer, L. (1962): *Physics of fully ionized gases*, Interscience, New York
- 3.22 Sweet, P.A. (1958): in *Electromagnetic phenomena in cosmical physics* (ed. B. Lehnert), Cambridge University Press
- 3.23 Treumann, R.A. and W. Baumjohann (1996): *Advanced space plasma physics*, Imperial College Press, London
- 3.24 Weiss, N.O. (1966): *Proc. R. Soc. London*, **A293**, 310

Chapter 4

- 4.1 Baumjohann, W., and R.A. Treumann (1996): *Basic space plasma physics*, Imperial College Press, London
- 4.2 Bohm, D. and E.P. Gross (1949): Theory of plasma oscillations, A. Origin of medium-like behavior, *Phys. Rev.* **75**, 1851
- 4.3 Chen, F.F. (1984): *Plasma physics and controlled fusion*, vol. 1, Plenum Press, New York
- 4.4 Goertz, C.K. and R.J. Strangeway (1995): Plasma waves, in *Introduction to space physics* (eds. M.G. Kivelson and C.T. Russell), Cambridge University Press, Cambridge
- 4.5 Goldston, R.J., and P.H. Rutherford (1995): *Plasma physics*, IOP, Bristol
- 4.6 Krall, N.A. and A.W. Trivelpiece (1986): *Principles in plasma physics*, San Francisco Press, San Francisco, CA.
- 4.7 Parks, G.K. (1991): *Physics of space plasmas – an introduction*, Addison-Wesley, Redwood City, CA
- 4.8 Stix, T.H. (1992): *Waves in plasmas*, American Institute of Physics, New York
- 4.9 Swanson, D.G. (1989): *Plasma waves*, Academic Press, Boston

Chapter 5

- 5.1 Benz, A.O. (1993): *Plasma astrophysics*, Kluwer, Dordrecht
- 5.2 Boyd, T.J.M. and J.J. Sanderson (1969): *Plasma dynamics*, Barnes and Noble, New York
- 5.3 Chen, F.F. (1984): *Plasma physics and controlled fusion*, vol. 1, Plenum Press, New York
- 5.4 Goldston, R.J. and P.H. Rutherford (1995): *Plasma physics*, IOP, Bristol
- 5.5 Krall, N.A. and A.W. Trivelpiece (1986): *Principles of plasma physics*, San Francisco Press Inc., San Francisco, CA
- 5.6 Sturrock, P.A. (1994): *Plasma physics*, Cambridge University Press, Cambridge

Chapter 6

- 6.1 Fisk, L.A. (1979): The interaction of energetic charged particles with the solar wind, in *Solar system plasma physics*, vol. I (eds. C.F. Kennel, L.J. Lanzerotti, and E.N. Parker), North-Holland, Amsterdam, p. 177

- 6.2 Forman, M.A., R. Ramaty, and E.G. Zweibel (1986): The acceleration and propagation of solar energetic flare particles, in *Physics of the Sun, vol. II* (ed. P.A. Sturrock), Reidel, Dordrecht
- 6.3 Hasselmann, K. and G. Wibberenz (1968): Scattering of charged particles by random electromagnetic fields, *Z. Geophys.* **34**, 353
- 6.4 Jokipii, J.R. (1966): Cosmic ray propagation. I: Charged particles in random electromagnetic fields, *Astrophys. J.* **146**, 480
- 6.5 Schlickeiser, R. (1984): Stochastic particle acceleration in astrophysical objects, in *Cosmic radiation in contemporary astrophysics* (ed. M.M. Shapiro), Reidel, Dordrecht
- 6.6 Toptygin, I.N. (1983): *Cosmic-rays in interplanetary magnetic fields*, Reidel, Dordrecht
- 6.7 Treumann, R.A. and W. Baumjohann (1996): *Advanced space plasma physics*, Imperial College Press, London
- 6.8 Tsurutani, B.T. and C.S. Lakhina (1997): Some basic concepts of wave-particle interactions in collisionless plasmas, *Rev. Geophys.* **35**, 491
- 6.9 Tverskoy, B.A., 1967: *Sov. Phys. JETP* **25**, 317
- 6.10 Walt, M. (1994): *Introduction to geomagnetically trapped radiation*, Cambridge University Press, Cambridge
- 6.11 Wibberenz, G. (1973): Propagation of cosmic rays in the interplanetary space, in *Lectures on space physics, vol. 1* (eds. A. Bruzek and H. Pilkuhn), Bertelsmann, Gütersloh

Chapter 7

- 7.1 Armstrong, T.P., M.E. Pesses, and R. B. Decker (1985): Shock drift acceleration, in [7.44], p. 271
- 7.2 Axford, W.I. (1982): Acceleration of cosmic rays by shock waves, in *Plasma Astrophysics* (eds. T.D. Tiyenen and T. Levy), ESA-SP 151
- 7.3 Axford, W.I., E. Leer, and G. Skadron (1977): The acceleration of cosmic rays by shock waves, *Proc. 15th Int. Cosmic Ray Conf.* **11**, 132
- 7.4 Bell, A.R.I. (1978): The acceleration of cosmic rays in shock fronts, *Mont. Not. R. Astron. Soc.* **182**, 147
- 7.5 Blandford, R.D., and J.P. Ostriker (1978): Particle acceleration by astrophysical shocks, *Astrophys. J. Lett.* **221**, L29
- 7.6 Boyd, T.J.M. and J.J. Sanderson (1969): *Plasma dynamics*, Barnes and Noble, New York
- 7.7 Burgess, D. (1995): Collisionless shocks, in *Introduction to space physics* (eds. M.G. Kivelson and C.T. Russel), Cambridge University Press, Cambridge, p. 129
- 7.8 Chao, J.K. and Y.H. Chen (1985): On the distribution of θ_{Bn} for shocks in the solar wind, *J. Geophys. Res.* **90**, 149
- 7.9 Chao, J.K. and B. Goldstein (1972): Observations of slow shocks in interplanetary space, *J. Geophys. Res.* **75**, 6394
- 7.10 Courant, R. and K.O. Friedrichs (1991): *Supersonic flow and shock waves*, Springer, Berlin, reprint of the second Interscience edition, 1948
- 7.11 Decker, R.B. (1981): The modulation of low-energy proton distributions by propagating interplanetary shock waves: a numerical simulation, *J. Geophys. Res.* **86**, 4537
- 7.12 Decker, R.B. (1983): Formation of shock spike events at quasi-perpendicular shocks, *J. Geophys. Res.* **88**, 9959

- 7.13 Decker, R.B. (1988): Computer modelling of test particle acceleration at oblique shocks, *Space Sci. Rev.* **48**, 195
- 7.14 Decker, R.B. and L. Vlahos (1985): Shock drift acceleration in the presence of waves, *J. Geophys. Res.* **90**, 47
- 7.15 Decker, R.B. and L. Vlahos (1986): Modeling of ion acceleration through drift and diffusion at interplanetary shocks, *J. Geophys. Res.* **91**, 13349
- 7.16 Decker, R.B. and L. Vlahos (1986): Numerical studies of particle acceleration at turbulent oblique shocks with an application to prompt ion acceleration during solar flares, *Astrophys. J.* **306**, 710
- 7.17 de Hoffman, F. and E. Teller (1950): Magnetohydrodynamic shocks, *Space Sci. Rev.* **80**, 692
- 7.18 Dorman, L.I. and G.I. Freidman (1959): *Problems of magnetohydrodynamics and plasma dynamics*, Zinātne, Riga
- 7.19 Drury, L.O'C. (1983): An introduction to the theory of diffusive shock acceleration of energetic particles in tenuous plasmas, *Rep. Prog. Phys.* **46**, 973
- 7.20 Dryer, M., S.T. Wu, G. Gislason, S.F. Han, Z.K. Smith, D.F. Smart, and M.A. Shea (1984): Magnetohydrodynamic modeling of interplanetary disturbances between sun and earth, *Astrophys. Space Sci.* **195**, 187
- 7.21 Forman, M.A. and G.M. Webb (1985): Acceleration of energetic particles, in [7.42], p. 91
- 7.22 Hundhausen, A.J. (1972): *Coronal expansion and the solar wind*, Springer, Berlin
- 7.23 Hundhausen, A.J. (1985): Some macroscopic properties of shock waves in the heliosphere, in [7.42]
- 7.24 Hundhausen, A.J., T.E. Holzer, and B.C. Low, 1987: Do slow shocks precede some coronal mass ejections?, *J. Geophys. Res.* **92**, 11173
- 7.25 Jones, F.C. and D.C. Ellison (1991): The plasma physics of shock acceleration, *Space Sci. Rev.* **58**, 259
- 7.26 Kennel, C.F., F.V. Coroniti, F.L. Scarf, W.A. Livesay, C.T. Russell, E.J. Smith, K.-P. Wenzel, and M. Scholer (1986): A test of Lee's quasi-linear theory of ion acceleration by traveling interplanetary shocks, *J. Geophys. Res.* **91**, 11917
- 7.27 Lee, M.A. (1982): Coupled hydromagnetic wave excitation and ion acceleration upstream of the earth's bow shock, *J. Geophys. Res.* **87**, 5063
- 7.28 Lee, M.A. (1983): Coupled hydromagnetic wave excitation and ion acceleration at interplanetary traveling shocks, *J. Geophys. Res.* **88**, 6109
- 7.29 Lee, M.A. (1986): Acceleration of energetic particles at solar wind shocks, in *The sun and the heliosphere in three dimensions* (ed. R.G. Marsden), Reidel, Dordrecht, p. 305
- 7.30 McKenzie, J.F. and H.J. Völk (1982): Non-linear theory of cosmic ray shocks including self-generated Alfvén waves, *Astron. Astrophys.* **116**, 191
- 7.31 Morfill, G. and M. Scholer (1977): Influence of interplanetary shocks on solar energetic particle events, *Astrophys. Space Sci.* **46**, 73
- 7.32 Pizzo, V.J. (1985): Interplanetary shocks on the large scale: a retrospective on the last decades theoretical efforts, in [7.44], p. 51
- 7.33 Richter, A.K. (1991): Interplanetary slow shocks, in *Physics of the inner heliosphere, vol. 2* (eds. E. Marsch and R. Schwenn), Springer, Berlin, p. 235
- 7.34 Russell, C.T., E.J. Smith, B.T. Tsurutani, J.T. Gosling, and S.J. Bame (1993): Multiple spacecraft observations of interplanetary shocks: characteristics of the upstream ULF turbulence, in *Solar Wind Five*, (ed. M. Neugebauer), NASA CP 2280, Washington, DC, p. 385
- 7.35 Saagdev, R.Z. and C.F. Kennel (1991): Collisionless shock waves, *Sci. Am.*, April 1991, 40

- 7.36 Schatzmann, E., 1963: On the acceleration of particles in shock fronts, *Ann. Astrophys.* **26**, 234
- 7.37 Scholer, M. (1985): Diffusive acceleration, in [7.44], p. 287
- 7.38 Scholer, M. (1987): Observational overview of energetic particle populations associated with interplanetary shocks, in *Solar Wind IV* (eds. V. Pizzo, T.E. Holzer, and D.G. Sime), NCAR/TN306+Proc., Nat. Center of Atm. Res., Boulder, CO
- 7.39 Schwenn, R. (1983): Direct correlation between coronal transients and interplanetary disturbances, *Space Sci. Rev.* **34**, 85
- 7.40 Smith, Z.K. and M. Dryer (1990): MHD study of temporal and spatial evolution of simulated interplanetary shocks in the ecliptic plane within 1 AU, *Solar Phys.* **129**, 387
- 7.41 Steinolfson, R.S. and A.J. Hundhausen (1990): MHD intermediate shocks in a coronal mass ejection, *J. Geophys. Res.* **95**, 6389
- 7.42 Stone, R.G. and B.T. Tsurutani (eds.) (1985): *Collisionless shocks in the heliosphere: a tutorial review*, Geophys. Monogr. 34, American Geophysical Union, Washington, DC
- 7.43 Toptygin, I.N. (1983): *Cosmic rays in interplanetary magnetic fields*, Reidel, Dordrecht
- 7.44 Tsurutani, B.T. and R.G. Stone (eds.) (1985): *Collisionless shocks in the heliosphere: reviews of current research*, Geophys. Monogr. 35, American Geophysical Union, Washington, DC
- 7.45 Völk, H.J. (1987): Particle acceleration in astrophysical shock waves, *Proc. 20th Internat. Cosmic Ray Conf.* **7**, 157

Chapter 8

- 8.1 Alfvén, H. (1981): *Cosmic plasmas*, Reidel, Dordrecht
- 8.2 Antonucci, E. and B.V. Somov (eds.) (1991): *Solar corona and solar wind*, Adv. Space. Res. 11, vol. 1, Pergamon Press, Oxford
- 8.3 Babcock, H. (1961): *Astrophys. J.* **133**, 572
- 8.4 Bahcall, J.N. (1979): Solar neutrinos: theory versus observation, *Space Sci. Rev.* **24**, 227
- 8.5 Balogh, A., A.J. Smith, B.T. Tsurutani, D.J. Southwood, R.J. Forsyth, and T.S. Horbury (1995): The heliospheric magnetic field over the south polar region of the sun, *Science* **268**, 1007
- 8.6 Barnes, A. (1992): Acceleration of the solar wind, *Rev. Geophys.* **30**, 43
- 8.7 Barnes, A., R.E. Hardtke, and J.H. Bredekamp (1971): On the energy transport in stellar winds, *Astrophys. J. Lett.* **166**, L53
- 8.8 Bath, M. (1974): *Spectral analysis in geophysics*, Elsevier, Amsterdam
- 8.9 Bendat, J.S. and A.G. Piersol (1971): *Random data: analysis and measurement procedures*, Wiley-Interscience, New York
- 8.10 Biermann, L. (1951): Kometenschweife und solare Korpuskularstrahlung, *Z. Astrophys.* **29**, 274
- 8.11 Birkeland, K. (1908): *The Norwegian aurora polaris expedition 1902-1903, vol. 1: On the cause of magnetic storms and the origin of terrestrial magnetism*, H. Ashebourg, Christiania
- 8.12 Bothmer, V., and R. Schwenn (1995): Eruptive prominences as sources of magnetic clouds in the solar wind, *Space Sci. Rev.* **70**, 215

- 8.13 Bougeret, J.-L. (1985): Observations of shock formation and evolution in the solar atmosphere, in *Collisionless shocks in the heliosphere: reviews of current research* (eds. B.T. Tsurutani and R.G. Stone), Geophys. Monogr. 35, American Geophysical Union, Washington, DC
- 8.14 Brückner, G.E., and 14 others (1995): The large angle spectroscopic coronograph (LASECO), in [8.33], p. 357
- 8.15 Bruno, R., B. Bavassano, and U. Villante (1985): Evidence for long-period Alfvén waves in the inner solar system, *J. Geophys. Res.* **90**, 4373
- 8.16 Burkepile, J. and O.C. St. Cyr (1993): *A revised and expanded catalogue of coronal mass ejections observed by the Solar Maximum Mission coronograph*, High Altitude Observatory, NCAR Technical Note, NCAR/TN-369+STR, Boulder, CO
- 8.17 Burlaga, L.F. (1991): Magnetic clouds, in [8.78], p. 1
- 8.18 Burlaga, L.F. (1995): *Interplanetary magnetohydrodynamics*, Oxford University Press, Oxford
- 8.19 Cane, H.V. (1988): The large-scale structure of flare-associated interplanetary shocks, *J. Geophys. Res.* **93**, 1
- 8.20 Chapman, S. (1919): An outline of a theory of magnetic storms, *Proc. R. Soc. Ser. A* **95**, 61
- 8.21 Chapman, S. (1957): Note on the solar corona and the terrestrial ionosphere, *Smithsonian Centr. Astrophys.* **2**, 1
- 8.22 Chapman, S. (1959): Interplanetary space and the earth's outermost atmosphere, *Proc. R. Soc. Ser. A* **253**, 462
- 8.23 Chapman, S. and V.c.A. Ferraro (1931): A new theory of magnetic storms, *Terr. Magn.* **36**, 77
- 8.24 Cox,A., W.C. Livingston, and M.S. Matthews (1991): *Solar interior and atmosphere*, University of Arizona Press, Tucson, AZ
- 8.25 Crooker, N., J.A. Joselyn, and J.A. Feynman (1997): *Coronal mass ejections*, Geophys. Monogr. 99, American Geophysical Union, Washington, DC
- 8.26 Culhane, J.L. and C. Jordan (eds.) (1991): *The physics of solar flares*, Proc. of a Royal Society Discussion Meeting, The Royal Society, London
- 8.27 Deinzer, W. (1991): Die Sonne, in *Plasmaphysik im Sonnensystem*, (eds. K.-H. Glassmeier and M. Scholer), BI, Mannheim
- 8.28 Denskat, K.U., H.J. Beinroth, and F.M. Neubauer (1983): Interplanetary magnetic field power spectra with frequencies from $2.4 \cdot 10^{-5}$ Hz to 470 Hz from Helios-observations during solar minimum conditions, *J. Geophys. Res.* **87**, 2215
- 8.29 Dryer, M. (1994): Interplanetary studies: propagation of disturbances between the sun and the magnetosphere, *Space Sci. Rev.* **67**, 363
- 8.30 Eddington, A.S. (1926): *The internal constitution of the stars*, Cambridge University Press, Cambridge
- 8.31 Eddy, J.A. (1976): The Sun since the Bronze Age, in *Physics of solar planetary environments, vol. II* (ed. D.J. Williams), p. 958, American Geophysical Union, Washington, DC
- 8.32 Feynman, J. and A.J. Hundhausen (1994): Coronal mass ejections and major solar flares: the great active center of March 1989, *J. Geophys. Res.* **99**, 8451
- 8.33 Fleck, B., V. Domingo, and A.I. Poland (eds.) (1995): *The SOHO mission*, Kluwer, Dordrecht
- 8.34 Gazis, P.R. (1996): Solar cycle variation in the heliosphere, *Rev. Geophys.* **34**, 379
- 8.35 Geiss, J., G. Gloeckler, R. von Steiger, H. Balsiger, L.A. Fisk, A.B. Galvin, F.M. Ipavich, S. Livi, J.F. McKenzie, K.W. Ogilvie, and B. Wilken (1995): The

- southern high-speed stream: results from the SWICS instrument on Ulysses, *Science* **268**, 1033
- 8.36 Gosling, J.T. (1992): In situ observations of coronal mass ejections in interplanetary space, in [8.89], p. 258
- 8.37 Gosling, J. (1993): The solar flare myth, *J. Geophys. Res.* **98**, 18 937
- 8.38 Gringauz, K.I., V.V. Bezrukikh, V.D. Ozerov, and R.E. Rybchinskii (1960): A study of the interplanetary gas, high-energy electrons, and corpuscular radiation from the sun by means of a three-electrode trap for charged particles on the Soviet cosmic rocket, *Sov. Phys. Dokl.*, Engl. transl., **5**, 361
- 8.39 Harrison, R.A. (1991): Coronal transients and their relation to solar flares, *Adv. Space Res.* **11** (1), 25
- 8.40 Hoeksema, J.T. (1991): Large-scale structure of the heliospheric magnetic field, *Adv. Space Res.* **11**, 15
- 8.41 Hoffmeister, C. (1944): Physikalische Untersuchungen auf Kometen II, Die Bewegung der Schweifmaterie und die Repulsivkraft der Sonne beim Kometen, *Z. Astrophys.* **23**, 1
- 8.42 Howard, R.A., N.R. Sheeley, Jr., M.J. Koomen, and D.J. Michels (1985): Coronal mass ejections 1979–1981, *J. Geophys. Res.* **90**, 8173
- 8.43 Hudson, H.S. and J. Ryan (1995): High energy particles in solar flares, *Annu. Rev. Astron. Astrophys.* **33**, 239
- 8.44 Hundhausen, A.J. (1988): The origin and propagation of coronal mass ejections, in *Proc. 6th Int. Solar Wind Conf.*, vol. I (eds. V.J. Pizzo, T.E. Holzer, and D.G. Sime), NCAR, Boulder, CO, p. 181
- 8.45 Hundhausen, A.J., T.E. Holzer, and B.C. Low (1987): Do slow shocks precede some coronal mass ejections?, *J. Geophys. Res.* **92**, 11173
- 8.46 Kahler, S.W. (1992): Solar flares and coronal mass ejections, *Annu. Rev. Astron. Astrophys.* **30**, 113
- 8.47 Kahler, S.W. and A.J. Hundhausen (1992): The magnetic topology of solar coronal structures following mass ejections, *J. Geophys. Res.* **97**, 1619
- 8.48 Kane, S.R. (1974): Impulsive (flash) phase of solar flares: hard X-ray, microwave, EUV, and optical observations, in *Coronal disturbances* (ed. G. Newkirk), Proc. 57th IAU Symp., Reidel, Dordrecht, p. 105
- 8.49 Kayser, S.E., J.L. Bougeret, J. Fainberg, and R.G. Stone (1987): Comparison of interplanetary type-III-storm footpoints with solar features, *Sol. Phys.* **109**, 107
- 8.50 Kippenhahn, R. and A. Schläuter (1957): *Z. Astrophys.* **43**, 36
- 8.51 Koutchmy, S., J.B. Zirker, R.S. Steinolfson, and J.D. Zhugzda (1991): Coronal avticity, in [8.24], p. 1044
- 8.52 Kundu, M.R. and L. Vlahos (1982): Solar microwave bursts – a review, *Space Sci. Rev.* **32**, 405
- 8.53 Lang, K. (1995): *Sun, earth and sky*, Springer, Berlin
- 8.54 Leer, E. and Ø. Sandbæk (1991): Solar wind models, *Adv. Space Res.* **11** (1), 197
- 8.55 Marsch, E. (1991): MHD turbulence in the solar wind, in [8.78], p. 159
- 8.56 Marsch, E. (1994): Theoretical models for the solar wind, *Adv. Space Phys.* **14** (4), 103
- 8.57 Marsden, R. (ed.) (1986): *The sun and the heliosphere in three dimensions*, Reidel, Dordrecht
- 8.58 McComas, D.J., P. Riley, J.T. Gosling, A. Balogh, and R. Forsyth: Ulysses' rapid crossing of the polar coronal hole boundary, *J. Geophys. Res.* **103**, 1955
- 8.59 McComas, D.J., J.L. Phillips, A.J. Hundhausen, and J.T. Burkepile (1991): Observation of disconnection of pen coronal magnetic structures, *Geophys. Res. Lett.* **18**, 73

- 8.60 McLean, D.J. and N.R. Labrum (eds.) (1985): *Solar radiophysics*, Cambridge University Press, Cambridge
- 8.61 Montgomery, D.C. (1983): Theory of hydrodynamic turbulence, in *Solar Wind 5* (ed. M. Neugebauer), NASA-E-CP-2280, p. 107
- 8.62 Montmerle, T. and M. Spiro (eds) (1986): *Neutrinos and the present day universe*, Commissariat à l'Energy Atomique, Saclay, France
- 8.63 Morrison, D.G.O. (1992): Updated review of solar models and solar neutrino experiments, *Proc. Europ. Cosmic Ray Symp.*, Geneva
- 8.64 Munro, R.H., J.T. Gosling, E. Hildner, R.M. MacQueen, A.I. Poland, and C.L. Ross (1979): The association of coronal mass ejection transients with other forms of solar activity, *Solar Phys.* **61**, 201
- 8.65 Pallavicini, R., S. Serio, and G.S. Vaiana (1977): A survey of soft X-ray limb flare images: the relation between their structure and other physical parameters, *Astrophys. J.* **216**, 108
- 8.66 Parker, E.N. (1958): Dynamics of the interplanetary magnetic field, *Astrophys. J.* **128**, 664
- 8.67 Parker, E.N. (1963): *Interplanetary dynamical processes*, Wiley, New York
- 8.68 Phillips, J.L., S.J. Bame, W.C. Feldman, B.E. Goldstein, J.T. Gosling, C.M. Hammond, D.J. McComas, M. Neugebauer, E.E. Scime, and S.T. Suess (1995): Ulysses solar wind plasma observations at high southerly latitudes, *Science* **268**, 1031
- 8.69 Pizzo, V.J. (1991): The evolution of corotating stream fronts near the ecliptic plane in the inner solar system, 2. Three-dimensional tilted dipole fronts, *J. Geophys. Res.* **96**, 5405
- 8.70 Priest, E.R. and A.W. Hood (eds.) (1991): *Advances in solar system magnetohydrodynamics*, Cambridge University Press, Cambridge
- 8.71 Raadu, A.M. and A. Kuperus (1973): Thermal instability of coronal neutral sheets and the formation of quiescent prominences, *Astrophys. J.* **344**, 1010
- 8.72 Raadu, A.M. and A. Kuperus (1974): The support of prominences formed in neutral sheets, *Astron. Astrophys.* **31**, 189
- 8.73 Ramaty, R. and R.E. Lingenfelter (1983): Gamma ray line astronomy, *Space Sci. Rev.* **36**, 305
- 8.74 Reames, D.V. (1990): Energetic particles from impulsive flares, *Astrophys. J. Suppl.* **73**, 253
- 8.75 Schatten, K.H., J.M. Wilcox, and N.F. Ness (1969): A model of interplanetary and coronal magnetic fields, *Solar Phys.* **6**, 442
- 8.76 Schwenn, R. (1990): Large-scale structure of the interplanetary medium, in [8.77], p. 99
- 8.77 Schwenn, R. and E. Marsch (eds.) (1990): *Physics of the inner heliosphere, vol. I*, Springer, Berlin
- 8.78 Schwenn, R. and E. Marsch (eds.) (1991): *Physics of the inner heliosphere, vol. II*, Springer, Berlin
- 8.79 Sheeley, N.R., Jr., R.T. Stewart, R.D. Robinson, R.A. Howard, M.J. Koomen, and D.J. Michels (1984): Association between coronal mass ejections and metric type II bursts, *Astrophys. J.* **279**, 839
- 8.80 Sheeley, N.R., Jr., R.A. Howard, J. Koomen, D.J. Michels, R. Schwenn, K.H. Mühlhäuser, and H. Rosenbauer (1985): Coronal mass ejections and interplanetary shocks, *J. Geophys. Res.* **90**, 163
- 8.81 Smith, E.J., R.G. Marsden, and D.E. Page (1995): Ulysses above the sun's south pole: an introduction, *Science* **268**, 1005
- 8.82 Smith, E.J., B.T. Tsurutani, and R.L. Rosenberg (1978): Observations of the interplanetary sector structure up to heliographic latitudes of 16°: Pioneer 11, *J. Geophys. Res.* **83**, 717

- 8.83 Sonett, C.P., M.S. Giampapa, and M.S. Matthews (eds.) (1991): *The sun in time*, University of Arizona Press, Tucson, AZ
- 8.84 Stix, M. (1987): in *Solar and stellar physics*, (eds. E.H. Schröter and M. Schüssler), Lect. Notes Phys., vol. 292, p. 15, Springer, Berlin
- 8.85 Stix, M. (1989): *The sun*, Springer, Berlin
- 8.86 Sturrock, P.A. (ed.) (1980): *Solar flares*, Reidel, Dordrecht
- 8.87 Suess, S.T. (1990): The heliopause, *Rev. Geophys.* **28**, 97
- 8.88 Suess, S.T. (1993): The relationship between coronal and interplanetary magnetic fields, *Adv. Space Res.* **13**, (9)31
- 8.89 Svestka, Z., B.V. Jackson, and M.E. Machado (eds.) (1992): *Eruptive solar flare*, Springer, Berlin
- 8.90 Tandberg Hanssen, E. and A.G. Emslie (1988): *The physics of solar flares*, Cambridge University Press, Cambridge
- 8.91 Tu, C.-Y. and E. Marsch (1995): MHD structures, waves and turbulence in the solar wind: observations and theories, *Space Sci. Rev.* **73**, 1
- 8.92 Vilmer, N. (1987): Hard X-ray emission processes in solar flares, *Sol. Phys.* **111**, 207
- 8.93 Webb, D.F. (1991): The solar cycle variation of rates of CMEs and related activity, *Adv. Space Res.* **11**, 1(37)
- 8.94 Zhao, X. and J.T. Hoeksema (1993): Unique determination of model coronal fields using photospheric observations, *Solar Phys.* **143**, 41
- 8.95 Zirin, H. (1988): *Astrophysics of the sun*, Cambridge University Press, Cambridge

Chapter 9

- 9.1 Barnes, C.W., and J.A. Simpson (1976): Evidence for interplanetary acceleration of nucleons in corotating interaction regions, *Astrophys. J.* **207**, 977
- 9.2 Bieber, J.W., W.H. Matthaeus, C.W. Smith, W. Wanner, M.-B. Kallenrode, and G. Wibberenz (1994): Proton and electron mean free paths: the Palmer consensus range revisited, *Astrophys. J.* **420**, 294
- 9.3 Bingham, R. (1993): Space plasma physics: microprocesses, in *Plasma physics* (ed. R. Dendy), Cambridge University Press, Cambridge, p. 233
- 9.4 Cane, H.V. (1995): The structure and evolution of interplanetary shocks and the relevance for particle acceleration, *Nucl. Phys. B (Proc. Suppl.)* **39A**, 35
- 9.5 Cane, H.V., R.E. McGuire, and T.T. von Rosenvinge (1986): Two classes of solar energetic particle events associated with impulsive and long-duration soft X-ray flares, *Astrophys. J.* **301**, 448
- 9.6 Cane, H.V., D.V. Reames, and T.T. von Rosenvinge (1988): The role of interplanetary shocks in the longitude distribution of energetic particles, *J. Geophys. Res.* **93**, 9555
- 9.7 Cane, H.V., I.G. Richardsen, and T.T. von Rosenvinge (1993): Cosmic ray decreases and particle acceleration in 1978 - 1982 and the associated solar wind structures, *J. Geophys. Res.* **98**, 13295
- 9.8 Chenette, D.L., T.F. Conlon, K.R. Pyle, and J.A. Simpson (1977): Observations of jovian electrons at 1 AU throughout the 13 month jovian synodic year, *Astrophys. J.* **215**, L95
- 9.9 Conlon, T.F. (1978): The interplanetary modulation and transport of jovian electrons, *J. Geophys. Res.* **83**, 541

- 9.10 Cornwall, J.M. (1986): Magnetospheric ion acceleration processes, in *Ion acceleration in the magnetosphere and ionosphere* (ed. T. Chang), Geophys. Monogr. 38, American Geophys. Union, Washington, DC
- 9.11 Dröge, W. (1993): Transport of solar energetic particles, *Astrophys. J. Suppl.* **90**, 567
- 9.12 Droege, W., and 11 others (1986): Effects of corotating interaction regions and Ulysses high energy particles, *Solar Wind Eight*, (eds. D. Winterhalter et al.), p. 515, AIP Press, Woodbury, NY, p. 515
- 9.13 Fisk, L.A. (1986): The anomalous component, its variation with latitude and related aspects of modulation, in *The sun and the heliosphere in three dimensions* (ed. R.G. Marsden), Reidel, Dordrecht, p. 401
- 9.14 Fisk, L.A. and M.A. Lee (1980): Shock acceleration of energetic particles in corotating interaction regions in the solar wind, *Astrophys. J.* **237**, 260
- 9.15 Flückiger, E. (1991): Solar and terrestrial modulation, *Proc. 22nd Int. Cosmic Ray Conf.* **5**, 273
- 9.16 Forbush, S.E. (1946): Three unusual cosmic ray increases possibly due to charged particles from the Sun, *Phys. Rev.* **70**, 771
- 9.17 Gloeckler, G. (1984): Characteristics of solar and heliospheric ion populations observed near Earth, *Adv. Space Res.* **4** (2-3), 127
- 9.18 Gosling, J., J.R. Asbridge, S.J. Bame, W.C. Feldman, R.D. Zwickl, G. Paschmann, N. Sckopke, and R.J. Hynds (1981): *J. Geophys. Res.* **86**, 547
- 9.19 Heber, B., A. Raviart, C. Paizis, M.S. Potgieter, W. Dröge, R. Ducros, P. Ferrando, L.J. Haasbroek, H. Kunow, R. Müller-Mellin, C. Rastoin, K. Röhres, H. Sierks, and G. Wibberenz (1995): Nucleon modulation from the Ulysses rapid latitude scan: COSPIN/KET results, *Proc. 24th Int. Cosmic Ray Conf.* **4**, 764
- 9.20 Heras, A.M., B. Sanahuja, Z.K. Smith, T. Detman, and M. Dryer (1992): The influence of the large-scale interplanetary shock structure on a low-energy particle event, *Astrophys. J.* **391**, 351
- 9.21 Hudson, H.S. and J. Ryan (1995): High energy particles in solar flares, *Annu. Rev. Astron. Astrophys.* **33**, 239
- 9.22 Jokipii, J.R. (1990): The anomalous component of cosmic rays, in *Physics of the outer heliosphere* (eds. E. Grzedzielsky and E. Page), COSPAR Colloq. Series Vol. 1, Pergamon Press, Oxford, p. 169
- 9.23 Kallenrode, M.-B. (1996): A statistical survey of 5 MeV proton events at transient interplanetary shocks, *J. Geophys. Res.* **101**, 24393
- 9.24 Kallenrode, M.-B. (1995): Acceleration and propagation of energetic charged particles in the inner heliosphere, *Nuclear Phys. B (Proc. Suppl.)* **39A**, 45
- 9.25 Kallenrode, M.-B., E.W. Cliver, and G. Wibberenz (1992): Composition and azimuthal spread of solar energetic particles from impulsive and gradual flares, *Astrophys. J.* **391**, 370
- 9.26 Kallenrode, M.-B., G. Wibberenz, and S. Hucke (1992): Propagation conditions for relativistic electrons in the inner heliosphere, *Astrophys. J.* **394**, 351
- 9.27 Kallenrode, M.-B. and G. Wibberenz (1997): Propagation of particles injected from interplanetary shocks: a black-box model and its consequences for acceleration theory and data interpretation, *J. Geophys. Res.* **102**,
- 9.28 Kunow, H., G. Wibberenz, G. Green, R. Müller-Mellin, and M.-B. Kallenrode (1991): Energetic particles in the inner heliosphere, in *Physics of the inner heliosphere, vol. 2* (eds. E. Marsch and R. Schwenn), Springer, Berlin, p. 235
- 9.29 Kunow, H., and 11 others (1995): High energy cosmic ray results on Ulysses: 2. Effects of a recurrent high-speed stream from the southern polar coronal hole, *Space Sci. Rev.* **72**, 397

- 9.30 Lee, M.A. (1980): Coupled hydromagnetic wave excitation and ion acceleration upstream of the Earth's bow shock, *J. Geophys. Res.* **87**, 5063
- 9.31 Lee, M.A. (1983): The association of energetic particles and shocks in the heliosphere, *Rev. Geophys. Space Phys.* **21**, 324
- 9.32 Lee, M.A. (1995): Ulysses race to the pole: symposium summary, *Space Sci. Rev.* **72**, 485
- 9.33 Luttrell, A.H. and A.K. Richter (1987): A study of MHD fluctuations upstream and downstream of quasi-parallel interplanetary shocks, *J. Geophys. Res.* **92**, 2243
- 9.34 Mandzhavidze, N. and R. Ramaty (1993): Particle acceleration in solar flares, *Nucl. Phys. B (Proc. Suppl.)* **33A,B**, 141
- 9.35 Matthaeus, W.H., M.L. Goldstein, and D.A. Roberts (1990): Evidence for the presence of quasi-two-dimensional nearly incompressible fluctuations in the solar wind, *J. Geophys. Res.* **95**, 20673
- 9.36 McKibben, R.B. (1987): Galactic cosmic rays and anomalous components in the heliosphere, *Rev. Geophys.* **25** (3), 711
- 9.37 Meyer, P., E.N. Parker, and J.A. Simpson (1956): Solar cosmic rays of February 1956 and their propagation through interplanetary space, *Phys. Rev.* **104**, 768
- 9.38 Meyer, P., R. Ramaty, and W.R. Webber (1974): Cosmic rays – astronomy with energetic particles, *Phys. Today* **27** (10),
- 9.39 Mihalov, J.D. (1987): Heliospheric shocks (excluding planetary bow shocks), *Rev. Geophys.* **25**, 697
- 9.40 Miller, J.A., P.J. Cargill, A.G. Emslie, G.D. Holman, B.R. Dennis, T.N. LaRosa, R.M. Winglee, S.G. Benka, and S. Tsuneta (1997): Critical issues for understanding particle acceleration in impulsive solar flares, *J. Geophys. Res.* **102**, 14 631
- 9.41 Orlowski, D.S., C.T. Russel, and R.P. Lepping (1992): Wave phenomena in the upstream region of Saturn, *J. Geophys. Res.* **97**, 19187
- 9.42 Palmer, I.D. (1982): Transport coefficients of low-energy cosmic rays in interplanetary space, *Rev. Geophys. Space Phys.* **20**, 335
- 9.43 Paschmann, G., N. Sckopke, I. Papamastorakis, J.R. Asbridge, S.J. Bame, and J.T. Gosling (1981): Characteristics of reflected and diffuse ions upstream of the Earth's bow shock, *J. Geophys. Res.* **86**, 4355
- 9.44 Pick, M. and M.E. Machado (eds.) (1993): *Fundamental problems in solar activity*, Adv. Space Res. 13, no. 9, Pergamon Press, Oxford
- 9.45 Potgieter, M. (1993): Modulation of cosmic rays in the heliosphere, *Proc. 23rd Int. Cosmic Ray Conf.*, Invitet Paper, p. 213
- 9.46 Ramaty, R., N. Mandzhavidze, B. Kozlovsky, and J.G. Skibo (1993): Acceleration in solar flares: interacting particles versus interplanetary particles, *Adv. Space Res.* **13** (9), 275
- 9.47 Reames, D.V. (1990): Acceleration of energetic particles by shock waves from large solar flares, *Astrophys. J. Lett.* **358**, L63
- 9.48 Reames, D.V., H.V. Cane, and T.T. von Rosenvinge (1990): Energetic particle abundances in solar electron events, *Astrophys. J.* **357**, 259
- 9.49 Reinhard, R., P. van Nes, T.R. Sanderson, K.-P. Wenzel, E.J. Smith, and B.T. Tsurutani (1983): A statistical study of interplanetary shock associated proton intensity increases, *Proc. 18th Int. Cosmic Ray Conf.* **3**, 160
- 9.50 Roelof, E.C. (1969): Propagation of solar cosmic rays in the interplanetary magnetic field, in *Lectures in high energy astrophysics* (eds. H. Ögelmann and J.R. Wayland), NASA SP-a99, p. 111
- 9.51 Russell, C.T. (ed.) (1995): *Physics of collisionless shocks*, Adv. Space Res. 15, no. 8/9, Pergamon Press, Oxford

- 9.52 Russell, C.T., and M.M. Hoppe (1983): Upstream waves and particles, *Space Sci. Rev.* **34**, 155
- 9.53 Russell, C.T., R.P. Lepping, and C.W. Smith (1990): Upstream waves at Uranus, *J. Geophys. Res.* **95**, 2273
- 9.54 Sanderson, T.R. (1984): ISEE-3 observations of energetic protons associated with interplanetary shocks, *Adv. Space Res.* **4**, (2-3)305
- 9.55 Scholer, M. (1987): Observational overview of energetic particle populations associated with interplanetary shocks, in *Solar Wind VI* (eds. V. Pizzo, T.E. Holzer, and D.G. Sime), NCAR/TN306+Proc., Nat. Center for Atm. Research, Boulder, CO
- 9.56 Scholer, M. and G.E. Morfill (1977): Modulation of energetic solar particle fluxes by interplanetary shock waves, in *Contributed papers to the study of traveling interplanetary phenomena* (eds. M.A. Shea, D.F. Smart, and S.T. Wu), Air Force Geophys. Lab., Hanscom AFB, Spec. Rep. 209, p. 221
- 9.57 Simpson, J.A. (1989): Evolution of our knowledge of the heliosphere, *Adv. Space Res.* **9** (4), 5
- 9.58 Simpson, J.A., and 19 others (1995): Cosmic ray and solar particle investigations over the south polar region of the sun, *Science* **268**, 1019
- 9.59 Smith, E.J. and J.H. Wolfe (1976): Observation of interaction regions and corotating shocks between one and five AU: Pioneers 10 and 11, *Geophys. Res. Lett.* **3**, 137
- 9.60 Tsurutani, B.T. and R.P. Lin (1985): Acceleration of >47 keV ions and >2 keV electrons by interplanetary shocks at 1 AU, *J. Geophys. Res.* **90**, 1
- 9.61 Völk, H.J. (1975): Cosmic ray propagation in interplanetary space, *Rev. Geophys. Space Phys.* **13**, 547
- 9.62 Wanner, W. and G. Wibberenz (1993): A study of the propagation of solar energetic protons in the inner heliosphere, *J. Geophys. Res.* **98**, 3513
- 9.63 Webber, W.R. (1987): Cosmic rays in the heliosphere, in *Essays in space science* (eds. R. Ramaty, T.L. Cline, and J.F. Ormes), NASA Goddard Space Flight Center, Greenbelt, MD, p. 125
- 9.64 Wenzel, K.-P., R. Reinhard, T.R. Sanderson, and E.T. Sarris (1985): Characteristics of energetic particle events associated with interplanetary shocks, *J. Geophys. Res.* **90**, 12
- 9.65 Zöllisch, F., G. Wibberenz, H. Kunow, and G. Green (1981): Corotating events in the energy range 4–13 MeV as observed on board Helios 1 and 2 in 1975 and 1976, *Adv. Space Res.* **1**, 89

Chapter 10

- 10.1 Anderson, K.A. and R.B. Lin (1969): Observation of interplanetary field lines in the magnetotail, *J. Geophys. Res.* **74**, 3953
- 10.2 Brekke, A. (1997): *Physics of the upper polar atmosphere*, Wiley, Chichester
- 10.3 Bryant, D. (1993): Space plasma physics I: basic processes in the solar system, in *Plasma physics* (ed. R. Dendy), Cambridge University Press, Cambridge, p. 209
- 10.4 Carpenter, D.L. (1966): Whistler studies of the plasmapause in the magnetosphere, *J. Geophys. Res.* **71**, 693
- 10.5 Chappell, R., K.K. Harris, and G.W. Sharp (1970): A study of the influence of magnetic activity on the location of the plasmapause as measured by OGO 5, *J. Geophys. Res.* **75**, 50

- 10.6 Chappell, C.R. (1972): Recent satellite measurements of the morphology and dynamics of the plasmasphere, *Rev. Geophys. Space Phys.* **10**, 951
- 10.7 Chang, T. (ed.) (1986): *Ion acceleration in the magnetosphere and ionosphere*, Geophys. Monogr. 38, American Geophys. Union, Washington, DC
- 10.8 Dungey, J.W. (1961): Interplanetary magnetic field and the auroral zones, *Phys. Rev. Lett.* **6**, 47
- 10.9 Dungey, J.W. (1967): The theory of the quiet magnetosphere, in *Solar-terrestrial physics* (ed. J.W. King and W.S. Newman), Academic Press, London, p. 91
- 10.10 Frank, L.A., et al. (1964): A study of charged particles in the Earth's outer radiation belts with Explorer 14, *J. Geophys. Res.* **69**, 2171
- 10.11 Ganguli, S.B. (1996): The polar wind, *Rev. Geophys.* **34**, 311
- 10.12 Gubbins, D. (1994): Geomagnetic polarity reversals: a connection with secular variation and core-mantle interaction, *Rev. Geophys.* **32**, 61
- 10.13 Hargreaves, J.K. (1992): *The solar-terrestrial environment*, Cambridge Atmospheric and Space Science Series 5, Cambridge University Press, Cambridge
- 10.14 Holzworth, R.H. and F.S. Mozer (1979): *J. Geophys. Res.* **81**, 2445
- 10.15 Hughes, W.J., (1995): The magnetopause, magnetotail, and magnetic reconnection, in [10.25], p. 227
- 10.16 Hultqvist, B. and M. Øieroset (eds.) (1997): *Transport across the boundaries of the magnetosphere*, Kluwer, Dordrecht
- 10.17 Iijima, T. and T.A. Potemra (1976): Field-aligned currents in the dayside cusps observed by Triad, *J. Geophys. Res.* **81**, 5971
- 10.18 Iijima, T. and T.A. Potemra (1978): Large-scale characteristics of field-aligned currents associated with substorms, *J. Geophys. Res.* **83**, 599
- 10.19 Inglis, D.R. (1955): Theories of the earth's magnetism, *Rev. Mod. Phys.* **27**, 212
- 10.20 International Association of Geomagnetism and Aeronomy (1985): International geomagnetic reference field revision 1985, *EOS Trans. AGU* **67**, 523
- 10.21 Jacobs (1984): *Reversals of the earth's magnetic field*, Hilger, Bristol
- 10.22 Johnson, C.Y. (1969): Ion and neutral composition of the ionosphere, *Ann. IQSY* **5**, MIT Press, Cambridge, MA
- 10.23 Jordan, C.E. (1994): Empirical models of the magnetospheric magnetic field, *Rev. Geophys.* **32**, 139
- 10.24 Kertz, W. (1971): *Einführung in die Geophysik*, BI, Mannheim
- 10.25 Kivelson, M.G. and C.T. Russell (eds.) (1995): *Introduction to space physics*, Cambridge University Press, Cambridge
- 10.26 Kelley, M.C. (1989) *The Earth's ionosphere*, Academic Press, San Diego
- 10.27 King, J.W. and W.S. Newman (eds.) (1967): *Solar-terrestrial physics*, Academic Press, London
- 10.28 Lemaire, J.F., D. Heynderickx, and D.N. Baker (eds.) (1997): *Radiation belts - models and standards*, Geophys. Monogr. 97, American Geophys. Union, Washington, DC
- 10.29 McCracken, K.G., et al. (1962): The cosmic ray flare effect: 3. Deductions regarding the interplanetary magnetic field, *J. Geophys. Res.* **67**, 447
- 10.30 McPherron, R.L. (1995): Magnetospheric dynamics, in [10.25]
- 10.31 Merrill, R.T. and M.W. McElhinny (1983): *The earth's magnetic field*, Academic Press, London
- 10.32 Merrill R.T., McElhinny, M.W., and P.L. McFadden (1996): *The magnetic field of the Earth: paleomagnetism, the core, and the deep mantle*, Academic Press, San Diego

- 10.33 Möbius, E. (1990): Energetic particles in the environment of the Earth's magnetosphere, *J. Atmos. Terr. Phys.* **52**, 1169
- 10.34 Moore, T.E. and D.C. Delcourt (1995): The geopause, *Rev. Geophys.* **33**, 175
- 10.35 Moore, T.E. and J.H. Waite, Jr. (eds.) (1988): *Modeling magnetospheric plasma*, Geophys. Monogr. 44, American Geophys. Union, Washington, DC
- 10.36 Ness, N.F., C.S. Scearce, and J.B. Seek (1964): Initial results of the IMP 1 magnetic field experiment, *J. Geophys. Res.* **69**, 3531
- 10.37 Olson, W.P. (ed.) (1979): *Quantitative modeling of magnetospheric processes*, American Geophys. Union, Washington, DC
- 10.38 Parks, G.K. (1991): *Physics of space plasmas*, Addison-Wesley, Redwood City, CA
- 10.39 Paulikas, G.A. and J.B. Blake (1979): Effects of the solar wind on magnetospheric dynamics: energetic electrons in the synchronous orbit, in [10.37], p. 180
- 10.40 Pilipp, W.G. and G. Morfill (1978): The formation of the plasma sheet resulting from plasma mantle dynamics, *J. Geophys. Res.* **83**, 5670
- 10.41 Reid, G.C. (1986): Solar energetic particles and their effect on the terrestrial environment, in *Physics of the Sun, vol. III* (eds. P.A. Sturrock, T.E. Holzer, D.M. Mihalas, and R.K. Ulrich), Reidel, Dordrecht
- 10.42 Roederer, J.G. (1970): *Dynamics of geomagnetically trapped radiation*, Springer, Berlin
- 10.43 Russell, C.T. and R.C. Elphic (1979): ISEE observations of flux transfer events at the dayside magnetopause, *Geophys. Res. Lett.* **6**, 33
- 10.44 Schulz, M. and L.J. Lanzerotti (1974): *Particle diffusion in radiation belts*, Springer, Berlin
- 10.45 Sheldon, R.B. and T.E. Eastman (1997): Particle transport in the magnetosphere: a new diffusion model, *Geophys. Res. Lett.* **24**, 811
- 10.46 Smith, M.F. and M. Lockwood (1996): Earth's magnetospheric cusps, *Rev. Geophys.* **34**, 233
- 10.47 Spreiter, J.R., A.L. Summers, and A.Y. Alksne (1966): Hydromagnetic flow around the magnetosphere, *Planet. Space Sci.* **14**, 223
- 10.48 Stephenson, F.R. and J.A. Wolfendale (eds.) (1988): *Secular solar and geomagnetic variations in the last 10000 years*, Kluwer Academic Publishers, Dordrecht
- 10.49 Strohbach, K. (1991): *Unser Planet Erde*, Borntraeger, Berlin
- 10.50 van Allen, J.A. (1959): The geomagnetically trapped corpuscular radiation, *J. Geophys. Res.* **64**, 1683
- 10.51 van Allen, J.A. (1991): Why radiation belts exist, *EOS* **72** (34), 361
- 10.52 Walker, R.J. (1979): Quantitative modeling of planetary magnetospheric magnetic fields, in *Quantitative modeling of magnetospheric processes* (ed. W.P. Olson), American Geophys. Union, Washington, DC, p. 1
- 10.53 Walt, M. (1994): *Introduction to geomagnetically trapped radiation*, Cambridge University Press, Cambridge
- 10.54 Williams, D.J., E.C. Roelof, and D.G. Mitchell (1992): Global magnetospheric imaging, *Rev. Geophys.* **30**, 183
- 10.55 Willis, D.M. (1971): Structure of the magnetopause, *Rev. Geophys. Space Phys.* **9**, 953
- 10.56 Wollin, G., D.B. Ericson, and W.B.F. Ryan (1971): Variations in magnetic intensity and climatic changes, *Nature* **232**, 549
- 10.57 Wolf, R.A. (1995): Magnetospheric configuration, in [10.25], p. 288

Chapter 11

- 11.1 Akasofu, S.-I. (1968): *Polar and magnetospheric substorms*, Reidel, Dordrecht
- 11.2 Akasofu, S.-I. (1989): Substorms, *EOS* **70**, 529
- 11.3 Akasofu, S.-I. and S. Chapman (1972): *Solar-terrestrial physics*, Oxford University Press, Oxford
- 11.4 Block, L.P. and C.G. Fälthammar (1990): The role of magnetic field aligned electric fields in auroral acceleration, *J. Geophys. Res.* **95**, 5877
- 11.5 Bone, N. (1991): *The aurora*, Ellis Horwood, New York
- 11.6 Brekke, A. (1997): *Physics of the upper polar atmosphere*, Wiley, Chichester
- 11.7 Brekke, A. and A. Egeland (1983): *Northern lights*, Springer, Berlin
- 11.8 Burlaga, L.F. and R.P. Lepping (1977): The causes of recurrent geomagnetic storms, *Planet. Space Sci.* **25**, 1151
- 11.9 Chapman, S. and J. Bartels (1940): *Geomagnetism*, Oxford University Press, Oxford
- 11.10 Crooker, N.U. and G.L. Siscoe (1986): The effect of the solar wind on the terrestrial environment, in *Physics of the Sun, vol. III* (eds. P.A. Sturrock, T.E. Holzer, D.M. Mihalas, and R.K. Ulrich), Reidel, Dordrecht, p. 193
- 11.11 Eather, R.H. (1980): *Majestic lights – the aurora in science, history, and arts*, American Geophys. Union, Washington, DC
- 11.12 Elphinstone, R.E., J.S. Murphree, and L.L. Cogger (1996): What is a global auroral substorm?, *Rev. Geophys.* **34**, 169
- 11.13 Frank, L.A. and D.J. Craven (1988): Imaging results from Dynamics Explorer 1, *Rev. Geophys.* **26**, 249
- 11.14 Fritz, H. (1881): *Das Polarlicht*, Brockhaus, Leipzig
- 11.15 Goertz, C.K. (1979): Double layers and electrostatic shocks in space, *Rev. Geophys. Space Phys.* **17**, 418
- 11.16 Gonzalez, W.D. and B.T. Tsurutani (1992): Terrestrial response to eruptive solar flares: geomagnetic storms, in *Eruptive solar flares* (eds. Z. Svestka, B.V. Jackson, and M.E. Machado), Springer, Berlin, p. 277.
- 11.17 Hones, E.W. (1984): *Magnetic reconnection*, Geophys. Monogr. 30, American Geophys. Union, Washington, DC
- 11.18 Kivelson, M.G. and C.T. Russell (1995): *Introduction to space physics*, Cambridge University Press, Cambridge
- 11.19 Kozyra, J.U., A.F. Nagy, and D.W. Slater (1997): High-latitude energy source(s) for stable auroral red arcs, *Rev. Geophys.* **35**, 155
- 11.20 Lyons, L.R. (1992): Formation of auroral arcs via magnetosphere–ionosphere coupling, *Rev. Geophys.* **30**, 93
- 11.21 McPherron, R.L. (1995): Magnetospheric dynamics, in [11.18], p. 400
- 11.22 Menvielle, M. and A. Berthelier (1991): The k-derived planetary indices: description and availability, *Rev. Geophys.* **29**, 413
- 11.23 Obabyashi, T. (1967): The interaction of the solar wind with the geomagnetic field during disturbed conditions, in *Solar-terrestrial physics* (eds. J.W. King and W.S. Newman), Academic Press, London
- 11.24 Reid, G.C. (1986): Solar energetic particles and their effect on the terrestrial environment, in *Physics of the Sun, vol. III* (eds. P.A. Sturrock, T.E. Holzer, D.M. Mihalas, and R.K. Ulrich), Reidel, Dordrecht
- 11.25 Schlegel, C. (1991): Das Polarlicht, in *Plasmaphysik im Sonnensystem* (eds. K.-H. Glassmeier and M. Scholer), BI, Mannheim
- 11.26 Silverman, S.M. (1992): Secular variations of the aurora for the past 500 years, *Rev. Geophys.* **30**, 333
- 11.27 Spreiter, J.R. and S.S. Stahara (1980): A new predictive model for determining solar wind–terrestrial planet interactions, *J. Geophys. Res.* **85**, 6769

- 11.28 Størmer, C. (1955): *The polar aurora*, Oxford University Press, London
- 11.29 Tsurutani, B.T. and W.D. Gonzalez (1997): The interplanetary causes of magnetic storms: a review, in [11.30]
- 11.30 Tsurutani, B.T., W.D. Gonzalez, Y. Kamide, and J.K. Arballo (eds.) (1997): *Magnetic storms*, Geophys. Monogr. 98, American Geophys. Union, Washington, DC

Chapter 12

- 12.1 Acuña, M.H. and 19 others (1998): Magnetic field and plasma observations at Mars: initial results from the Mars Global Surveyor Mission, *Science* **279**, 1676
- 12.2 Bagenal, F. (1992): Giant planet magnetospheres, *Annu. Rev. Earth Planet Sci.* **20**, 289
- 12.3 Bergstrahl, J.T., E.D. Miner, and M.S. Matthews (1991): *Uranus*, University of Arizona Press, Tucson, AZ
- 12.4 Bryant, D.A. (1993): Space plasma physics: basic processes in the solar system, in *Plasma physics: an introductory course* (ed. R. Dendy), Oxford University Press, Oxford, p. 209
- 12.5 Dessler, J.A. (eds.) (1983): *Physics of the Jovian magnetosphere*, Cambridge University Press, Cambridge
- 12.6 Encrenaz, T. and J.J. Bibring, with M. Blanc (1990): *The solar system*, Springer, Berlin
- 12.7 Gehrels, T. and M.S. Matthews (1984): *Saturn*, University of Arizona Press, Tucson, AZ
- 12.8 Gurnett, D.A., R.R. Shaw, R.R. Abderson, and W.S. Kurth (1979): Whistlers observed by Voyager 1: detection of lightning on Jupiter, *Geophys. Res. Lett.* **6**, 511
- 12.9 Lang, K.R. and C.A. Whitney (1991): *Wanderers in space*, Cambridge University Press, Cambridge
- 12.10 Neubauer, F.M. (1991): Die Magnetosphären der Planeten, in *Plasmaphysik im Sonnensystem* (eds. K.-H. Glassmeier and M. Scholer), Spektrum, Heidelberg, p. 184
- 12.11 Ness, N.F., M.H. Acuna, L.F. Burlaga, J.E.P. Connerney, R.P. Lepping, and F.M. Neubauer (1989): Magnetic fields at Neptune, *Science* **246**, 1473
- 12.12 Russell, C.T. and M.M. Hoppe (1983): Upstream waves and particles, *Space Sci. Rev.* **34**, 155
- 12.13 Russell, C.T., R.P. Lepping, and C.W. Smith (1990): Upstream waves at Uranus, *J. Geophys. Res.* **95**, 2273
- 12.14 Russell, C.T. and R.J. Walker (1995): The magnetospheres of the outer planets, in *Introduction to space physics* (eds. M.G. Kivelson and C.T. Russell), Cambridge University Press, Cambridge, p. 503
- 12.15 Simpson, J.A. and B. McKibben (1976): Dynamics of the jovian magnetosphere and energetic particle radiation, in *Jupiter* (ed. T. Gehrels), University of Arizona Press, Tucson, AZ

Chapter 13

- 13.1 Bucha, V. (1980): Influence of solar activity on atmospheric circulation types, *Ann. Geophys.* **6**, (5)513
- 13.2 Calder, N. (1997): *The manic Sun – weather theories confounded*, Pilkington, Yelvertoft
- 13.3 Crutzen, P.J., I.S.A. Isaksen, and G.C. Reid (1975): Solar proton event – stratospheric source of nitric oxide, *Science* **189**, 457
- 13.4 Dameris, M. and S. Pawson (1994): Modeling the effects of solar variability in the middle atmosphere: a review, *Adv. Space Res.* **14** (9), 211
- 13.5 Eddy, J.A. (1976): The Maunder-minimum, *Science* **192**, 1189
- 13.6 Eddy, J.A. (1976): The Sun since the Bronze Age, in *Physics of solar planetary environments, vol. II* (ed. D.J. Williams), American Geophysical Union, Washington, DC
- 13.7 Gorney, D.J. (1990): Solar cycle effects on the near-earth space environment, *Rev. Geophys.* **28**, 315
- 13.8 Hargreaves, J.K. (1992): *The solar-terrestrial environment*, Cambridge University Press, Cambridge
- 13.9 Heath, D.F., A.J. Krueger, and P.J. Crutzen (1977): Solar proton event: influence in stratospheric ozone, *Science* **197**, 886
- 13.10 Herman, J.R. and R.A. Goldberg (1978): *Sun, weather, and climate*, NASA-SP-426-1978
- 13.11 Hoyt, D.V. and K.H. Schatten (1997): *The role of the sun in climate change*, Oxford University Press, Oxford
- 13.12 Joselyn, J.A. (1992): The impact of solar flares and magnetic storms on humans, *EOS Trans. Am. Geophys. Union*, **73/7**, 81
- 13.13 Joselyn, J.A. (1995): Geomagnetic activity forecasting: the state of the art, *Rev. Geophys.* **33**, 383
- 13.14 Labitzke, K. and H. van Loon (1989): Recent work correlating the 11-year solar cycle with the atmospheric elements grouped according to the phase of the quasi-biennial oscillation, *Space Sci. Rev.* **49**, 239
- 13.15 Lean, J. (1992): Variations in the sun's radiative output, *Rev. Geophys.* **29**, 505
- 13.16 McCormac, B.M. and T.A. Seliga (eds.) (1979): *Solar-terrestrial influences on weather and climate*, Reidel, Dordrecht
- 13.17 Reid, G.C. (1986): Solar energetic particles and their effect on the terrestrial environment, in *Physics of the Sun, Vol. III* (eds. P.A. Sturrock, T.E. Holzer, D.M. Mihalas, and R.K. Ulrich), Reidel, Dordrecht
- 13.18 Roberts, W.O. and H. Lansford (1979): *The climate mandate*, Freeman, San Francisco
- 13.19 Roble, R.G. (1977): *The upper atmosphere and magnetosphere*, Nat. Acad. Sciences, Washington, DC
- 13.20 Siskind, D.E. and D.W. Rusch (1992): Nitric oxide in the middle and upper thermosphere, *J. Geophys. Res.* **97**, 3209
- 13.21 Villoresi, G., Y.A. Kopytenko, N.G. Ptitsyna, M.I. Tyasto, E.A. Kopytenko, N. Iucci, and P.M. Voronov (1994): The influence of geomagnetic storms and man-made magnetic field disturbances on the incidence of myocardial infarction in St. Petersburg (Russia), *Physica Medica* **10**, (4), 107
- 13.22 Weeks, L.H., R.S. Cuikay, and J.R. Orbin (1972): Ozone measurements in the mesosphere during the solar proton event of 2 November 1969, *J. Atmosph. Sci.* **29**, 1138
- 13.23 Xanthakos, J. (ed.) (1973): *Solar activity and related interplanetary and terrestrial phenomena*, Springer, Berlin

Chapter 14

- 14.1 Dynamics Explorer (1981): *Space Sci. Instr.* **5** (4)
- 14.2 Galileo Instruments (1992): *Space Sci. Rev.* **60**, 79
- 14.3 Fleck, B.,V. Domingo, and A.I. Poland (1995): *The SOHO mission*, Kluwer, Dordrecht; reprinted from *Solar Phys.* **162**
- 14.4 Hunsucker, R.D. (1991): *Radio techniques for probing the terrestrial atmosphere*, Springer, Berlin
- 14.5 Lin, R.P. (1989): New techniques for charged particle measurements in interplanetary space, in *Solar system plasma physics*, Geophys. Monogr. 54, American Geophys. Union, Washington, DC, p. 333
- 14.6 McIntyre, S.A. (1991): Magnetic files sensor design, *Sensor Rev.* **11**, 7
- 14.7 Polar Instrumentation (1995): *Space Sci. Rev.* **71**
- 14.8 Primdahl, F. (1998): Scalar magnetometers for space applications, *IEEE Trans.*, in press
- 14.9 Russell, C.T. and D.J. Southwood (1982): *The IMS source book*, American Geophys. Union, Washington, DC
- 14.10 Tait, W.H. (1980): *Radiation detection*, Butterworth, London
- 14.11 Ulysses Instruments (1992): *Astron. Astrophys.* **92** (2)
- 14.12 Wind Instrumentation (1995): *Space Sci. Rev.* **71**

Index

- $\alpha\Omega$ dynamo 61
- α -effect 61
- absorption cross section 229
- acceleration
 - diffusive shock 134
 - first-order Fermi 28
 - shock drift 131
 - stochastic 117, 200
- action integral 23
- adiabatic compression 38
- adiabatic exponent 38
- adiabatic invariant 22
 - equation of state 39
 - first 17, 23, 25
 - flux 249
 - flux invariant 29
 - longitudinal 27
 - magnetic moment 23, 25
 - second 27, 28, 248
 - third 29, 249
- advection 33
- AE index 258
- Alfvén shock 128
- Alfvén speed 72, 74
- Alfvén wave 69, 114, 155, 163
 - compressive 74
 - dispersion relation 72
 - speed 72
- alfvénic fluctuations 164
- Ampere's law 11
- anisotropic diffusion 107
- anisotropy 195
- anomalous component 188
- anomalous cosmic rays 188
- Archimedean spiral 156
 - path length 157
- astronomical unit 3
- AU 3
- aurora 6, 224, 239
- auroral electrojet 256, 264
- auroral oval 236
- Ballerina model 159
- barometric height formula 49, 229
- bell curve 106
- betatron effect 248
- Birkeland current 239, 271
- blast wave 174
- blast wave shock 121
- Bohm–Gross equation 77
- Boltzmann equation 94
 - collision term 94, 113
 - collisionless 94
- Bouguer–Lambert–Beer law 229
- bow shock 79, 189, 204, 225, 227
- Brownian motion 99
- centrifugal force 36, 37
- cgs system 9
- Chapman layer 296
- Chapman profile 229
- Chapman–Ferraro current 222–224
- chromosphere 147
- cold plasma 31, 76
- collision frequency 100
- collision term 97, 113
- collisionless shock 120, 122
- collisionless viscosity 36
- collisions
 - large-angle 101
- compression
 - adiabatic 38
 - isothermal 38
- compression ratio 131, 136, 200
- conductivity 12, 43
 - field-aligned 231
 - Hall 231
 - Pederson 231
 - thermal 52
- connection longitude 186
- conservation of mass 38

- continuity
 - equation of 11, 38
 - electric charge 38
 - electromagnetic field 13
 - phase space 94
 - mass 38
- coplanarity normal 130
- coplanarity theorem 130
- Coriolis force 36, 37, 302
- corona 147
 - hydrodynamic model 152
 - hydrostatic model 151
- coronal hole 148, 150
- coronal mass ejection (CME) 6, 47, 151, 171, 176
 - and coronal shock 178
 - disconnection event 180
 - filament 179
 - flare 178, 182
 - magnetic cloud 183
 - magnetic field configurations 180
 - metric type II burst 178
 - reconnection 180
 - speed 178
- coronograph 176
- corotating interaction region (CIR) 160, 189, 208
 - forward shock 161
 - interface 160
 - reverse shock 161
- cosmic ray albedo neutron decay 246
- cosmogenic nuclide 217, 246, 305, 307
- Cowling's theorem 59, 60
- CRAND 246
- cross-section 99, 100, 246
 - Coulomb scattering 100
 - large-angle deflection 100
- current acceleration 44
- current sheet 149
- cut-off rigidity 250
- cyclotron frequency 15
- cyclotron resonance 85

- D_{st} index 260
- D_{st} index 256, 258
- dawn-to-dusk electric field 223, 235
- de Hoffmann-Teller frame 125
- de Hoffmann-Teller speed 125
- Debye length 62, 63, 319
- declination 5, 213
- derivative
 - convective 33
 - local 33

- partial 32
- total 32, 33
- differential flux 93, 94
- differential rotation 60, 169
- diffusion 103
 - in momentum space 111
 - isotropic 108
 - L-shell 249
 - magnetic fields 53
 - pitch angle 110
 - radial 249
- diffusion coefficient 105
 - in momentum space 112
- diffusion equation 108
 - solution 109
- diffusion tensor 107
- diffusion-convection equation 110
- diffusive shock acceleration 134, 135, 202, 204
 - acceleration time 136
 - energy spectrum 136
 - ESP event 199
 - self-generated turbulence 138
 - transport equation 135
 - upstream intensity increase 137
- dimensionless variables 55
- dip-equator 213
- disappearing filament 176, 185
- disconnection event 180
- discontinuity 126
 - contact 126
 - rotational 127
 - tangential 126
- dispersion relation 68
 - Alfvén wave 72
 - electromagnetic wave 81
 - electron plasma wave 77
 - electrostatic ion wave 80
 - extraordinary wave 84
 - ion cyclotron wave 80
 - ion wave 78
 - ion-acoustic wave 78
 - Langmuir wave 77
 - lower hybrid wave 81
 - magneto-sonic wave 74
 - ordinary wave 83
 - R and L waves 85
 - sound wave 78
 - upper hybrid oscillation 80
 - Whistler 85
- dissipation of fields 52, 53
- dissipation range 163
- dissipation time 52

- distribution function 33, 90, 94
- double layer 271
- drift 17
 - $\mathbf{E} \times \mathbf{B}$ 18, 316
 - $E_{\text{grad}}B$ 21
 - along neutral line 20
 - curvature 20
 - curved magnetic field 21
 - energy change 21
 - gradient 19, 21
 - gravitational 19
 - heliospheric current sheet 20, 207
- drift current 21
- drift energy 25
- drift velocity
 - derivation 18
 - general 19
- driven shock 121
- drunkards walk 104
- dynamo
 - $\alpha\Omega$ 61
 - homogeneous 59
 - MHD 58
- easy access 209
- electromagnetic wave 81
- electron plasma frequency 76, 77
- electron plasma wave 77
- electronvolt 15
- electrostatic ion wave 80
- electrostatic shock 272
- electrostatic wave 75
- energetic particles 3
- energetic storm particles 189, 198
- energy density
 - electromagnetic field 13
- energy equation
 - electromagnetic field 13, 44
- equation of state 38
- double adiabatic 39
- Euler's description 32
- eV 15
- expected distance 104, 105
- extraordinary wave 84
- Faraday's law 10
- Fermi acceleration 248
 - first order 135
 - second-order 139
- Fermi effect
 - first order 28
- filament
 - configuration 179
- coronal mass ejection 179
- disappearing 176
- inverse configuration 180
- K-S configuration 180
- normal configuration 179
- R-K configuration 180
- solar 2, 47, 171, 179
- flare 6, 171
 - and coronal mass ejection 178, 182
 - classes of 175, 191
 - confined 176, 182
 - electromagnetic radiation 182
 - energetic particles 191
 - eruptive 176, 182
 - gradual 175
 - impulsive 175
 - model 182
 - nano-flare 155
 - preflare phase 171
 - selective heating 183
- flux tubes 169
- flux-transfer event 227, 237
- focused transport 194
- Fokker–Planck equation 97
- Forbush decrease 203
- foreshock
 - planetary 280
- forward shock 174
- Fourier transformation 72
- frequency 68
- Friedrichs diagram 74
- frozen-in field 40, 45, 50, 52, 237
 - reconnection 53
- frozen-out field 51
- galactic cosmic rays 187
- anomalous component 188
- Galton board 106
- garden-hose effect 156
- gauss
 - unit 156
- Gauss's distribution 106
- Gauss's law
 - electric field 10
 - magnetic field 10
- geomagnetic coordinates 212
- geomagnetic disturbance 6
- geomagnetic reference field 215
- geosphere 229, 233
- gradient drift 28
- ground level event 188, 251
- group velocity 68
- guiding center 17

- Hale cycle 167, 308
 Hale's polarity law 169
 Hall current 12
 Hall effect 44
 heat conduction equation 52
 heliopause 153
 heliosphere 3, 143
 heliospheric current sheet (HCS) 20, 149, 159, 170, 207
 helmet streamer 147, 148
 herringbone burst 174
 hodograph 74
 hydrostatic equation 46
 impact parameter 100, 101
 impedance probe 320
 inclination 213
 inertial range 163
 information horizon 120
 interplanetary propagation 194
 interplanetary shock 262
 ion cyclotron wave 80
 ion plasma frequency 78
 ion trap 319
 ion wave 78
 ion-acoustic wave 78
 ionopause 291
 ionosphere 1, 7, 116, 229, 297
 – reflection of radio waves 83
 isochasm 6
 isoclines 213
 isothermal compression 38
 Jeans' theorem 95
 jet stream 296
 jovian electrons 189
 kappa-distribution 93, 197
 L-shell diffusion 249
 L-shell parameter 214
 Lagrange description 32
 Langmuir oscillation 76, 174
 Langmuir probe 318
 Langmuir wave 77
 large-angle collisions 101
 Larmor radius 16
 Legendre expansion 214
 Liouville's theorem 94
 lobe 224
 local geometry 127, 129
 Lorentz force 15
 loss cone 27, 116, 281
 loss-cone distribution 116, 244
 loss-cone scattering 115
 low-latitude boundary layer (LLBL) 223
 lower hybrid frequency 116
 lower hybrid wave 81
 Mach number 122, 128
 – Alfvén 128, 129, 185
 – critical 129
 – sonic 128
 magnetic bottle 23, 27
 magnetic buoyancy 169
 magnetic cloud 183, 262
 magnetic compression 133, 200
 magnetic field dissipation
 – time scale 52
 magnetic flux 10, 25
 – constancy 25
 magnetic mirror 23, 25, 26, 200
 – restoring force 26
 magnetic moment 16, 212, 271
 – adiabatic invariant 23
 – constancy 17
 magnetic pressure 40, 45, 69, 169, 221
 magnetic Reynolds number 56
 magnetic stream function 57
 magnetic tension 47, 53, 69
 magnetic viscosity 56
 magneto-sonic wave 69
 – dispersion relation 74
 – fast 74
 – phase speed 74
 magnetohydrodynamics (MHD), *see MHD* 31
 magnetohydrokinematics 31, 49
 magnetohydrostatics 31, 44
 magnetometer
 – alkali-vapor 315
 – fluxgate 314
 – proton-precession 315
 – pulsation 314
 magnetopause 8, 221, 224, 227
 magnetosheath 225, 235
 magnetosphere 3
 – definition 3
 – pole-on 288
 magnetotail 223, 224
 Mars Global Surveyor 278
 mass spectrometer 320
 Maunder Minimum 168
 Maxwell distribution 34, 35, 77, 91, 97, 123, 197

- Maxwell's equations 10
- mean free path 36, 99, 100, 111
 - and rigidity 115
 - estimate 110
 - solar energetic particles 195
- merged interaction region 161, 208
- mesosphere 297
- metric burst 174
- MHD
 - one fluid description 32
 - two-fluid description 41
- MHD dynamo 37, 58, 278
 - solar 169
 - terrestrial 218
- MHD wave
 - fast 74
- mirror point 25, 26
- mirror point
 - energy conservation 25
- mirror ratio 27
- modulation 187, 206
- momentum balance 32, 37
 - electromagnetic force 33
 - fictitious forces 36
 - pressure-gradient force 33
 - stress tensor 35
 - two-fluid description 41
 - macroscopic 43
 - viscosity 36
- nano-flares 155
- Navier–Stokes equation 37
- neutrino problem 146
- normal incidence frame 124
- nuclear fusion 145
- oblique shock 127
- Ohm's law 12
 - current acceleration 44
 - generalized 12
 - Hall term 44
 - pressure diffusion 44
 - two-fluid description 41, 42
 - two-fluid description
 - macroscopic 44
- ohmic losses 13, 44
- ordinary wave 83
- ozone layer 296, 298
- Palmer consensus range 195
- parallel shock 127
- perpendicular shock 127
- perturbation theory 69, 112
- phase space 89
- phase space density 90, 111
- phase velocity 68
- photosphere 146
- pitch angle 16
- pitch-angle diffusion 110
- pitch-angle diffusion coefficient 111, 114
- pitch-angle scattering 97
- plane wave 68
- plasma 1
- plasma frequency
 - electron 77
 - ion 78
- plasma oscillation 76, 77
- plasma sheet 224
- plasma- β 40
- plasmapause 232
- plasmasphere 229, 231
- Poisson's equation 10
- polar cap 223, 224
- polar cap absorption (PCA) 241, 252, 298
 - and ozone 300
- polar cusp 223, 235, 247, 252
- polar glow 273
- polar glow aurora (PGA) 253
- polar wind 232
- post-shock increase 200
- power density spectrum
 - ipl. magnetic field 114
- Poynting vector 13, 44
- precursor 171
- pressure diffusion 44
- pressure-gradient force 33, 35
- proton–proton cycle 145
- protuberance 47
- quasi-linear theory 69, 112, 196
- quasi-neutral 1
- quasi-parallel shock 127
- quasi-perpendicular shock 127
- radiation belt 7
 - losses 116
 - new 250
 - planetary 281
- radio burst 77, 174
 - coronal mass ejection 178
 - kilometric type II 174, 178
 - kilometric type III 174
 - metric type I 174
 - metric type II 121, 174, 178
 - metric type III 174

- metric type IV 174
- metric type V 174
- Rankine–Hugoniot equations 123
 - gasdynamic 124
 - magnetohydrodynamic shock 125
 - shock speed 124
- reconnection 53, 127, 172, 180, 224, 234, 235, 263, 264
 - Petschek 54
 - Sweet–Parker 54
- refraction index 68
 - light wave 82
 - R-waves 85
 - Whistler 85
 - X wave 84
- region 1 current 239
- region 2 current 239
- resonance condition 114
- resonance interaction 113
- resonance probe 320
- resonance scattering 114
- retarding potential analyzer 319
- reverse shock 120, 174
- Reynolds number 56
 - magnetic 56
- rigidity
 - cut-off 250
 - magnetic 16
- ring current 22, 28, 216, 257
- rotational discontinuity 235
- scale height 49
- scatter-free shock acceleration 131
- scattering cross-section 99
- selective heating 183, 192
- self-consistent fields 95
- sfe variation 256
- shock
 - Alfvén 128
 - blast wave 121
 - collisionless 120, 122
 - compression ratio 184
 - coronal 178
 - coronal and coronal mass ejection 178, 192
 - de Hoffmann–Teller frame 125
 - definition 119, 122
 - diffusive acceleration 134
 - driven 121
 - fast 128
 - frame of reference 125
 - intermediate 128
 - interplanetary and particles 198
 - kilometric type II burst 174
 - magnetic compression 184
 - metric type II burst 174
 - normal incidence frame 124, 125
 - oblique 127
 - parallel 127
 - perpendicular 127
 - quasi-parallel 127
 - quasi-perpendicular 127
 - reconnection 53
 - rest frame 122
 - reverse 120
 - scatter-free acceleration 131
 - slow 129
 - standing 119, 122
 - traveling 119, 122
- shock drift acceleration 131, 199, 202, 204
 - energy gain 133
- shock normal
 - calculation 130
- shock speed 124, 130, 200
- shock wave
 - reconnection 53
- shock-spike 199
- slab model 114, 196
- slot 242
- solar breeze 154
- solar constant 144, 293
- solar cycle 167
 - Babcock model 169
 - magnetic 167
 - model 168
 - sunspot numbers 168
- solar electromagnetic radiation
 - γ -rays 173
 - elementary flare bursts 173
 - flare 171
 - H α 173
 - hard X-rays 173
 - metric type I 174
 - metric type II burst 174
 - metric type III 174
 - microwaves 173
 - preflare phase 171
 - radio bursts 174
 - soft X-rays 173
 - type IV burst 174
 - type V burst 174
- solar energetic particles 171, 188
 - transport 194
- solar wind 6, 150
 - acceleration 155

- early evidence 150
- fast 150, 259
- heating 155
- hydrodynamic model 153
- properties 150
- slow 151
- solar-terrestrial relationships 4
- sound speed 73, 78, 121
- sound wave 78
- source surface 148
- South Atlantic Anomaly 215, 244
- Spörer's law 169
- specific heat ratio 38
- sputtering 279
- Sq variation 216, 255
- Størmer orbits 250
- Størmer unit 251
- stand-off distance 222
- standard deviation 106
- standard QLT 196, 202
- standing shock 119, 122
- Stefan-Boltzmann's law 144
- stochastic acceleration 117, 139, 200, 202
- stop band 84
- stratosphere, 296
- stream function
 - magnetic 57
- streamer belt 148, 151
- streaming
 - diffusive 107
- stress tensor 35
- subsolar point 204, 221, 222
- substorm 116, 256
- substorm current 240
- substorm current wedge 264
- sudden commencement 256
- sudden ionospheric disturbance (SID) 231, 252, 295
- sunspot 1, 46, 169
- sunspot cycle 5
- tail current 224
- tangential discontinuity 235
- temperature
 - effective 144
- termination shock 153, 188
- thermal speed 77, 91
- thermosphere 297
 - circulation 301
- theta aurora 274
- tilt angle 159, 170
- time to maximum 109
- time-of-flight 321
- transformation
 - electromagnetic field 11
- transition region 147
- traveling shock 119, 122
- tropopause 296
- troposphere 295
- Ulysses 165, 208
 - cosmic rays 209
 - magnetic field 166
 - solar wind 166
 - trajectory 165
- upper hybrid frequency 80
- upper hybrid oscillation 79
- viscosity 36
 - collisionless 36
 - kinematic 36
 - magnetic 56
- viscosity coefficient 36
- Vlasov equation 95, 112
- vorticity area index 308
- wave
 - Alfvén 69, 71, 155, 163
 - - compressive 74
 - electromagnetic 81
 - electron plasma 77
 - electrostatic 75
 - electrostatic ion 80
 - extraordinary 84
 - fast 74
 - ion 78
 - ion acoustic 163
 - ion cyclotron 80
 - ion-acoustic 78
 - ion-cyclotron 163
 - Langmuir 77
 - lower hybrid 81
 - magnet-sonic 74
 - magneto-sonic 69
 - magnetohydrodynamic 74, 155
 - MHD 69
 - ordinary 83
 - plasma 77
 - R and L 85
 - sound 78
 - upper hybrid 80
 - Whistler 86, 116, 163
- wave vector 68
- Whistler 85, 86, 115, 116, 232, 286
- X-point 53, 180, 224, 235, 264