PLASMA ASTROPHYSICS

D.B. Melrose

School of Physics
University of Sydney

Lecture Notes

prepared for Physics IV 2005

PLASMA ASTROPHYSICS

Lecture 1 : Fast Particles in Astrophysical Plasmas
Galactic Cosmic Rays, Energetic Particles from the Sun, The Earth's Magnetosphere.

Lecture 2 : Orbit Theory
Motion of a Charged Particle, Drift Motions, Adiabatic Invariants, Coulomb Interactions.

Lecture 3 : Magnetic Field Lines
Descriptions of Magnetic Fields, Equations of a Magnetic Field Line, Dipolar Magnetic Field, Frozen-in Magnetic Field Lines, Wandering of Magnetic Field Lines.

Lecture 4 : Magnetohydrodynamics
Derivation of the MHD Equations, Energy Density and Energy Flow, Magnetic Pressure and Magnetic Tension, MHD Equilibrium.

Lecture 5 : Stellar Winds and Rotating Magnetospheres
Equations for a Wind, Sonic Point, Magnetic Winds, Rotating Magnetospheres.

Lecture 6 : MHD Waves
Alfvén waves, Sound and Magnetoacoustic Waves, The Fast and Slow Modes, Alfvén Waves in Astrophysics.

Lecture 7 : MHD Shock Waves
Boundary Conditions at a Discontinuity, Shocks in Unmagnetized Plasmas, Magnetohydrodynamic Shocks, Shock Waves in Astrophysical Plasmas.

Lecture 8 : Wave-Particle Resonance
Resonance in Unmagnetized Plasma, Resonance in Magnetized Plasma, The Einstein Coefficients, The Quasilinear Equations.

Lecture 9 : Resonant Scattering
Pitch Angle Scattering, Scattering of Ions by MHD Waves, Spatial Diffusion, Growth of Waves due to a Spatial Gradient.

Lecture 10 : Applications of Pitch Angle Scattering
Generation of Alfvén Waves by Streaming Cosmic Rays, Reduction of the Streaming Speed, Scattering of Electrons by Whistlers, Scattering of Particles in the Magnetosphere.

Lecture 11 : Stochastic Acceleration
Energy Changing Mechanisms, Diffusion in Momentum Space, Stochastic Acceleration due to MHD Turbulence, The Spectrum of Magnetic Turbulence, Applications of Stochastic Acceleration.

Lecture 12 : Diffusive Shock Acceleration
Qualitative Introduction to Diffusive Shock Acceleration, Diffusive Acceleration at a Shock Front, Alternative Treatment of the Acceleration, Acceleration at Multiple Shocks, Applications of Diffusive Shock Acceleration.

Lecture 13 : Shock Drift Acceleration
Subluminal and Superluminal Shocks, Reflection and Transmission at a Shock Front, Maximum Energy Change on Reflection, The Foreshock and Interplanetary Shocks, Shock Drift Acceleration in the Crab Nebula.

Lecture 14 : Acceleration by Parallel Electric Fields
Acceleration by Langmuir Waves and Ion Sound Waves, Runaway Electrons, Acceleration of Impulsive Solar Electrons, Electrostatic Double Layers in the Auroral Zones.

Lecture 15 : Bremsstrahlung
Treatment Based on the Larmor Formula, Emission and Absorption in a Thermal Plasma, Bremsstrahlung at High Frequencies, Optically Thin and Optically Thick Sources, Applications of Bremsstrahlung.

Lecture 16 : Gyromagnetic Emission
Gyromagnetic Emission in Vacuo, Cyclotron Emission and Absorption by Thermal Electrons, Cyclotron Maser Emission, Gyrosynchrotron Emission.

Lecture 17 : Synchrotron Radiation
Forward Emission by Relativistic Particles, Semiquantitative Treatment of Synchrotron Emission, The Emissivity for Synchrotron Emission, Emission by a Power-Law Distribution, Synchrotron Absorption, The Razin Effect.

Lecture 18 : Interpretation of Synchrotron Sources
Shape of Synchrotron Spectra, Quantitative Interpretation of Synchrotron Emission, Polarization of Sources, Evolution of Synchrotron Spectra, Discussion of Synchrotron Sources.

Lecture 19 : Plasma Emission
Qualitative Discussion of Plasma Emission, Weak Turbulence Theory, Growth of the Langmuir Waves, Saturation Model for Plasma Emission, Strong Turbulence and Plasma Emission.

Lecture 20 : Pulsars
Goldreich–Julian Density, Vacuum Gaps, The Pair Plasma, The Pulsar Wind, Radiation from Pulsar Magnetospheres.

Lecture 21 : Active Galactic Nuclei
Very hot plasmas in AGN, Synchrotron-self-Compton model, Relativistic beaming effects, Extragalactic jets.

Appendices :
Physical, Astronomical and Plasma Constants, Conversions Between Units, MHD Equations in Cylindrical and Spherical Polar Coordinates

Exercise Sets

Update

The original version of these lecture notes is over a decade old. I have made a few revisions for the course in 2005, and plan to update parts of the course during the semester. This is more a course in theoretical physics and plasma physics than in astrophysics *per se*, and the motivation for the theoretical developments remain essentially unchanged. Although a few of the astrophysical points need updating, the required updates are mostly minor. I attach the original **Preface**.

Don Melrose
July 2005

Preface

Plasma astrophysics, which became recognized as a branch of astrophysics in the 1960s, is concerned with processes in ionized gases of astrophysical interest. An earlier, related branch of astrophysics, called *cosmical electrodynamics*, was concerned primarily with astrophysical plasmas treated as magnetized fluids. An important step in the development of plasma astrophysics was the recognition of the importance of processes that cannot be described in terms of a fluid approach. The most characteristic of these processes arise from resonant wave-particle interactions between suprathermal particles and various types of plasma waves, leading to the scattering and acceleration of the particles or to the generation of nonthermal radio emission. Plasma astrophysics has its origins in two separate branches of astrophysics: cosmic ray physics and radio astronomy. The relevant aspects of cosmic ray physics are the acceleration of the cosmic rays and their propagation in the interstellar and interplanetary magnetic fields. In the earliest years of radio astronomy it was already obvious that most radioastronomical sources are nonthermal. That is, unlike the optical emission from stars, the radio emission usually does not come from a system in or near thermal equilibrium. Most nonthermal radio emission is due to synchrotron emission by relativistic electrons. These electrons are thought to be accelerated by shock waves.

This lecture course is being given for the first time in 1992. It occurs at the same time as a compulsory lecture course in kinetic theory, and as a lecture course in electromagnetic theory. This has influenced the sequence and arrangement of these lectures. The concept of a single-particle distribution function is introduced and used on the assumption that the reader has met this previously. For example, MHD theory is justified in lecture 4 using kinetic theory. No prior knowledge of fluid mechanics is assumed because not all students have taken a course in fluid mechanics. The discussion of radiation processes is left to near the end of this course. Logically, it could be argued that radiation processes should be discussed first. However, the relevant sections of the electromagnetic theory course occur near the end of the semester, and it seems more sensible to discuss astrophysical applications of radiation processes at about the same time.

An unfortunate fact of life in astrophysics is that one needs to be familiar with several sets of units. Although all undergraduate teaching is in SI units, and the use of SI is recommended by the International Astronomical Union, the use of SI units in astrophysics is far less common than the use of rationalized gaussian units. In addition, there are many astrophysical units that are standard. For example, units of length include the parsec (pc), the light year and the Astronomical Unit (AU) (the the mean radius of the Earth's

orbit), a standard unit of mass is the solar mass, $M_\odot$, and a standard unit of luminosity (power) is the solar luminosity, $L_\odot$. The situation concerning electromagnetic units is far worse. The gaussian unit for magnetic field (the gauss) is standard in astrophysics, except in the Russian literature where it is replaced by the oersted; however, in the space physics literature the SI unit (the tesla) is used by a significant fraction of authors. (In the space physics literature one still sometimes encounters the gamma, which is a nanotesla, that is, $1\gamma = 10^{-9}\,\mathrm{T} = 10^{-5}\,\mathrm{G}$.) Electric units are confusing in that while gaussian units are commonly used for charge and charge density in most astrophysical discussion, electric potentials are usually expressed in volts (the gaussian unit is the statvolt). For electric field strengths the mixed unit of a volt per centimeter tends to used, rather than the gaussian unit (statvolt per centimeter) or the SI unit (volt per meter). In these lecture notes, all the theory is in SI units, but the discussion is in units that the reader is likely to encounter in the astrophysical literature. It is an important part of a training in astrophysics to be able to use dimensional arguments, to make simple semiquantitative estimates, and to change freely from one set of units to another. Some appendices containing physical astronomical and plasma physics constants and conversion factors between different systems of units are included.

These notes are intended to complement lectures. Some of the chapters, referred to as lectures, are in the nature of reference material, and these will not be covered in detail. Notably lectures 15, 16 and 19 follow closely relevant chapters in the book Melrose and McPhedran (1991), which is the basis for the lecture course on electromagnetic theory given at the same time as this lecture course, and these chapters are more detailed than is appropriate for this course. It is anticipated that there will be insufficient time to cover all the material in the remaining chapters in these notes. As a result certain chapters may be covered only briefly or not at all. In particular, lectures 13, 14 and 19 will be omitted if time is short.

It is intended that about five exercise sets will included as part of this course, and will be part of the assessment for this course. However, the preparation of these has not been completed at this time, and the exercise sets will be added to these notes during the semester.

In preparing these notes I have not followed any textbook closely. The books to which I have referred most are:

"The Origin of Cosmic Rays", V.L. Ginzburg and S.I. Syrovatskii, Macmillan (1964).

"High Energy Astrophysics", M.S. Longair, Cambridge University Press (1981).

"Plasma Astrophysics, Volumes 1 & 2", D.B. Melrose, Gordon & Breach (1980).

"Electromagnetic Processes in Dispersive Media", D.B. Melrose and R.C. McPhedran, Cambridge University Press (1991).

"Radio Astrophysics", A.G. Pacholczyk, Freeman (1970).

"Solar Magnetohydrodynamics", E.R. Priest, Reidel (1982).

"Physics of High Temperature Plasmas", G. Schmidt, Academic Press (1966).

"The Physics of Astrophysics Volume I: Radiation", F.H. Shu, University Science Books (1991).

D.B. Melrose
January 1992

The University of Sydney
AUSTRALIA

School of Physics

PHYSICS HONOURS, SEMESTER 2, 2005
PLASMA ASTROPHYSICS

LECTURE ARRANGEMENTS

There will be 20 lectures on Plasma Astrophysics:

Lecturer: Don Melrose

Times and place of lectures: noon Tues, noon Wed, 11am Thurs LT4

No lectures 2,3,4 & 10 August; 6,7,8 & 13 September

LECTURE NOTES

The course will be based on lecture notes that will be handed out and will be available in electronic form.

ASSIGNMENTS AND EXAM

Assignments will contribute 25% of the final mark, with an exam contributing 75%. These arrangements are open to negotiation with the class at the start of the course. Graduate students will be expected to hand in the assignments and to sit the exam, and will be graded only pass or fail.

COURSE DEFINITION – PLASMA ASTROPHYSICS

Extract from Preample to lecture notes:

Plasma astrophysics became recognized as a branch of astrophysics in the 1960s, is concerned with processes in ionized gases of astrophysical interest. An earlier, related branch of astrophysics, called *cosmical electrodynamics*, was concerned primarily with astrophysical plasmas treated as magnetized fluids. An important step in the development of plasma astrophysics was the recognition of the importance of processes that cannot be described in terms of a fluid approach. The most characteristic of these processes arise from resonant wave-particle interactions between suprathermal particles and various types of plasma waves, leading to the scattering and acceleration of the particles or to the generation of nonthermal radio emission. Plasma astrophysics has its origins in two separate branches of astrophysics: cosmic ray physics and radio astronomy. The relevant aspects of cosmic ray physics are the acceleration of the cosmic rays and their propagation in the interstellar and interplanetary magnetic fields. In the earliest years of radio astronomy it was already obvious that most radioastronomical sources are nonthermal. That is, unlike the optical emission from stars, the radio emission usually does not come from a system in or near thermal equilibrium. Most nonthermal radio emission is due to synchrotron emission by relativistic electrons. These electrons are thought to be accelerated by shock waves.

End of Extract

The course will emphasize plasma physics that is either motivated by astrophysical applications or is essential to understanding astrophysical theories. Basic plasma physics will include orbit theory, MHD (including MHD waves and shocks), kinetic theory and emission and absorption processes. Important parts of the course will involve the scattering and acceleration of fast particles and radiation processes in astrophysical plasmas. Some specific astrophysical applications will be discussed in more detail. The titles of the lectures are listed below.

PLASMA ASTROPHYSICS
Lecture 1 : Fast Particles in Astrophysical Plasmas
Lecture 2 : Orbit Theory
Lecture 3 : Magnetic Field Lines
Lecture 4 : Magnetohydrodynamics
Lecture 5 : Stellar Winds and Rotating Magnetospheres
Lecture 6 : MHD Waves
Lecture 7 : MHD Shock Waves
Lecture 8 : Wave-Particle Resonance
Lecture 9 : Resonant Scattering
Lecture 10 : Applications of Pitch Angle Scattering
Lecture 11 : Stochastic Acceleration
Lecture 12 : Diffusive Shock Acceleration
Lecture 13 : Shock Drift Acceleration
Lecture 14 : Acceleration by Parallel Electric Fields
Lecture 15 : Bremsstrahlung
Lecture 16 : Gyromagnetic Emission
Lecture 17 : Synchrotron Radiation
Lecture 18 : Interpretation of Synchrotron Sources
Lecture 19 : Plasma Emission
Lecture 20 : Pulsars
Lecture 21 : Active Galactic Nuclei

Plasma Astrophysics

Appendices

These appendices contain numerical values for relevant constants, some information of electromagnetic units and some details of the vector algebra in cylindrical polar and spherical polar coordinates.

1. Physical, Astronomical and Plasma Constants

Table A1: Physical Constants

physical quantity	symbol	SI units	gaussian units
speed of light	c	$3.0 \times 10^{8}\,\mathrm{m\,s^{-1}}$	$3.0 \times 10^{10}\,\mathrm{cm\,s^{-1}}$
fundamental charge	e	$1.6 \times 10^{-19}\,\mathrm{C}$	$4.8 \times 10^{-10}\,\mathrm{esu}$
electron mass	m_e	$9.1 \times 10^{-31}\,\mathrm{kg}$	$9.1 \times 10^{-28}\,\mathrm{g}$
proton mass	m_p	$1.67 \times 10^{-27}\,\mathrm{kg}$	$1.67 \times 10^{-24}\,\mathrm{g}$
mass ratio	m_p/m_e	1837	1837
Boltzmann constant	κ_B	$1.38 \times 10^{-23}\,\mathrm{J\,K^{-1}}$	$1.38 \times 10^{-16}\,\mathrm{erg\,K^{-1}}$
electron volt	eV	$1.6 \times 10^{-19}\,\mathrm{J}$	$1.6 \times 10^{-12}\,\mathrm{erg}$
gravitational constant	G	$6.67 \times 10^{-11}\,\mathrm{N\,m^2\,kg^{-2}}$	$6.67 \times 10^{-8}\,\mathrm{dyne\,cm^2\,g^{-2}}$
(Planck's constant)$/2\pi$	$\hbar$	$1.05 \times 10^{-34}\,\mathrm{J\,s}$	$1.05 \times 10^{-27}\,\mathrm{erg\,s}$
classical e^- radius	r_0	$2.8 \times 10^{-15}\,\mathrm{m}$	$2.8 \times 10^{-13}\,\mathrm{cm}$
Thomson cross section	σ_T	$6.65 \times 10^{-29}\,\mathrm{m^2}$	$6.65 \times 10^{-25}\,\mathrm{cm^2}$
critical B field	B_c	$1.44 \times 10^{9}\,\mathrm{T}$	$1.44 \times 10^{13}\,\mathrm{G}$
permittivity of free space	ε_0	$8.85 \times 10^{-12}\,\mathrm{F\,m^{-1}}$	
permeability of free space	μ_0	$1.23 \times 10^{-6}\,\mathrm{H\,m^{-1}}$	

Table A2: Astronomical Constants

astronomical quantity	symbol	SI units	gaussian units
astronomical unit	AU	$1.5 \times 10^{11}\,\mathrm{m}$	$1.5 \times 10^{13}\,\mathrm{cm}$
parsec	pc	$3.09 \times 10^{16}\,\mathrm{m}$	$3.09 \times 10^{18}\,\mathrm{cm}$
light year	ly	$0.95 \times 10^{16}\,\mathrm{m}$	$0.95 \times 10^{18}\,\mathrm{cm}$
angstrom unit	Å	$10^{-10}\,\mathrm{m}$	$10^{-8}\,\mathrm{cm}$
solar radius	$R_\odot$	$7.0 \times 10^{8}\,\mathrm{m}$	$7.0 \times 10^{10}\,\mathrm{cm}$
Earth's radius	R_E	$6.4 \times 10^{6}\,\mathrm{m}$	$6.4 \times 10^{8}\,\mathrm{cm}$
Jupiter's radius	R_J	$7.1 \times 10^{7}\,\mathrm{m}$	$7.1 \times 10^{9}\,\mathrm{cm}$
solar mass	$M_\odot$	$2.0 \times 10^{30}\,\mathrm{kg}$	$2.0 \times 10^{33}\,\mathrm{g}$
solar luminosity	$L_\odot$	$3.9 \times 10^{26}\,\mathrm{W}$	$3.9 \times 10^{33}\,\mathrm{erg\,s^{-1}}$

The following plasma formulas give expression in SI units and gaussian units when the physical quantity is given in the relevant units. Thus the formula for v_A in SI units involves B in tesla and n_e per cubic meter, formula in gaussian units involves B in gauss and n_e per cubic centimeter. Formulas involving the mass density are contain a factor μ defined by $\eta = \mu m_p n_e$.

Table A3: Plasma Physics Quantities

astronomical quantity	symbol	SI units	gaussian units
plasma frequency	ω_p	$56.4 n_e^{1/2}\,\mathrm{s}^{-1}$	$5.64 \times 10^4 n_e^{1/2}\,\mathrm{s}^{-1}$
	ν_p	$9.0 n_e^{1/2}\,\mathrm{Hz}$	$9.0 \times 10^3 n_e^{1/2}\,\mathrm{Hz}$
electron gyrofrequency	Ω_e	$1.76 \times 10^{11} B\,\mathrm{s}^{-1}$	$1.76 \times 10^7 B\,\mathrm{s}^{-1}$
	ν_B	$2.8 \times 10^{10} B\,\mathrm{Hz}$	$2.8 \times 10^6 B\,\mathrm{Hz}$
proton gyrofrequency	Ω_p	$0.98 \times 10^7 B\,\mathrm{s}^{-1}$	$0.98 \times 10^4 B\,\mathrm{s}^{-1}$
Alfvén speed	v_A	$2.2 \times 10^{16} B(\mu n_e)^{-1/2}\,\mathrm{m\,s}^{-1}$	$2.2 \times 10^{11} B(\mu n_e)^{-1/2}\,\mathrm{cm\,s}^{-1}$
sound speed	c_s	$1.5 \times 10^2 T^{1/2}\,\mathrm{m\,s}^{-1}$	$1.5 \times 10^4 T^{1/2}\,\mathrm{cm\,s}^{-1}$
ion sound speed	v_s	$91 T_e^{1/2}\,\mathrm{m\,s}^{-1}$	$9.1 \times 10^3\,\mathrm{cm\,s}^{-1}$
Debye length	λ_D	$69 T_e^{1/2} n_e^{-1/2}\,\mathrm{m}$	$6.9 T_e^{1/2} n_e^{-1/2}\,\mathrm{cm}$
thermal e^- speed	V_e	$3.9 \times 10^3\,\mathrm{m\,s}^{-1}$	$3.9 \times 10^5\,\mathrm{cm\,s}^{-1}$
collision frequency	ν_0	$1.37 \times 10^{-5} (\ln\Lambda) n_e T_e^{-3/2}\,\mathrm{s}^{-1}$	$13.7 (\ln\Lambda) n_e T_e^{-3/2}\,\mathrm{s}^{-1}$

2. Conversions Between Units

Table A4: Conversion Factors between SI and gaussian units

quantity	SI/gaussian	gaussian/SI
length	$10^{-2}\,\mathrm{m/cm}$	$10^2\,\mathrm{cm/m}$
mass	$10^{-3}\,\mathrm{kg/g}$	$10^3\,\mathrm{g/kg}$
energy	$10^{-7}\,\mathrm{J/erg}$	$10^7\,\mathrm{erg/J}$
power	$10^{-7}\,\mathrm{W/erg\,s}^{-1}$	$10^7\,\mathrm{erg\,s}^{-1}\mathrm{/W}$
force	$10^{-5}\,\mathrm{N/dyne}$	$10^5\,\mathrm{dyne/N}$
charge	$\frac{1}{3} \times 10^{-9}\,\mathrm{statcoul/C}$	$3 \times 10^9\,\mathrm{C/statcoul}$
electric field	$3 \times 10^4\,\mathrm{V\,m}^{-1}\mathrm{/statvolt\,cm}^{-1}$	$\frac{1}{3} \times 10^{-4}\,\mathrm{statvolt\,cm}^{-1}\mathrm{/V\,m}^{-1}$
current	$\frac{1}{3} \times 10^{-9}\,\mathrm{A/statamp}$	$3 \times 10^9\,\mathrm{statamp/A}$
current density	$\frac{1}{3} \times 10^{-5}\,\mathrm{A\,m}^{-2}\mathrm{/statamp}$	$3 \times 10^5\,\mathrm{statamp\,cm}^{-2}\mathrm{/A\,m}^{-2}$
magnetic induction	$10^{-4}\,\mathrm{T/G}$	$10^4\,\mathrm{G/T}$

Conversion factors between different systems of units are made most conveniently by introducing conversion factors and regarding units as algebraic symbols. For example, consider conversion from meters to centimeters, or vice versa. Given that $10^2\,\mathrm{cm} = 1\,\mathrm{m}$, one may introduce conversion factors by writing $1 = 10^2\,\mathrm{cm\,m^{-1}}$ or $1 = 10^{-2}\,\mathrm{m\,cm^{-1}}$, both of which follow directly from the basic relation. Then if one is given a formula in meters, as $L = x\,\mathrm{m}$, where x is a number, then the quantity in centimeters, that is, $L = y\,\mathrm{cm}$ where y is another number, is given by $L = x\,\mathrm{m}(10^2\,\mathrm{cm\,m^{-1}}) = \mathrm{y\,cm}$, so that one identifies $y = 10^2 x$. Thus, using the conversion formula in this case, one regards the units m as algebraic symbols and cancels them.

The detailed formulas in these notes are given in SI units. To convert any formula to gaussian units involves no change if no electromagnetic quantities are involved. If electromagnetic quantities appear, no change is made to the following quantities

charge q,	electric field $\mathbf{E}$,	current density $\mathbf{J}$.
charge density ρ,	electric potential Φ,	

The following changes are made

magnetic field $\mathbf{B} \to \mathbf{B}/c$,	permittivity of free space $\varepsilon_0 \to 1/4\pi$,
vector potential $\mathbf{A} \to \mathbf{A}/c$,	permeability of free space $\mu_0 \to 4\pi/c^2$.

Other conversion factors include the following.

Table A5: Other Coversion Factors

quantity	factor	inverse
temperature	$8.6 \times 10^{-5}\,\mathrm{eV/K}$	$1.16 \times 10^{4}\,\mathrm{K/eV}$
X-ray energy	$4.1 \times 10^{-15}\,\mathrm{eV/Hz}$	$2.4 \times 10^{14}\,\mathrm{Hz/eV}$
angle	$2.06 \times 10^{5}\,\mathrm{arcsec/rad}$	$4.85 \times 10^{-6}\,\mathrm{rad/arcsec}$
time	$3.16 \times 10^{7}\,\mathrm{s/yr}$	$3.17 \times 10^{-8}\,\mathrm{yr\,s}$

3. MHD Equations in Cylindrical and Spherical Polar Coordinates

The following equations give explicit formulas for the differential vector operators in cylindrical polar coordinates ϖ, ϕ, z. The component of a vector $\mathbf{V}$ are written (V_ϖ, V_ϕ, V_z), where these are components along unit vectors in the radial, azimuthal and axial directions, respectively.

The differential vector operators are

$$\mathrm{grad}\,\chi = \left(\frac{\partial\chi}{\partial\varpi}, \frac{1}{\varpi}\frac{\partial\chi}{\partial\phi}, \frac{\partial\chi}{\partial z}\right), \tag{A.1}$$

$$\mathrm{div}\,\mathbf{F} = \frac{1}{\varpi}\frac{\partial(\varpi F_\varpi)}{\partial\varpi} + \frac{1}{\varpi}\frac{\partial F_\phi}{\partial\phi} + \frac{\partial F_z}{\partial z}, \tag{A.2}$$

$$\mathrm{curl}\,\mathbf{F} = \left(\frac{1}{\varpi}\frac{\partial F_z}{\partial\phi} - \frac{\partial F_\phi}{\partial z}, \frac{\partial F_\varpi}{\partial z} - \frac{\partial F_z}{\partial\varpi}, \frac{1}{\varpi}\frac{\partial(\varpi F_\phi)}{\partial\varpi} - \frac{1}{\varpi}\frac{\partial F_\varpi}{\partial\phi}\right), \tag{A.3}$$

$$\begin{aligned}(\mathbf{A}\cdot\mathrm{grad}\,)\mathbf{B} = \Bigg(&A_\varpi\frac{\partial B_\varpi}{\partial\varpi} + \frac{A_\phi}{\varpi}\frac{\partial B_\varpi}{\partial\phi} + A_z\frac{\partial B_\varpi}{\partial z} - \frac{A_\phi B_\phi}{\varpi}, A_\varpi\frac{\partial B_\phi}{\partial\varpi} + \frac{A_\phi}{\varpi}\frac{\partial B_\phi}{\partial\phi}\\ &+ A_z\frac{\partial B_\phi}{\partial z} + \frac{A_\phi B_\varpi}{\varpi}, A_\varpi\frac{\partial B_z}{\partial\varpi} + \frac{A_\phi}{\varpi}\frac{\partial B_z}{\partial\phi} + A_z\frac{\partial B_z}{\partial z}\Bigg).\end{aligned} \tag{A.4}$$

In using these equations below axial symmetry implies $\partial/\partial\phi = 0$.

Consider a flow velocity $\mathbf{u}$ and a magnetic field $\mathbf{B}$ in an axially symmetric system. Both may be written in similar forms:

$$\mathbf{u} = \left(-\frac{1}{\varpi}\frac{\partial\psi}{\partial z}, v_\phi, \frac{1}{\varpi}\frac{\partial\psi}{\partial\varpi}\right), \tag{A.5}$$

$$\mathbf{B} = \left(-\frac{1}{\varpi}\frac{\partial\chi}{\partial z}, B_\phi, \frac{1}{\varpi}\frac{\partial\chi}{\partial\varpi}\right). \tag{A.6}$$

The function χ may be interpreted in terms of the ϕ component of the vector potential, with $\chi = \varpi A_\phi$. A similar interpretation may be given to ψ. Taking the curl of (A.6) gives

$$\mathrm{curl}\,\mathbf{B} = \left(-\frac{\partial B_\phi}{\partial z}, -\frac{1}{\varpi}\frac{\partial^2\chi}{\partial z^2} - \frac{\partial}{\partial\varpi}\left(\frac{1}{\varpi}\frac{\partial\chi}{\partial\varpi}\right), \frac{1}{\varpi}\frac{\partial(\varpi B_\phi)}{\partial\varpi}\right). \tag{A.7}$$

Then, neglecting the displacement current, the azimuthal component of the current is given by

$$\mu_0 J_\phi = -\frac{1}{\varpi}\frac{\partial^2\chi}{\partial z^2} - \frac{\partial}{\partial\varpi}\left(\frac{1}{\varpi}\frac{\partial\chi}{\partial\varpi}\right). \tag{A.8}$$

The magnetic force is

$$\mathbf{J}\times\mathbf{B} = \frac{1}{\mu_0}\left(j_\phi\frac{\partial\chi}{\partial\varpi} - \frac{B_\phi^2}{\varpi} - \frac{1}{2}\frac{\partial B_\phi^2}{\partial\varpi}, -\frac{1}{\varpi^2}\frac{\partial\chi}{\partial z}\frac{\partial\varpi B_\phi}{\partial\varpi} + \frac{1}{\varpi}\frac{\partial\chi}{\partial\varpi}\frac{\partial B_\phi}{\partial z}, -\frac{1}{2}\frac{\partial B_\phi^2}{\partial z} + j_\phi\frac{\partial\chi}{\partial z}\right), \tag{A.9}$$

with $j_\phi = \mu_0 J_\phi/\varpi$. The convection electric field is given by

$$-\mathbf{u}\times\mathbf{B} = -\left(\frac{v_\phi}{\varpi}\frac{\partial\chi}{\partial\varpi} - \frac{B_\phi}{\varpi}\frac{\partial\psi}{\partial\varpi}, -\frac{1}{\varpi^2}\frac{\partial\psi}{\partial\varpi}\frac{\partial\chi}{\partial z} + \frac{1}{\varpi^2}\frac{\partial\psi}{\partial z}\frac{\partial\chi}{\partial\varpi}, -\frac{B_\phi}{\varpi}\frac{\partial\psi}{\partial z} + \frac{v_\phi}{\varpi}\frac{\partial\chi}{\partial z}\right). \quad (A.10)$$

The associated charge density is

$$-\varepsilon_0 \mathrm{div}\,\mathbf{u}\times\mathbf{B} = \frac{\varepsilon_0}{\varpi}\left[\frac{\partial}{\partial\varpi}\left(B_\phi\frac{\partial\psi}{\partial\varpi}\right) + \frac{\partial}{\partial z}\left(B_\phi\frac{\partial\psi}{\partial z}\right) - \frac{\partial}{\partial\varpi}\left(v_\phi\frac{\partial\chi}{\partial\varpi}\right) - \frac{\partial}{\partial z}\left(v_\phi\frac{\partial\chi}{\partial z}\right)\right]. \quad (A.11)$$

The Poynting flux $\mathbf{F}_{\mathrm{EM}} = \mathbf{E}\times\mathbf{B}/\mu_0$ for the case with $\mathbf{E} = -\mathbf{u}\times\mathbf{B}$ reduces to

$$\begin{aligned}\mathbf{F}_{\mathrm{EM}} = \frac{1}{\mu_0}\big(&v_\varpi(B_\phi^2 + B_z^2) - B_\varpi(v_z B_z + v_\phi B_\phi),\\ &v_\phi(B_\varpi^2 + B_z^2) - B_\phi(v_\varpi B_\varpi + v_z B_z), v_z(B_\varpi^2 + B_\phi^2) - B_z(v_\varpi B_\varpi + v_\phi B_\phi)\big).\end{aligned} \quad (A.12)$$

The divergence of the stress tensor is

$$\begin{aligned}(\mathrm{grad}\,\mathbf{T})_\varpi &= \frac{1}{\varpi}\frac{\partial(\varpi T_{\varpi\varpi})}{\partial\varpi} + \frac{1}{\varpi}\frac{\partial T_{\phi\varpi}}{\partial\phi} + \frac{\partial T_{z\varpi}}{\partial z} - \frac{T_{\phi\phi}}{\varpi},\\ (\mathrm{grad}\,\mathbf{T})_\phi &= \frac{1}{\varpi}\frac{\partial(\varpi T_{\varpi\phi})}{\partial\varpi} + \frac{1}{\varpi}\frac{\partial T_{\phi\phi}}{\partial\phi} + \frac{\partial T_{z\phi}}{\partial z} + \frac{T_{\phi\varpi}}{\varpi},\\ (\mathrm{grad}\,\mathbf{T})_z &= \frac{1}{\varpi}\frac{\partial(\varpi T_{\varpi z})}{\partial\varpi} + \frac{1}{\varpi}\frac{\partial T_{\phi z}}{\partial\phi} + \frac{\partial T_{zz}}{\partial z}.\end{aligned} \quad (A.13)$$

The following equations give explicit formulas for the differential vector operators in spherical polar coordinates r, θ ϕ. The component of a vector $\mathbf{V}$ are written (V_r, V_θ, V_ϕ), where these are components along unit vectors in the radial, polar and azimuthal directions, respectively:

$$\mathrm{grad}\,\chi = \left(\frac{\partial\chi}{\partial r}, \frac{1}{r}\frac{\partial\chi}{\partial\theta}, \frac{1}{r\sin\theta}\frac{\partial\chi}{\partial\phi}\right), \quad (A.14)$$

$$\mathrm{div}\,\mathbf{F} = \frac{1}{r^2}\frac{\partial(r^2F_r)}{\partial r} + \frac{1}{r\sin\theta}\frac{\partial(\sin\theta F_\theta)}{\partial\theta} + \frac{1}{r\sin\theta}\frac{\partial F_\phi}{\partial\phi}, \quad (A.15)$$

$$\begin{aligned}\mathrm{curl}\,\mathbf{F} = \Bigg(&\frac{1}{r\sin\theta}\left[\frac{\partial(\sin\theta F_\theta)}{\partial\phi} - \frac{\partial F_\theta}{\partial\phi}\right],\\ &\frac{1}{r\sin\theta}\frac{\partial F_r}{\partial\phi} - \frac{1}{r}\frac{\partial(rF_\phi}{\partial r}, \frac{1}{r}\left[\frac{\partial(rF_\theta)}{\partial r} - \frac{\partial F_r}{\partial\theta}\right]\Bigg),\end{aligned} \quad (A.16)$$

$$\begin{aligned}(\mathbf{A}\cdot\mathrm{grad}\,)\mathbf{B} = \Bigg(&A_r\frac{\partial B_r}{\partial r} + \frac{A_\theta}{r}\frac{\partial B_\varpi}{\partial\theta} + \frac{A_\theta}{r\sin\theta}\frac{\partial B_r}{\partial\phi} - \frac{A_\theta B_\theta + A_\phi B_\phi}{r},\\ &A_r\frac{\partial B_\theta}{\partial r} + \frac{A_\theta}{r}\frac{\partial B_\theta}{\partial\theta} + \frac{A_\phi}{r\sin\theta}\frac{\partial B_\theta}{\partial\phi} + \frac{A_r B_\theta}{r} - \cos\theta\frac{A_\phi B_\varpi}{r\sin\theta},\\ &\frac{A_r B_\phi}{r} + A_r\frac{\partial B_\phi}{\partial r} + \frac{A_\phi}{r\sin\theta}\frac{\partial B_\phi}{\partial\phi} + \frac{A_\theta}{r\sin\theta}\frac{\partial(\sin\theta B_\phi)}{\partial\theta}\Bigg).\end{aligned} \quad (A.17)$$

Lecture 1

Fast particles in astrophysical plasmas

The presence of populations of fast particles in astrophysical plasmas is almost universal. In this lecture a qualitative description is given of the particle distributions in several different examples of astrophysical plasmas. Some of the features of the particle motions can be understood adequately in terms of orbit theory, but there are clear examples where orbit theory leads to false predictions, and the motion of the particles is evidently diffusive.

1.1 Galactic cosmic rays

Cosmic rays are high energy particles, mostly protons, that continuously bombard the Earth. The energy, ε, of a cosmic ray is usually measured in electron volts, $1\,\mathrm{eV} = 1.602 \times 10^{-19}\,\mathrm{J}$. The energy spectrum, $N(\varepsilon)$, of cosmic rays is defined such that the number density in the energy range between ε and $\varepsilon + d\varepsilon$ is $N(\varepsilon)d\varepsilon$. The observed energy spectrum has a peak at about $10^9\,\mathrm{eV} = 1\,\mathrm{GeV}$. Below this peak the spectrum of cosmic rays varies with the 11-year solar cycle, and on shorter time scales it varies in association with solar flares. Above this peak the spectrum is unaffected by the Sun, and falls off as a power law

$$N(\varepsilon)d\varepsilon = K\varepsilon^{-a}d\varepsilon, \tag{1.1}$$

where K is a constant and where a is the spectral index. One finds $a = 2.6$ for $\varepsilon \lesssim 10^{15}\,\mathrm{eV}$, and $a \approx 3$ at $\varepsilon \gtrsim 10^{15}\,\mathrm{eV}$.

The ionic composition of the cosmic rays is similar to, but not exactly the same as the composition of the Sun and other stars, and in the interplanetary medium. In particular, the light elements (Li, Be, B) are relatively overabundant, and there is very strong evidence that these are spallation products, that

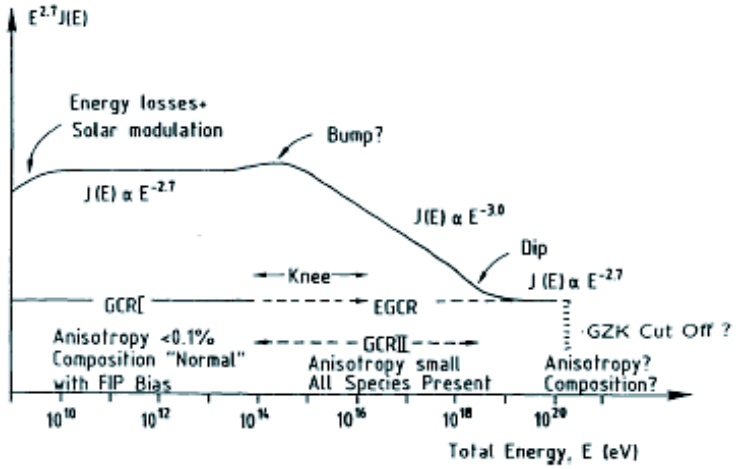


Figure 1.1: A schematic representation of the Galactic cosmic ray differential energy spectrum, multiplied by $E^{1.7}$, showing the three main component, GCRI, GCRII and EGCR. [from Axford, W.I., 1994, ApJS, 90, 937]

is, the daughter nuclei produced as the result of nuclear interactions between the cosmic rays and particles in the interstellar medium. The relative abundance of the light elements (when combined with the known cross sections for spallation interactions) then provides information on the column density (mass per unit area) of matter that a typical cosmic ray has passed through ($\approx 5\,\mathrm{g\,cm^{-2}}$). An estimate of the density of interstellar matter ($\approx 10^{-25}\,\mathrm{g\,cm^{-3}}$), together with the fact that the cosmic rays are propagating at close to the speed of light, then provides an estimate of the distance that a typical cosmic ray has propagated through the interstellar medium ($\approx 5 \times 10^{24}\,\mathrm{cm} \approx 2 \times 10^{7}\,\mathrm{ly}$), and hence of the time that it has been propagating. The particles propagate at the speed of light, and hence this implies a life time of order $10^7\,\mathrm{yr}$, which is much shorter than the age of the galaxy and much longer than the propagation time to the edge of the Galactic disk. The propagation time may be estimated by noting that the particle needs to propagate several hundred parsecs ($1\,\mathrm{pc} = 3.02\,\mathrm{ly}$) along the Galactic magnetic field, which is approximately in the plane of the disk, cf. Figure 1.1, in order to escape, and propagating at the speed of light this takes $\approx 10^3\,\mathrm{yr}$. Data from the effect of cosmic ray tracks in meteorites imply that the cosmic ray flux has not changed significantly over the age of the solar system ($> 10^9\,\mathrm{yr}$). This suggests that cosmic rays are trapped in the Galaxy for about $10^7\,\mathrm{yr}$, and are in balance between a steady input and a slow escape. This confinement time is so long that the only plausible explanation for it is that the cosmic rays diffuse very slowly, due to an efficient form of scattering.

Of the many aspects of cosmic rays that are inadequately understood, three

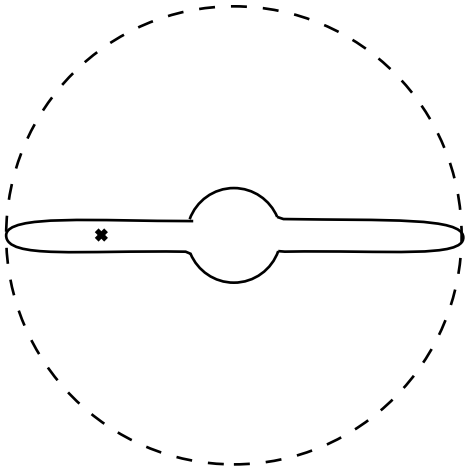


Figure 1.2: The gross structure of the Galaxy is indicated schematically. The Galactic disk is of radius $\approx 15\,\mathrm{kpc}$ and of half-thickness $\approx 200\,\mathrm{pc}$. The Sun, indicated by the cross, is $\approx 10\,\mathrm{kpc}$ from the Galactic center. The Galactic halo is roughly spherical, indicated by the dashed line.

are mentioned here. First, the mechanism for the acceleration of the cosmic rays is still not clear. On energetic grounds it seems that only supernova explosions can provide sufficient energy to maintain the cosmic ray spectrum. The favored theory is that shock waves, generated by supernova explosions, accelerate cosmic rays in the interstellar medium, with the particles themselves coming from the interstellar medium and not directly from supernovas. Second, the mechanism by which cosmic rays are confined for such a long time in the Galaxy are partly but not completely understood. It is thought that cosmic rays are scattered by magnetic fluctuations in the interstellar medium, through a process called *resonant scattering*, and that as a result they diffuse slowly along magnetic field lines. Resonant scattering of cosmic rays also occurs in the interplanetary medium and causes the lower-energy part of the cosmic ray spectrum to be modulated by the 11-year cycle of solar activity. Another process that may play a significant role in the propagation of cosmic rays is the diffusion of magnetic field lines, which is discussed in the next lecture. Third, the ratio of electrons to ions, which is about three percent at a given energy, is not understood. The presence of the electrons is important from an observational viewpoint because of the synchrotron radiation that they emit. By mapping the distribution of radio emission, one can map the distribution of cosmic rays throughout the Galaxy, and also the distribution of cosmic rays in neighboring Galaxies. The radio emission is slightly stronger in the spiral arms of the Galaxy than between the arms.

At the highest energy the gyroradii of cosmic rays appears to be larger than the thickness of the Galactic disk, suggesting that the highest energy cosmic rays may be of extragalactic origin. The gyroradius of a relativistic particle is given by

$$R = 1.7 \times 10^{-3} \, \frac{\gamma\beta \sin\alpha}{B} \, \mathrm{m} = 1.0 \times 10^{-15} \left(\frac{\varepsilon \sin\alpha}{1 \, \mathrm{eV}} \right) \left(\frac{B}{1 \, \mu\mathrm{G}} \right)^{-1} \, \mathrm{pc} \qquad (1.2)$$

where the first expression is in SI units with $v = \beta c$, and where the second expression applies to highly relativistic particle ($\beta \approx 1$) in units most often used in astrophysics. The strength of the interstellar magnetic field is probably about $5 \, \mu\mathrm{G} = 5 \times 10^{-10} \, \mathrm{T}$. Hence cosmic rays with energies $\gtrsim 10^{18} \, \mathrm{eV}$ would appear to have gyroradii larger than the thickness of the Galactic disk, and so could not be confined to the disk.

1.2 Energetic particles from the Sun

The Sun is a sporadic source of cosmic rays, which correlate with large solar flares. The solar energetic particles (SEPs), as they are now often called, may be accelerated by shock waves. A possible alternative is stochastic acceleration in turbulent magnetic fields, which is sometimes also called Fermi acceleration. Besides SEPs, there are other contexts where particles are accelerated in the solar atmosphere, most notably in solar flares.

In solar flares it seems that much of the energy released in a flare goes initially into electrons with energy $\approx 10 \, \mathrm{keV}$. Direct signatures of these electrons are hard X rays, generated by bremsstrahlung when the electrons precipitate into the denser regions of the solar atmosphere. Some of the higher energy electrons trapped in the magnetic field emit radio waves in the microwave range ($\approx 1 \, \mathrm{GHz}$) due to so-called gyrosynchrotron radiation, which is the counterpart of synchrotron radiation for mildly relativistic electrons. Temporal variations in the X-ray and microwave fluxes imply that the acceleration time for the electrons is very short, $\lesssim 100 \, \mathrm{ms}$. The precipitating electrons also cause the optical and UV radiation from solar flares. Some of the electrons escape from the solar atmosphere as electron beams, and these generate radio emission, called type III solar radio bursts, as they propagate through the solar corona. The electron beams can be studied directly by spacecraft in the solar wind. In some large flares, data on $\gtrsim 10 \, \mathrm{MeV}$ gamma-rays imply that ions and relativistic electrons are accelerated promptly ($\lesssim 1 \, \mathrm{s}$). The gamma rays are produced by nuclear reactions as these energetic particles penetrate into the denser regions of the solar atmosphere.

Using various spacecraft in the interplanetary medium, the spectrum of energetic particles has been studied in detail over a wide range of energies. Besides the particles already mentioned, that is, Galactic cosmic rays, SEPs and type III emitting electrons, there are three other notable sources of particles. One is the

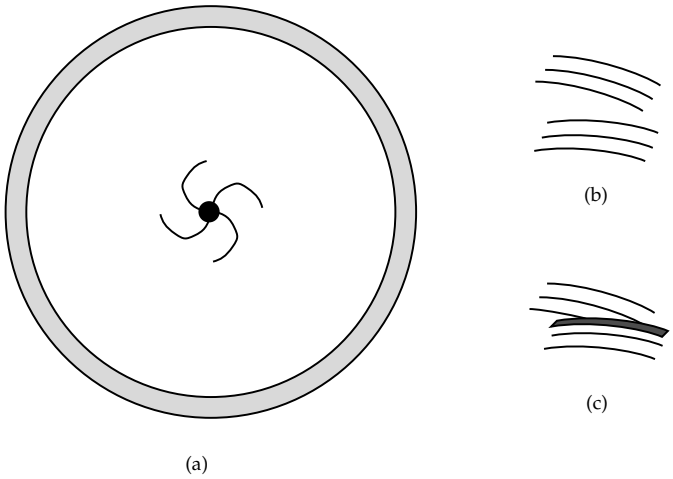


Figure 1.3: (a) The heliosphere is defined as the region magnetically connected to the Sun, it includes the interplanetary medium and extends out to the heliopause, shown by shaded region, where there must be a shock transition at which the solar wind terminates. The central region shows how the solar wind draws the magnetic field (four field lines shown) into an Achimedian spiral pattern. (b) Far from the Sun the field is nearly tangential, and faster streams (less tightly wound) in the solar wind overtake slow streams (more tighly wound), leading (c) to a shock transition at the interface between the two streams.

Earth itself, especially the bow shock of the Earth. A second source is interplanetary shock waves: the intensity of energetic ions tends to rise as a shock wave approaches and fall again after the shock wave passes. Acceleration occurs at interplanetary shocks and at planetary bow shocks, due to so-called shock drift acceleration. Another source of energetic particles is the planet Jupiter, which produces electrons in the range 1–10 MeV that populate the entire heliosphere, cf. Figure 1.3.

1.3 The Earth's magnetosphere

The magnetosphere of the Earth is the region in which the magnetic field lines are connected to the Earth. Far from the Earth ($r \gg 10R_E$) in the solar wind, the magnetic field lines are connected to the Sun, and the principal energy density is that in the kinetic energy of the solar wind, which may be interpreted as a ram pressure and written as $\frac{1}{2}\eta_{\rm sw}v_{\rm sw}^2$, where $\eta_{\rm sw}$ is the mass density and $v_{\rm sw}$ is the speed of the solar wind. Near the Earth ($r \ll 10R_E$) the Earth's magnetic field is nearly dipolar, and the dominant pressure in that in the magnetic field,

which may be written as $B^2/2\mu_0$. A dipolar field falls off as r^{-3}, so that the magnetic pressure falls off as r^{-6}. The two pressure balance at $r \approx 10R_E$. The solar wind slows down at a *bow shock* on the sunward side of the Earth, and then flows around the magnetosphere, which forms a teardrop-like cavity in the solar wind. The magnetosphere is illustrated schematically in Figure 1.4.

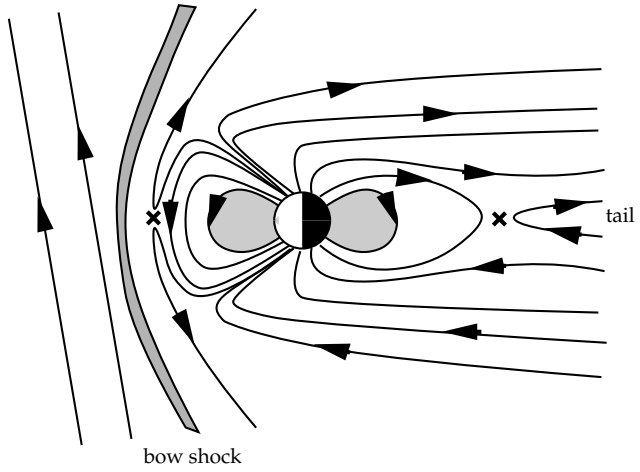


Figure 1.4: The Earth's magnetosphere is indicated schematically. The Sun is to the left of the figure. The ram pressure of the solar wind draws the magnetic field lines out into a magnetotail in the antisunward direction. The bow shock on the sunward side is indicated. The shaded region in the inner magnetosphere is the plasmasphere.

Acceleration of fast particles occurs in several different regions of the magnetosphere, most notably in the auroral zones. Electrons precipitating from the tail into the auroral zones excite O and N atoms at altitudes around 100 km and the decay of the excited atoms produces the auroral lights. However, here I discuss another population of fast particles, specifically those trapped in the so-called radiation (or Van Allen) belts. The dipolar-like field lines of the Earth form magnetic bottles that can trap fast particles. In the early days of space research, before 1960, the existence of trapped distributions of electrons (40 keV–2 MeV) and of trapped ions ($\gtrsim 100$ keV) was known. The life time of the trapped particles was not known, and some authors thought that it could be very long. For example, for a 100 keV electron in a plasma with $n_e = 1\,\mathrm{cm}^{-3}$, (1.25) implies a collision time, $1/\nu(v)$, of about 3×10^8 s, that is, about ten years. In 1962 there was an experiment, called STARFISH, in which a nuclear bomb was set off in the magnetosphere, and the decay of the radioactive products was studied. It was found that the life times vary from tens of minutes to a few weeks. Clearly the particles are scattered much more efficiently than they would be through

Coulomb interactions alone, and this is attributed to resonant scattering.

An obvious question is how the particles in the radiation belts are accelerated. A mechanism that seems to explain most of the features of the acceleration can be understood in terms of adiabatic invariants. Recall that there are three invariants in an idealized magnetosphere. Let these be denoted $M = p_\perp^2/B$, $J = \oint ds\, p_\parallel$ and $\Phi = \oint ds\, A_\phi$. In a dipolar field, with $r = LR_E$, one has $B \propto L^{-3}$, and hence a vector potential $A \propto L^{-2}$. Then because the lengths in the integrals are both proportional to L, one finds $p_\perp^2 \propto ML^{-3}$, $p_\parallel \propto JL^{-1}$ and $\Phi \propto L^{-1}$. If the third adiabatic invariant is conserved, then the particles move on a single L shell and their energy is conserved. However, it turns out that Φ is not conserved, because there are significant fluctuations in the structure of the magnetosphere on time scales in which particles drift around the equator. The fluctuations cause a diffusion in Φ values, which implies a diffusion in L. Hence, particles injected from the solar wind at the boundary of the magnetosphere tend to diffuse inward. As they do so, M and J are conserved, and this implies that their pitch angle increases and their energy increases. Suppose that we average over pitch angle, and assume

$$\left\langle p_\perp^2 p_\parallel \right\rangle \propto p^3 \propto L^{-4}, \tag{1.3}$$

for fixed M and J. Hence the kinetic energy of a particle goes up proportional to $L^{-8/3}$ as it diffuses inward. For example, a 2 keV particle injected from the solar wind at $L = 10$ would have an energy of 150 keV when it reaches $L = 2$. The inward diffusing particles are ultimately scattered into the loss cone, precipitate into the denser regions of the atmosphere and are lost.

One suggested extension of this idea is to a model for the acceleration of electrons by Jupiter is illustrated in Figure 1.5. Electrons are assumed to be injected at the point labeled 1, and to diffuse into the point labeled 2 due to diffusion at constant M and J. At point 2 the electron distribution is highly anisotropic, specifically being a pancake-like distribution about $\alpha = \pi/2$. The anisotropic electrons generate whistlers which scatter the electrons, so that diffusion in M occurs. This makes the electrons distribution more isotropic and drives the average height of mirroring towards the surface of the planet. It is postulated that electrons mirroring close to the planet are subject to meridional diffusion due to low-frequency waves, in bounce resonance with the electrons, violating the second adiabatic invariant, J. This drives electrons onto higher L-shells taking them back towards point 1. The cycle can then repeated. In the simplest version of the theory, the energy of an electron increases by a factor $(L_1/L_2)^b$ in each cycle, with $b = 8/3$ for nonrelativistic electrons and $b = 4/3$ for relativistic electrons.

This is not the only possible acceleration mechanism for the MeV-electrons from Jupiter, and it is not the most plausible mechanism in my opinion. There is direct evidence for the acceleration of electrons in association with the motion of the innermost Galilean satellite, Io, through the Jovian magnetosphere. This

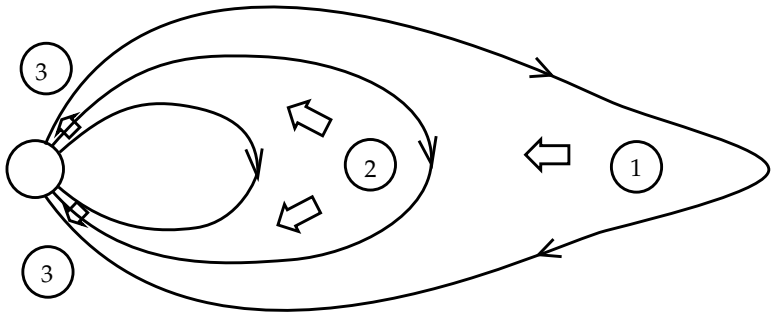


Figure 1.5: A "recirculation" model for electron acceleration in the Jovian magnetosphere: the arrows indicate the direction of diffusive transport.

evidence is from the Jovian decametric radio emission, that is attributed to electron cyclotron maser emission, and that correlates with the position of Io. Also, during the encounters of the Pioneer and Voyager spacecraft with Jupiter it was found that the distribution of such electrons peaks at Io's orbit. A more plausible acceleration mechanism for these electrons is due to parallel electric fields generated in association with Io's motion.

Lecture 2

Orbit theory

A characteristic feature of astrophysical plasmas is that they are populated by distributions of energetic particles confined by a magnetic field. The study of the motion of individual particles in magnetic and electric fields is often referred to as *orbit theory*. In this lecture a summary of the important features of orbit theory is given.

2.1 Motion of a charged particle

Consider a particle with charge q and mass m moving in a magnetic field $\mathbf{B}$ and an electric field $\mathbf{E}$. Newton's equation of motion is

$$\frac{d\mathbf{p}}{dt} = q\big[\mathbf{E} + \mathbf{v} \times \mathbf{B}\big], \tag{2.1}$$

where $\mathbf{v}$ is the velocity of the particle and where

$$\mathbf{p} = \gamma m \mathbf{v}, \quad \varepsilon = \gamma m c^2, \quad \gamma = (1 - v^2/c^2)^{-1/2} \tag{2.2}$$

are the momentum, energy and Lorentz factor of the particle, respectively.

2.1.1 Uniform B, E

$\mathbf{E} = \mathbf{0}$

In a uniform magnetic field with no electric field, the following quantities are constants of the motion: the energy $\varepsilon = \gamma m c^2$, and the components $p_\perp = \gamma m v_\perp$ and $p_\parallel = \gamma m v_\parallel$ of the momentum perpendicular and parallel to the magnetic field, respectively. The motion of the particle may be decomposed into a motion at constant velocity along the field lines plus a circular motion perpendicular to

the field lines. Thus the particle exhibits a spiralling motion with the pitch of the spiral defining the *pitch angle* α:

$$v_\perp = v\sin\alpha, \qquad v_\parallel = v\cos\alpha. \tag{2.3}$$

The frequency of the circular motion is called the *gyrofrequency* Ω, and the radius is called the *radius of gyration* R (or sometimes the *Larmor radius*):

$$\Omega = \frac{\Omega_0}{\gamma}, \quad \Omega_0 = \frac{|q|B}{m}, \quad R = \frac{v_\perp}{\Omega} = \frac{p_\perp}{|q|B}. \tag{2.4}$$

The sense of gyration, which is the handedness of the circular motion in a screw sense relative to $\mathbf{B}$, depends on the sign of the charge

$$\eta = q/|q|. \tag{2.5}$$

Positively charged particles ($\eta = +1$) gyrate in a left hand screw sense relative to $\mathbf{B}$, and negatively charged particles ($\eta = -1$) gyrate in a right hand screw sense relative to $\mathbf{B}$.

The orbit of the particle is described by an equation that gives the position $\mathbf{x}$ of the particle as a function of time, and can be written as $\mathbf{x} = \mathbf{X}(t)$. Solving (2.1) in this case gives

$$\mathbf{X}(t) = \mathbf{x}_0 + \big(R\sin(\phi_0 + \Omega t), \eta R\cos(\phi_0 + \Omega t), v_\parallel t\big), \tag{2.6}$$

where ϕ_0 and $\mathbf{x}_0$ are determined by the position of the particle at $t = 0$, and where the z axis is chosen along the direction of $\mathbf{B}$. The instantaneous velocity of the particle is given by

$$\mathbf{v}(t) = \dot{\mathbf{X}}(t) = (v_\perp\cos(\phi_0 + \Omega t), -\eta v_\perp\sin(\phi_0 + \Omega t), v_\parallel). \tag{2.7}$$

The sense of gyration is such that the magnetic field produced by the spiralling charge opposes the externally applied field. Plasmas are *diamagnetic* – the motions of the individual particles always tend to reduce the applied magnetic field.

$\mathbf{E} \neq \mathbf{0}$

Now consider the effect of inclusion of a uniform, nonzero electric field. In most applications it is assumed that there is no parallel component, i.e. $\mathbf{E}\cdot\mathbf{B} = 0$. The reason is that plasmas are highly electrically conducting, and charges can flow freely along magnetic field lines to short out any parallel component of electric field. However, particles do not flow freely across field lines and so an electric field perpendicular to the magnetic field is not shorted out and can persist.

So far we have chosen the z axis to be along the direction of $\mathbf{B}$, We are still free to rotate our coordinate system about this axis, so let's choose this rotation

such that $\mathbf{E} = (E, 0, 0)$. For nonrelativistic particles the different components of equation (2.1) are then

$$\begin{aligned} \frac{dv_x}{dt} &= \frac{q}{m}E + \eta\,\Omega\,v_y\,, \\ \frac{dv_y}{dt} &= -\eta\,\Omega\,v_x\,, \\ \frac{dv_z}{dt} &= 0\,. \end{aligned} \tag{2.8}$$

The equation for the z component is trivial. Differentiating the other two, using the fact that $\mathbf{E}$ is constant, we obtain

$$\begin{aligned} \frac{d^2v_x}{dt^2} &= \eta\,\Omega\frac{dv_y}{dt} = -\Omega^2 v_x\,, \\ \frac{d^2v_y}{dt^2} &= -\eta\,\Omega\frac{dv_x}{dt} = -\Omega^2(v_y + E/B)\,. \end{aligned} \tag{2.9}$$

The equation for v_y can then be rewritten as

$$\frac{d^2}{dt^2}\left(v_y + \frac{E}{B}\right) = -\Omega^2\left(v_y + \frac{E}{B}\right) \tag{2.10}$$

and then if we make the replacement

$$v_y' = v_y + E/B \tag{2.11}$$

equation (2.10) reduces to the same form as the equation for v_x in (2.9). Furthermore, these are simply the equations for a simple harmonic oscillator which we solved for the case where $\mathbf{E} = 0$. So by analogy with equations (2.6) and (2.7) we have the following orbit equations:

$$\begin{aligned} \mathbf{v}(t) &= (v_\perp\cos(\phi_0 + \Omega t), -\eta v_\perp \sin(\phi_0 + \Omega t) - E/B, v_\parallel)\,, \\ \mathbf{X}(t) &= \mathbf{x}_0 + \big(R\sin(\phi_0 + \Omega t), \eta R\cos(\phi_0 + \Omega t) - Et/B, v_\parallel t\big)\,. \end{aligned} \tag{2.12}$$

An alternative way of understanding the effect of a perpendicular electric field is to note that the field may be removed by making a Lorentz transformation. The quantities $\mathbf{E}\cdot\mathbf{B}$ and $B^2 - E^2/c^2$ are Lorentz invariants, and provided $E < B$ one may transform to a (primed) frame with $E' = 0$ and $B' = (B^2 - E^2/c^2)^{1/2}$. The velocity of the transformation is in the direction perpendicular to both $\mathbf{E}$ and $\mathbf{B}$, and is of magnitude $v_E = E/B$. The motion of the particles in the primed frame is simply a spiral around the magnetic field $\mathbf{B}'$. The fact that the primed frame drifts relative to the unprimed frame implies that the particle motion in the unprimed frame is a simple spiral around a guiding center (or gyrocenter) which is drifting perpendicular to both the electric and the magnetic fields. Note that in a uniform electric field *all* particles drift with the same drift velocity, $\mathbf{v}_E$.

Aside: The Lorentz transformation argument suggests that the gyroradius of a particle in an electric and a magnetic field is smaller than if it were in the magnetic field alone, since $\Omega' = e(B - E/c)/m$, but we did not find this in the derivation of equation (2.12). The reason is that in deriving (2.12) we neglected relativistic effects, which is equivalent to requiring $E/c \ll B$ in which case $\Omega' \approx \Omega$.

2.2 Drift motions

In many situations where the field structure is more complicated than a uniform magnetic field and zero electric field, but where the departure from this situation is small in the sense that the forces on a particle change only slightly over a gyroradius, the motion of charged particles may be separated into two parts: spiralling about the gyrocenter plus motion of the gyrocenter. The motion of the gyrocenter is described in terms of drift motions across the field lines. The following drifts are commonly identified:

$$\text{electric drift :} \quad \mathbf{v}_E = \frac{\mathbf{E} \times \mathbf{B}}{B^2}, \tag{2.13}$$

$$\text{gravitational drift :} \quad \mathbf{v}_g = \frac{m\mathbf{g} \times \mathbf{B}}{qB^2}, \tag{2.14}$$

$$\text{gradient drift :} \quad \mathbf{v}_B = \frac{p_\perp v_\perp}{2qB} \frac{\mathbf{B} \times \nabla B}{B^2}, \tag{2.15}$$

$$\text{inertial drift :} \quad \mathbf{v}_\mathrm{i} = \frac{\mathbf{B} \times (d\mathbf{p}/dt)^0}{qB^2}, \tag{2.16}$$

$$\text{centrifugal drift :} \quad \mathbf{v}_c = \frac{p_\parallel v_\parallel}{qB^2} \frac{\mathbf{B} \times (\mathbf{B} \cdot \nabla)\mathbf{B}}{B^2}, \tag{2.17}$$

$$\text{polarization drift :} \quad \mathbf{v}_P = \frac{m}{q} \frac{\dot{\mathbf{E}}_\perp}{B^2}. \tag{2.18}$$

A physical explanation of the electric drift (2.13) is given above. The gravitational drift (2.14) may be derived from the electric drift by replacing $q\mathbf{E}$ in (2.1) by $m\mathbf{g}$, and thence in (2.13) to obtain (2.14). Note that the drift is perpendicular to both the gravitational field and to the magnetic field. Also, it is in opposite senses for charges of opposite signs. Charges of opposite signs flowing in opposite directions imply an electric current. In a fluid description, the current density $\mathbf{J}$ implies a force per unit volume $\mathbf{J} \times \mathbf{B}$ that opposes the gravitational force density $\eta\mathbf{g}$ on the fluid of mass density η. The fact that the force due to the current opposes the initial force that drives the current is an example of Lenz' law. Physically, the gravitational drift may be understood as illustrated in Figure 2.1. The gravitational force accelerates particles downward, so that they have higher perpendicular momenta near the bottom of their

orbits, and so, according to (2.4), have larger gyroradii there than near the top of their orbits.

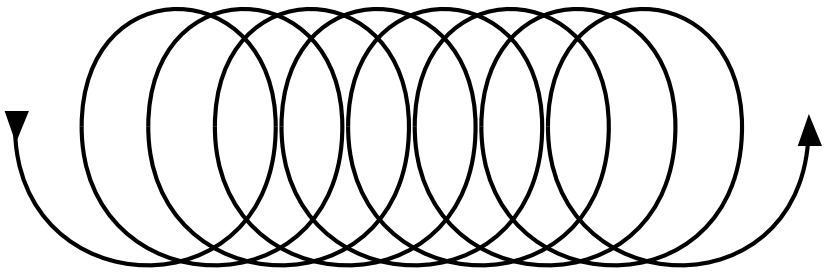


Figure 2.1: The gravitational drift of a positively-charged particle is illustrated for the case where the magnetic field is into the page and the gravitational force is directed downward.

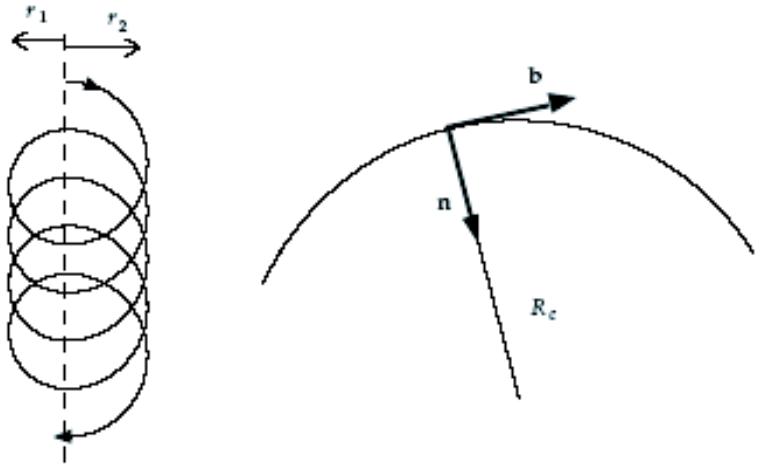
The gradient drift (2.15) is associated with a change in the strength of the magnetic field. The drift is perpendicular to both the direction of the magnetic field and to the direction of ∇B, and is opposite for oppositely charged particles. In this case the implied current generates a magnetic field that is such as to oppose the gradient in B. The field generated by the induced current satisfies $\nabla\times\mathbf{B} = \mu_0\mathbf{J}$ which is Ampere's Law in the case of constant or zero $\mathbf{E}$. Physically, the gradient drift may be understood in terms of an argument similar to that used to explain the gravitational drift. As illustrated in Figure 2.2, if one takes an idealized case in which the magnetic field changes abruptly at a surface that passes through the center of gyration of the particle, then the gyroradius is different in two halves of the orbit. On joining a sequence of semicircles with radii that alternate between two values, one obtains the orbit illustrated in Figure 2.2, which shows that a drift motion results.

The inertial drift (2.16) is attributed to the coordinate frame in which the spiralling motion is described not being an inertial frame. For example, in a rotating plasma, the frame in which the plasma is momentarily at rest is not an inertial frame. The quantity $(d\mathbf{p}/dt)^0$, which is the time derivative of the momentum relative to an inertial frame (e.g., the instantaneous rest frame), is the inertial force.

The centrifugal drift (2.17) and the polarization drift (2.18) are both specific examples of inertial drift. Centrifugal or curvature drift is associated with curvature of the magnetic field. Introducing the unit vector $\mathbf{b} = \mathbf{B}/B$, one has

$$\frac{\mathbf{B}\times(\mathbf{B}\cdot\nabla)\mathbf{B}}{B^3} = \mathbf{b}\times(\mathbf{b}\cdot\nabla\mathbf{b}), \quad \mathbf{b}\cdot\nabla\mathbf{b} = \frac{\mathbf{n}}{R_c}, \tag{2.19}$$

where R_c is the radius of curvature of the field lines, and where $\mathbf{n}$ is a unit vector along the direction toward the center of gyration, as illustrated in Figure 2.2b. Polarization drift is associated with a time-varying electric field, in which case the inertial drift is $(d\mathbf{p}/dt)^0 = md\mathbf{v}_E/dt$.



(a) (b)

Figure 2.2: (a) An idealized case which demonstrates how the gradient drift occurs: the magnetic field is into the page and its strength increases abruptly at a surface (dashed line) that passes through the gyrocenter of the negatively-charged particle so that the gyroradii are different on the two sides. The curve is drawn by joining semicircles with radii r_1 and r_2.
(b) A curved magnetic field line may be approximated by the arc of a circle: the radius of curvature R_c is the radius of this circle.

An alternative way of writing the drift motions is in terms of the average (over the spiralling motion) position $\mathbf{R} = \langle \mathbf{X} \rangle$, with $\mathbf{X}$ given by (2.6) in the case of a uniform field. Retaining only the electric, gradient and curvature drift one has

$$\dot{\mathbf{R}} = v_\parallel \mathbf{b} + \frac{\mathbf{E} \times \mathbf{b}}{B} + \frac{v_\perp p_\perp}{2qB} \left[\frac{\mathbf{b} \times \nabla B}{B} \right] + \frac{v_\parallel p_\parallel}{2qB} \mathbf{b} \times (\mathbf{b} \cdot \nabla)\mathbf{b}. \tag{2.20}$$

It is not at all obvious but it can be shown that the equation of motion for the gyrocenter has the following components

$$\dot{p}_\parallel = q\mathbf{b} \cdot \mathbf{E} + \frac{1}{2} v_\perp p_\perp \nabla \cdot \mathbf{b}, \quad \dot{p}_\perp = -\frac{1}{2} v_\perp p_\parallel \nabla \cdot \mathbf{b}. \tag{2.21}$$

Magnetic fields do no work (since the force is always perpendicular to the motion) so although the magnetic field can induce drifts we expect any change

in the particle energy to be determined solely by the electric field. Moreover, the perpendicular component of the electric field can be removed by a Lorentz transformation, and so in the simplest approximation it too does no work. If one writes $\varepsilon = (p_\perp^2 + p_\parallel^2)/(2m)$ then $\dot{\varepsilon} = v_\perp \dot{p}_\perp + v_\parallel \dot{p}_\parallel$ and the substitution of (2.21) leads to the expected result $\dot{\varepsilon} = qE_\parallel$.

2.3 Adiabatic invariants

Any mechanical system that has one or more nearly periodic motions has a nearly conserved quantity corresponding to each such motion; these conserved quantities are called *adiabatic invariants*. Formally, this may be seen simply in terms of Lagrangian or Hamiltonian dynamics. In terms of Lagrangian dynamics, let us choose one of the generalized coordinates to be the angle, θ, corresponding to the quasiperiodic motion, so that the Lagrangian for the system is $L(\theta, \dot{\theta})$, where the dependence on other variables is of no interest. The Lagrangian equation of motion is

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\theta}}\right) - \frac{\partial L}{\partial \theta} = 0. \tag{2.22}$$

Suppose one integrates (2.22) over one period of the motion, say over $0 < \theta < 2\pi$. The configuration of the system is the same at $\theta = 2\pi$ as at $\theta = 0$ so that the final term in (2.22) integrates to zero. Thus (2.22) implies that the time derivative of a quantity is zero and hence that the quantity is conserved. Thus one finds

$$\oint d\theta \, \frac{\partial L}{\partial \dot{\theta}} = \text{constant}, \qquad \oint dQ \, P = \text{constant}, \tag{2.23}$$

which is the desired adiabatic invariant. The second form in (2.23) is the corresponding form in Hamiltonian dynamics, when the periodic motion is in an arbitrary generalized coordinate Q with conjugate momentum P.

Let us apply this to a spiralling charge in a magnetic field, B, that is changing slowly (compared to a gyroradius) in space. A subtle point that is not important here but is important later is that the conjugate 3-momentum (conjugate to $\mathbf{x}$) is $\mathbf{p} + q\mathbf{A}$, where $\mathbf{A}$ is the vector potential of the magnetic field. The angle with respect to which the motion is periodic is the angle ϕ known as the *gyrophase*, which is equal to $\phi_0 + \Omega t$ in the case (2.6) of a uniform field. The conjugate momentum is the angular momentum of the motion, that is the ϕ-component of $\mathbf{p} + q\mathbf{A}$ times the radius, R, of gyration. The ϕ-component of $\mathbf{p}$ is equal to $p_\perp$ and so gives a contribution $2\pi p_\perp R$ to the integral in (2.23). Since $R \propto p_\perp/B$, this contribution is proportional to $p_\perp^2/B$. The term $q\mathbf{A}$ gives a contribution of the same form. Hence, one finds that

$$\frac{p_\perp^2}{B} = \text{constant} \tag{2.24}$$

is an adiabatic invariant, sometimes called the first adiabatic invariant, and sometimes referred to an the magnetic moment of the particle. (A charge moving in a circle corresponds to a current loop and the magnetic moment is that associated with this current loop.)

The adiabatic invariant (2.24) may be shown to be an invariant using the result (2.21) from the theory of drift motions. First note the following result:

$$\frac{d}{dt}\,\frac{1}{B} = v_\parallel \mathbf{b}\cdot\nabla\,\frac{1}{B} = \frac{v_\parallel \nabla\cdot\mathbf{b}}{B}, \tag{2.25}$$

where the first identity follows for $\partial B/\partial t = 0$, and where $\mathbf{b} = \mathbf{B}/B$ and $\nabla\cdot\mathbf{B} = 0$ are used in the second identity. Then using (2.21) and (2.25), one finds

$$\frac{d}{dt}\left(\frac{p_\perp^2}{B}\right) = \frac{2p_\perp \dot{p}_\perp}{B} + p_\perp^2\,\frac{d}{dt}\,\frac{1}{B} = 0. \tag{2.26}$$

One implication of the conservation of the first adiabatic invariant is the reflection of a particle from a magnetic compression. In the absence of any field other than an inhomogeneous magnetic field, one has $p =$ constant, and hence (2.24) implies $\sin^2\alpha/B =$ constant, where α is the pitch angle of the particle, cf. (2.3). Hence, as a particle propagates in a direction of increasing $\mathbf{b}\cdot\nabla B$, $\sin^2\alpha$ increases $\propto B$. As $\sin^2\alpha$ increases, $|\cos\alpha|$ decreases and so $|p_\parallel|$ decreases. If $\sin^2\alpha$ reaches unity then the particle motion is strictly circular, with $p_\parallel = 0$, and the particle reflects at that point and moves back in the direction of decreasing $\mathbf{b}\cdot\nabla B$. This leads to the concept of a *magnetic bottle*, which is a region of weak B between regions of stronger B such that particle can reflect at either end, as illustrated in Figure 2.3. Note that particles with sufficiently small α in the center of the bottle are not reflected, and escape from the ends of the bottle. The range $\alpha < \alpha_0$ for which particles are not trapped is called the *loss cone*.

A particle trapped in a magnetic bottle has a quasiperiodic motion corresponding to its bounce motion between the reflection points. There is an adiabatic invariant corresponding to this motion, sometimes called the second adiabatic invariant. Let the distance, s, along the field lines be a generalized coordinate, whose conjugate momentum is $p_\parallel$. (The component of $\mathbf{A}$ along $\mathbf{b}$ is zero.) Then (2.23) implies

$$\oint ds\, p_\parallel = \text{constant}, \tag{2.27}$$

where the integral is along the orbit of the gyrocenter between the reflection points.

There is a third adiabatic invariant for particles trapped in a magnetic field which is roughly dipolar, as is the case for the Earth's magnetic field within several Earth radii, R_E. Then the curvature drift causes particles to drift (in magnetic longitude) around the Earth. As this drift is quasiperiodic there is an

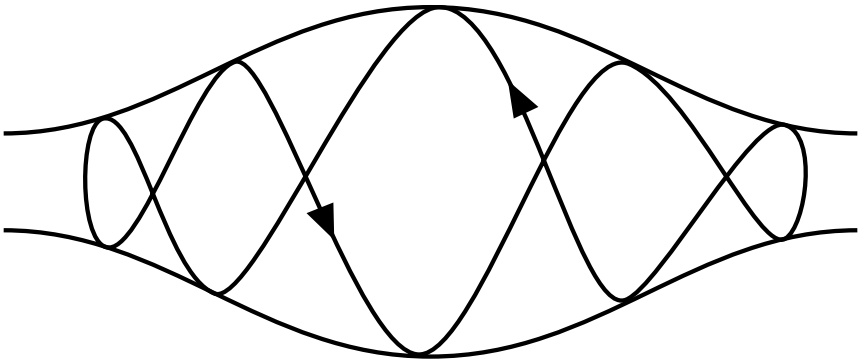


Figure 2.3: The motion of a trapped particle inside a magnetic bottle is illustrated schematically.

adiabatic invariant associated with it. This invariant is given by the integral of the component of $q\mathbf{A}$ in the direction of the drift around the closed orbit, which integral involves the radial distance $r = LR_{\rm E}$, implying that the orbit (rather the center of the bounce motion) of the particle is confined to a given r or, as is standard jargon in this context, to a given L shell.

2.4 Coulomb interactions

Interparticle collisions are often negligible for fast particles in astrophysical plasmas. The scattering of fast particles in plasmas is discussed in a later lecture, and for present purposes let us simply note the expression for the typical collision frequency for electrons with speed v due to Coulomb interactions with ions

$$\nu(v) = \frac{Z^2 e^4 n_e}{4\pi\varepsilon_0^2 m_e^2 v^3} \ln\Lambda, \tag{2.28}$$

where n_e is the electron number density and Ze is the charge on the ions. The final quantity in (2.28) is the Coulomb logarithm, which is defined only to within a factor of order unity inside the logarithm. Numerically one has

$$\ln\Lambda = \begin{cases} 16.0 - \frac{1}{2}\ln n_e + \frac{3}{2}\ln T_e - \ln Z & \text{for } T_e \lesssim 1.4\times 10^5\,\text{K}, \\ 22.0 - \frac{1}{2}\ln n_e + \ln T_e & \text{for } T_e \gtrsim 1.4\times 10^5\,\text{K}, \end{cases} \tag{2.29}$$

where n_e is per cubic meter and T_e is in kelvin. Typically in astrophysical plasmas $\ln\Lambda$ is between 10 and 20. The collision frequency (2.28) may be written in the form

$$\nu(v) = \nu_0 \left(V_e/v\right)^3, \tag{2.30}$$

where V_e is the thermal speed of the electrons. Numerical values for ν_0 in SI and gaussian units are given in an Appendix.

Additional reading

Introduction to Plasma Physics and Controlled Fusion, Vol. 1, F. F. Chen, Plenum (1983).

Lecture 3

Magnetic field lines

The concept of a magnetic field line is a familiar one, which may be given a formal definition by writing down differential equations for the field line. An important special case is that of a dipolar field, for which the equations for the field lines may be solved exactly. In a perfectly conducting medium, which is a good approximation for plasmas in many circumstances, the magnetic field lines are frozen in to the plasma. Nevertheless, the field lines may not be perfectly defined because magnetic fluctuations introduce a randomness into the direction of a field line.

3.1 Descriptions of magnetic fields

The magnetic field is usually described in terms of the magnetic induction, $\mathbf{B}$, whose SI unit is the tesla (T). In much of the astrophysical literature gaussian units are used: the gaussian unit for B is the gauss (G); $1\,\mathrm{G} = 10^{-4}\,\mathrm{T}$. Sometimes the magnetic field is described in terms of the magnetic field strength, $\mathbf{H}$, whose SI unit is weber per square meter ($\mathrm{Wb\,m^{-2}}$) and whose gaussian unit is the oersted (Oe), which is often used in the Russian literature. In a non-magnetic medium, which applies to most plasmas, one has $\mathbf{H} = \mathbf{B}/\mu_0$, with $\mu_0 = 4\pi \times 10^{-7}$ in SI units and $\mu_0 = 1$ in gaussian units.

Another mathematical description of a magnetic field is in terms of the vector potential, $\mathbf{A}$, which is such that one has

$$\mathbf{B} = \nabla \times \mathbf{A}. \tag{3.1}$$

The description ensures $\nabla \cdot \mathbf{B} = 0$ due to $\nabla \cdot \nabla \times$ giving identically zero when operating any vector. The vector potential is defined only to within a gauge transformation $\mathbf{A} \rightarrow \mathbf{A}'$,

$$\mathbf{A}' = \mathbf{A} + \nabla\psi, \tag{3.2}$$

where ψ is an arbitrary scalar quantity. On inserting (3.2) into (3.1), the fact that $\nabla \times \nabla\psi$ is identically zero implies that the value of $\mathbf{B}$ is unaffected by the choice of ψ.

Any magnetic field may be described in terms of a vector potential $\mathbf{A}$. Two other mathematical descriptions of magnetic fields are useful under restricted circumstances. A *potential* field may be defined as a magnetic field in a region where no currents are flowing. Then Ampère's equation implies $\nabla \times \mathbf{B} = 0$, and this may be satisfied by writing $\mathbf{B}$ as the gradient of a magnetic potential, ψ_M:

$$\mathbf{B} = -\nabla\psi_M \qquad \text{(potential field)}. \tag{3.3}$$

The other description is in terms of *Euler potentials*, α and β. These are surfaces whose intersection defines a magnetic field line. Such a description is of restricted validity in topologically complicated magnetic structures. (Specifically it cannot be used, at least without modification, to describe a magnetic structure that has separate intertwined regions like the links in a chain; the mathematical description of this topological restriction involves a quantity called the magnetic helicity.) The normals to the surfaces are along $\nabla\alpha$ and $\nabla\beta$, and the field line is orthogonal to both these normals. Hence in this description one has

$$\mathbf{B} = F(\alpha, \beta)\, \nabla\alpha \times \nabla\beta. \tag{3.4}$$

By suitable choice of α and β, $F(\alpha, \beta)$ can always be chosen equal to unity; for example, define $\alpha' = F(\alpha, \beta)\alpha$ and then (3.4) implies $\mathbf{B} = \nabla\alpha' \times \nabla\beta$.

3.2 Equations of a magnetic field line

Qualitatively, a field line for any vector field $\mathbf{V}$ is a curve that is tangential to $\mathbf{V}$ at every point along the line. The concept of a field line may be given a mathematical description by writing down the equation for the field line. A convenient starting point is the parametric equations for the field line in cartesian coordinates:

$$\frac{dx}{B_x} = \frac{dy}{B_y} = \frac{dz}{B_z}. \tag{3.5}$$

By way of illustration as to how such parametric equations are used, consider a magnetic field confined to the x-y plane, such that one has $B_z = 0$. Then (3.5) reduces to two equations: $dx/dy = B_x/B_y$ and $dz = 0$. Once B_x and B_y are given as functions of x and y, the first of these is a differential equation that may be integrated to find x as a function of y or to find y as a function of x, or more generally to find a solutions of the form $f(x, y) =$ constant. In this case the equations of the field lines are $f(x, y) =$ constant, $z =$ constant.

The parametric equations (3.5) may be written in terms of any other orthogonal coordinate system. The two most widely used alternatives are spherical

polar coordinates, r, θ, ϕ and cylindrically polar coordinates, ϖ, ϕ, z. The relation between spherical polar and cartesian coordinates is

$$\left(\begin{array}{c} x = r\sin\theta\cos\phi, \\ y = r\sin\theta\sin\phi, \\ z = r\cos\theta, \end{array}\right) \qquad \left(\begin{array}{c} r = (x^2+y^2+z^2)^{1/2}, \\ \theta = \arccos(z/r), \\ \phi = \arctan(y/x), \end{array}\right) \tag{3.6}$$

and the radial cylindrical coordinate is

$$\varpi = r\sin\theta = (x^2+y^2)^{1/2}. \tag{3.7}$$

In terms of these coordinate systems, (3.4) is replaced by

$$\frac{dr}{B_r} = \frac{rd\theta}{B_\theta} = \frac{r\sin\theta d\phi}{B_\phi}, \qquad \frac{d\varpi}{B_\varpi} = \frac{\varpi d\phi}{B_\phi} = \frac{dz}{B_z}, \tag{3.8}$$

respectively.

3.3 Dipolar magnetic field

A magnetic dipole is strictly a moment of a current distribution. It is the lowest moment, the next highest being the octupole, and an arbitrary moment being the 2^l-multipole, with $l = 1$ for the dipole and $l = 2$ for the octupole. To define the multipoles, let $\mathbf{J}(\mathbf{x})$ be a current distribution confined to a volume V, and consider $\mathbf{A}$ in the Coulomb gauge $\nabla\cdot\mathbf{A} = 0$. Then one has $\mu_0\mathbf{J} = \nabla\times\nabla\times\mathbf{A} = -\nabla^2\mathbf{A}$, whose solution is

$$\mathbf{A}(\mathbf{x}) = \frac{\mu_0}{4\pi}\int_V d^3\mathbf{x}'\,\frac{\mathbf{J}(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}. \tag{3.9}$$

The multipole expansion follows from the Taylor series expansion

$$\frac{1}{|\mathbf{x}-\mathbf{x}'|} = \sum_{n=0}^{\infty}\frac{1}{n!}\left(-\mathbf{x}'\cdot\frac{\partial}{\partial\mathbf{x}}\right)^n\frac{1}{r} \tag{3.10}$$

in the integrand in (3.9). For static fields the term $n = 0$ gives zero (there are no magnetic monopoles), as may be shown by noting that one has $\nabla\cdot\mathbf{J} = 0$ for a static field, and hence, for a current confined to a volume V with surface S,

$$0 = \int_V d^3\mathbf{x}'\,\mathbf{x}'\,\nabla\cdot'\,\mathbf{J}(\mathbf{x}') = \int_S d^2\mathbf{x}'\cdot\mathbf{x}'\,\mathbf{J}(\mathbf{x}') - \int_V d^3\mathbf{x}'\,\mathbf{J}(\mathbf{x}'), \tag{3.11}$$

where a partial integration is performed. The surface terms vanishes because the current is assumed to be confined to the volume V, and so vanishes on the surface. The next order term is the magnetic dipole term:

$$\mathbf{A}(\mathbf{x}) = \frac{\mu_0}{4\pi}\,\nabla\times\frac{\mathbf{m}}{r}, \qquad \mathbf{m} = \frac{1}{2}\int_V d^3\mathbf{x}'\,\mathbf{x}'\times\mathbf{J}(\mathbf{x}'). \tag{3.12}$$

A dipolar magnetic field is given by

$$\mathbf{B} = \frac{\mu_0}{4\pi} \nabla \times \nabla \times \frac{\mathbf{m}}{r} = \frac{\mu_0}{4\pi} \nabla\nabla \cdot \frac{\mathbf{m}}{r}, \tag{3.13}$$

where $\nabla^2(1/r) = 0$ for $r \neq 0$ is used. It follows that the magnetic potential, cf. (3.3), is

$$\psi_M = -\frac{\mu_0}{4\pi} \nabla \cdot \frac{\mathbf{m}}{r} = \frac{\mu_0}{4\pi} \frac{\mathbf{m} \cdot \mathbf{x}}{r^3}. \tag{3.14}$$

On evaluating (3.13), one finds

$$\mathbf{B} = \frac{\mu_0}{4\pi} \frac{3\mathbf{x}\, \mathbf{m} \cdot \mathbf{x} - r^2 \mathbf{m}}{r^5}. \tag{3.15}$$

Assuming that $\mathbf{m}$ is along the polar axis, in spherical polar coordinates (3.3) with (3.14) implies

$$(B_r, B_\theta, B_\phi) = -\frac{\mu_0}{4\pi} \left(\frac{\partial}{\partial r}, \frac{1}{r} \frac{\partial}{\partial \theta}, \frac{1}{r \sin\theta} \frac{\partial}{\partial \phi} \right) \frac{m \cos\theta}{r^2} = \frac{\mu_0}{4\pi} \frac{m}{r^3} (2\cos\theta, \sin\theta, 0). \tag{3.16}$$

A dipolar magnetic field of an astrophysical object, such as a planet or star, is often expressed in terms of the polar magnetic field, B_p. If R is the radius of the planet or star, then (3.16) becomes

$$(B_r, B_\theta, B_\phi) = \frac{B_p R^3}{r^3} (\cos\theta, \frac{1}{2}\sin\theta, 0), \qquad B_p = \frac{\mu_0}{4\pi} \frac{2m}{R^3}. \tag{3.17}$$

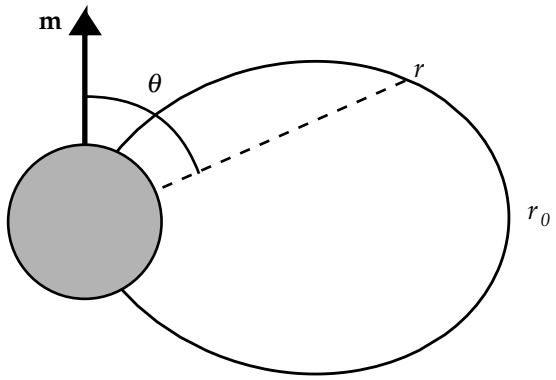


Figure 3.1: A dipolar magnetic field line that satisfies (3.18) is illustrated.

The equations for a dipolar magnetic field line may be found by substituting (3.16) or (3.17) into (3.8). One finds

$$\frac{dr}{2\cos\theta} = \frac{r d\theta}{\sin\theta}, \qquad d\phi = 0,$$

which integrates to

$$r = r_0 \sin^2 \theta, \qquad \phi = \phi_0, \tag{3.18}$$

where r_0 and ϕ_0 are constants of integration. The parameter r_0 may be interpreted as the distance at which the field line crosses the magnetic equator. A dipole field line is illustrated in Figure 3.1. .

3.4 Frozen-in magnetic field lines

In a perfectly conducting fluid, magnetic field lines and fluid move together: one says that the magnetic field is *frozen in* to the fluid. Thus one can give physical meaning to motion of field lines in a perfectly conducting fluid: the velocity of an element of fluid across **B** may be interpreted as the velocity of the field line. This should be contrasted with a magnetic field line in a vacuum or in an insulator, where the concept of a moving magnetic field line is meaningless. This has important physical implications. An example is illustrated in Figure 3.2: the Earth's field lines are frozen into the solid Earth and frozen into the magnetosphere, but are not frozen in to the neutral atmosphere. This allows motions, such as that indicated, which do not change **B** and which would be impossible in a conducting atmosphere like that of a star.

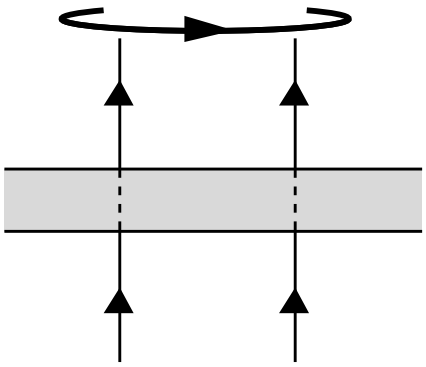


Figure 3.2: An unionized neutral atmosphere, shown shaded, allows the indicated as rotational motion of the frozen-in magnetic field in the magnetosphere independent of any motions in the underlying solid planet.

The proof of the frozen flux theorem is in two parts. First, one derives an equation of motion for a magnetic field line by combining Maxwell's equations with Ohm's law for the fluid. Second, one derives the condition for the magnetic flux threading an element of fluid to remain constant as the fluid moves. Comparing the two results establishes the theorem.

Ohm's law in a conducting fluid may be written in the form $\mathbf{J} = \sigma\mathbf{E}'$ or $\mathbf{E}' = \eta\mathbf{J}$, where σ is the electrical conductivity, $\eta = 1/\sigma$ is the resistivity, and $\mathbf{E}'$ is the electric field in the rest frame of the fluid. If the fluid is moving with velocity $\mathbf{u}$, then one has $\mathbf{E}' = \mathbf{E} + \mathbf{u} \times \mathbf{B}$. Hence, Ohm's law takes the form

$$\mathbf{J} = \sigma(\mathbf{E} + \mathbf{u} \times \mathbf{B}), \tag{3.19}$$

which may be rewritten as

$$\mathbf{E} = -\mathbf{u} \times \mathbf{B} + \mathbf{J}/\sigma. \tag{3.20}$$

On inserting (3.20) in Faraday's equation, $\partial\mathbf{B}/\partial t = -\nabla \times \mathbf{E}$, one obtains the equation of motion for a magnetic field:

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times \mathbf{u} \times \mathbf{B} - \nabla \times \frac{\mathbf{J}}{\sigma}. \tag{3.21}$$

Perfect conductivity corresponds to $\sigma = \infty$, and then the final term in (3.21) is absent.

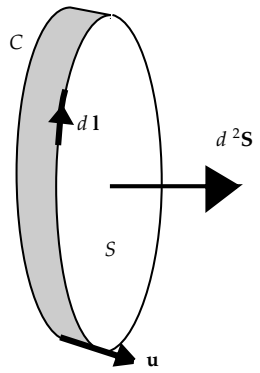


Figure 3.3: A surface C bounding an area S is convected with the fluid at velocity $\mathbf{u}$; $d^2\mathbf{S}$ is an element of surface area, and $d\mathbf{l}$ is an element of the curve C.

The magnetic flux passing through an element $d^2\mathbf{S}$ of a surface is $d\Phi_M = \mathbf{B} \cdot d^2\mathbf{S}$ surface. As the fluid element moves with velocity $\mathbf{u}$, there is an extra area, $\mathbf{u} \times d\mathbf{l}$ swept up per unit time, as illustrated in Figure 3.3, where $d\mathbf{l}$ is an element of length around the boundary of the fluid element. The rate of change of magnetic flux consists of two parts: the change due to a temporal change in $\mathbf{B}$ plus the flux that passes through the additional area swept up. For a macroscopic fluid element one has

$$\frac{d\Phi_M}{dt} = \int_S d^2\mathbf{S} \cdot \frac{\partial \mathbf{B}}{\partial t} - \oint_C d\mathbf{l} \cdot \mathbf{u} \times \mathbf{B}. \tag{3.22}$$

On transforming the final term in (3.22) into a surface integral using Stokes' theorem, one obtains

$$\frac{d\Phi_M}{dt} = \int_S d^2\mathbf{S} \cdot \left(\frac{\partial \mathbf{B}}{\partial t} - \nabla \times \mathbf{u} \times \mathbf{B} \right). \tag{3.23}$$

It follows that the magnetic flux flows with the fluid provided that the integrand in (3.22) vanishes. This corresponds to (3.21) for $\sigma = \infty$, establishing the theorem.

Conversely, in a current-carrying plasma with $\sigma \neq \infty$, (3.21) implies that magnetic field lines have an additional motion. For σ = constant, on using $\nabla \times \mathbf{B} = \mu_0 \mathbf{J}$ and $\sigma = 1/\eta$, (3.21) gives

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times \mathbf{u} \times \mathbf{B} + \mu_0 \eta \nabla^2 \mathbf{B}. \tag{3.24}$$

Thus the additional motion corresponds to a diffusion of the magnetic field lines, with spatial diffusion coefficient $D = \mu_0 \eta = \mu_0 / \sigma$.

3.5 Wandering of magnetic field lines

In the presence of magnetic fluctuations, due to a spectrum of Alfvén waves, the direction of a magnetic field line wanders due to the magnetic field, $\delta\mathbf{B}$, associated with the fluctuations. Consider two particles that start at the same point at two different times. Each particle sees a different detailed spectrum of fluctuations, and so its guiding center follows a slightly different path. Hence, a large number of particles starting from the same point at different times spread out from the mean direction of the magnetic field as a function of distance, z, along the average direction of the magnetic field, as illustrated in Figure 3.4. This effect is referred to as wandering of magnetic field lines.

To treat field line wandering in a quantitative way, one needs a statistical description of the fluctuations. Let the average field be $\mathbf{B}_0$. A statistical average over the fluctuations gives

$$\langle \delta \mathbf{B} \rangle = 0, \qquad \langle \mathbf{B} \rangle = \mathbf{B}_0,$$
$$\langle B_i(z) B_j(z+\zeta) \rangle = \langle \delta B_i(z) \delta B_j(z+\zeta) \rangle = S_{ij}(\zeta), \tag{3.25}$$

where $S_{ij}(\zeta)$ is the correlation of the fluctuations. The spectrum of fluctuations can be measured, usually in terms of the spatial Fourier transform of the correlation function:

$$R_{ij}(k) = \int d\zeta \, e^{ik\zeta} S_{ij}(\zeta). \tag{3.26}$$

In the absence of detailed information on the spectrum, one assumes that the fluctuations are gaussian, $S_{ij}(\zeta) \propto \exp(-\zeta^2/2\zeta_0^2)$, where ζ_0 is a correlation length.

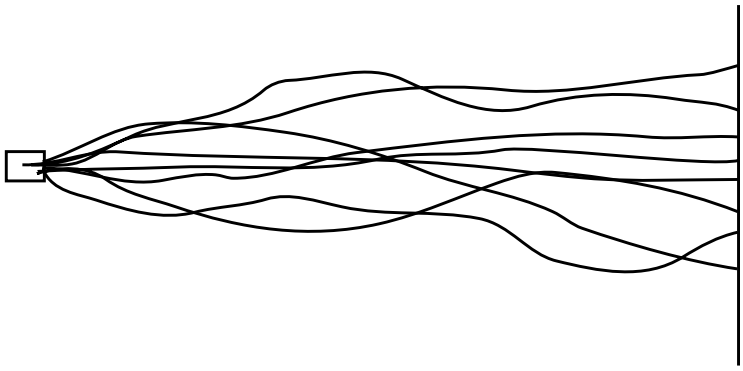


Figure 3.4: Field line wandering due to a spectrum of magnetic fluctuations is illustrated. The lines represent the paths of the guiding centers of particles starting from the origin at different times.

The other ingredient in a theory for field line wandering is a statistical description of the field lines. Let $\mathcal{N}(x, y)$ denote the number of field lines per unit area in the x-y plane. Assuming that the spectrum has cylindrical symmetry about $\mathbf{B}_0$, the field line wandering may be descried in terms of a spatial diffusion of the field lines in the x-y plane:

$$\frac{\partial \mathcal{N}}{\partial z} = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) (D_\perp \mathcal{N}). \tag{3.27}$$

The spatial diffusion coefficient may be evaluated in terms of Fokker-Planck coefficients:

$$D_\perp = \frac{1}{2} \left\langle \frac{(\Delta x)^2}{\Delta z} \right\rangle = \frac{1}{2} \left\langle \frac{(\Delta y)^2}{\Delta z} \right\rangle . \tag{3.28}$$

These coefficient are found by integrating the parametric equations (3.5), with $B_z = B_0$ here, in the presence of the fluctuations:

$$\Delta x = \frac{1}{B_0} \int_0^{\Delta z} dz\, \delta B_x(z), \qquad (\Delta x)^2 = \frac{1}{B_0^2} \int_0^{\Delta z} dz \int_0^{\Delta z} dz'\, \langle \delta B_x(z) \delta B_x(z') \rangle . \tag{3.29}$$

Provided that Δz is much larger than the correlation length ζ_0, the integrals in (3.29) may be rewritten in terms of integrals over $z + z'$ and $z - z'$, and the latter extended to infinity and performed using (3.26). This gives

$$\left\langle \frac{(\Delta x)^2}{\Delta z} \right\rangle = \frac{1}{B_0^2} \int d\zeta\, S_{xx}(\zeta) = \frac{R_{xx}(0)}{B_0^2}. \tag{3.30}$$

Thus the spatial diffusion coefficient in (3.27) is determined by the long-wavelength ($k \to 0$) limit of the spectrum of fluctuations.

Applied to cosmic rays or to energetic particles in any astrophysical magnetic field, the wandering of field lines implies that particles diffuse slowly across the field lines. In practice this turns out to be relatively unimportant in comparison to diffusion along the field lines, which is due to a different physical effect called resonant scattering.

Lecture 4

Magnetohydrodynamics

The most general theoretical description of a plasma is in terms of kinetic theory. However, for many purposes a fluid description suffices, and the fluid theory for a magnetized plasma is *magnetohydrodynamics*, usually abbreviated to MHD. Here the derivation of the MHD equations from the kinetic theory is indicated and then some applications of MHD theory to astrophysical plasmas are discussed briefly.

4.1 Derivation of the MHD equations

A fully ionized astrophysical plasma consists of distributions of electrons and various species of positively charged ions. Let an arbitrary species of particles be denoted by α, with $\alpha = e$ for electrons $\alpha = i$ for ions and $\alpha = i = p$ for protons. Particles of species α have mass m_α and charge q_α, with $q_e = -e$, $q_i = Z_i e$. The distribution function $f_\alpha(\mathbf{x}, \mathbf{p})$ for species α satisfies Boltzmann's equation

$$\left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \frac{\partial}{\partial \mathbf{x}} + \mathbf{F}_\alpha \cdot \frac{\partial}{\partial \mathbf{p}}\right) f_\alpha = \left(\frac{\partial f_\alpha}{\partial t}\right)_{\text{coll}}, \tag{4.1}$$

where $\mathbf{F}_\alpha$ is the force acting on a particle of species α, and where the right hand term describes the effect of collisions on particles of species α. The MHD equations are derived by taking moments of (4.1). The formal derivation is only presented in outline here.

The distributions are assumed to be normalized according to

$$\int d^3\mathbf{p}\, f_\alpha = n_\alpha, \qquad \int d^3\mathbf{p}\, K_\alpha\, f_\alpha = n_\alpha \langle K_\alpha \rangle . \tag{4.2}$$

where n_α is the number density for species α, and where $\langle K_\alpha \rangle$ is the moment of the quantity K_α. The zeroth moment of (4.1) for species α corresponds to $K_\alpha = 1$ in (4.2). On multiplying by m_α and summing over all α the zeroth moment

of the distribution gives the mass density, η, and similarly, on multiplying by q_α and summing over all α, (4.2) gives the charge density, ρ. The form of the collisional term is such that its zeroth moment is identically zero, and using this property the zeroth moment of (4.1) gives equations for the evolution of the mass and charge density. These are

$$\frac{\partial \eta}{\partial t} + \nabla \cdot (\eta \mathbf{u}) = 0, \tag{4.3}$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0, \tag{4.4}$$

where $\mathbf{u}$ is the fluid velocity, and $\mathbf{J}$ is the electric current density, which are defined by

$$\sum_\alpha m_\alpha n_\alpha \langle \mathbf{v}_\alpha \rangle = \eta \mathbf{u}, \qquad \mathbf{J} = \sum_\alpha q_\alpha n_\alpha \langle \mathbf{v}_\alpha \rangle . \tag{4.5}$$

In most cases, the plasma is charge-neutral, because if this were not so to a high approximation, the electric fields due to the charges would exert enormous unbalanced forces on the plasma. In MHD theory one usually assumes $\rho = 0$, so that (4.4) reduces to $\nabla \cdot \mathbf{J} = 0$.

Both (4.3) and (4.4) are examples of *continuity equations*, which express a conservation law. They express conservation of mass and charge, respectively. More generally, a continuity for a quantity Q is of the form $\partial(\text{density of } Q)/\partial t + \nabla \cdot (\text{flux of } Q) = (\text{source of } Q)$. There are no sources of mass and charge, so that there is no source term in either (4.3) or (4.4).

The first moment of (4.1), which involves setting $K_\alpha = \mathbf{v}$ in (4.2) leads to two further equations. One, obtained by multiplying by m_α and summing over α, is the equation of motion for the fluid, and the other, obtained by multiplying by q_α and summing over α, is the generalized Ohm's law. Let us first suppose that the system is not changing as a function of time, that is, let us consider the static case. Then the first term in (4.1) does not contribute. The right hand term may also be neglected when considering the momentum equation; this is because collisions lead to a change of momentum between different species of particles, but no net loss of momentum by the plasma, so that the frictional forces between different species cancel each other. The middle term on the left hand side of (4.1) has a second moment of the form $\partial(n_\alpha \langle v_i v_j \rangle)/\partial x_j$. In the static case the fluid velocity is zero, and one has $\langle v_i v_j \rangle = \frac{1}{3}\delta_{ij} \langle v^2 \rangle$, with $\langle v^2 \rangle = 3\kappa_B T_\alpha / m_\alpha$ for a thermal distribution with temperature T_α, where κ_B is Boltzmann's constant. Hence this term may be expressed in terms of the pressure gradient, where the partial pressure due to species α is $P_\alpha = n_\alpha \kappa_B T_\alpha$, and the total pressure is $P = \sum_\alpha P_\alpha$. The final term on the right hand side depends on the forces acting. If the total force on particles of species α is

$$\mathbf{F}_\alpha = q_\alpha(\mathbf{E} + \mathbf{v} \times \mathbf{B}) + m_\alpha \mathbf{g}, \tag{4.6}$$

then the momentum equation leads to the following equation for magnetohydrostatic equilibrium:

$$\nabla P = \rho\mathbf{E} + \mathbf{J} \times \mathbf{B} + \eta\mathbf{g}. \tag{4.7}$$

As already remarked, one can usually assume that the charge density ρ is negligible.

The equation of motion for the fluid is

$$\eta \left[\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u} \right] = -\nabla P + \rho\mathbf{E} + \mathbf{J} \times \mathbf{B} + \eta\mathbf{g}. \tag{4.8}$$

The right hand side is the force per unit volume. Note that the fluid velocity, $\mathbf{u}$, is an Eulerian velocity, so that it is a function of both space and time. Its total time-derivative is $d\mathbf{u}/dt = \partial\mathbf{u}/\partial t + (dx_i/dt) \cdot (\partial\mathbf{u}/\partial x_i)$, with $d\mathbf{x}/dt = \mathbf{u}$ by definition. The term $(\mathbf{u} \cdot \nabla)\mathbf{u}$ is called the convective derivative. It is strictly absent in a Lagrangian approach, such as Newton's equation of motion for a single particle, because $\mathbf{x}$ and $\mathbf{v}$ are then independent variables by definition, so that $\partial\mathbf{v}/\partial x_i$ is zero by definition. In the Eulerian (fluid) approach, $\mathbf{u}$ is the velocity of a fluid element, and not the velocity of the individual particles that constitute that fluid element at any one time.

The derivation of Ohm's law is similar to the derivation of the momentum equation. The general form is unnecessarily cumbersome for most purposes. The simplest useful form is derived by adopting the frame in which the ions are at rest, so that the current is carried entirely by the electrons, and assuming that the collisional term is of the form

$$\left(\frac{\partial f_e}{\partial t} \right)_{\text{coll}} = -\nu_e f_e, \tag{4.9}$$

where ν_e is the collision frequency for electrons. On multiplying (4.1) for electrons by $-en_e$ and integrating over momentum for a Maxwellian distribution drifting at $\mathbf{u}_e = \langle \mathbf{v}_e \rangle$, one identifies $\mathbf{J} = -en_e\mathbf{u}_e$, where $\mathbf{u}_e$ is the fluid velocity of the electrons. This gives

$$\frac{\partial \mathbf{J}}{\partial t} - \frac{e}{m_e}\nabla P_e + \frac{e^2 n_e}{m_e}(\mathbf{E}' + \mathbf{u}_e \times \mathbf{B}) = -\nu_e \mathbf{J}, \tag{4.10}$$

with $\mathbf{E}' = \mathbf{E} + \mathbf{u} \times \mathbf{B}$, where $\mathbf{E}$ is the electric field in the laboratory frame, and where the relative velocity of the laboratory frame and the rest frame of the ions is approximated by $\mathbf{u}$. The resistivity and the conductivity are identified by writing

$$\eta = \frac{\nu_e m_e}{e^2 n_e} = \frac{1}{\sigma}, \tag{4.11}$$

and then (4.10) may be written

$$\mathbf{J} + \frac{\sigma m_e}{e^2 n_e}\frac{\partial \mathbf{J}}{\partial t} + \frac{\sigma}{en_e}\mathbf{J} \times \mathbf{B} = \sigma \left(\mathbf{E} + \mathbf{u} \times \mathbf{B} + \frac{\nabla P_e}{en_e} \right). \tag{4.12}$$

For most astrophysical purposes, (4.12) is approximated by $\mathbf{J} = \sigma(\mathbf{E} + \mathbf{u} \times \mathbf{B})$, which is often simplified further to $\mathbf{E}+\mathbf{u}\times\mathbf{B} = 0$ in the perfectly conducting limit $\sigma = \infty$. The electrical conductivity σ is an example of a transport coefficient in a plasma.

Higher moments of (4.1) lead to further transport equations. The equation for heat transport is important in some astrophysical applications. However, the discussion of such transport processes is usually not considered as part of MHD theory, but as a complement to it. As with any set of moment equations, one needs to cut off the expansion (here in powers of v) at some order, and the lowest nontrivial order for $\sigma = \infty$ corresponds to (4.3), (4.4), (4.8) and (4.12).

In addition to these equations, the MHD equations include two further sets of equations. One is Maxwell's equations

$$\nabla \times \mathbf{E} = -\partial \mathbf{B}/\partial t, \tag{4.13a}$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + (1/c^2)\, \partial \mathbf{E}/\partial t, \tag{4.13b}$$

$$\nabla \cdot \mathbf{E} = \rho/\varepsilon_0, \tag{4.13c}$$

$$\nabla \cdot \mathbf{B} = 0, \tag{4.13d}$$

with the displacement current usually omitted from (4.13b). It is conventional in MHD theory to eliminate the electric field using Ohm's law in the form $\mathbf{J} = \sigma(\mathbf{E} + \mathbf{u} \times \mathbf{B})$. Then (4.13a) reduces to the *induction equation*, written down in (4.14) below. One regards $\mathbf{J} = \nabla \times \mathbf{B}/\mu_0$ as a subsidiary vector, so that (4.13b) is no longer relevant. In addition, ρ is assumed to be negligibly small, so that (4.4) reduces to $\nabla \cdot \mathbf{J} = 0$, which follows trivially from $\mathbf{J} = \nabla \times \mathbf{B}/\mu_0$. Now $\mathbf{E}$ is eliminated and ρ is neglected, so that (4.13c) is not relevant. Hence, our MHD equations reduce to the induction equation with (4.3) and (4.8):

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B}) + \eta \nabla^2 \mathbf{B}, \tag{4.14}$$

$$\frac{\partial \eta}{\partial t} + \nabla \cdot (\eta \mathbf{u}) = 0, \tag{4.15}$$

$$\eta \left[\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u} \right] = -\nabla P + \mathbf{J} \times \mathbf{B} + \mathbf{f}, \tag{4.16}$$

where $\mathbf{f}$ is the force per unit volume on the fluid. For $\eta = 0$, (4.14) describes a frozen-in magnetic field that is convected along with the fluid, and for $\mathbf{u} = 0$ (4.14) describes diffusion of magnetic field lines.

The status of (4.13d) is now effectively reduced to a boundary condition. If one has $\nabla \cdot \mathbf{B} = 0$ everywhere initially, then the divergence of (4.14) implies that $\nabla \cdot \mathbf{B}$ remains zero everywhere thereafter. Thus (4.13d) is regarded as a constraint on an acceptable initial $\mathbf{B}$.

The set of equations (4.14)–(4.16) involves two vector equations and one scalar equation, but we have two vectorial variables ($\mathbf{B}$, $\mathbf{u}$) and two scalar

variables (η, P), so that we need one more equation to close our set. The additional equation is an equation of state for the fluid. Three alternative equations of state are used in different contexts: the incompressible, isothermal and adiabatic equations of state. The incompressible conditions corresponds to $\eta =$ constant, and then (4.3) implies $\nabla \cdot \mathbf{u} = 0$. The isothermal equation of state is $T =$ constant. The adiabatic equation is $P \propto \eta^\Gamma$, with Γ the adiabatic index, which has the specific value $\Gamma = 5/3$ for a monatomic gas. Thus one has

$$\nabla \cdot \mathbf{u} = 0, \qquad \text{(incompressible)}, \tag{4.17a}$$

$$T = \text{constant}, \qquad \text{(isothermal)}, \tag{4.17b}$$

$$P \propto \eta^\Gamma, \qquad \text{(adiabatic)}. \tag{4.17c}$$

The incompressible equation of state (with $\nabla P = 0$) is appropriate when MHD theory is applied to a conducting liquid, such as a column of mercury. It is sometimes useful for gaseous plasmas, specifically when the effects of interest are known to involve no compressions, as in an Alfvén wave. The isothermal equation of state is appropriate for processes that occur on such a long time scale that there is time for thermal conduction to maintain a constant temperature. This condition is not satisfied in most applications of interest in this lecture course. In the following discussion the adiabatic equation (4.17c) is assumed to apply, except where stated otherwise.

4.2 Energy density and energy flow

An additional equation that may be derived from the foregoing equations is the continuity equation for energy. First note that from Maxwell's equations alone, one has the continuity equation

$$\frac{\partial}{\partial t}\left(\tfrac{1}{2}\varepsilon_0|\mathbf{E}|^2 + \tfrac{1}{2}|\mathbf{B}|^2/\mu_0\right) + \nabla\cdot(\mathbf{E}\times\mathbf{B}/\mu_0) = -\mathbf{J}\cdot\mathbf{E}. \tag{4.18}$$

The terms inside the derivatives on the left hand side are interpreted as the electric energy density (usually neglected in MHD theory), the magnetic energy density and the electromagnetic energy flux (the Poynting vector), respectively. The right hand term gives the power input per unit volume into the electromagnetic field.

In a perfectly conducting medium, such that one has $\mathbf{E} + \mathbf{u}\times\mathbf{B} = 0$, and with only electromagnetic forces acting, (4.18) implies the equation of energy continuity in MHD theory:

$$\frac{\partial}{\partial t}\left(\frac{1}{2}\eta u^2 + \frac{P}{\Gamma-1} + \frac{B^2}{2\mu_0} + \tfrac{1}{2}\varepsilon_0 E^2\right) + \nabla\cdot\left(\frac{1}{2}\eta u^2\mathbf{u} + \frac{\Gamma}{\Gamma-1}P\mathbf{u} + \frac{1}{\mu_0}\mathbf{E}\times\mathbf{B}\right) = 0. \tag{4.19}$$

The three terms in the energy density are the kinetic energy density, the thermal energy density (the internal energy), the magnetic energy density and the electric energy density, respectively. The three terms in the energy flux are the kinetic energy flux, the thermal energy flux ($(\Gamma/(\Gamma-1)P$ is the enthalpy) and the the Poynting vector, respectively.

The derivation of (4.19) from (4.18) proceeds as follows. The power input per unit volume into the plasma may be obtained from the work done per unit volume on the plasma, which follows from $\mathbf{u}\cdot$(4.8), with the terms $\rho\mathbf{E}$ and $\eta\mathbf{g}$ omitted here. The term $\mathbf{u}\cdot\mathbf{J}\times\mathbf{B}$ may be rewritten as $-\mathbf{J}\cdot\mathbf{u}\times\mathbf{B}=\mathbf{J}\cdot\mathbf{E}$. This term is identified as the source term for the electromagnetic field, and is substituted for the right hand term in (4.18). After rearrangements, this leads to (4.19).

Energy flow in astrophysical plasmas can take several familiar forms: conduction, convection and radiation. However, there are three additional forms of energy propagation, included in (4.19), that are important in different contexts. One is energy transport through a flux of particles, which corresponds to the flux of kinetic energy in (4.19). However, such energy flow is not usually treated as an MHD process. A second form of energy propagation is in terms of Alfvén and other MHD waves, which are discussed in the next lecture. The third form of energy propagation is as a Poynting flux involving static fields.

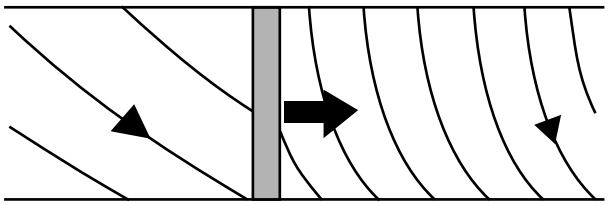


Figure 4.1: A front (shaded) region is shown propagating to the right (solid arrow) into a twisted cylindrical magnetic flux tube; at the front the twist (B_ϕ) in the magnetic field is reduced releasing magnetic energy. Behind the front the plasma is rotating, corresponding to the unwinding motion (u_ϕ), and there is a Poynting flux ($u_\phi B_\phi B_z/\mu_0$) directed to the right transporting the energy released away from the front.

Let us see how energy propagation occurs through a Poynting vector in the static limit. Rewriting the Poynting vector using $\mathbf{E}+\mathbf{u}\times\mathbf{B}=0$ gives

$$\frac{1}{\mu_0}\mathbf{E}\times\mathbf{B}=-\frac{1}{\mu_0}(\mathbf{u}\times\mathbf{B})\times\mathbf{B}=\frac{1}{\mu_0}[B^2\mathbf{u}-(\mathbf{u}\cdot\mathbf{B})\mathbf{B}]. \qquad (4.20)$$

The right hand side includes one term, $(B^2/2\mu_0)\mathbf{u}$, that has a simple interpretation: it corresponds to magnetic energy density being transported at the

fluid velocity. However, clearly this is not the entire contribution. It follows that the Poynting vector in MHD is not just the flux of magnetic energy at the fluid velocity. An example of another kind of energy flux is illustrated in Figure 4.1: an unwinding motion can release magnetic energy that goes into a Poynting vector. In this case the energy flow is along the direction of the mean background magnetic field and is orthogonal to the fluid velocity. This type of energy propagation is included in (4.20) but cannot be interpreted as magnetic energy being transported at the fluid velocity. There is no subtlety here: the Poynting vector (4.20) implies that a form of energy propagation exists that is unfamiliar. Energy propagation in this way is involved in solar flares and in other applications where magnetic energy is stored and released.

4.3 Magnetic pressure and magnetic tension

The force (per unit volume) $\mathbf{J} \times \mathbf{B}$ in the equation of fluid motion (4.8) is the due to the magnetic stresses, sometimes called the Maxwell stress. Using $\mathbf{J} = \nabla \times \mathbf{B}/\mu_0$ and a vector identity, it may be written in the form

$$\mathbf{J} \times \mathbf{B} = (\nabla \times \mathbf{B}) \times \mathbf{B}/\mu_0 = -\nabla(B^2/2\mu_0) + (\mathbf{B} \cdot \nabla)\mathbf{B}/\mu_0. \tag{4.21}$$

The two terms in (4.21) may be interpreted as a magnetic pressure and a magnetic tension, respectively.

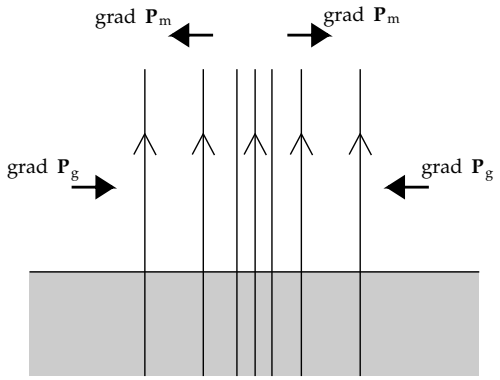


Figure 4.2: A vertical magnetic flux tube above a stellar surface is illustrated schematically. The magnetic pressure, P_m, has a gradient directed away from the center of the flux tube, and is balanced by a gradient in the gas pressure, P_g, toward the center of the flux tube.

The pressure is isotropic, that is, it acts equally in all directions, just like the pressure in a gas. The force due to a gradient in the magnetic pressure tends to push plasma from regions of high magnetic energy density to regions

of low magnetic energy density. The total pressure in a gas is the sum of the magnetic and gas pressures. Consider, for example, a vertical magnetic flux tube on the surface of a star, as illustrated in Figure 4.2. Gravity and the magnetic tension then both act in the vertical direction. Consider pressure balance in the horizontal direction. Pressure balance requires $P + B^2/2\mu_0 =$ constant in any horizontal plane. Hence the gas pressure, $P = P_{\rm out}$, outside the flux tube must be greater than the gas pressure, $P = P_{\rm in}$, inside the flux tube by an amount equal to the magnetic pressure: $P_{\rm out} - P_{\rm in} = B^2/2\mu_0$. This difference in gas pressure may be due to a difference in temperature or to a difference in density, or to a combination of the two. For example, sunspots are regions of strong magnetic field ($B \approx 0.15\,{\rm T}$) on the solar surface, and one might speculate that they are darker than their surroundings because of the gas pressure deficiency within the region of strong magnetic field. In one sense this must be the correct explanation for why sunspots are dark, but a detailed justification is much more complicated than this simple suggestion would indicate.

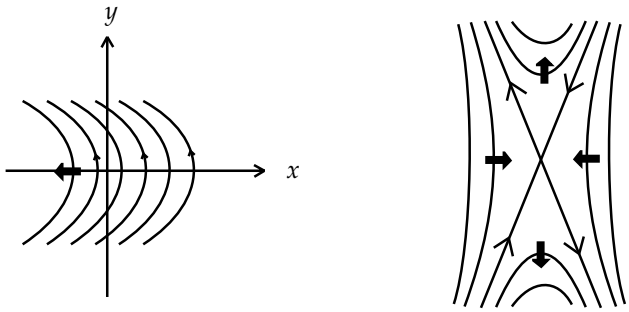
Magnetic tension is closely analogous to the tension in a stretched string. Thus, the effect of the tension is to tend to reduce the length of the field line. The tension force may be evaluated in terms of two unit vectors: $\mathbf{b} = \mathbf{B}/B$ along the magnetic field lines, and $\mathbf{n}$ pointing towards the center of curvature of a curved field line, and defined by writing $\mathbf{b}\cdot\nabla\mathbf{b} = \mathbf{n}/R_c$, where R_c is the radius of curvature of the field line, cf. Figure 2.2b. The final term in (4.21) includes the magnetic tension. This term may be written in the form

$$\frac{1}{\mu_0}(\mathbf{B}\cdot\nabla)\mathbf{B} = \frac{B}{\mu_0}\mathbf{b}\cdot\nabla(B\mathbf{b}) = \mathbf{b}\,\mathbf{b}\cdot\nabla\left(\frac{B^2}{2\mu_0}\right) + \frac{B^2}{\mu_0}\,\frac{\mathbf{n}}{R_c}. \tag{4.22}$$

The first term on the right hand side of (4.22) is equal but opposite to the magnetic pressure gradient in (4.21) along the field lines. Thus (4.21) and (4.22) imply that the magnetic tension has a component along the field line which exactly cancels the pressure gradient along $\mathbf{B}$. The other term in (4.22) implies a tension force directed toward the center of curvature of the field line.

An example of the tension force for a curved field line is illustrated in Figure 4.3a. The tension force on the plasma is like that in a slingshot, forcing the plasma in the direction that tends to reduce the length of the field lines. The particular field illustrated in Figure 4.3a corresponds to $\mathbf{B} = -\alpha y\hat{\mathbf{x}} + \beta\hat{\mathbf{y}}$, where α and β are constants. This corresponds to a current along the z-direction: $\mathbf{J} = (\alpha/\mu_0)\hat{\mathbf{z}}$. The field lines are then the surfaces $\beta x + \frac{1}{2}\alpha y^2 =$ constant, $z =$ constant. In this case the magnetic pressure $P_{\rm m} = (\alpha^2 y^2 + \beta^2)/2\mu_0$ has a gradient only in the y-direction, and the magnetic tension balances this force in the y-direction and produces a net tension force along the negative x-axis, given by $\mathbf{J}\times\mathbf{B} = -(\alpha\beta/\mu_0)\hat{\mathbf{x}}$.

Another example is illustrated in Figure 4.3b. This corresponds to a magnetic field $\mathbf{B} = ay\hat{\mathbf{x}} + bx\hat{\mathbf{y}}$, where a and b are constants. The analysis of this example is left as an exercise. Physically this example may be used to illustrate



(a) (b)

Figure 4.3: (a) The magnetic field lines for the field described in the text are illustrated; the solid arrow gives the direction of the net tension force. (b) A magnetic configuration containing an X-type neutral point is illustrated.

magnetic reconnection: the central point in Figure 4.3b is an X-type magnetic neutral point where magnetic field lines on opposite sides of a vertical line through the center can drift toward the vertical line, reconnect and drift out as field lines on opposite sides of a horizontal line through the center. The net tension force is in the directions indicated, favoring this reconnection. Magnetic reconnection is widely believed to be the basis of the energy release in solar and stellar flares, and in other contexts where magnetic energy release occurs explosively in astrophysical plasmas.

4.4 MHD equilibrium

An MHD configuration is said to be in equilibrium if the forces on it are in balance. In most cases this corresponds to

$$-\nabla P + \mathbf{J} \times \mathbf{B} + \eta \mathbf{g} = 0. \tag{4.23}$$

In modeling a magnetic structure one seeks to find an appropriate solution of (4.23). Examples include a magnetic loop on the Sun, a magnetically supported interstellar cloud, and a magnetically confined astrophysical jet. However, a structure that satisfies (4.23) is not necessarily stable. As in mechanics, an equilibrium may be either stable or unstable. There is a large variety of MHD instabilities. Many of these are given descriptive names, such as the kink instability, the flute instability, the interchange instability, the sausage instability, the disruptive instability, the sawtooth instability and the ballooning instability. Such instabilities have been investigated in detail in connection with laboratory plasmas, where the essential problem is to confine the plasma, and so to prevent

instabilities from occurring. Plasmas tend to be unstable to the simplest and fastest growing instabilities when the magnetic field confining the plasma has convex field lines as viewed from the plasma. Concave field lines tend to lead to a more stable configuration. Twisting of field lines tends to make a configuration more stable, basically by complicating the topology and so eliminating instabilities that rely on a simple geometric structure.

The situation with astrophysical plasmas is quite different from laboratory plasmas. We see the astrophysical structures, and so we know that they must be stable in some meaningful sense. The problem is to explain why the structure is stable, or rather to infer from the incomplete astrophysical data what the structure might be, granted that it must be relatively stable. Another type of problem that arises in connection with the stability of astrophysical structures is how to account for systems that clearly become unstable. A notable case is that of solar magnetic flux tubes, forming coronal filaments or prominences, becoming unstable and leading to the eruption of the filament or prominence. An eruption involves the filament or prominence lifting off from the Sun and moving outward through the solar corona and the solar wind. There is a vast literature on MHD instabilities in general, and of MHD stability in astrophysical plasmas in particular. However, this topic is not discussed in detail in this lecture course.

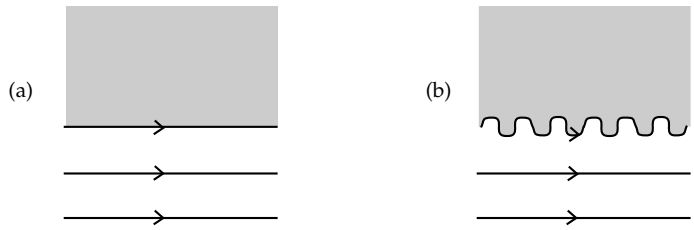


Figure 4.4: (a) A plasma supported against gravity by a magnetic field is subject to the Rayleigh–Taylor instability, analogous to a denser liquid supported by a less dense liquid. (b) The instability develops through ripples at the interface that grow without limit.

Two specific examples of an unstable equilibrium that should be mentioned are the Kelvin–Helmholtz instability and the Rayleigh–Taylor instability. The Kelvin–Helmholtz instability occurs in a boundary layer between two media in relative motion, and is an important source of MHD turbulence. The simplest example of a Kelvin–Helmholtz instability is that of wind over water generating water waves (which are surface gravity waves). The *Rayleigh–Taylor instability* is illustrated in Figure 4.4. The simplest example of such an instability occurs when a less dense fluid (e.g., oil) supports a more dense fluid (e.g., water). The final result of the instability is that buoyancy causes the less dense fluid to flow to above the more dense fluid, and the more dense fluid to sink below the less

dense fluid.

There are several classes of so-called ideal MHD instabilities, which exist in the absence of any dissipative processes such as electrical and thermal conductivity. Even when an MHD structure is stable in a nondissipative medium, it may be unstable when dissipative effects are taken into account. Such instabilities are referred to as resistive instabilities. The important feature that is permitted when dissipation is included is a relative motion between the fluid and the magnetic field lines, cf. (4.14). In particular this allows the magnetic topology to change, which is the magnetic reconnection discussed in connection with Figure 4.3b.

In some astrophysical applications the pressure and gravity forces may be neglected to a first approximation, and then (4.23) requires $\mathbf{J} \times \mathbf{B} = 0$. A magnetic structure satisfying this condition is said to be *force free*. Thus in a force-free configuration, currents flow along magnetic field lines. Somewhere the current much close by flowing across field lines, and hence a force-free structure must be bounded by a non-force-free region where the currents can close.

Lecture 5

MHD waves

Plasmas support a large variety of different types of wave motion. One refers to waves in specific *wave modes*. Low frequency waves in plasmas may be treated using MHD theory, and hence are called MHD waves.

5.1 Alfvén waves

Alfvén waves are shear waves that rely on the tension of the magnetic field lines. They are closely analogous to waves on a stretched string. Alfvén waves exist in both compressible and incompressible fluids; they do not involve any compression and hence their properties are independent of the equation of state assumed for the fluid. As a consequence, it is simplest to treat them by making the incompressible assumption.

Consider the following set of ideal MHD equations:

$$\eta \frac{d\mathbf{u}}{dt} = \mathbf{J} \times \mathbf{B}, \qquad \mathbf{E} + \mathbf{u} \times \mathbf{B} = 0, \tag{5.1a,b}$$

$$\nabla \cdot \mathbf{u} = 0, \qquad \nabla \cdot \mathbf{B} = 0, \tag{5.1c,d}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \qquad \nabla \times \mathbf{B} = \mu_0 \mathbf{J}. \tag{5.1e,f}$$

In treating small-amplitude waves, one linearizes these equations keeping only terms linear in the amplitude of the wave, and then one seeks solutions that oscillate as a plane wave, that is, solutions that vary in space and time as $\exp[-i(\omega t - \mathbf{k} \cdot \mathbf{x})]$.

It is assumed that in the absence of the waves, the background magnetic field, $\mathbf{B}_0$, and the background plasma density, η_0, are uniform and that there is no current, no fluid motion and no electric field. The terms linear in the wave amplitude are $\mathbf{B}_1$, η_1, $\mathbf{J}_1$, and $\mathbf{E}_1$. Now linearize (5.1) and assume a solution

$\propto \exp[-i(\omega t - \mathbf{k}\cdot\mathbf{x})]$, so that one makes the replacements $\partial/\partial t \to -i\omega$, $\nabla \to i\mathbf{k}$. This gives

$$-i\omega\eta_0\mathbf{u}_1 = \mathbf{J}_1 \times \mathbf{B}_0, \qquad \mathbf{E}_1 + \mathbf{u}_1 \times \mathbf{B}_0 = 0, \tag{5.2a,b}$$

$$\mathbf{k}\cdot\mathbf{u}_1 = 0, \qquad \mathbf{k}\cdot\mathbf{B}_1 = 0, \tag{5.2c,d}$$

$$\mathbf{k}\times\mathbf{E}_1 = \omega\mathbf{B}_1, \qquad i\mathbf{k}\times\mathbf{B}_1 = \mu_0\mathbf{J}_1. \tag{5.2e,f}$$

A convenient way of analysing (5.2) is in terms of the fluid displacement, ξ_A, which is defined such that its time derivative is the fluid velocity. That is, one writes $\mathbf{u} = -i\omega\xi_A$. One finds a solution of the set of equations (5.2) only for

$$\kappa\cdot\xi_A = 0, \qquad \mathbf{b}\cdot\xi_A = 0, \qquad \xi_A \propto \kappa\times\mathbf{b}. \tag{5.3}$$

The equations then imply the *dispersion relation*

$$\omega^2 = k^2 v_A^2 \cos^2\theta, \tag{5.4}$$

where the notations

$$v_A^2 = \frac{B_0^2}{\eta_0\mu_0}, \qquad \kappa = \frac{\mathbf{k}}{k}, \qquad \mathbf{b} = \frac{\mathbf{B}_0}{B_0}, \qquad \kappa\cdot\mathbf{b} = \cos\theta \tag{5.5}$$

are introduced. The quantity v_A is the *Alfvén speed*, κ is the *wave normal direction*, and θ is the *wave angle*. The other quantities in (5.2) are given by

$$\mathbf{u}_1 = -i\omega\xi_A \qquad \mathbf{E}_1 = i\omega B_0 \xi_A \times \mathbf{b},$$

$$\mathbf{B}_1 = ikB_0\cos\theta\,\xi_A, \qquad \mathbf{J}_1 = -\frac{k^2 B_0\cos\theta}{\mu_0}\,\kappa\times\xi_A. \tag{5.6}$$

Note that in an Alfvén wave, the fluid velocity, the electric field, the magnetic fluctuation and the current are all in the plane orthogonal to the background magnetic field. The wave motion in an Alfvén wave may be attributed to an interplay between magnetic tension and plasma inertia. When fluid is displaced perpendicular to $\mathbf{B}_0$, the magnetic field is displaced with the fluid, the field line becomes locally curved, and this leads to a tension tending to straighten the field line by pulling the fluid element back toward the unperturbed configuration. The inertia of the plasma causes it to overshoot, setting up an oscillatory motion. The speed of wave propagation may be understood by analogy with the speed of energy propagation for a stretched string, which is [(tension)/(mass per unit length)]$^{1/2}$, which may be rewritten as [(tension per unit area)/(mass per unit volume)]$^{1/2}$. The magnetic tension per unit area is B_0^2/μ_0 and the mass per unit volume is η_0, implying that the wave speed is v_A, as given by (5.5).

5.2 Sound and magnetoacoustic waves

In the absence of a magnetic field, the waves that can exist in a gas are sound waves. If the gas pressure is much greater than the magnetic pressure, that is, in a high-β plasma, with

$$\beta = \frac{P}{B^2/2\mu_0}, \tag{5.7}$$

then the wave properties may be approximated by those of sound waves in an unmagnetized gas. On the other hand, in a low-β plasma, the sound-like waves are magnetoacoustic in character. Before considering the general case let us discuss these two limiting cases separately.

In a high-β plasma, the relevant hydrodynamic equations are those that do not involve the electromagnetic field, together with the adiabatic equation of state:

$$\eta\frac{d\mathbf{u}}{dt} = -\nabla P, \qquad \frac{\partial\eta}{\partial t} + \nabla\cdot(\eta\mathbf{u}) = 0, \qquad P \propto \eta^{\Gamma}. \tag{5.8a,b,c}$$

On linearizing, expressing $\mathbf{u}$ as the time-derivative of the fluid displacement ξ_S, and assuming variation $\propto \exp[-i(\omega t - \mathbf{k}\cdot\mathbf{x})]$, one finds

$$-\omega^2\eta_0\xi_S = -i\mathbf{k}P_1, \qquad \eta_1 = \eta_0 i\mathbf{k}\cdot\xi_S, \qquad \frac{P_1}{P_0} = \Gamma\frac{\eta_1}{\eta_0}. \tag{5.9a,b,c}$$

A solution exists only for

$$\xi_S \propto \kappa, \qquad P_1 \propto \eta_1 \propto \mathbf{k}\cdot\xi_S. \tag{5.10}$$

Then the equations imply the dispersion relation

$$\omega^2 = k^2c_s^2, \qquad c_s^2 = \Gamma P_0/\eta_0, \tag{5.11}$$

where c_s is the (adiabatic) sound speed. These are the familiar sound waves, as in an unionized gas.

In a low-β plasma, the properties of magnetoacoustic waves may be derived from the set of equations (5.1) used to treat Alfvén waves, but with the incompressible condition relaxed. Consider the set (5.2), with (5.2c) omitted. On writing $\mathbf{u} = -i\omega\xi$ in (5.2a), one infers that all solutions must have $\mathbf{b}\cdot\xi = 0$. According to (5.3), Alfvén waves have $\xi_A \propto \kappa\times\mathbf{b}$. There is another possible solution that is orthogonal to this, specifically, one with $\xi = \xi_M$,

$$\xi_M \propto \kappa - \kappa\cdot\mathbf{b}\,\mathbf{b}. \tag{5.12}$$

The solution corresponding to (5.12) describes magnetoacoustic waves, which satisfy the dispersion relation

$$\omega^2 = k^2v_A^2. \tag{5.13}$$

Unlike Alfvén waves, these waves involve a magnetic compression; that is, $\mathbf{B}_1 = i\omega\times(\xi_M\times\mathbf{B}_0)$ has a component along $\mathbf{B}_0$.

5.3 The fast and slow modes

The full set of equations for MHD waves in a compressible plasma is

$$\eta\frac{d\mathbf{u}}{dt} = -\nabla P + \mathbf{J}\times\mathbf{B}, \qquad \frac{\partial\eta}{\partial t} + \nabla\cdot(\eta\mathbf{u}) = 0, \tag{5.14a,b}$$

$$\mathbf{E} + \mathbf{u}\times\mathbf{B} = 0, \qquad P \propto \eta^{\Gamma}, \tag{5.14c,d}$$

$$\nabla\times\mathbf{E} = -\frac{\partial\mathbf{B}}{\partial t}, \qquad \nabla\times\mathbf{B} = \mu_0\mathbf{J}. \tag{5.14e,f}$$

The linearized forms of these equations, after assuming variation $\propto \exp[-i(\omega t - \mathbf{k}\cdot\mathbf{x})]$, lead to an MHD wave equation which may written in the form

$$\Gamma_{ij}(\omega,\mathbf{k})\xi_i = 0, \tag{5.15}$$

where ξ is the fluid displacement, and with

$$\Gamma_{ij}(\omega,\mathbf{k}) = \omega^2\delta_{ij} - k^2c_s^2\kappa_i\kappa_j \\ -k^2v_A^2[\kappa_i\kappa_j - \cos\theta(\kappa_ib_j + b_i\kappa_j) + \cos^2\theta\,\delta_{ij}], \tag{5.16}$$

where θ is the angle between κ (in the x–z plane) and $\mathbf{b}$ (along the z axis).

The dispersion relation for MHD waves may be found by writing Γ_{ij} as a matrix and setting its determinant equal to zero. The details are omitted here. There are three solutions of the resulting dispersion equation. Solving for the square of the phase speed, $v_\phi^2 = \omega^2/k^2$, we find that one of the solutions gives $v_\phi^2 = v_A^2\cos^2\theta$, corresponding to the Alfvén wave mode discussed in § 4.1. The other two solutions are

$$v_\phi^2 = v_\pm^2, \qquad v_\pm^2 = \tfrac{1}{2}(v_A^2 + c_s^2) \pm \tfrac{1}{2}\left[(v_A^2 + c_s^2)^2 - 4v_A^2c_s^2\cos^2\theta\right]^{1/2}. \tag{5.17}$$

These correspond to the fast (+) and slow (−) magnetoacoustic modes. These two modes may be interpreted in terms of gas sound waves, satisfying (5.11) in the absence of a magnetic field, being modified by the contribution from the magnetic pressure, and in terms of magnetoacoustic waves, satisfying (5.13) in the absence of gas pressure, being modified by the contribution from the gas pressure. This interpretation is based on the fact that(5.11) and (5.13) are two limiting cases of (5.17), corresponding to the fast mode for $c_s^2 \gg v_A^2$ and $c_s^2 \ll v_A^2$, respectively. Note that the fast mode is gas-sound-like for $c_s^2 \gg v_A^2$ and magnetoacoustic-like for $c_s^2 \ll v_A^2$. Conversely, the slow mode for propagation nearly along the field lines is magnetoacoustic-like for $c_s^2 \gg v_A^2$ and gas-sound-like for $c_s^2 \ll v_A^2$, and is quite strongly modified from these properties at large angles of propagation.

The direction of the fluid displacement may be found from the wave equation (5.15). For Alfvén waves one finds $\xi_A = (0,1,0)$, as given by (5.3). The direction

of the fluid displacement in the magnetoacoustic waves is given by

$$\xi_\pm = (\sin\psi_\pm, 0, \cos\psi_\pm), \qquad \tan\psi_\pm = \frac{c_s^2 \sin\theta\cos\theta}{v_\pm^2 - v_A^2 - c_s^2\sin^2\theta}. \tag{5.18}$$

Using these properties one may find the other fields in the waves from the linearized forms of (5.14), but this is not done here.

In the low-β limit, (5.17) may be approximated by

$$v_+^2 \approx v_A^2 + c_s^2\sin^2\theta, \qquad v_-^2 \approx c_s^2\cos^2\theta. \tag{5.19}$$

Thus the fast mode is a slightly modified magnetoacoustic mode and the slow mode reduces to a sound wave for propagation along the magnetic field lines, but is increasingly modified by the magnetic field as the direction of propagation approaches perpendicular to the field lines.

5.4 Alfvén waves in astrophysics

The three MHD waves play quite different physical roles. In a low-β plasma fast mode waves play the role that sound waves play in an ordinary gas. That is, they carry away any pressure fluctuation, and maintain pressure equilibrium. Hence, pressure balance in a magnetized plasma is set up and maintained on an Alfvén crossing time, which is the time it takes for a wave to propagate across the plasma at v_A. The fast mode propagates almost isotropically, and so maintains pressure balance both across and along the magnetic field lines. The slow mode is important in specific contexts, but is generally of less importance than the other two modes.

Alfvén waves, which are shear waves in the magnetic field, rely on the magnetic tension, as opposed to fast mode waves which rely on the magnetic pressure. Thus, when a magnetic field line is wiggled, the wiggles propagate away as Alfvén waves. An important property of Alfvén waves is that their group velocity is along the direction of the magnetic field lines. This may be seen in the following two ways. First, consider the Poynting vector in an Alfvén wave. The Poynting vector in a waves is proportional to the real part of $\mathbf{E}_1^* \times \mathbf{B}_1$, where the complex conjugate is needed in a theory in which the amplitude is a complex quantity. On using (5.2) and (5.3), one finds $\mathbf{E}_1 \propto (\mathbf{k}\times\mathbf{b})\times\mathbf{b}$ and $\mathbf{B}_1 \propto \mathbf{k}\times\mathbf{b}$, which imply $\mathbf{E}_1^*\times\mathbf{B}_1 \propto \mathbf{b}$. Second, consider the group velocity in Alfvén waves. The dispersion relation (5.4) may be written as $\omega = |k_z|v_A$, where the z-axis is along $\mathbf{b}$. The group velocity in Alfvén waves is, in cartesian coordinates

$$\mathbf{v}_{gA} = \left(\frac{\partial\omega}{\partial k_x}, \frac{\partial\omega}{\partial k_y}, \frac{\partial\omega}{\partial k_z}\right). \tag{5.20}$$

It follows from the fact that the dispersion relation for Alfvén waves involves k_z, but not the other components of $\mathbf{k}$, that the group velocity is along the

z-axis, that is, along $\mathbf{b}$. It follows that, irrespective of the direction of wave propagation (i.e., the value of θ), the energy in Alfvén waves propagates only along the magnetic field lines.

Three specific applications that illustrate the role that Alfvén waves can play in astrophysical plasmas are: the heating of stellar coronas, the loss of angular momentum by stars with winds (called magnetic braking, and discussed in the next lecture) and the interaction of Io with Jupiter's magnetic field. Let me comment on the first and last of these briefly.

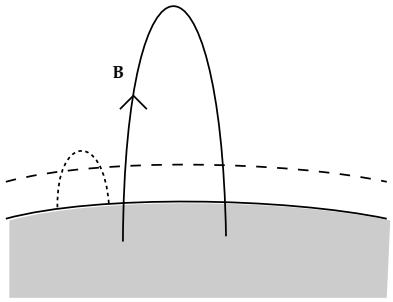


Figure 5.1: A magnetic field line is illustrated passing from below the photosphere (shaded region) across the layer $c_s = v_A$ (dashed line) into the main part of the corona. Alfvén waves follow the field lines. However, fast mode waves are refracted back to the photosphere, following a path illustrated by the dotted line.

The solar corona is much hotter ($> 10^6\,\mathrm{K}$) than the photosphere ($\approx 6 \times 10^3\,\mathrm{K}$), and it must be continuously heated by energy propagating from below the photosphere. This energy is widely believed to be in the form of waves; the only alternative that cannot be ruled out is that it is in the form of electric currents. An early theory, from before 1950, favored energy transport via sound waves generated by the turbulence seen at the photosphere in the form of the solar granulation. At the photosphere one has $c_s^2 \gg v_A^2$, and sound waves are fast mode waves. However, the plasma density falls off rapidly with height, and hence the Alfvén speed $v_A \propto \eta^{-1/2}$ increases rapidly with height, and low in the corona the inequality reverses to $c_s^2 \ll v_A^2$. As the surface defined by $c_s^2 = v_A^2$ is crossed, fast mode waves change from being sonic in character to being magnetoacoustic in character. Hence, as the waves generated by the turbulence associated with the solar granulation enter the corona, they are essentially magnetoacoustic waves, which can damp quickly and heat the plasma. This early theory is now regarded as unacceptable because the fast mode waves are strongly refracted and cannot reach the corona. This may be seen by noting that the refractive

index for fast mode waves for $c_s^2 \ll v_A^2$ is $c^2/v_A^2 \propto \eta/B^2$, which decreases rapidly with height. Waves refract toward the direction of increasing refractive index, and away from the direction of decreasing refractive index. Fast mode waves are refracted back toward the Sun and do not reach into the regions of the corona where the heating is required, cf. Figure 5.1. On the other hand, Alfvén waves have their group velocity along the magnetic field lines and so can reach all regions that are connected magnetically to the Sun. This is the reason that the solar corona is thought to be heated by Alfvén waves. However, the suggestion is not without difficulties: there is no clear observational evidence for it, and there is a difficulty in understanding how the Alfvén waves damp, which they must do to heat the plasma. Other stars have coronas and there are coronas above accretion disks around some compact stars and black holes, and most are thought to be heated by similar mechanisms to the solar corona. Heating by dissipation of electric currents remains an alternative to heating by Alfvén waves.

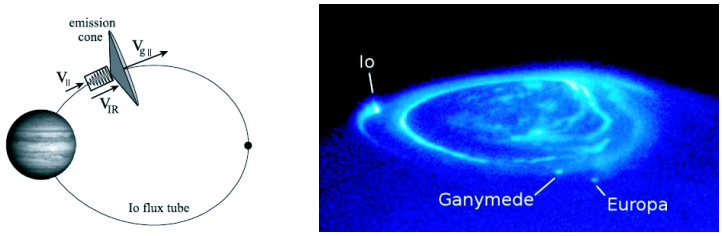


Figure 5.2: (Left: The Io flux tube is illustrated, indicating the radio emission mechanism [after Willes 2002]; Right: Observations of the northern Jovian aurora showing signature of the footpoint of the Io flux tube.

A further example mentioned here is the interaction of the innermost Galilean satellite of Jupiter with Jupiter's magnetic field. The motivation for understanding this interaction arose originally, in the 1960s, from the observational fact that the position of Io affects radio bursts from near Jupiter's surface. The idea proposed for understanding how this might occur is as follows. The rotating magnetosphere sweeps by Io, so that Io is completely surrounded by plasma threaded by magnetic field lines attached to Jupiter. Io is a good electrical conductor, due to its ionosphere, and hence there is a tendency for field lines to get frozen into Io, as illustrated in Figure 5.2. The magnetic flux tube frozen into Io therefore moves relative to the background plasma. Consider the electric fields in two regions moving at velocities $\mathbf{u}_1$ and $\mathbf{u}_2$, respectively. The electric fields are $\mathbf{E}_{1,2} = -\mathbf{u}_{1,2} \times \mathbf{B}$. If the velocity difference is $\Delta\mathbf{u} = \mathbf{u}_1 - \mathbf{u}_2$, then the potential drop across the flux tube threading Io is, cf. (6.17), $\Phi_{\rm Io} = \Delta\mathbf{u} \cdot \mathbf{A}$, where $\mathbf{A}$ is the vector potential of Jupiter's magnetic field at Io (at $r = 5.9R_{\rm J}$, where $R_{\rm J} = 7.1 \times 10^6\,{\rm m}$ is the radius of Jupiter). The radius of Io, $R_{\rm Io} = 1.7 \times 10^6\,{\rm m}$, the magnetic field at the northern magnetic pole $B_{\rm J} = 1.4 \times 10^{-3}\,{\rm T}$ and the

relative velocity, $\Delta u = 6 \times 10^4\,\mathrm{m\,s^{-1}}$ of Io to the rotating magnetosphere are known. The implied potential drop is $\Phi \approx 5 \times 10^5\,\mathrm{V}$. It was argued that this potential drives currents along the magnetic field lines so that they close in the Jovian ionosphere, as illustrated in Figure 5.2. This potential was also assumed to accelerate the energetic electrons that cause the radio emission.

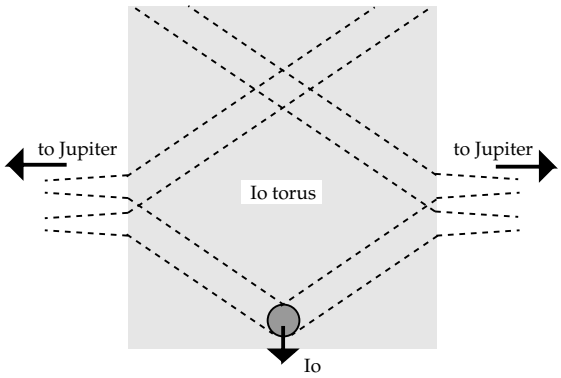


Figure 5.3: The path of Alfvén waves (dashed lines) generated by Io is illustrated schematically. Inside the Io torus the v_A is relatively slow, and it increases abruptly at the edge of the torus. The returning waves, after reflection from Jupiter's ionosphere, do not intersect Io.

Although this theory seems capable of accounting for the required coupling between Io and Jupiter's radio emission, the theory is unacceptable for a reason that only became clear a decade or so after the theory was proposed. Around 1980 it became clear that there is relatively dense plasma ($n_e = 10^2$–$10^4\,\mathrm{cm^{-3}}$) near Io's orbit. This *Io plasma torus* results from plasma escaping from Io's ionosphere. Because of the relatively dense plasma, the Alfvén speed is relatively slow. The foregoing theory applies only if Alfvén waves can transmit the stresses back to Jupiter fast enough, specifically the return travel time of an Alfvén wave from Io to Jupiter must be less than the time it takes Io to move through a distance equal to its diameter. If v_A were constant, this would require $v_A/5.9R_\mathrm{J} < \Delta u/R_\mathrm{Io}$. This condition is not satisfied; that is, when the Alfvén waves generated by Io return to Io's orbit, Io has moves away, as illustrated in Figure 5.3. This complicates the model considerably, to the extent that the acceleration of electrons due to the passage of Io through the magnetosphere is not adequately understood.

Lecture 6

Stellar winds and rotating magnetospheres

The problem of how the solar wind develops and becomes supersonic was solved in the 1950s, and the theory was generalized to include a magnetic field about a decade later. The following is a brief outline of the theory in which two points are emphasized: the influence of the magnetic field on the angular momentum loss rate, and the importance of so-called *critical points*. Somewhat related phenomena, from a theoretical viewpoint, are astrophysical jets and rotating magnetospheres; the properties of the latter are also discussed here.

6.1 Equations for a wind

A stellar wind forms when the atmosphere of a star is driven off, either by radiation pressure or by coronal heating. An important feature of a wind is that it becomes supersonic, or superalfvénic in the case of a magnetized wind. How this occurs is one of the main points discussed here.

The theory for stellar winds is based on the MHD equations. In a steady state ($\partial/\partial t = 0$), the equation of motion (4.8) for the fluid takes the form

$$\eta(\mathbf{u}\cdot\nabla)\mathbf{u} = -\nabla\left(P + \frac{B^2}{2\mu_0}\right) + \frac{1}{\mu_0}(\mathbf{B}\cdot\nabla)\mathbf{B} - \frac{GM_*\eta}{r^2}\,\hat{\mathbf{r}}, \tag{6.1}$$

where G is the gravitational constant, and where the gravitational field is that due to the central star of mass M_*.

A somewhat artificial geometry was adopted in the early theories of magnetized winds. Spherical symmetry is assumed in the gravitational force in (6.1), and spherical symmetry is also assumed in identifying the mass loss rate as

$$\dot{M} = 4\pi\eta r^2 u_r, \tag{6.2}$$

where u_r is the radial component of the wind velocity. However, the fluid motion and the magnetic field are considered only in the equatorial plane, and are effectively assumed to have cylindrical symmetry. Specifically, both $\mathbf{u}$ and $\mathbf{B}$ are assumed to have only components u_r, u_ϕ and B_r, B_ϕ, respectively. Azimuthal symmetry is assumed, that is, no quantity depends explicitly on ϕ. Then $\nabla\cdot\mathbf{B} = 0$ reduces (in spherical polar coordinates) to $(1/r^2)d(r^2B_r)/dr$, which integrates to give

$$B_r r^2 = B_* R_*^2. \tag{6.3}$$

where B_* is the magnetic field at the radius $r = R_*$ of the star.

Two further equations assumed in the theory of winds are the adiabatic equation of state $P \propto \eta^\Gamma$ and the frozen-in field condition, which with $\nabla\times\mathbf{E} = 0$ for a static magnetic field ($\partial\mathbf{B}/\partial t = 0$) implies

$$\nabla \times (\mathbf{u} \times \mathbf{B}) = 0. \tag{6.4}$$

6.2 Sonic point

Before discussing the effect of rotation and of the magnetic field, let us consider the wind solution in the absence of these effects. Then the fluid flow is purely radial, and the radial component of (6.1) reduces to

$$\eta u \frac{du}{dr} = -\frac{dP}{dr} - \frac{GM_*\eta}{r^2}. \tag{6.5}$$

The pressure gradient may be eliminated by using the adiabatic equation of state (4.17c) to write $dP/dr = (dP/d\eta)d\eta/dr$, with $dP/d\eta = c_s^2$. The fact that the mass flux is independent of radius (i.e., that mass is conserved) requires $\eta u r^2 =$ constant, which may be used to eliminate $d\eta/dr$. In this way (6.5) reduces to

$$\frac{u^2 - c_s^2}{u}\frac{du}{dr} = \frac{2c_s^2}{r} - \frac{GM_*}{r^2}. \tag{6.6}$$

The flow can become supersonic, which requires that it pass from $u < c_s$ to $u > c_s$. Such a sonic transition is possible only if the solution of (6.6) passes through the singular point where the left and right hand sides of (6.6) vanish simultaneously.

A simple example of a situation where a flow changes from subsonic flow to supersonic flow is in a de Laval nozzle, as illustrated in Figure 6.1. In this case conservation of mass requires $A\eta u =$ constant, where A is the cross-sectional area. This implies a differential equation $dA/A + d\eta/\eta + du/u = 0$. Bernouilli's equation, $P + \frac{1}{2}\eta u^2 =$ constant when gravity is neglected, requires $dP + \eta u du = 0$, and using $dP = c_s^2 d\eta$, this differential equation reduces to $dA/A = [(u^2 - c_s^2)/c_s^2]du/u$. It follows that the sonic point, $u = c_s$, must occur at the neck of the nozzle where dA is zero. The gravitational field in (6.6) effectively plays the same role as a nozzle.

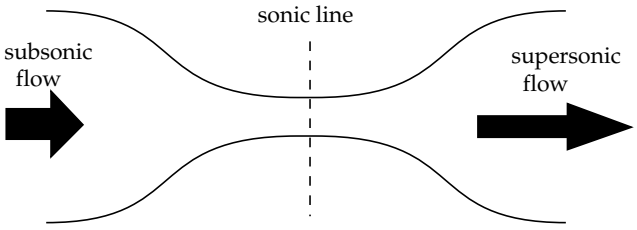


Figure 6.1: At a de Laval nozzle an incoming subsonic flow becomes supersonic.

The wind solution of (6.6) starts with $u = 0$ at the surface of the star, passes through the singular point where the flow passes from subsonic to supersonic, and has a finite (supersonic) flow velocity at infinity. Note that the r-dependence on the right hand side of (6.6) must be such that the right hand side passes through zero at some point outside the star for a wind solution to exist.

6.3 Magnetic winds

Now consider the general case described by (6.1), with the fluid velocity allowed to have a radial component u_r and an azimuthal component u_ϕ. At the surface of the star one has $u_\phi = \Omega R_*$, where Ω is the angular velocity of rotation of the star. Conservation of mass leads to the conditions (6.2). Conservation of magnetic flux determines the radial component of the magnetic field through (6.3). In addition, it is assumed that the field is radial at $r = R_*$, that is, $B_\phi = 0$ at $r = R_*$. The frozen-in field condition leads to (6.4), the ϕ-component of which requires

$$\frac{1}{r}\frac{\partial}{\partial r}(u_\phi B_r - u_r B_\phi) = 0. \tag{6.7}$$

On integrating (6.7), the constant of integration is found by considering the boundary condition at $r = R_*$. Hence one finds

$$B_\phi = \frac{u_\phi - \Omega R_*}{u_r}\, B_r. \tag{6.8}$$

The ϕ-component of (6.1) gives

$$\eta u_r \frac{d}{dr}(r u_\phi) = \frac{B_r}{\mu_0}\,\frac{d}{dr}(r B_\phi). \tag{6.9}$$

It follows from (6.2) and (6.3) that $\eta u_r / B_r = \dot{M}/B_* R_*^2$ is a constant, and hence (6.9) may be integrated trivially, giving

$$r\left(u_\phi - \frac{B_r B_\phi}{\mu_0 \eta u_r}\right) = \ell, \tag{6.10}$$

where the constant of integration ℓ may be interpreted as the angular momentum per unit mass. Then (6.9) with (6.10) may be rewritten in the form

$$ru_\phi = \frac{M_A^2\ell - \Omega r^2}{M_A^2 - 1}, \qquad M_A^2 = \frac{u_r^2}{B_r^2/\mu_0\eta}, \tag{6.11}$$

where M_A is the radial Mach number.

The Alfvénic critical point is defined as the point at which the radial flow is equal to the Alfvén speed, that is, $M_A = 1$. Now u_ϕ must remain finite everywhere, and hence the numerator in (6.11) must vanish at the point where M_A passes through unity. Let the Alfvénic critical point be at $r = r_A$, with $M_A = 1$ there by definition. The vanishing of the numerator in (6.11) at this critical point gives $\ell = \Omega r_A^2$, and thus determines the angular momentum per unit mass at $r = r_A$. The angular momentum loss rate is then $\dot{M}\ell = \dot{M}\Omega r_A^2$. Thus (6.11) implies that, from the viewpoint of the amount of angular momentum carried off, the plasma is effectively thrown off with a lever-arm of length r_A. Specifically, the angular momentum carried off by a rotating, magnetized stellar wind is equivalent to that implied by rigid rotation out to the Alfvén radius r_A.

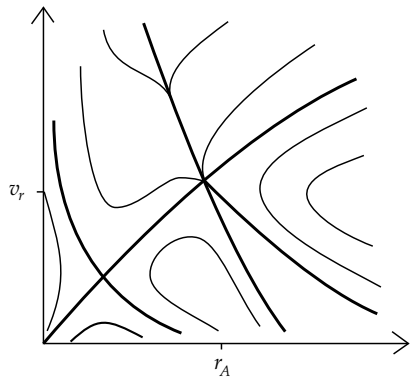


Figure 6.2: Solutions of (6.12) are plotted. There are three critical points, corresponding to the flow speed being equal to the speed of the slow, Alfvén and fast modes, but the latter two are so close together that they are indistinguishable in the figure. The only acceptable solution is the central solid curve that passes through all three critical points and gives a supersonic flow at infinity.

The radial component of the equation of motion (6.1) then becomes

$$\eta\left(u_r\frac{du_r}{dr} - \frac{u_\phi^2}{r}\right) = -\frac{dP}{dr} - \frac{B_\phi}{\mu_0 r}\frac{d}{dr}(rB_\phi) - \frac{GM_*\eta}{r^2}, \tag{6.12}$$

which is the generalization of (6.5) to include rotation and a magnetic field. A wind solution is found by integrating (6.12). Examples of solutions are illustrated in Figure 6.2. There is only one solution, corresponding to the central

solid line, that has $u_r = 0$ at the surface of the star and is supersonic at large distances. There are actually three critical points in the magnetized case, and the physically acceptable wind solution passes through all of them.

6.4 Magnetic braking

Stars like the Sun are thought to be rapidly rotating when they first form, and to lose most of their angular momentum through magnetic braking by a stellar wind. The solar wind, and the stellar wind from other stars similar to the Sun, results from the heating of the corona. The hot gas cannot radiate away its energy fast enough, and the input of energy is balance by adiabatic losses and outward transport in a continuously expanding corona.

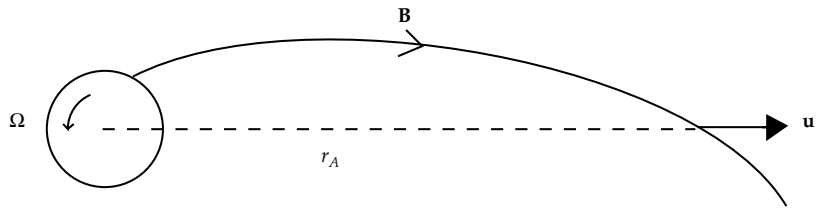


Figure 6.3: The drawing out of a magnetic field line by the solar wind, with radial velocity **u**, is illustrated schematically. The distance r_A is that at which the solar wind speed equals the Alfvén speed.

Near the Sun, the expansion speed is low, and as the density falls off the radial component v_r of the expansion speed increases rapidly to maintain a constant rate of mass loss with radius, that is, $\dot{M} = 4\pi r^2 \eta v_r$ independent of r. Near the Earth, the flow speed is known to be superalfvénic, $v_r \approx 500\,\mathrm{km\,s^{-1}}$, which is two to three times v_A. The Alfvén radius, r_A, where the flow speed passes through the Alfvén speed, is thought to occur at $r_A = 10$–$20R_*$, as illustrated in Figure 6.3. The solar wind carries away angular momentum, and the amount of angular momentum carried away is strongly dependent on the magnetic field.

To understand the effect of the magnetic field, suppose that the solar wind were unmagnetized. Then it would carry off angular momentum at the rate, $\dot{M}\Omega R_*^2$, which is determined by the angular momentum per unit mass of the wind as it leaves the solar surface. For the actual value of $\dot{M}$, this rate is too small to lead to any significant reduction of the angular momentum of the Sun in its life time on the main sequence. Now consider a magnetized wind, for which the angular momentum loss rate is $\dot{M}\Omega r_A^2$, and so is larger than for an unmagnetized wind by the factor $r_A^2/R_*^2 \sim 10^2$–10^3. This does allow a solar-like star to slow down significantly in its life time. The Sun appears to have

lost nearly all its angular momentum through a wind (nearly all the angular momentum of the solar system is in the planetary motions, overwhelmingly dominated by Jupiter's contribution). To account for this, it appears that the solar wind must have been stronger (larger $\dot{M}$) or the magnetic field was stronger (larger r_A) early in the life time of the Sun than at present, and this is plausible if the Sun was born with a faster rotation than at present.

This enhanced loss of angular momentum can be understood in two ways that complement each other. One viewpoint is based on the magnetic stresses. As indicated in Figure 6.3, the magnetic field lines are drawn out into a spiral structure, and so are convex relative to the direction of rotation of the Sun (counterclockwise in the figure). The magnetic tension, tending to straighten the field lines, therefore exerts a force in the azimuthal direction that tends to speed up the rotation of the other regions of wind. In effect, the magnetic field is attempting to enforce corotation, that is, it is attempting to make the plasma in the solar wind rotate as a rigid body with the Sun. This tension force transfers angular momentum from the footpoint of the magnetic field line, where it is tied to the Sun, to more distant regions of the solar wind. As a result, the magnetized solar wind carries off more angular momentum than an unmagnetized solar wind would.

The other, complementary, viewpoint is based on the transmission of stresses by Alfvén waves. Alfvén waves can propagate from the solar wind to the Sun and back, thereby transferring stresses, but this is possible only out to the Alfvénic point, that is, only in the region where the flow speed is less than v_A. This implies a strong magnetic coupling between the solar wind and the Sun within this region. Thus the foregoing argument relating to the transfer of angular momentum outward in the solar wind applies only out to the Alfvénic point. Beyond that point, Alfvén waves cannot propagate back to the Sun, and so no stresses can be transferred between the plasma and the solar surface.

As already shown, the angular momentum loss rate is $\dot{M} r_A^2 \Omega$, where Ω is the angular frequency of rotation of the Sun. This angular momentum loss rate is the value that one would estimate by assuming rigid rotation out to r_A. The foregoing argument suggests that the angular momentum loss rate should be of order this value, but it is a little surprising that an idealized model of rigid rotation out to r_A and free expansion beyond r_A gives the angular momentum loss rate exactly. In fact, nothing notable happens at the Alfvénic point—there is no kink in the field and no abrupt change in the flow velocity—and the tension force continues to accelerate the plasma beyond r_A.

Magnetic braking is thought to be important in less massive stars (F, G and M stars) that have convective zones and hence, presumably, dynamo-generated magnetic fields. However, there is no evidence for significant braking of more massive stars without convective zones, most of which are rapidly rotating.

6.5 Planetary and stellar magnetospheres

A magnetosphere is a region around a solid object, such as a planet or star, where the motion of particles is dominated by the magnetic field of the solid object. The more familiar examples are the magnetospheres of the Earth and Jupiter, and those of pulsars.

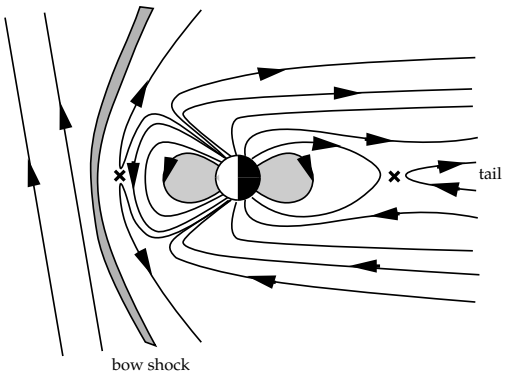


Figure 6.4: The Earth's magnetosphere is indicated schematically. The Sun is to the left of the figure. The ram pressure of the solar wind draws the magnetic field lines out into a magnetotail in the antisunward direction. The bow shock on the sunward side is indicated. The shaded region in the inner magnetosphere is the plasmasphere.

The Earth's magnetosphere forms a cavity in the solar wind, as illustrated in Figure 6.4. On the sunward side, the magnetic pressure associated with the Earth's magnetic field causes the solar wind to slow down abruptly at a bow shock, with the solar wind then flowing around the magnetospheric cavity. The region inside the magnetospheric cavity is populated by plasma from two sources: an outer source from the solar wind and an inner source from the ionosphere. The inner part of the magnetosphere, called the plasmasphere, is populated by plasma is diffusive equilibrium with the ionosphere. The giant planets (Jupiter, Saturn, Uranus and Neptune) have magnetospheres, which are populated by plasma from the ionospheres of the planets and, most notably in the case of Io in the jovian magnetosphere, the ionospheres of moons. Mercury also has a magnetosphere populated by plasma from the solar wind (Mercury has negligible atmosphere and ionosphere).

There are also magnetospheres around some strongly magnetized stars. The solar wind is not a magnetosphere because the plasma drags out the magnetic field lines, rather than the magnetic field dominating the motion of the plasma. However, the solar wind does have some magnetosphere-like properties. In particular, the solar wind forms a magnetosphere-like cavity in the interstellar

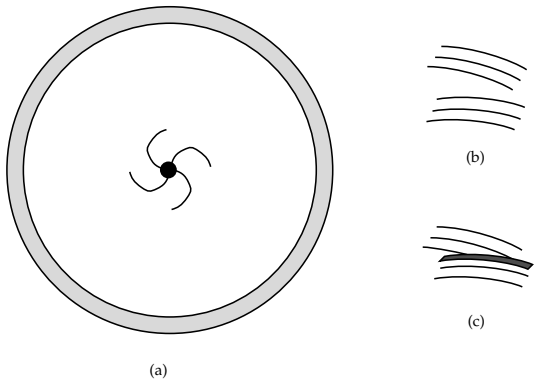


Figure 6.5: (a) The heliosphere is defined as the region magnetically connected to the Sun, it includes the interplanetary medium and extends out to the heliopause, shown by shaded region, where there must be a shock transition at which the solar wind terminates. The central region shows how the solar wind draws the magnetic field (four field lines shown) into an Achimedian spiral pattern. (b) Far from the Sun the field is nearly tangential, and faster streams (less tightly wound) in the solar wind overtake slower streams (more tightly wound), leading to (c) a shock transition at the interface between the two streams.

medium. This cavity is called the heliosphere, as illustrated schematically in Figure 6.5. Also, within few solar radii there are closed magnetic field structures, variously referred to a prominences, loops and filaments, where the plasma is confined by the magnetic field. The solar wind flows out from so-called coronal holes, where the magnetic field lines are open. Of course, $\nabla \cdot \mathbf{B} = 0$ requires that all magnetic field lines are closed somewhere, and by "open" field lines one means field lines that close along way from the Sun, presumably in the heliospheric boundary.

In stars with stronger magnetic fields and stronger winds than the Sun, the closed and open regions can become clearly separated. Assuming a dipole-like field, the field lines near the equator are much more able to confine the plasma than the field lines near the pole, which extend to large distances. This leads to the concept of a so-called dead zone in the inner region of a magnetized star with a wind, as illustrated in Figure 6.6.

Rotation plays an important role in magnetospheres. As indicated above in connection with the solar wind, a magnetic field tends to enforce corotation of a wind, and in the case of a magnetosphere, where magnetic forces dominate, corotation is maintained to a first approximation. This requires an electric field such that the electric drift velocity is equal to the corotation velocity. The theory of rotating magnetospheres was developed originally in connection with planetary magnetospheres: the inner part of the magnetosphere of the Earth

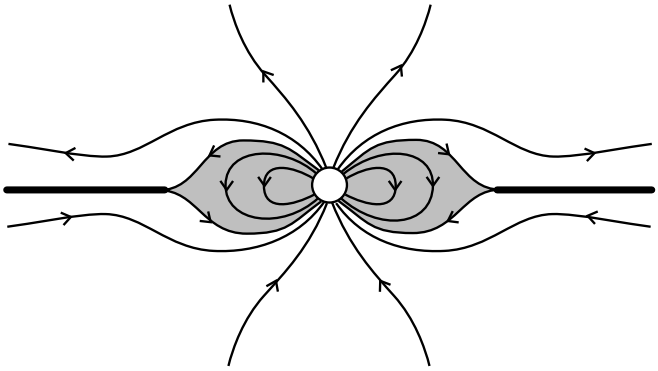


Figure 6.6: The model for a magnetized star with a wind. At large distances the wind drags the magnetic field lines towards the radial direction. A current sheet, indicated by the solid line, forms in the equatorial plane. In the "dead zone", shown shaded, the magnetic field lines are closed and confine the plasma.

corotates with the Earth (out to $\approx 3R_E$), and rotation plays a particularly important role in the magnetosphere of Jupiter, where the centrifugal force on a corotating particle exceeds the gravitational force beyond a few planetary radii. An exotic example of a rotating magnetosphere is that around a neutron star, that is, a pulsar magnetosphere. It is thought that pulsar magnetospheres are populated by a plasma consisting of electron positron pairs, as discussed in detail in a later lecture.

Lecture 7

MHD shock waves

In this lecture I outline the theory of shock waves and discuss some astrophysical applications briefly. A *shock wave* is a discontinuity between two media moving relative to each other. There are several types of discontinuity, and only one is called a shock wave; others include a *tangential discontinuity* and a *contact surface*. A distinguishing feature of a shock is that irreversible processes occur at the shock front, that is, entropy is generated at the boundary or interface between the two media. Also, in a shock there is a mass flux across the front.

7.1 Boundary conditions at a discontinuity

The theory of shock waves relates the plasma properties behind the shock, which is referred to as the *downstream* region, to the plasma properties ahead of the shock, which is referred to as the *upstream* region. The relevant relations depend on the variables that describe the shock. In the absence of a magnetic field there is only one independent variable, and this is usually chosen to be the Mach number, M, which is the ratio of the flow speed to the sound speed in the upstream region. In the presence of a magnetic field, the angle, θ, between the flow velocity and the magnetic field is another independent parameter. One refers to shocks with $\theta = \pi/2$ as *perpendicular* and shocks with $\theta = 0$ (or $\theta = \pi$) as *parallel*. The ratio of the sound speed, c_s, to the Alfvén speed, v_A, in the unshocked plasma is another important independent parameter. In the magnetized case one usually defines an Alfvénic Mach number, M_A, in terms of the Alfvén speed rather than the sound speed. Thus, our first objective is to derive a set of equations that determine the ratio of various plasma parameters across the shock in terms of M, or M_A, and θ.

Consider a boundary (an interface) between two semi-infinite plasmas with different densities, η, temperatures, T, and magnetic fields, $\mathbf{B}$. The normal to the boundary is denoted by the unit vector $\mathbf{n}$. Let the two regions be labeled 1

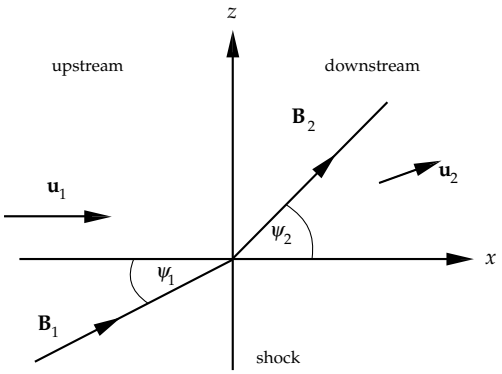


Figure 7.1: The changes in velocity and magnetic field across a shock front are illustrated in the shock frame. In this frame the shock is at rest, the upstream plasma is approaching with velocity $\mathbf{u}_1$ perpendicular to shock front, and the downstream plasma is receding at a velocity $\mathbf{u}_2$, which is at an oblique angle to the shock normal in general. The magnetic field in the upstream region is at an angle ψ_1 to the shock normal, and this changes to ψ_2 in the downstream region.

and 2, as illustrated in Figure 7.1. Such a boundary may be used to describe a shock in the *shock frame*, that is, in a frame in which the shock is at rest. One is always free to make a Lorentz transformation to the shock frame. Having found the properties of shocks in this shock frame, one may then make a Lorentz transformation back to the frame of physical interest, usually called the *laboratory frame*, in which the shock is propagating from region 2 (*downstream* of the shock) into region 1 (*upstream* of the shock).

Let the change in any quantity across the boundary be denoted by square brackets. Thus the changes in a scalar quantity, A, and in the normal component V_n of some vector, $\mathbf{V}$, are

$$[A] = A_1 - A_2, \qquad [V_n] = (\mathbf{V}_1 - \mathbf{V}_2) \cdot \mathbf{n} = 0. \tag{7.1}$$

The normal and tangential components of $\mathbf{V}$ are, respectively,

$$V_n = \mathbf{V} \cdot \mathbf{n}, \qquad \mathbf{V}_t = \mathbf{V} - (\mathbf{V} \cdot \mathbf{n})\,\mathbf{n}. \tag{7.2}$$

Boundary conditions that relate parameters across the interface between the two media are often called *jump conditions*.

Of particular interest are the changes in quantities that satisfy a conservation law. The conservation law for a scalar quantity, Q, with associated flux, $\mathbf{F}$, is

$$\frac{\partial Q}{\partial t} + \nabla \cdot \mathbf{F} = S_Q, \tag{7.3}$$

where S_Q is a source term of Q. Relevant examples of conserved quantities are mass, momentum, energy and charge. Suppose that our boundary is not moving, and that the source term is zero. Then on integrating (7.3) over all space, only the term involving $\nabla \cdot \mathbf{F}$ contributes, and it may be converted into a surface integral. Thus (7.3) gives

$$\int_S d^2\mathbf{x}\, \mathbf{n} \cdot \mathbf{F} = 0. \tag{7.4}$$

This expresses the obvious requirement for a conserved quantity that the incoming fluence from region 1 is equal to the outgoing fluence in region 2. For example, the fluence of mass is the number of kilograms per second flowing into the boundary from region 1, and this is obviously equal to the number of kilograms per second flowing out of the boundary into region 2.

For our stationary boundary with no source terms, (7.3) has the following explicit form for mass ($Q \to \eta$ = mass density, $\mathbf{F} \to \eta\mathbf{u}$, $\mathbf{u}$ = flow velocity):

$$\nabla \cdot (\eta \mathbf{u}) = 0. \tag{7.5}$$

The corresponding expression for the energy is, cf. (4.19),

$$\nabla \cdot \left(\frac{1}{2}\eta u^2 \mathbf{u} + \frac{\Gamma P}{\Gamma - 1}\mathbf{u} + \frac{\mathbf{E} \times \mathbf{B}}{\mu_0} \right) = 0, \tag{7.6}$$

where the first term describes the kinetic energy, the second term describes the thermal energy, with Γ the ratio of the specific heats ($\Gamma = 5/3$ is assumed below), and the final term is the electromagnetic energy flux, which is the Poynting vector. The conservation equation for momentum has a different form because the momentum density $\eta\mathbf{u}$ is a vector, and the counterpart of the flux is the stress tensor. Let $S_{ij} = S_{ji}$ be the ij-component of the stress tensor, with i and j equal to x, y or z. The counterpart of (7.5) or (7.6) for the ith component of the momentum flux is then of the form $\partial S_{ij}/\partial x_j$, with the sum over $j = x$, y, z understood. Specifically, one has

$$\frac{\partial}{\partial x_i} \left\{ \eta u_i u_j + P\delta_{ij} + \left(\frac{\varepsilon_0 E^2}{2} + \frac{B^2}{2\mu_0} \right) \delta_{ij} - \varepsilon_0 E_i E_j - \frac{1}{\mu_0} B_i B_j \right\} = 0. \tag{7.7}$$

The first term in (7.7) describes the stress due to the mass motion, the second term is due to an isotropic pressure, which describes the stress due to random thermal motions, the next two terms are the electric and magnetic pressures, respectively, and the final terms describe the stresses associated with the tensions along the electric and magnetic field lines, respectively. Whilst, these boundary conditions suffice for an unmagnetized shock (when one has both $\mathbf{B} = 0$ and $\mathbf{E} = 0$); they need to be complemented by boundary conditions for $\mathbf{B}$ and $\mathbf{E}$ in the magnetized case.

The plasma is assumed to be infinitely conducting, so that one has

$$\mathbf{E} = -\mathbf{u} \times \mathbf{B}. \tag{7.8}$$

Using (7.8) one eliminates $\mathbf{E}$ from (7.6) and (7.7). For example, the Poynting vector is then given by

$$\frac{\mathbf{E} \times \mathbf{B}}{\mu_0} = \frac{1}{\mu_0}(B^2\mathbf{u} - \mathbf{u} \cdot \mathbf{B}\,\mathbf{B}).$$

The boundary condition that expresses conservation of mass follows from (7.5). The normal component of the mass flux must satisfy

$$[\eta u_n] = 0. \tag{7.9}$$

The normal component of the mass flux is continuous across the boundary. The conservation law (7.6) for energy gives, after eliminating $\mathbf{E}$ using (7.8),

$$\left[\frac{1}{2}\eta u^2 u_n + \frac{\Gamma}{\Gamma - 1} P u_n + \frac{B^2}{\mu_0}\, u_n - \frac{\mathbf{u} \cdot \mathbf{B}}{\mu_0} B_n\right] = 0. \tag{7.10}$$

The final conservation law (7.7) for momentum has both normal and tangential components, which give, respectively,

$$\left[\eta u_n^2 + P + \frac{B_t^2}{2\mu_0}\right] = 0, \tag{7.11}$$

$$\left[\eta u_n \mathbf{u}_t - \frac{B_n}{\mu_0}\mathbf{B}_t\right] = 0. \tag{7.12}$$

The boundary conditions that follow from Maxwell's equations are that the normal component of $\mathbf{B}$ and the tangential component of $\mathbf{E}$ be continuous. The first of these gives

$$[B_n] = 0. \tag{7.13}$$

The continuity of the tangential component of $\mathbf{E}$ is equivalent to the continuity of $\mathbf{E} \times \mathbf{n}$; on inserting (7.8) and using the identity

$$(\mathbf{u} \times \mathbf{B}) \times \mathbf{n} = \mathbf{u} \cdot \mathbf{n}\,\mathbf{B} - \mathbf{B} \cdot \mathbf{n}\,\mathbf{u},$$

one requires

$$[u_n \mathbf{B}_t - B_n \mathbf{u}_t] = 0. \tag{7.14}$$

We now face the task of combining (7.8)–(7.14) into a single equation.

7.2 Shocks in unmagnetized plasmas

To understand how one is to proceed, it is helpful to consider the example of an unmagnetized plasma, that is, the case $\mathbf{B} = 0$. We are free to make a Lorentz transformation to the frame in which the tangential velocity, $\mathbf{u}_t$, is zero on one side of the boundary. Then (7.12) implies

$$\mathbf{u}_t = 0 \tag{7.15}$$

on both sides of the boundary. (This is not generally the case for $\mathbf{B} \neq 0$.) Then, introducing the mass flux, j, and the compression ratio, r, (7.9) implies

$$j = \eta_1 u_{1n} = \eta_2 u_{2n}, \qquad r = \frac{\eta_2}{\eta_1} = \frac{u_{1n}}{u_{2n}}. \tag{7.16}$$

On introducing the (adiabatic) sound speed, $c_s = (\Gamma P_1/\eta_1)^{1/2}$, (7.10) gives

$$P_2 = \eta_1 \left[\frac{u_{1n}^2 (r-1)}{r} + \frac{c_s^2}{\Gamma} \right], \qquad P_1 = \frac{\eta_1 c_s^2}{\Gamma}. \tag{7.17}$$

Using (7.16) and (7.17), (7.11) may be reduced to

$$u_{1n}^2 + \frac{2c_s^2}{\Gamma - 1} = \frac{u_{1n}^2}{r^2} + \frac{2\Gamma}{\Gamma - 1} \left[\frac{u_{1n}^2 (r-1)}{r^2} + \frac{c_s^2}{r\Gamma} \right]. \tag{7.18}$$

Of the remaining equations, (7.12) leads to (7.15), and (7.13) and (7.14) are identically zero. Hence we have only the three equations (7.16)–(7.18).

One has the following six variables: η_1, η_2, u_{1n}, u_{2n}, P_1, P_2, where the label n on u_{1n}, u_{2n} is retained despite the fact that the tangential component is zero due to choice of inertial frame. Of these η_1, u_{1n}, P_1 refer to the properties of the upstream medium ahead of the shock, and we wish to determine the properties η_2, u_{2n}, P_2 of the downstream medium in terms of them. We may eliminate P_1 using (7.17), so that the upstream medium is described in terms of η_1, u_{1n}, c_s. Similarly, we may eliminate P_2 using (7.17). Thus we reduce the problem to finding η_2, u_{2n} in terms of η_1, u_{1n}. In (7.18), u_{2n} is already eliminated using (7.16), and so our downstream medium is described in terms of only the one variable, η_2. It is traditional to introduce the Mach number, M, as the independent variable:

$$M = \frac{u_{1n}}{c_s}. \tag{7.19}$$

Then (7.18) reduces to

$$M^2 = \frac{r}{a}, \qquad a = \frac{1}{2}[\Gamma + 1 - r(\Gamma - 1)]. \tag{7.20}$$

The other variables may be expressed in terms of r and M, specifically,

$$\frac{P_2}{P_1} = \frac{2\Gamma M^2}{\Gamma + 1} - \frac{\Gamma - 1}{\Gamma + 1} = \frac{5M^2 - 1}{4}. \tag{7.21}$$

Note that the density ratio r cannot be arbitrarily large. The maximum value for a *strong shock*, $M \to \infty$, corresponds to $a = 0$ in (7.20). For $\Gamma = 5/3$ (7.20) gives $M^2 = 5r/(4-r)$, and the maximum density ratio is $r = 4$. Thus, for $\Gamma = 5/3$, no shock can lead to a compression in the density by a factor greater than four. On the other hand, (7.21) implies that the pressure ratio becomes arbitrarily large for arbitrarily strong shocks. This may be attributed to the temperature $T_2 \propto P_2/\eta_2$ becoming arbitrarily large behind the shock for $M^2 \to \infty$.

7.3 MHD shocks

When the magnetic field is retained, we are still free to choose the frame in which $\mathbf{u}_1$ is parallel to $\mathbf{n}$, but unlike the unmagnetized case it does not follow that $\mathbf{u}_2$ is also parallel to $\mathbf{n}$. It is convenient to introduce the parameters

$$\beta = \frac{c_{1s}^2}{v_{1A}^2}, \qquad v_{1A}^2 = \frac{B^2}{\mu_0 \eta_1}, \qquad M_A = \frac{u_1}{v_{1A}}. \tag{7.22}$$

Then the jump conditions lead to the following equation for the Alfvén Mach number, M_A:

$$\left(\frac{aM_A^2}{r} - \beta\right)\left(\frac{M_A^2}{r} - \cos^2\theta\right)^2 - \frac{M_A^2}{r}\sin^2\theta\left[\frac{M_A^2}{r}\left(\frac{a}{r} - \frac{1-r}{2}\right) - a\cos^2\theta\right] = 0, \tag{7.23}$$

with a given by (7.20). There is a variety of different solutions of (7.23), each of which describes a shock-like transition, but not all of which are shocks. It is impractical to summarize the wide range of different solutions of (7.23) and their interpretations. The following summary includes solutions of astrophysical interest.

First note that (7.23) is cubic in M_A^2, and so it has three solutions in general. In the limit of a weak discontinuity, that is, for $r \to 1$, $a \to 1$, (7.23) reduces to the dispersion equation for MHD waves. There are three MHD wave modes, the Alfvén mode, and the fast and slow magnetoacoustic modes, as discussed in the previous lecture. The solutions of (7.23) may be classified according to their relation to the MHD modes, that is, as Alfvénic, fast or slow mode "shocks". The description "shock" is usually restricted to a compressive shock, and the Alfvénic solution is noncompressive; nevertheless it remains of physical interest as a contact discontinuity.

For parallel propagation ($\sin\theta = 0$), (7.23) has solutions

$$M_A^2 = r, \qquad M_A^2 = r\beta/a, \tag{7.24}$$

with the solution at $M_A^2 = r$ being a double solution. The latter solution $M_A^2 = r\beta/a$ corresponds to the shock solution in the unmagnetized case, that

is, to a slow mode shock for $v_{1A} \gg c_{1s}$, and in this case the magnetic field does not change across the front. Now consider the other solution $M_A^2 = r$. Recall that we have chosen the frame such that $\mathbf{u}_1$ is along $\mathbf{n}$, and then $\sin\theta = 0$ implies that $\mathbf{B}_1$ is also along $\mathbf{n}$. However, neither $\mathbf{u}_2$ nor $\mathbf{B}_2$ are along $\mathbf{n}$. This case is called a *switch-on shock*, reflecting the fact that the tangential component of $\mathbf{B}$ switches on as the front passes.

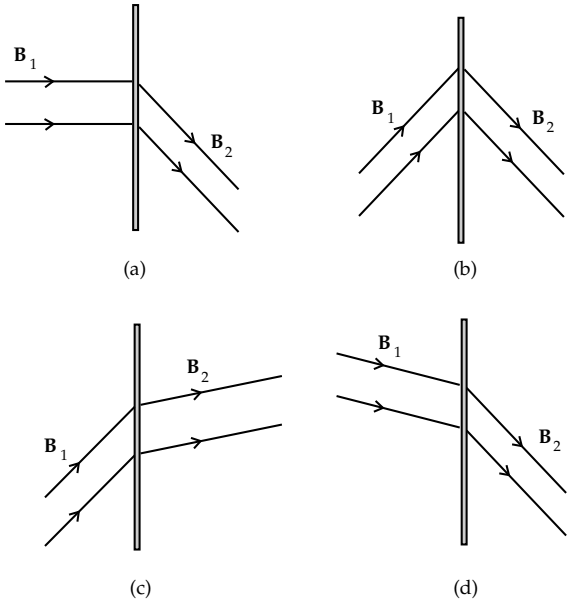


Figure 7.2: Shocks and shock-like discontinuities are illustrated with the upstream (undisturbed) medium on the left and the downstream (shocked) medium on the right: (a) a switch-on shock, (b) a tangential discontinuity, (c) a slow mode shock, and (d) a fast mode shock. A tangential discontinuity, also called a rotational discontinuity or an intermediate wave, is not a shock in the sense that there is no compression of the plasma. In a slow mode shock the perpendicular component of the magnetic field decreases across the shock, and in a fast mode shock the perpendicular component of the magnetic field increases across the shock. A switch-on shock is a special case of a fast mode shock in which the perpendicular component of the magnetic field is initially zero; there is an analogous switch-off shock which is a special case of a slow mode shock.

An Alfvén wave is a torsional wave and involves no compression of the magnetic field. Hence one would not expect the generalization of the Alfvén wave solution to give $r > 1$ and so to describe a compressive shock. One finds that $r = 1$, $M_A^2 = \cos^2\theta$ is always a solution of (7.23), and this includes the solu-

tion corresponding to small amplitude Alfvén waves. The generalization of this solution to large amplitudes is not a compressive shock because $r = 1$ implies $\eta_2 = \eta_1$ and hence no compression. In this case $r = 1$ implies $[\eta] = 0$, this in (7.9) implies $[u_n] = 0$, then (7.10) and (7.13) imply $[P] = 0$, and (7.11) implies $[B_t^2] = 0$. Hence this case seems trivial because nothing seems to change. However, this is not correct, because $[B_t^2] = 0$ has two solutions: one is the trivial solution $\mathbf{B}_{1t} = \mathbf{B}_{2t}$, in which the tangential components of the magnetic field do not change, and so nothing changes, and the other is $\mathbf{B}_{1t} = -\mathbf{B}_{2t}$, in which the tangential components of the magnetic field reverse sign. This type of solution, with

$$u_{2n} = u_{1n}, \quad B_{2n} = B_{1n}, \qquad \mathbf{u}_{2t} = -\mathbf{u}_{1t}, \quad \mathbf{B}_{2t} = -\mathbf{B}_{1t}, \tag{7.25}$$

is called a *rotational discontinuity*.

The change in the magnetic field across a shock is determined by

$$B_{2n} = B_{1n}, \qquad \mathbf{B}_{2t} = \mathbf{B}_{1t}\left[\frac{r(M_A^2 - \cos^2\theta)}{M_A^2 - r\cos^2\theta}\right]. \tag{7.26}$$

For a strong shock, the maximum compression of the tangential components of the magnetic field is $|\mathbf{B}_{2t}|/|\mathbf{B}_{1t}| \to 4$ for $M_A^2 \to \infty$. A distinction between a fast mode shock and a slow mode shock is the change in the magnitude of the tangential component of the magnetic field: it increases ($|\mathbf{B}_{2t}| > |\mathbf{B}_{1t}|$) in a fast mode shock, and decreases ($|\mathbf{B}_{2t}| < |\mathbf{B}_{1t}|$) in a slow mode shock. These various cases are illustrated in Figure 7.2.

7.4 Shock waves in astrophysical plasmas

Before discussing some specific astrophysical examples of shock waves, it is relevant to point out the distinction between a *driven shock* and a *blast wave*. An analogy that is easily visualized concerns surface water waves. When a stone in dropped into a pond, a spherical front is generated. This is the counterpart of a blast wave, for example, as generated by an explosion at some point. The important feature is that energy is released only at one point. In contrast, the bow shock of a ship is analogous to a driven shock. One refers to the driver, in this case the ship, as a *piston*. If the driver changes, for example, if the ship turns or speeds up, the properties of the shock adjust accordingly.

As a first example, consider interplanetary shock waves. These can be driven by matter being thrown off from the solar corona, called *coronal mass ejections* (CMEs). A CME acts as a driver which is moving supersonically (which means superalfvenically in this case) relative to the upstream plasma. The position of the shock adjusts so that it is ahead of the CME at a standoff distance. At the shock the flow speed of the upstream plasma is slowed, relatively to the CME, from supersonic to subsonic, so that it can flow around the driver, as

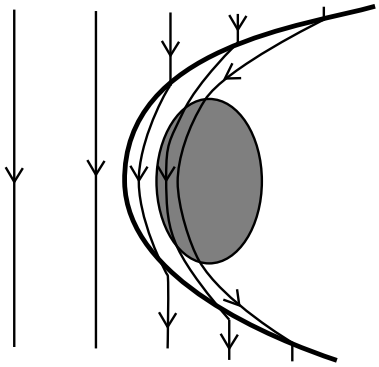


Figure 7.3: A dense cloud of plasma, shown shaded, moving supersonically through a background plasma act as a piston in generating a bow shock. Magnetic field lines are deflected at the shock. Such a shock is often referred to as a *stand-off shock*, or sometimes a *bow shock*, although the latter is more appropriate for a solid object such as a planet or a comet.

illustrated in Figure 7.3. Note that the flow speed must be zero at the surface of the driver. Whenever a driver moves supersonically, there must be a transition from supersonic to subsonic flow somewhere ahead of the driver, so that the flow speed can be zero on the surface of the driver. In general this requires a shock transition. There are also interplanetary shocks and tangential discontinuities associated with interfaces between streams of plasma flowing away from the Sun at different speeds.

Interplanetary shock are "collisionless", which means that interparticle collisions play no role in determining their structure. Collisionless processes determine the shock structure, that is, the structure of the interface itself. The detailed physical processes involved do not seem to have a significant effect on the jump conditions. However, it is clear that in some shocks a significant fraction of the energy goes into energetic particles, rather than into heat. Another example of a shock is the bow shock of the Earth, cf. Figure 2.4. This arises due to the superalfvenic motion of the solar wind relative to the Earth. There are analogous bow shocks for most of the other planets and for comets in the solar wind. There are also analogous shocks associated with winds from other stars and other astrophysical objects.

Cases of considerable astrophysical interest involve shocks generated by supernova explosions. The matter thrown off, typically at several times $10^4\,\mathrm{km\,s^{-1}}$, during the explosion acts as a piston driving a shock. The material ahead of this shock may be the interstellar plasma or it may be a slow wind from the presupernova star. Specifically, the shock in SN 1987a in the Large Magellenic Cloud is now believed to be propagating through a wind generated in a phase

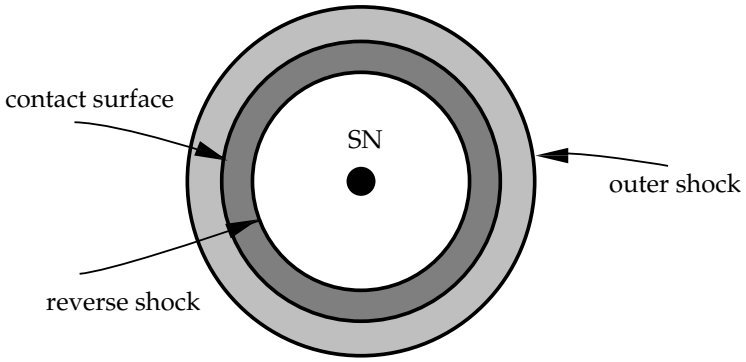


Figure 7.4: A supernova explosion generates an outgoing spherical shock which, even in an idealized case, consists of three discontinuities. There is an outer shock, which separates the undisturbed interstellar ahead of it and the shocked interstellar matter (shown lightly shaded) behind it. There is an inner shock propagating back toward the source of the explosion (the enlarged point denoted SN) through the rapidly outflowing ejected material, so that the material (shown darkly shaded) downstream of this shock has a lower outflow speed than the material (shown in white) inside the shock. Between these two shocks is a contact surface which separates the shocked interstellar matter (lightly shaded) inside the outer shock from the compressed ejected material (darkly shaded) outside the reverse shock.

in which the presupernova star was a blue giant. Within a few years the shock should encounter the much denser, slower wind associated with an earlier phase of the presupernova star in which it was a red giant. The structure of a supernova shock consists of three regions, as illustrated in Figure 7.4. There is an outer shock propagating into the undisturbed interstellar plasma or presupernova wind. At this shock the upstream plasma is compressed and heated. Behind this shock, that is, on the side nearer the point of the explosion, the matter is compressed interstellar or wind plasma. There is a contact surface that defines the inner boundary of this region. Inside the contact surface, all the matter is ejecta from the supernova explosion. Finally, there is also a so-called *reverse shock* on the inner side of the contact surface. The reverse shock is propagating radially outward but at a slower speed than the contact surface. At the reverse shock, plasma ejected by the supernova is heated and compressed. As time evolves, the total amount of matter between the two shocks increases, as more and more interstellar or wind plasma is swept up by the outer shock, and as more and more supernova ejecta crosses the reverse shock.

The ultimate fate of supernova shocks depends on the properties of the interstellar medium (ISM). The ISM is now known to consist of a hot ($> 10^6$ K),

fully ionized, low density phase, in pressure equilibrium with cooler, denser, weakly ionized clouds. Consider the fate of a shock propagating into the latter kind of plasma. As more and more mass is swept up, eventually there is insufficient energy in the ejecta to keep driving the shock. However, the shock does not stop propagating: it continues to sweep up interstellar gas, conserving momentum. This final phase of a supernova shock is sometimes referred to as the snowplough phase, in which the shock slows down as a fixed amount of momentum is distributed to a larger and larger net mass as matter is swept up.

On the other hand, if the shock is in the hot region of the ISM it can propagate effectively indefinitely. It is now thought that the hot phase is generated and maintained by supernova shocks, and that it forms a connected region in which the cooler denser clouds are embedded. The acceleration of Galactic cosmic rays is thought to be due to the effect of the many supernova shocks propagating through this region, so that any given cosmic ray encounters shocks regularly. Acceleration of cosmic rays, and of the relativistic electrons in synchrotron sources is attributed to diffusive acceleration at shocks, and is discussed in a later lecture.

Lecture 8

Wave-particle resonance

The interaction between a wave and a particle can be decomposed into two parts: the part of interest here is the *resonant part*. The other part is the "sloshing about" of a particle in the wave. Wave-particle resonance occurs when a resonance condition is satisfied, and then the particle remains in phase with the wave for a long time. A simple physical picture of these two types of interaction is for water waves: a surfer who has caught the wave before it breaks is in resonance with it, and a floating piece of wood that bobs up and down as a wave crest passes is sloshing about in the wave.

8.1 Resonance in unmagnetized plasma

The resonance condition for a particle with velocity $\mathbf{v}$ in an unmagnetized plasma is

$$\omega - \mathbf{k} \cdot \mathbf{v} = 0, \tag{8.1}$$

which is called the *Cerenkov condition*. This condition corresponds to a resonance in the sense that when (8.1) is satisfied the particle sees the electric field of the wave as a static electric field in its (the particle's) rest frame and so is systematically accelerated by it. To show this, consider a wave that has a frequency ω and wave vector $\mathbf{k}$ in the laboratory frame. By applying a Lorentz transformation from the laboratory frame to the rest frame of the particle one obtains the corresponding quantities ω', $\mathbf{k}'$ that describe the wave in the rest frame. One finds

$$\omega' = \gamma(\omega - \mathbf{k} \cdot \mathbf{v}), \tag{8.2}$$

where γ is the Lorentz factor of the particle. Resonance, in the sense that the fields appear static in the rest frame of the particle, corresponds to $\omega' = 0$, implying (8.1) in the laboratory frame.

Particles close to resonance with a wave tend to be dragged into resonance, like a surfer catching a wave. Hence if the particle is slightly slower than the wave

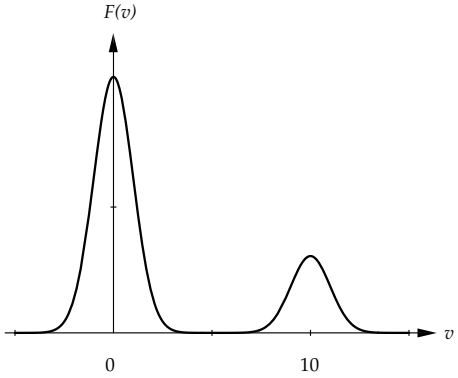


Figure 8.1: A bump-in-tail distribution $F(v)$. Waves with phase speed equal to the particle speed either grow or damp due to the effect of the wave-particle resonance; growth occurs in the range $dF(v)/dv > 0$.

it tends to be accelerated, gaining energy from the wave. On the other hand, a particle that is slightly faster than the wave tends to be slowed down and so to give energy to the wave. It follows that if the number of particles is a decreasing function of velocity at the phase velocity, $\mathbf{v}_\phi = \omega\mathbf{k}/k^2$ of the wave, then the resonant particles extract energy from the wave causing it to damp. This is called *Landau damping*, and it is the dominant damping mechanism for waves in an unmagnetized, thermal plasma. On the other hand, if the number of particles increases with increasing velocity at the resonant velocity, then the energy flow is from the particles to the waves and the waves grow. This is a particular example of a microinstability, usually called the bump-on-tail instability. Due to this type of instability, a beam of electrons with speed much greater than the thermal speed of electrons in the plasma causes Langmuir waves (longitudinal electron plasma waves) to grow, as illustrated in Figure 8.1. Langmuir waves generated in this way play an essential role in the theory of solar radio bursts. The Langmuir waves can be partially converted to transverse waves by nonlinear processes in the plasma. This so-called *plasma emission* leads to emission at the fundamental and second harmonic of the plasma frequency.

The Cerenkov condition can be satisfied only for waves with phase speed less than the speed of light, which corresponds to a refractive index greater than unity. This is seen by writing the Cerenkov condition (8.1) in the form

$$1 - n\,\boldsymbol{\kappa}\cdot\boldsymbol{\beta} = 1 - n\beta\cos\chi = 0, \tag{8.3}$$

with $\beta = \mathbf{v}/c$, and where χ is the angle between $\mathbf{k}$ and $\mathbf{v}$. It follows from $\beta < 1$ and $\cos\chi \leq 1$ that (8.3) can be satisfied only for $n \geq 1/\beta > 1$. Thus emission is possible only for particles with $\beta \geq 1/n$. Emission at the threshold $\beta = 1/n$ is at $\chi = 0$ and the angle χ increases with the speed of particles to a

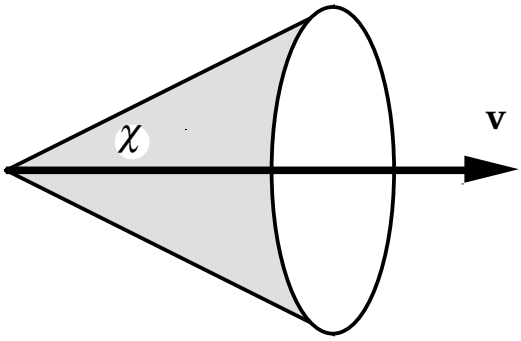


Figure 8.2: The Cerenkov cone half angle χ is between the direction of motion of the particle and the wave normal direction.

maximum value $\chi_{\max} = \arccos(1/n)$ for highly relativistic particles. It follows that Cerenkov emission is confined to the surface of a cone, with the cone half-angle χ determined by the speed of the electron, as illustrated in Figure 8.2. The Cerenkov emission pattern may also be viewed as analogous to the bow wave of a ship or a supersonic aircraft, as illustrated in Figure 8.3. The half-angle of the bow wave is $\theta' = \pi/2 - \chi = \arcsin(1/\beta n)$.

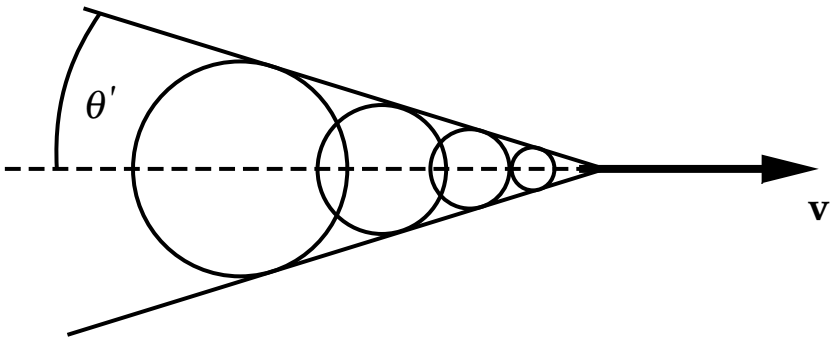


Figure 8.3: Cerenkov emission forms a bow shock with half-angle $\theta' = \pi/2 - \chi$.

A different interpretation of the Cerenkov condition emerges from a quantum mechanical viewpoint. In this case the condition is interpreted as a consequence of conservation of energy and momentum on a microscopic scale. The semi-classical interpretation of emission is as a process in which a particle emits a quantum of energy. This quantum has energy $\hbar\omega$ and momentum $\hbar\mathbf{k}$. Let the

initial energy and momentum of the particle be ε, $\mathbf{p}$ and let the corresponding quantities after emission of the wave quantum be ε', $\mathbf{p}'$. Conservation of energy and momentum then require

$$\varepsilon' = \varepsilon - \hbar\omega, \quad \mathbf{p}' = \mathbf{p} - \hbar\mathbf{k}. \tag{8.4}$$

The relativistically correct formula for the energy is

$$\varepsilon = \left(m^2c^4 + \mathbf{p}^2c^2\right)^{1/2}, \quad \varepsilon' = \left(m^2c^4 + \mathbf{p}'^2c^2\right)^{1/2}. \tag{8.5}$$

On inserting $\mathbf{p}' = \mathbf{p} - \hbar\mathbf{k}$ into the expression for ε' in (8.5) one expands in powers of $\hbar$. This gives

$$\omega = \mathbf{k}\cdot\mathbf{v} - \frac{\hbar}{2\varepsilon}\left[k^2c^2 - (\mathbf{k}\cdot\mathbf{v})^2\right] + \cdots. \tag{8.6}$$

The lowest order terms reproduce the Cerenkov condition (8.1). Thus the Cerenkov condition is interpreted as the requirement that energy conservation be compatible with momentum conservation on a microscopic level. It is only when both energy and momentum are conserved that the particle can emit a wave quantum. The next order term in $\hbar$ in (8.6) is interpreted as the quantum recoil of the emitting particle.

8.2 Resonance in magnetized plasma

The resonance condition in the presence of a magnetic field is the Doppler condition

$$\omega - s\Omega - k_\parallel v_\parallel = 0, \tag{8.7}$$

where s is an integer, $\Omega = |q|B/m\gamma$ is the relativistic gyrofrequency, and $k_\parallel$, $v_\parallel$ are the components of $\mathbf{k}$, $\mathbf{v}$ parallel to $\mathbf{B}$. Resonances at $s > 0$ are said to be via the *normal Doppler effect*, and those at $s < 0$ are said to be via the *anomalous Doppler effect*. Resonances at $s \leq 0$ are possible only for waves with refractive index greater than unity.

From a classical viewpoint, (8.7) corresponds to the wave frequency being an integral multiple of the gyrofrequency in the inertial frame in which the gyrocenter of the particle is at rest. In a quantum mechanical treatment the energy eigenvalues of a relativistic particle (ignoring its spin) are

$$\varepsilon = \varepsilon_n(p_\parallel) = (m^2c^4 + p_\parallel^2c^2 + 2n|q|B\hbar c^2)^{1/2}. \tag{8.8}$$

The quantum numbers are the continuous parameter $p_\parallel$ and the discrete parameter $n = 0, 1, 2, \ldots$. When a particle emits a wave quantum in a magnetic field, momentum perpendicular to the magnetic field is not conserved in general.

Let $p'_\parallel$ and n' be the quantum numbers after emission of the wave quantum. Conservation of parallel momentum and of energy require

$$p'_\parallel = p_\parallel - \hbar k_\parallel, \qquad \varepsilon' = \varepsilon_{n'}(p'_\parallel) = \varepsilon_n(p_\parallel) - \hbar\omega. \tag{8.9}$$

Now write $n' = n - s$ and assume that the changes in $p_\parallel$, ε, n are all small. Then an expansion in the small quantities gives

$$\varepsilon - \hbar\omega = \left(1 - \hat{D}_s + \tfrac{1}{2}\hat{D}_s^2 + \cdots\right)\varepsilon, \qquad \hat{D}_s = \hbar\left(\frac{s\Omega}{v_\perp}\frac{\partial}{\partial p_\perp} + k_\parallel \frac{\partial}{\partial p_\parallel}\right). \tag{8.10}$$

To lowest order in $\hbar$ this reproduces (8.7).

The resonance condition (8.7) is amenable to a graphical interpretation. It is convenient to plot the resonance curve in $v_\perp$–$v_\parallel$ space for given ω and $k_\parallel$. The resonance condition for each harmonic defines a *resonance ellipse*. The resonance ellipse corresponds to all the values of $v_\perp$ and $v_\parallel$ for which resonance with a wave at given ω, $k_\parallel$ and s is possible. That is, a given wave resonates with all particles that lie on the resonance ellipse that it defines. Similarly, a given particle resonates with all waves that define resonance ellipses that pass through the representative point of the particle in $v_\perp$–$v_\parallel$ space.

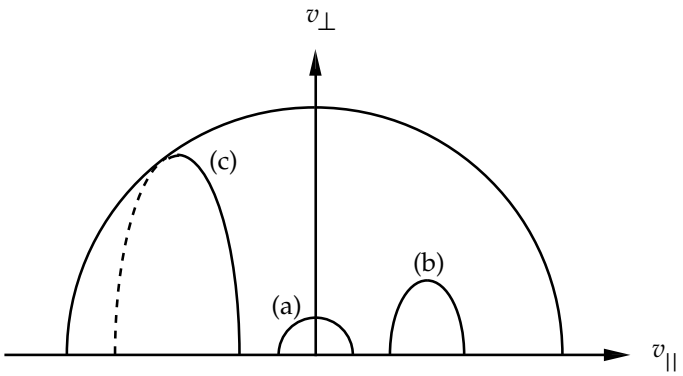


Figure 8.4: Examples of resonance ellipses: (a) a semicircle centered on the origin, (b) an ellipse inside $v = c$, (c) an ellipse touching $v = c$.

The resonance ellipse is centered on the $v_\parallel$ axis at $v_\parallel = v_\mathrm{c}$, with semi-major axis v_R perpendicular to the $v_\parallel$ axis, and with eccentricity e. These parameters are given by

$$\frac{v_\mathrm{c}}{c} = \frac{\omega k_\parallel c}{s^2\Omega_0^2 + k_\parallel^2 c^2}, \quad \left(\frac{v_R}{c}\right)^2 = \frac{s^2\Omega_0^2 + k_\parallel^2 c^2 - \omega^2}{s^2\Omega_0^2 + k_\parallel^2 c^2}, \quad e^2 = \frac{k_\parallel^2 c^2}{s^2\Omega_0^2 + k_\parallel^2 c^2}. \tag{8.11}$$

Some examples of resonance ellipses are illustrated in Figure 8.4.

8.3 The Einstein coefficients

Efficient scattering of particles is a characteristic feature of astrophysical plasmas, and this scattering is attributed to the effects of resonant interactions between ions and Alfvén waves and between electrons and whistler waves. The relevant scattering may be treated using *quasilinear theory*, which describes the backreaction on a distribution of particles of the effects of wave damping or growth by those particles.

When the resonance condition (8.7) is satisfied, the particles emit or absorb the waves. A convenient general way of describing the emission and absorption is in terms of the probability per unit time of emission of a wave quantum. Specifically consider waves in a certain wave mode, labeled M. For example, in resonant scattering one is concerned with Alfvén waves ($M \to A$) and whistler waves ($M \to W$), and in other contexts one may be interested in Langmuir waves ($M \to L$) or the more familiar transverse electromagnetic waves ($M \to T$). To write down an explicit expression one needs a general description of the waves properties. Three characteristic wave properties suffice. These are chosen to be the dispersion relation, the polarization vector and the ratio of electric to total energy in the waves. In specifying a wave mode it is to be understood that one is specifying particular forms for these three quantities.

The dispersion relation is an algebraic relation between ω and $\mathbf{k}$. For the arbitrary mode M let this be $\omega = \omega_M(\mathbf{k})$, where ω_M is a function that is different for each wave mode. The polarization vector, $\mathbf{e}_M(\mathbf{k})$, which is complex in general, is a unimodular vector along the direction of the electric vector in the wave. The ratio of electric to total energy, $R_M(\mathbf{k})$, is the fraction of the (time-averaged) wave energy in the electric field in the waves. Note that only the electric field does work, and so only the electric field is involved in the energy exchange between particles and waves. This leads to a simple interpretation of the fact that the probability is proportional to $R_M(\mathbf{k})$: the power emitted by a particle (which is equal to the power gained by the waves due to this emission) is proportional to the electric energy in the waves and is independent of the energy in magnetic fluctuations and in the kinetic energy in forced particle motions associated with the waves.

The "probability of emission" is understood to mean the probability per unit time and per unit volume of $\mathbf{k}$-space (element $d^3\mathbf{k}/(2\pi)^3$) for spontaneous emission by the particle of a wave quantum in the wave mode M, with energy $\hbar\omega_M(\mathbf{k})$ and momentum $\hbar\mathbf{k}$. A detailed calculation for a particle of charge q and mass m and momentum components $p_\parallel$ and $p_\perp$ in a uniform magnetic field $\mathbf{B}$ gives

$$w_M(\mathbf{p},\mathbf{k},s) = \frac{2\pi q^2 R_M(\mathbf{k})}{\varepsilon_0 \hbar \omega_M(\mathbf{k})} \left|\mathbf{e}_M^*(\mathbf{k})\cdot\mathbf{V}(\mathbf{k},\mathbf{p};s)\right|^2 \delta\big(\omega_M(\mathbf{k}) - s\Omega - k_\parallel v_\parallel\big),$$

$$\mathbf{V}(\mathbf{k},\mathbf{p};s) = \Big(v_\perp \frac{s}{k_\perp R} J_s(k_\perp R), -i\eta v_\perp J_s'(k_\perp R), v_\parallel J_s(k_\perp R)\Big),$$

$$\Omega = \frac{\Omega_0}{\gamma}, \quad \Omega_0 = \frac{|q|B}{m}, \quad R = \frac{v_\perp}{\Omega} = \frac{p_\perp}{|q|B}. \tag{8.12}$$

The emission and absorption may be treated in terms of the probability by appealing to the Einstein coefficients. The Einstein coefficients relate spontaneous emission, stimulated emission and true absorption on a microscopic level. Consider a quantum description of the states of the particle, as used in (8.9). Let the state of the particle before emission of the wave quantum have a set of quantum numbers $\{q\}$, which changes to $\{q'\}$ after emission of the wave quantum, as in (8.9). Suppose that the quantum mechanical occupation numbers of these states are n_q and $n_{q'}$, respectively. Let $w_{qq'}(\mathbf{k})$ be the probability of spontaneous emission $\{q\} \to \{q'\}$ for a single particle. Then, according to the Einstein coefficients, the probabilities of stimulated emission $\{q\} \to \{q'\}$ and of true absorption $\{q'\} \to \{q\}$ for the distribution of particles are

$$w_{qq'}(\mathbf{k})n_q, \quad w_{qq'}(\mathbf{k})n_q N_M(\mathbf{k}), \quad w_{qq'}(\mathbf{k})n'_q N_M(\mathbf{k}),$$

respectively, where $N_M(\mathbf{k})$ is the occupation number of the wave quantum. The wave energy density in the mode M, W_M, is related to the occupation number by multiplying by the energy per wave quantum, $\hbar\omega_M(\mathbf{k})$, and integrating over $\mathbf{k}$-space. Thus one has

$$W_M = \int \frac{d^3\mathbf{k}}{(2\pi)^3}\, \hbar\omega_M(\mathbf{k}) N_M(\mathbf{k}). \tag{8.13}$$

The rate of change of the occupation number for the wave quanta may be found by adding the contributions from spontaneous and stimulated emission, and subtracting the contribution from true absorption. This gives

$$\frac{dN_M(\mathbf{k})}{dt} = \sum_{q,q'} w_{qq'}(\mathbf{k}) \left\{ n_q[1 + N_M(\mathbf{k})] - n_{q'} N_M(\mathbf{k}) \right\}. \tag{8.14}$$

The rate of change of the occupation number n_q for the particles may be found by adding the rates of increase due to absorption $\{q'\} \to \{q\}$ from lower energy states and emission $\{q''\} \to \{q\}$ from higher energy states, and subtracting the rates of decrease due to the inverse processes. This gives

$$\frac{dn_q}{dt} = \int \frac{d^3\mathbf{k}}{(2\pi)^3} \Big(\sum_{q''} w_{q''q}(\mathbf{k}) \left\{ n_{q''}[1 + N_M(\mathbf{k})] - n_q N_M(\mathbf{k}) \right\} - \sum_{q'} w_{qq'}(\mathbf{k}) \left\{ n_q[1 + N_M(\mathbf{k})] - n_{q'} N_M(\mathbf{k}) \right\} \Big). \tag{8.15}$$

The quasilinear equations are obtained by rewriting (8.14) and (8.15) in a more appropriate notation. This involves replacing the occupation number n_q for the particles by the classical distribution function, and replacing the sum over quantum states by an integral over momentum space.

8.4 The quasilinear equations

The kinetic equation (8.14) for the waves may be written in the form

$$\frac{dN_M(\mathbf{k})}{dt} = \left(\frac{dN_M(\mathbf{k})}{dt}\right)_{\rm spont} - \gamma_M(\mathbf{k})N_M(\mathbf{k}), \tag{8.16}$$

where the first term on the right hand side describes the effect of spontaneous emission, and is of no particular interest here. The other terms define the absorption coefficient. An explicit value for the absorption coefficient follows by rewriting the relevant terms in (8.14) and using the differential operator (8.10) to take the classical limit. This gives

$$\gamma_M(\mathbf{k}) = -\sum_{s=-\infty}^{\infty}\int d^3\mathbf{p}\, w_M(\mathbf{p},\mathbf{k},s)\hat{D}_s f(\mathbf{p}),$$

$$\hat{D}_s = \hbar\left(\frac{s\Omega}{v_\perp}\frac{\partial}{\partial p_\perp} + k_\parallel\frac{\partial}{\partial p_\parallel}\right) = \frac{\hbar\omega}{v}\left(\frac{\partial}{\partial p} + \frac{\cos\alpha - n_M\beta\cos\theta}{p\sin\alpha}\frac{\partial}{\partial\alpha}\right), \tag{8.17}$$

where in the final form a change from cylindrical to polar coordinates is made in momentum space.

The quasilinear diffusion equation for the particles follows from (8.15) in a similar way. Only the induced terms are retained on the right hand side; these are the terms proportional to $N_M(\mathbf{k})$. One finds that the particle distribution function evolves according to

$$\begin{aligned}\frac{df(\mathbf{p})}{dt} &= \frac{1}{p_\perp}\frac{\partial}{\partial p_\perp}\left\{p_\perp\left[D_{\perp\perp}(\mathbf{p})\frac{\partial}{\partial p_\perp} + D_{\perp\parallel}(\mathbf{p})\frac{\partial}{\partial p_\parallel}\right]f(\mathbf{p})\right\}\\ &\quad + \frac{\partial}{\partial p_\parallel}\left\{\left[D_{\parallel\perp}(\mathbf{p})\frac{\partial}{\partial p_\perp} + D_{\parallel\parallel}(\mathbf{p})\frac{\partial}{\partial p_\parallel}\right]f(\mathbf{p})\right\}\\ &= \frac{1}{\sin\alpha}\frac{\partial}{\partial\alpha}\left\{\sin\alpha\left[D_{\alpha\alpha}(\mathbf{p})\frac{\partial}{\partial\alpha} + D_{\alpha p}(\mathbf{p})\frac{\partial}{\partial p}\right]f(\mathbf{p})\right\}\\ &\quad + \frac{1}{p^2}\frac{\partial}{\partial p}\left\{p^2\left[D_{p\alpha}(\mathbf{p})\frac{\partial}{\partial\alpha} + D_{pp}(\mathbf{p})\frac{\partial}{\partial p}\right]f(\mathbf{p})\right\},\end{aligned} \tag{8.18}$$

with the diffusion coefficients in either set of coordinates given by

$$D_{QQ'}(\mathbf{p}) = \sum_{s=-\infty}^{\infty}\int\frac{d^3\mathbf{k}}{(2\pi)^3}\, w_M(\mathbf{p},\mathbf{k},s)\,\Delta Q\,\Delta Q'\,N_M(\mathbf{k}), \qquad \Delta Q = \hat{D}_s Q,$$

$$\Delta p_\perp = \frac{s\Omega}{v_\perp}, \quad \Delta p_\parallel = \hbar k_\parallel, \quad \Delta\alpha = \frac{\hbar(\omega\cos\alpha - k_\parallel v)}{pv\sin\alpha}, \quad \Delta p = \frac{\hbar\omega}{v}. \tag{8.19}$$

The set of equations (8.17)–(8.19) is quite general, and these equations are used in several different ways in later lectures. The first application is to pitch angle scattering, when the most important term is that involving $D_{\alpha\alpha}$ in (8.18). This application is discussed in the next lecture.

Lecture 9

Resonant scattering

Resonant scattering causes Galactic cosmic rays to diffuse through the interstellar plasma and solar cosmic rays to diffuse through the interplanetary plasma. It also limits the trapping of particles in the Earth's radiation belts, and plays an essential ingredient in both diffusive shock acceleration and stochastic or Fermi acceleration. Indeed, resonant scattering is an important feature of the physics in any situation involving fast particles in a magnetic field, and this includes a wide class of situations of astrophysical interest. A feature of resonant scattering is that the particles themselves may generate the waves that cause the scattering.

9.1 Pitch angle scattering

Transfer of momentum between waves and particles occurs in resonant scattering, and under conditions identified below, the change in momentum can be primarily in direction rather than in magnitude. In this case, resonant scattering of fast particles may be described in terms of a diffusion in pitch angle α. Thus pitch angle scattering is described by an equation of the form

$$\frac{df}{dt} = \frac{1}{\sin\alpha}\frac{\partial}{\partial\alpha}\left(\sin\alpha\, D_{\alpha\alpha}\,\frac{\partial f}{\partial\alpha}\right), \tag{9.1}$$

where $D_{\alpha\alpha}$ is a pitch-angle diffusion coefficient. Scattering tends to make the particles isotropic. A distribution of particle, $f(p,\alpha)$, is *isotropic* for $\partial f(p,\alpha)/\partial\alpha = 0$, and for $\partial f(p,\alpha)/\partial\alpha \neq 0$ the distribution is *anisotropic*.

Resonant scattering of fast ions and relativistic electrons is due to Alfvén and magnetoacoustic waves, and resonant scattering of nonrelativistic electrons is due to whistlers. The scattering is due primarily to resonances at $s = \pm 1$. In these cases, the wave frequency is much less than the cyclotron frequency, so

that in the resonance condition (8.7) one has

$$s\Omega \approx -k_\parallel v_\parallel, \qquad \omega \ll \Omega. \tag{9.2}$$

These conditions require $|v_\parallel| \gg \omega/|k_\parallel|$, which requires $v \gg v_\phi$, where $v_\phi = \omega/k$ is the phase speed of the wave. Consequently for ions to resonate with Alfvén or magnetoacoustic waves one requires $v \gtrsim v_A$, where $v_\phi = v_A$ is the phase speed in this case.

One of the reasons that resonant scattering is so effective is that it allows transfer of momentum with little change in energy. To see this, consider the changes in $p_\parallel$ and $p_\perp$ when a particle emits a wave quantum. Let the change any quantity Q be denoted ΔQ. Using the semi-classical formalism, the changes are

$$\Delta p_\parallel = \hbar k_\parallel, \quad \Delta p_\perp = \hbar s\Omega/v_\perp, \quad \Delta\varepsilon = \hbar\omega. \tag{9.3}$$

The diffusion coefficient, D_{QQ}, in Q is $\propto (\Delta Q)^2$. The ratio of the pitch-angle scattering rate to the rate of energy transfer may be estimated from the ratio

$$D_{\alpha\alpha} : D_{pp}/p^2 = (\Delta\alpha)^2 : (\Delta p)^2/p^2. \tag{9.4}$$

Small changes in $p_\parallel = p\cos\alpha$ and $p_\perp = p\sin\alpha$ satisfy

$$\Delta p = \cos\alpha\Delta p_\parallel + \sin\alpha\Delta p_\perp, \quad p\Delta\alpha = -\sin\alpha\Delta p_\parallel + \cos\alpha\Delta p_\perp. \tag{9.5}$$

Then (9.3)–(9.5) imply

$$D_{\alpha\alpha} : D_{pp}/p^2 = (\omega\cos\alpha - k_\parallel v)^2 : (\omega\sin\alpha)^2. \tag{9.6}$$

One already requires $k_\parallel v \gg \omega$, and the ratio (9.6) is of order $(k_\parallel v/\omega)^2$, so that it is necessarily large. Thus efficient transfer of momentum, in the form of pitch-angle scattering, can occur with small transfer of energy between waves and particles.

Physically, large momentum transfer with small energy transfer occurs when the Lorentz force $q\mathbf{v}\times\mathbf{B}_\mathrm{w}$ due to the magnetic field in the wave is much larger than the electric force $q\mathbf{E}_\mathrm{w}$ due to the wave. The Lorentz force does no work, and only causes the direction of the momentum to change. With $\nabla\times\mathbf{E}_\mathrm{w} = -\partial\mathbf{B}_\mathrm{w}/\partial t$ implying $\mathbf{k}\times\mathbf{E}_\mathrm{w} = \omega\mathbf{B}_\mathrm{w}$ for a plane wave, this condition requires $kv \gg \omega$, which is essentially the same as the condition for the ω-term to be omitted from the resonance condition (9.2).

9.2 Scattering of ions by MHD waves

Resonant scattering of ions by Alfvén waves and by magnetoacoustic waves (assuming $v_A \gg c_s$) is possible for particles whose speed exceeds the phase speed of the waves, that is, for $v > v_A$. However, this is a necessary but not

a sufficient condition for resonance. To see that is is not a sufficient condition, consider the case of Alfvén waves, for which the dispersion relation is $\omega = |k_\parallel| v_A$. Then (9.2), with $\Omega = \Omega_i$ for the ion and with $s = \pm 1$, implies

$$|v_\parallel| = \frac{\Omega_i}{|k_\parallel|} = \frac{\Omega_i}{\omega}\, v_A. \tag{9.7}$$

The condition $v > v_A$ can only be satisfied for $\omega < \Omega_i$, and then (9.7) requires the stronger condition $v > (\Omega_i/\omega) v_A$. In fact Alfvén waves only exist for $\omega < \Omega_i$. As ω approaches Ω_i from below the dispersion relation becomes strongly modified from $\omega = |k_\parallel| v_A$ due to the strong interaction between the waves and the ions near the ion cyclotron frequency, as illustrated in Figure 9.1.

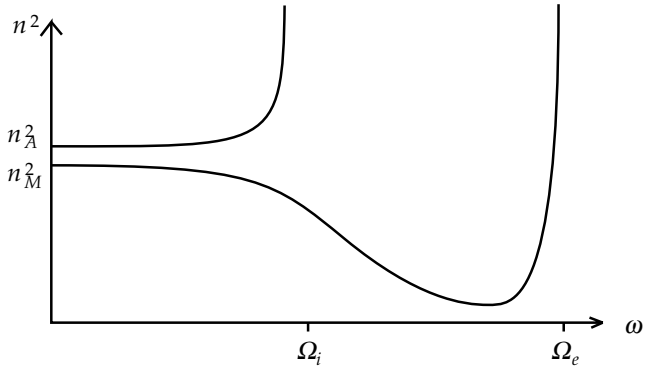


Figure 9.1: Schematic plots showing the dependence of the square of the refractive index on frequency for the Alfvén mode (upper curve) and the magnetoacoustic mode (lower curve). The Alfvén mode only exists for $\omega < \Omega_i$, and the magnetoacoustic mode at $\omega \lesssim \Omega_i$ becomes the whistler mode for $\Omega_i \lesssim \omega < \Omega_e$.

Relevant wave properties are for Alfvén waves (A) and magnetoacoustic waves (M) for $\omega \ll \Omega_i$ are

$$n_A^2 = \frac{c^2}{v_A^2 \cos^2\theta}, \quad \mathbf{e}_A = (1, 0, 0), \quad R_A = \frac{v_A^2}{2c^2}, \tag{9.8}$$

$$n_M^2 = \frac{c^2}{v_A^2 \cos^2\theta}, \quad \mathbf{e}_M = (0, i, 0), \quad R_M = \frac{v_A^2}{2c^2}. \tag{9.9}$$

These dispersion relations follow from (6.4) and (6.17) for $c_s^2 \ll v_A^2$, rewritten in terms of $n^2 = k^2 c^2/\omega^2$.

Quite generally, for waves in a magnetic field with $\mathbf{b}$ along the z-axis and $\mathbf{k}$ in the x-z plane, the y-component of the polarization vector is imaginary when the x- and z-components are real, and it is to maintain this general property that $\mathbf{e}_M$ is written as an imaginary vector in (9.9).

In treating the emission and absorption, a relevant simplifying assumption is the small gyroradius approximation, such that the argument of the Bessel functions is $\ll 1$ in the probability (8.12). (This assumption is not necessarily well justified, but adopting it should not lead to serious error.) Making this approximation, only $s = 0$ and $s = \pm 1$ give nonzero probabilities, cf. (8.12), and $s = 0$ is not relevant here. For $s = \pm 1$ (8.12) reduces to

$$\mathbf{V}(\mathbf{k}, \mathbf{p}; \pm 1) \approx \pm \tfrac{1}{2} v_\perp (1, \mp i\eta, 0), \tag{9.10}$$

$$w_M(\mathbf{p}, \mathbf{k},) = \frac{2\pi q^2 R_M(\mathbf{k})}{\varepsilon_0 \hbar \omega_M(\mathbf{k})} \left| \mathbf{e}_{Mx}(\mathbf{k}) \mp i\eta \mathbf{e}_{My}(\mathbf{k}) \right|^2 \delta[\omega_M(\mathbf{k}) - s\Omega - k_\parallel v_\parallel] \tag{9.11}$$

With these approximations, the simplest useful approximation for the probability for Alfvén waves is

$$w_{A,M}(\mathbf{p}, \mathbf{k}, s) = \frac{\pi q^2}{4\varepsilon_0} \frac{v_A^2}{c^2} \frac{v \sin^2 \alpha}{\hbar\omega |\cos\alpha \cos\theta|} \delta(k + s\Omega_0 / \gamma v \cos\alpha \cos\theta). \tag{9.12}$$

Then the absorption coefficient (8.17) for Alfvén waves reduces to

$$\gamma_A(\omega, \theta) = \frac{2\pi q^2 v_A^2}{\varepsilon_0 kc} \int_{-1}^{+1} d\cos\alpha \frac{\sin^2\alpha}{|\cos\alpha|} \times \left[\frac{p^2 v_A}{c} \left(\frac{\cos\theta}{|\cos\theta|} \frac{v}{v_A \sin\alpha} \frac{\partial}{\partial\alpha} - p \frac{\partial}{\partial p} \right) f(p, \alpha) \right]_{p=p_R} . \tag{9.13}$$

The absorption coefficient, $\gamma_M(\omega, \theta)$, for magnetoacoustic waves differs from (9.13) only in the omission of the factor $|\cos\theta|$.

The pitch angle diffusion coefficient (8.19) takes a simple form when evaluated for Alfvén waves with the foregoing approximation. Writing $W_A(k_\parallel) dk_\parallel$ for the energy density in the Alfvén waves in the range $k_\parallel$ to $k_\parallel + dk_\parallel$, one finds

$$D_{\alpha\alpha} = \frac{\pi q^2}{4\varepsilon_0 \gamma m p^2 |\cos\alpha|} W_A(k_R), \tag{9.14}$$

where $k_R = \Omega_i / \gamma p |\cos\alpha|$ is the resonant wavenumber.

9.3 Spatial diffusion

Resonant scattering causes diffusion in pitch angle. The relation between diffusion in pitch angle and spatial diffusion along the magnetic field lines includes some subtleties.

The left hand side of the equation (9.1) that describes diffusion in pitch angle may be written in the form

$$\left(\frac{\partial}{\partial t} + v\cos\alpha \frac{\partial}{\partial z} \right) f = \frac{1}{\sin\alpha} \frac{\partial}{\partial\alpha} \left(\sin\alpha \, D_{\alpha\alpha} \frac{\partial f}{\partial\alpha} \right), \tag{9.15}$$

where any force (other than due to the waves) on the particles is neglected. It is assumed that the only relevant spatial variation is along the z axis, and we seek a spatial diffusion equation of the form

$$\frac{df}{dt} = \frac{\partial}{\partial z}\left(\kappa \frac{\partial f}{\partial z}\right). \tag{9.16}$$

To relate (9.16) to (9.15) in a general way, one expands the pitch angle distribution in Legendre polynomials:

$$f(p,\alpha) = \sum_{n=0}^{\infty} f_n(p) P_n(\cos\alpha), \quad P_0(x) = 1, \quad P_1(x) = x, \ldots. \tag{9.17}$$

Here only the two leading terms in the expansion are retained. The evolution of $f_n(p)$ follows from (9.15) by inserting the expansion (9.17), multiplying by $P_n(\cos\alpha)$, integrating over $\cos\alpha$, and using the orthogonality relations for the Legendre polynomials. One finds

$$\frac{\partial f_0}{\partial t} + \frac{v}{3}\frac{\partial f_1}{\partial z} = 0, \quad \frac{\partial f_1}{\partial t} + \frac{v}{3}\frac{\partial f_0}{\partial z} = -\frac{1}{2} f_1 \int_{-1}^{+1} d\cos\alpha \, \sin^2\alpha \, D_{\alpha\alpha}, \tag{9.18}$$

and so on. Ignoring the term $\partial f_1/\partial t$, the second equation determines f_1. On substituting for f_1 in the first equation one obtains an equation of the form (9.16) with

$$\kappa = \frac{v^2}{9}\left[\frac{1}{2}\int_{-1}^{+1} d\cos\alpha \, \sin^2\alpha \, D_{\alpha\alpha}\right]^{-1}, \quad f_1 = -\frac{3\kappa}{v}\frac{\partial f_0}{\partial z}. \tag{9.19}$$

One subtle feature of resonant scattering causing spatial diffusion is that the pitch angle and spatial diffusion coefficients are essentially inverses of each other. This may be understood qualitatively by comparing a case where pitch angle scattering is weak with one where it is strong. For strong pitch angle scattering the particles repeatedly turn around on a short timescale, due to their diffusing from $0 < \alpha < \pi/2$ to $\pi/2 < \alpha < \pi$. In the presence of such scattering, particles can be confined to a relatively small region for a relatively long time. Hence a large spatial gradient may be maintained for a relatively long time. Now consider this case from the viewpoint of spatial diffusion. A large spatial gradient can persist against spatial diffusion only if the diffusion is weak. Strong spatial diffusion smears out gradients rapidly, and large gradients imply weak diffusion. Hence strong pitch angle diffusion implies weak spatial diffusion, and the converse is also valid. Another way of understanding this is in terms of the scattering mean free path. The mean free path is related to the distance in which the particles turn around, in the sense defined above. When pitch angle scattering is strong, this distance is short, and when pitch angle scattering is weak this distance is large. The spatial diffusion coefficient is proportional to the square of the mean free path, again implying weak spatial diffusion when the pitch angle scattering is strong.

9.4 Growth of waves due to a spatial gradient

As already remarked, one of the important features of resonant scattering is that the anisotropic particles themselves may generate the waves required to scatter them. An example that is relevant to diffusive shock acceleration involves a simple extension of the foregoing theory of spatial diffusion. This relates to the generation of the waves by particles with a spatial gradient.

That the resonant wave grow due to the effect of a spatial gradient may be seen as follows. On substituting the expansion (9.17) into (8.17), retaining only the two leading term in (9.17) and only $s = \pm 1$ in (8.17), and inserting the expression (9.19) for f_1, one finds

$$\gamma_M(\mathbf{k}) = -\sum_{s=\pm 1} \int d^3\mathbf{p}\; w_M(\mathbf{p},\mathbf{k},s)\,\hat{D}_s f(\mathbf{p}),$$

$$\hat{D}_s f(\mathbf{p}) = \frac{\hbar\omega}{v}\left(\frac{\cos\alpha - n_M\beta\cos\theta}{p\sin\alpha}\,\frac{3\kappa}{v}\,\frac{\partial f_0}{\partial z} + \frac{\partial f_0}{\partial p}\right). \tag{9.20}$$

In practice $\partial f_0/\partial p$ is negative and leads to a positive contribution to absorption; that is, the final term in (9.20) opposes growth of the waves that cause the resonant scattering. The relevant waves have large refractive indices, and for $n_M\beta \gg 1$ the term involving the spatial derivative contributes to wave growth for $\cos\theta(\partial f_0/\partial z) < 0$. Thus the waves that tend to grow propagate in the direction opposite to the gradient ($\cos\theta < 0$ for $\partial f_0/\partial z > 0$).

This example may be used to illustrate a general property of resonant scattering when the waves are generated by the anisotropic particles themselves: the momentum transfer from the particles to the waves is always such as to reduce the anisotropy of the particles. In the present case, the waves generated are propagating in the direction of the spatial gradient. Now, according to (9.19), a positive spatial gradient implies a negative f_1, and this implies that $f \approx f_0 - |f_1|\cos\alpha$ has more particles in the backward hemisphere ($\cos\alpha < 0$) than in the forward hemisphere ($\cos\alpha > 0$). The gradient is reduced by any process that reduces this excess, that is, any process that transfers particles from the backward to the forward hemisphere. Emission of waves in the forward direction reduces the backward momentum of the particles, and emission of waves in the backward direction reduces the forward momentum of the particles. Hence, the inferred backward emission reduces the anisotropy, and hence the spatial gradient. More generally, in any case where a maser operates, the emission tends to reduce the feature that is causing the wave growth. This reduction may be treated quantitatively using the quasilinear equation, and hence is sometimes called *quasilinear relaxation*.

Another subtle feature of resonant scattering and spatial diffusion relates to scattering between hemispheres. For a range of pitch angles about $\alpha = \pi/2$ the resonance condition cannot be satisfied for $s = \pm 1$. Hence, the mechanism as outlined above cannot account for particles reversing direction relative to the

magnetic field. This problem has long been recognized, and several different solutions to it have been suggested. One suggestion is that in the presence of variations in the magnitude of B (associated with fluctuations at much lower frequencies than are relevant for resonant scattering) the adiabatic invariant $p^2 \sin^2 \alpha / B$ allows particles with pitch angles close to $\alpha = \pi/2$ to be reflected by the familiar magnetic mirror effect. Irrespective of the detailed suggestion as to how this difficulty is overcome, observational data on ions and electrons in the solar wind suggest that there is little impediment to the particles being scattered in pitch angle across $\alpha = \pi/2$.

Lecture 10

Applications of pitch angle scattering

Some examples of astrophysical situations where efficient scattering occurs are discussed qualitatively in lecture 2. Here quantitative theories for resonant scattering in these situations are outlined.

10.1 Generation of Alfvén waves by streaming cosmic rays

As discussed in lecture 2, observations of galactic cosmic rays imply that they diffuse through the galaxy. This diffusion is attributed to resonant scattering by Alfvén and magnetoacoustic waves. These waves are generated by the streaming cosmic rays themselves.

A streaming motion may be described in terms of the $n = 1$ term in the expansion (10.17) of the distribution function in Legendre polynomials. Let $v_{\rm CR}$ be the streaming velocity for the cosmic rays; in general $v_{\rm CR}$ is a function of p. Then, on retaining only the terms $n \leq 1$ in (10.17), one has

$$f(p,\alpha) = f_0(p) + f_1(p)\cos\alpha, \quad f_0(p) = K_{\rm CR}\left(\frac{p}{p_0}\right)^{-b}, \quad f_1(p) = \frac{v_{\rm CR}}{v}\, f_0(p). \tag{10.1}$$

The spectrum $f_0(p)$ in (10.1) is the same as the spectrum (2.1), but written in a different notation. The relation between the energy spectrum and the distribution function is

$$N(\varepsilon)d\varepsilon = 4\pi f_0(p)\, p^2 dp. \tag{10.2}$$

Consequently, for relativistic particles, the power law indices a in (2.1) and b in (10.1) are related by $b = a + 2$, implying $b = 4.6$ for galactic cosmic rays. The

normalization constant $K_{\rm CR}$ is related to the number density of cosmic rays above a given energy, say above p_0:

$$n_{\rm CR}(p > p_0) = 4\pi \int_{p_0} p^2 dp\, f_0(p) = \frac{4\pi K_{\rm CR}\, p_0^2}{b-3}. \tag{10.3}$$

If one chooses $p_0 = m_p c = 0.94 \times 10^9\,{\rm eV}/c$, observations give

$$n_{\rm CR}(p > p_0) = 2 \times 10^{-10}\,{\rm cm}^{-3}, \qquad K_{\rm CR} = 2.5 \times 10^{-11} p_0^2\,{\rm cm}^{-3}. \tag{10.4}$$

The absorption coefficient (9.14) for Alfvén waves may be evaluated directly for the distribution (10.1). The absorption coefficient for magnetoacoustic waves differs from (9.14) only in the omission of the factor $|\cos\theta|$, and for present purposes this difference is unimportant and is ignored. The absorption coefficient reduces to

$$\begin{aligned}\gamma_{A,M}(\omega,\theta) &= -\frac{3(b-3)}{2b(b-2)}\,\frac{\pi q^2 v_A}{\varepsilon_0 k_0 p_0 c^2}\, n_{\rm CR}(p > p_0) \left(\frac{k|\cos\theta|}{k_0}\right)^{b-3} \left(\frac{\cos\theta}{|cos\theta|}\, v_{\rm CR} - \frac{b}{3}\, v_A\right) \\ &= -2.7 \times 10^{-7} \left(\frac{n_e}{1\,{\rm cm}^{-3}}\right)^{-1/2} \left(\frac{p}{p_0}\right)^{-1.6} \left(\frac{\cos\theta}{|cos\theta|}\,\frac{v_{\rm CR}}{c} - 1.5\,\frac{v_A}{c}\right). \end{aligned} \tag{10.5}$$

where in the latter form numerical values are inserted for galactic cosmic rays, and where $|\cos\theta|$ is approximated by unity.

It follows from (10.5) that the waves grow in the forward streaming direction ($\cos\theta > 0$ for $v_{\rm CR} > 0$) provided that the streaming speed exceeds a *neutral streaming speed*, $1.5 v_A$. Qualitatively, the neutral streaming speed may be attributed to the fact that the scattering tends to isotropize the particles in a frame in which the scattering centers are at rest. This frame corresponds to one moving with v_A when the scattering is due to MHD waves.

The growth rate (10.5) is fast enough to account for growth of the waves that resonate with lower energy cosmic rays. For example, for $n_e = 10^{-2}\,{\rm cm}^{-3}$ and $v_{\rm CR} - 1.5 v_A = 30\,{\rm km\,s}^{-1}$, one finds a growth time $1/\gamma_{A,M} \approx 200 (p/p_0)^{1.6}\,{\rm yr}$. Thus the waves needed to scatter the cosmic rays with $p \approx 10^9\,{\rm eV}/c$ grow in a few hundred years, which is less than the propagation time of free-streaming cosmic rays to the edge of the galactic disc. However, the growth time increases rapidly with p, and there is an approximate one-to-one relation between the resonant wave number k and the momentum of the resonant particles $kp = |q|B$. Hence, one would predict that there is insufficient time for the resonant waves to grow for particles with sufficiently high energy. For example, for $p \gg 10^{15}\,{\rm eV}/c$ the growth time exceeds the life time of the cosmic rays. It may be that there are other adequate sources for the resonant waves for the higher energy cosmic rays, and that it is not necessary for the cosmic rays to generate the waves themselves. However, the observational evidence is consistent with the cosmic rays of all energies being scattered efficiently, although the direct evidence for

this (from the observed isotropy of cosmic rays) is restricted to $p \lesssim 10^{15}\,\mathrm{eV}/c$. Hence one concludes that resonant waves are evidently present for all cosmic rays with $p \lesssim 10^{15}\,\mathrm{eV}/c$, although it clear that the waves have time to grow due to the streaming particles themselves only for low energy cosmic rays.

A point that should be mentioned concerns the structure of the interstellar medium, which consists of a hot ($> 10^6\,\mathrm{K}$), low density $n_e \approx 10^{-2}\,\mathrm{cm}^{-3}$ interconnected region, plus cooler, denser clouds of various types. The scattering of cosmic rays is usually assumed to occur only in the hot region. In cool, dense clouds, MHD waves damp rapidly (due to charge exchange between neutrals and ions), and the foregoing theory cannot account for scattering in such clouds.

10.2 Reduction of the streaming speed

The scattering of the cosmic rays by the resonant waves tends to reduce the streaming speed. This may be treated using quasilinear theory. The rate of change of the streaming velocity is given by

$$\frac{dv_{\mathrm{CR}}}{dt} = \frac{1}{2f_0(p)} \int_{-1}^{+1} d\cos\alpha \, \frac{df(p,\alpha)}{dt} \, v\cos\alpha, \tag{10.6}$$

with $df(p,\alpha)/dt$ given by (8.18). In pitch angle scattering only the term involving $D_{\alpha\alpha}$ is retained, as in (10.1), but it is relevant to retain the term involving $D_{\alpha p}$ in order to discuss the effect of the neutral streaming speed.

If one evaluates the coefficient $D_{\alpha\alpha}$, ignoring the term proportional to ω in the expression (8.19) for $\Delta\alpha$, the result (9.14) is obtained. On evaluating the coefficient $D_{\alpha p}$, one obtains

$$D_{\alpha p} = -\frac{\zeta - 1}{2} \, \frac{\cos\theta}{|\cos\theta|} \, \frac{v_A p \sin\alpha}{v} \, D_{\alpha\alpha}, \tag{10.7}$$

where ζ is the fraction of the waves propagating in the forward direction, with $1 - \zeta$ propagating in the backward direction. The waves generated by the streaming particles themselves are propagating in the forward direction.

Using (10.7), (10.6) reduces to

$$\frac{dv_{\mathrm{CR}}}{dt} = \nu_s \left(v_{\mathrm{CR}} - \frac{\zeta - 1}{2} \, \frac{|\cos\theta|}{cos\theta} \, \frac{b}{3} \, v_A \right), \tag{10.8}$$

with the effective scattering frequency given by

$$\nu_s = 3 \int_{-1}^{+1} d\cos\alpha \, \sin^2\alpha \, D_{\alpha\alpha}, \tag{10.9}$$

with $D_{\alpha\alpha}$ given by (9.14).

On applying (10.8) with (10.9) to the scattering of galactic cosmic rays, one encounters a practical difficulty in that one has no information on the spectrum MHD waves in the interstellar medium. In this case one expects the waves to grow until the scattering rate is fast enough to reduce the streaming speed and hence the growth rate. Thus, saturation should occur when $\gamma_{A,M}$ and ν_s are comparable. This may be used to estimate the expected level of the turbulence.

On the other hand, on applying this theory to scattering of fast particles in the interplanetary plasma, one has quite detailed knowledge of the spectrum of MHD turbulence. It is a power law spectrum, with power-law index approximately equal to 5/3, corresponding to a *Kolmogorov spectrum*:

$$W(k) \propto k^{-5/3}. \tag{10.10}$$

In this case there is reasonable quantitative agreement between the predictions of quasilinear theory and the observations of diffusing fast particles. However, the comparison is not straightforward from an observational viewpoint because the scattering waves and the scattered particles can be sampled only at isolated points (the locations of spacecraft), and this precludes a detailed comparison.

10.3 Scattering of electrons by whistlers

Another application of resonant scattering discussed in lecture 2 is to the electrons and ions trapped in the Earth's radiation belts. In order to develop a quantitative treatment appropriate to this application, we need to extend the theory of resonant scattering to include electrons.

Consider the possibility of an electron resonating with an Alfvén wave. Repeating the argument leading to (10.7) for an electron rather than an ion, one finds that resonance is possible only for $v > (\Omega_e/\omega)v_A$. Then $\omega < \Omega_i$ for Alfvén waves requires $v > (\Omega_e/\Omega_i)v_A$, that is $v > (m_i/m_e)v_A \approx (43)^2 v_A$. It follows that only very fast electrons can resonate with Alfvén waves. However, it is possible for slower electrons, in the range $43v_A \lesssim v \lesssim (43)^2 v_A$ to resonate with whistlers. As shown schematically in Figure 9.1, the magnetoacoustic mode joins on continuously to the whistler mode which exists in the frequency range $\Omega_i < \omega < \Omega_e$.

The properties of whistler waves may be derived using the magnetoionic theory, in which the electrons are treated as a cold gas. (The ions are usually assumed to form a uniform positively charged background, but in fact they play no role whatsoever.) The magnetoionic theory gives the following dispersion relation for whistler waves:

$$n_W^2 = 1 + \frac{\omega_p^2}{\omega(\Omega_e|\cos\theta| - \omega)}, \tag{10.11}$$

where ω_p is the plasma frequency,

$$\omega_p = \left(\frac{e^2 n_e}{\varepsilon_0 m_e}\right)^{1/2}. \tag{10.12}$$

For present purposes, only the regime $\omega \ll \Omega_e|\cos\theta|$ is considered, and in the plasmas of interest the unit term in (10.11) is then negligible. With these approximations, the properties of whistlers reduce to

$$n_W^2 = \frac{\omega_p^2}{\omega\Omega_e|\cos\theta|}, \quad \mathbf{e}_W = \frac{(1, i|\cos\theta|, 0)}{(1+\cos^2\theta)^{1/2}}, \quad R_W = \frac{\omega\Omega_e}{2\omega_p^2}\frac{1+\cos^2\theta}{|\cos\theta|}. \tag{10.13}$$

The probability of emission for electrons and whistlers is obtained by inserting (10.13) in (8.12), with $\eta = -1$ for electrons, this gives

$$w_W(\mathbf{p}, \mathbf{k}, s) = \frac{\pi q^2}{4\varepsilon_0}\frac{\Omega_e}{\hbar\omega_p^2}\frac{v\sin^2\alpha}{|\cos\alpha|}\frac{(1+s|\cos\theta|)^2}{\cos^2\theta}\,\delta(k + s\Omega_e/\gamma v\cos\alpha\cos\theta). \tag{10.14}$$

The minimum speed for an electron to resonate with a whistler follows from the δ-function in (10.14), together with the dispersion relation (10.13), which determined $k = n_W\omega/c$. One finds that resonance is possible only for $v > 43v_A$. For such electrons, growth is possible provided that the absorption coefficient is negative. The specific form of the absorption coefficient for whistlers due to anisotropic electrons follows from (8.17). After summing over $s = \pm 1$ this gives

$$\gamma_W(\omega) = \frac{\pi^2 k}{2n_e}\int_{-1}^{+1} d\cos\alpha\,\sin^2\alpha\left[p^3 v\left((1+\cos^2\theta) - 2\cos\theta\frac{\cos\alpha}{|\cos\alpha|}\right)\right.$$
$$\left.\times\left(\frac{\cos\theta}{\sin\alpha}\frac{\partial}{\partial\alpha} - \frac{43v_A}{v|\cos\alpha|}p\frac{\partial}{\partial p}\right)f(p,\alpha)\right]_{p=p_R}, \tag{10.15}$$

The conditions for growth due to a streaming distribution are similar to those for streaming ions generating Alfvén waves. However, these is a subtle restriction on growth of whistlers due to anisotropic non-streaming distribution of electrons, that is, one that is an even function of $\cos\alpha$. Such distributions may be classified into two types, a *positive anisotropy*, with $\partial f/\partial\sin\alpha > 0$, and a *negative anisotropy*, with $\partial f/\partial\sin\alpha < 0$. For anisotropic ions and MHD waves both types of distribution lead to wave growth, but for electrons and whistlers only a positive anisotropy can lead to wave growth. This is due to a combination of two effects. First, the contribution from $s = -1$ is always smaller than that due to $s = +1$, because of the factor $(1 + s|\cos\theta|)^2$ in (10.14). Second, emission at $s > 0$ reduces $p_\perp$, and so can be driven by a distribution with and excess (over an isotropic distribution) of $p_\perp$, and emission at $s < 0$ increases $p_\perp$, and so can be driven by a distribution with a deficiency of $p_\perp$. Thus, a positive anisotropy leads to instability only through emission at $s = +1$, and a negative anisotropy leads to instability only through emission at $s = -1$, but the latter is not possible for electrons interacting with whistlers.

10.4 Scattering of particles in the magnetosphere

Historically, resonant scattering of particles in the magnetosphere was the first (in the early 1960s) suggested application of resonant scattering. The scattering of low energy ions was attributed to ion cyclotron waves, and the scattering of electrons to whistlers. (The ion cyclotron waves are on the rising section of the dispersion curve for Alfvén waves in Figure 9.1.)

Consider the model for the acceleration of particles in the Earth's radiation belts described in lecture 2. The particles diffuse inward due to violation of the third adiabatic invariant. As they do so they become increasingly anisotropic, due to $p_\perp$ increasing $\propto L^{-3/2}$, and $p_\parallel$ increasing $\propto L^{-1}$ as L decreases. (Recall that L is the distance to the center of the field line in units of the planetary radius.) Hence the particles develop a positive anisotropy. The number density of the energetic particles also tends to increase as L decreases. The reason for this is that the diffusion tends to make the particle distribution, f, uniform in the phase space, implying $n \propto p_\perp^2 p_\parallel f$; conservation of $M \propto p_\perp^2/B$ and $J \propto p_\parallel L$ implies $p_\perp^2 p_\parallel \propto MJB/L \propto L^{-4}$ for M and J constant. Thus the density tends to build up close to the Earth toward a distribution of particles with number density $\propto L^{-4}$. As the growth rate of the resonant waves is proportional to the number density, the growth rate must also increase with decreasing L. Hence, at sufficiently small L one expects the resonant waves to grow fast enough to scatter the particles efficiently. Particles must be scattered to smaller $\sin\alpha$, because this reduces the anisotropy that causes the wave growth. Hence particles are scattered into the loss cone, precipitate into the neutral atmosphere of the Earth and are lost.

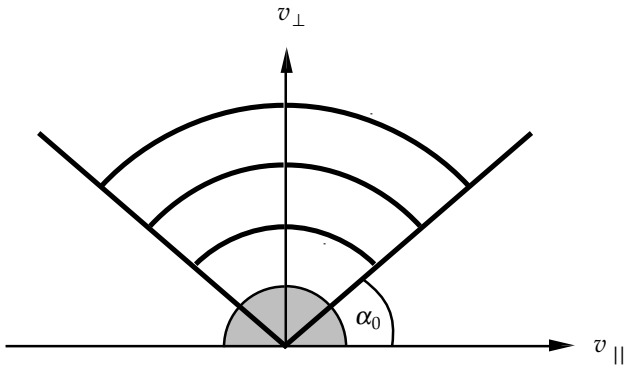


Figure 10.1: The idealized loss-cone distribution (10.16) is illustrated; α_0 is the loss cone angle, the shaded region is filled by isotropic thermal particles, and the circular arcs denote contours of constant f.

This wave growth and associated scattering may be treated semiquantita-

tively. Consider an electron distribution with an idealized loss cone of the form

$$f(p,\alpha) = f_0(p)[H(\alpha-\alpha_0) + H(\pi-\alpha_0-\alpha)], \tag{10.16}$$

where $H(x)$ is the step function. That is, (10.16) corresponds to no particles in loss cones $0 < \alpha < \alpha_0$ and $\pi/2 - \alpha_0 < \alpha < \pi/2$, with the distribution independent of α for $\alpha_0 < \alpha < \pi/2 - \alpha_0$, as illustrated in Figure 10.1. In this case the growth comes from the derivatives of (10.16) with respect of α:

$$\frac{\partial f}{\partial \alpha} = f_0\,[\delta(\alpha-\alpha_0) - \delta(\pi-\alpha_0-\alpha)], \tag{10.17}$$

The first of these δ-functions causes waves at $\theta > \pi/2$ to grow, and the second causes waves at $\theta < \pi/2$ to grow. In either case, the absorption coefficient may be approximated by

$$\gamma_W \approx \frac{n_1}{n_e}\,\Omega_e \left[\alpha_0^2 - b\left(\frac{43 v_A}{v}\right)^2\right], \tag{10.18}$$

where n_1 is the number density of the fast particles. The waves grow provided the growth time is shorter than the effective loss time for the whistlers. The main loss mechanism is simply propagation of the whistlers out of the region where the growth occurs. The distance they need to propagate to escape from the growth region is of order $r = LR_E$. The group velocity $\mathbf{v}_{gW}$ for whistlers is approximately along the field lines, so that the escape time for the whistlers is

$$T_W \approx \frac{LR_E}{v_{gW}}, \qquad v_{gW} \approx c\left(\frac{\omega_p^2}{\omega\Omega_e}\right)^{1/2} \approx \frac{\omega_p^2}{\Omega_e^2}\,v, \tag{10.19}$$

where in the final expression the resonance condition and the dispersion relation are used to eliminate ω in favor of the speed v of the resonant electrons. Growth is effective only for $\gamma_W T_W \gg 1$.

Once the whistlers have been generated they scatter particles into the loss cone, so that the loss cone is no longer empty. Suppose that the escape rate for particles in the loss cone is ν_L, with $\nu_L \approx v/\ell$, where ℓ is characteristic of the length of the flux tube. Then pitch angle diffusion within the loss cone may be described by (9.1) for small α. On balancing the diffusion by the loss term one obtains, writing $D_{\alpha\alpha} = D$,

$$\frac{1}{\alpha}\frac{\partial}{\partial\alpha}\left(\alpha D\frac{\partial f}{\partial\alpha}\right) - \nu_L f = 0, \tag{10.20}$$

which applies for $\alpha < \alpha_0$. The general solution of (10.20) may be expressed in terms of modified Bessel functions $I_\nu(x)$ of order $\nu = 0$ and argument $x =$

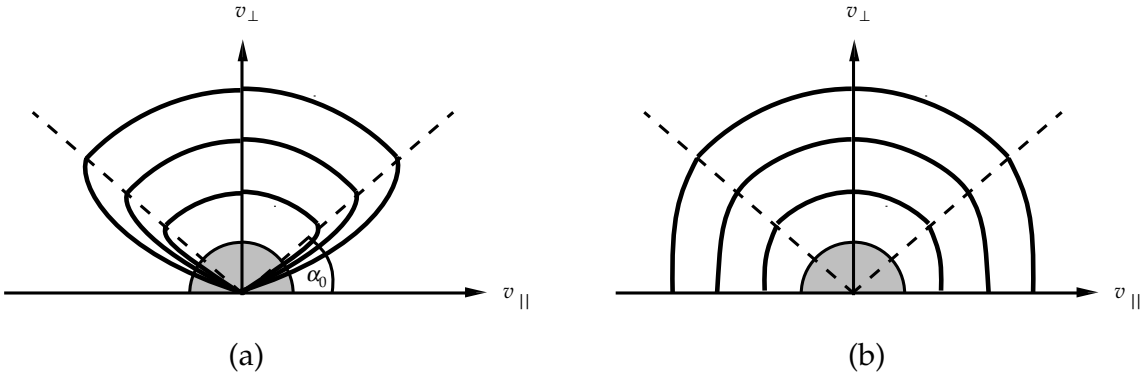


Figure 10.2: The distribution inside the loss cone is illustrated for (a) weak diffusion and (b) strong diffusion. In weak diffusion particles scattered into the loss cone have a high probability of being lost due to precipitation, so that the loss cone is almost empty. In strong diffusion particles that diffuse into the loss cone have a higher probability of diffusing out again than of precipitating, so that the loss cone is almost full.

$\alpha[\nu_L/D]^{1/2}$. On requiring that the distribution function be equal to f_0 at $\alpha = \alpha_0$, one obtains

$$f = f_0 \frac{I_0(\alpha[\nu_L/D]^{1/2})}{I_0(\alpha_0[\nu_L/D]^{1/2})}. \tag{10.21}$$

The solution (10.21) is qualitatively different in two limiting cases

$$\begin{aligned} &\text{weak diffusion:} && D \ll \nu_L \alpha_0^2, \\ &\text{strong diffusion:} && D \gg \nu_L \alpha_0^2, \end{aligned} \tag{10.22}$$

which correspond to the limits of large and small arguments for the modified Bessel function, with $I_0(x) \approx e^x/(2\pi x)^{1/2}$ for $x \gg 1$ and $I_0(x) = 1 + x^2/4 + \ldots$ for $x \ll 1$. These two cases are illustrated in Figure 10.2. In the case of weak diffusion the loss cone is nearly empty, and in the case of strong diffusion the loss cone is almost full. In effect, in weak diffusion, particles diffuse slowly, and when a particle diffuses into the loss cone it is much more likely to precipitate than to diffuse out again. On the other hand, in the case of strong diffusion, particles diffuse in and out of the loss cone on a time scale shorter than the loss time $1/\nu_L$.

This theory is satisfactory from a semiquantitative viewpoint in that it accounts reasonably for the properties of the electrons trapped in the radiation belts. The whistler waves are also observed. However, there is a qualitative difference between theory and observation. The theory implies that random phase whistler noise should be generated. Observationally, the precipitating electrons

correlate with narrow band whistlers, referred to an VLF emissions. These have surprising properties, one notable one being that they can be triggered by terrestrial radio noise. They can be triggered by a Morse code 'dash' (150 ms) but not by a a Morse code 'dot' (50 ms). They appear to be continuously triggered by high harmonics of the AC frequency (50 Hz or 60 Hz in North America) produced as noise by large electricity generating plants. The interpretation of such narrow band, triggered VLF emissions remains a challenging problem in magnetospheric physics.

Lecture 11

Stochastic acceleration

Acceleration of fast particles occurs in a wide variety of contexts in astrophysical plasmas and there is a correspondingly wide variety of possible acceleration mechanisms. Eight different mechanisms that will be discussed in this and later lectures are:

1. Stochastic Acceleration
2. Diffusive Shock Acceleration
3. Shock Drift Acceleration
4. Resonant Acceleration by MHD Waves
5. Resonant Acceleration by Longitudinal Waves
6. Runaway Acceleration
7. Acceleration by Potential Double Layers
8. Acceleration during Magnetic Reconnection

The first of these was the first discussed from an historical viewpoint, and is the subject of this lecture.

11.1 Energy changing mechanisms

A *stochastic acceleration* mechanism involves three ingredients. The first ingredient is an energy changing mechanism, that is, a process that causes the energy of a particle to increase or decrease. All relevant energy changing processes may be attributed to electric fields, and different energy changing processes may be interpreted in terms of different ways in which the localized electric fields are set up. In stochastic acceleration, as opposed to systematic acceleration, both increases and decreases in energy occur, and the net acceleration is due to a small difference between the gains and losses. The second ingredient is an effective scattering process. This is needed because, in practice, the energy changes tend to make the particle distribution anisotropic in such a way as to impede

and then prevent further net energy gains. The third ingredient is a statistical theory to describe the net effect of these processes.

There are two important energy changing mechanisms, and a third that needs to be considered for completeness. These three may be described as the betatron effect, the Fermi mechanism and transit time damping. These were identified before the modern theory of stochastic acceleration was developed in the late 1960s.

The betratron effect, which is also known as magnetic pumping, was first discussed by Swann in the 1933. The energy changes result from changes in the magnitude of the magnetic field. Such changes imply an induced electric field through $\nabla \times \mathbf{E} = -\partial \mathbf{B}/\partial t$, and it is this electric field that does work in accelerating particles. It is simpler to interpret the acceleration in terms of the conservation of the adiabatic invariant $p_\perp^2/B$ than in terms of the electric field. For purely temporal variations in B, as B increases $p_\perp^2$ increases, and as B decreases $p_\perp^2$ decreases, with $p_\parallel$ unchanged. Hence as $B(t)$ varies relative to a fiducial value B_0, the energy of the particle varies according to

$$\varepsilon(t) = [m^2c^4 + p_\parallel^2 c^2 + p_{\perp 0}^2 c^2 B(t)/B_0]^{1/2}, \tag{11.1}$$

where one has $p_\perp = p_{\perp 0}$ at $B(t) = B_0$. Note however that if the variations in the magnetic field are periodic then the changes in energy are also periodic. Swann's original idea involved a single large increase in B, but this is not a realistic mechanism in most circumstances, and magnetic pumping is attributed to small amplitude quasiperiodic changes in B. A scattering mechanism is then required to prevent the energy changes being quasiperiodic.

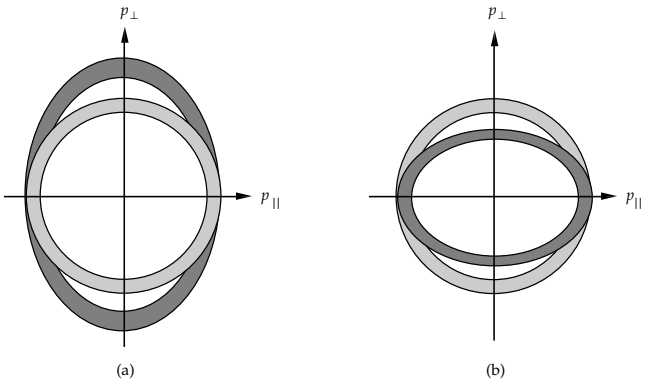


Figure 11.1: The effect of magnetic pumping on a distribution of particles is illustrated. The lightly shaded region is a section of an isotropic distribution that is deformed into the darkly shaded region by (a) a compression, and (b) a rarefaction.

Consider the effect of pitch angle scattering on a distribution of particles

subject to magnetic pumping. At $t = 0$, writing $B(0) = B_0$, suppose that the distribution function $f(p_\perp, p_\parallel)$ is isotropic, so that the contours of f in momentum space are spheres, corresponding to circles in $p_\perp$-$p_\parallel$ space. As $B(t)$ increases the distribution function becomes anisotropic; the circles are deformed by being elongated along the $p_\perp$-axis, as illustrated in Figure 11.1. Any scattering tends to reduce this anisotropy, making the contours of constant f closer to circles, with a larger radius than the initial circle. Now consider a later time when $B(t)$ has gone through its maximum and is decreasing. In the absence of scattering the contours of constant f would become more circular, and then for $B(t) < B_0$ would become shortened along the $p_\perp$-axis. However, in the presence of scattering, the contours remain nearly circular, and pass from being slightly elongated along the $p_\perp$-axis as B approaches maximum to slightly shortened along the $p_\perp$-axis after B passes maximum. One then finds that the increase in average particle energy as B increases is greater in magnitude than the decrease in average particle energy as B decreases. Over one cycle of the changes in B there is a net energy gain by the particles.

The energy transfer to the particles in magnetic pumping may be attributed to scattering causing a viscous or frictional drag on the particle distribution in momentum space. As a result of this drag pulsations in the magnetic field tend to damp. The pulsations need to do work on the particles when compressing them, and as a result of the scattering some of this energy is randomized in particle motions and cannot be regained by the magnetic field in a subsequent rarefaction phase.

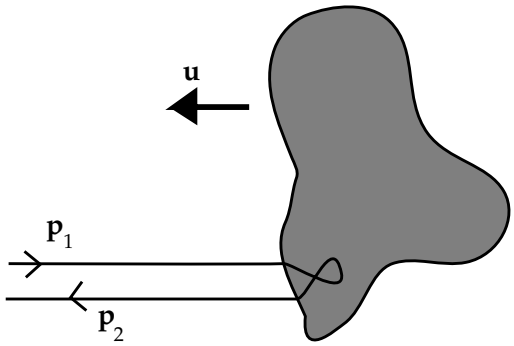


Figure 11.2: Fig. 11.2
A cosmic ray moving along a uniform intercloud magnetic field in the horizontal direction is reflected by the tangled magnetic field inside the cloud (shaded region).

The other important energy changing mechanism was first pointed out by Fermi in 1949 and is called the Fermi mechanism. The specific problem dis-

cussed by Fermi was the acceleration of cosmic rays by bouncing off magnetized interstellar clouds. Consider a cloud with velocity $\mathbf{u}$ relative to an average rest frame for the interstellar medium. The magnetic field in the cloud can reflect a cosmic ray, as illustrated in Figure 11.2. The energy changes in this case may be attributed to the fact that a particle reflecting from an approaching surface gains energy, and a particle reflecting from a receding surface loses energy. This may be treated quantitatively by noting that there is no energy change in a frame in which the reflecting surface is at rest, and then making a Lorentz transformation to the laboratory frame in which the surface is moving with velocity $\mathbf{u}$. Let the particle properties be labeled 1 before and 2 after reflection, and let the properties in the rest frame of the reflecting surface be denoted by primes. Then one has $\varepsilon_2' = \varepsilon_1'$, $\mathbf{p}_2' \cdot \mathbf{u} = -\mathbf{p}_1' \cdot \mathbf{u}$. One then has

$$\varepsilon_{1,2} = \Gamma(\varepsilon_{1,2}' - \mathbf{p}_{1,2}' \cdot \mathbf{u}), \qquad \mathbf{p}_{1,2} \cdot \mathbf{u} = \Gamma(\mathbf{p}_{1,2}' \cdot \mathbf{u} - \varepsilon_{1,2}' u^2/c^2), \tag{11.2}$$

with $\Gamma = (1 - u^2/c^2)^{1/2}$. The energy change is then

$$\Delta\varepsilon = \varepsilon_2 - \varepsilon_1 = 2\Gamma(\varepsilon_1 u^2/c^2 + \mathbf{p}_1 \cdot \mathbf{u}). \tag{11.3}$$

Energy gains result from head on collisions ($\mathbf{p}_1 \cdot \mathbf{u} > 0$) and energy losses result from overtaking collisions ($\mathbf{p}_1 \cdot \mathbf{u} < 0$). However, the second order (in u/c) changes do not average to zero. One second order change is due to the term $\varepsilon_1 u^2/c^2$ in (11.3).

There is another second order change due to the probability of a head-on reflection being higher than the probability of an overtaking reflection. Let the mean distance traveled between clouds be L. Then the rate of collisions is

$$\nu = \frac{|\mathbf{v} - \mathbf{u}|}{L}, \tag{11.4}$$

which is larger for a head-on collision ($\mathbf{v} \cdot \mathbf{u} < 0$) than for an overtaking collision ($\mathbf{v} \cdot \mathbf{u} > 0$). A statistical treatment shows that the net effect of the second order changes is to cause diffusion in momentum space, as shown in the next section.

The third energy changing mechanism is transit time damping. One form of transit acceleration is when particles diffuse along magnetic field lines, due to resonant scattering as described by (9.16), through a spatially varying magnetic field on a time scale shorter than that for temporal variation in the magnetic field. The diffusing particles gain and lose energy in a manner similar to the betatron effect.

These various processes were identified originally as distinct processes, and were initially considered in contexts where the energy change of a particle in any given interaction is small but significant. However, it turns out that all these processes are related when one considers acceleration by MHD turbulence. In the presence of MHD turbulence all the energy changes are minuscule, but they are also frequent, and it turns out that the net acceleration due to frequent small changes provides more effective acceleration in practice than fewer, larger but less frequent energy changes.

11.2 Diffusion in momentum space

The characteristic feature of a stochastic acceleration mechanism is that the acceleration may be described by a diffusion equation in momentum space. That is, the average (over pitch angle) distribution function evolves according to an equation of the form

$$\frac{df(p)}{dt} = \frac{1}{p^2}\frac{\partial}{\partial p}\left[p^2 D(p)\frac{\partial f(p)}{\partial p}\right], \tag{11.5}$$

where $D(p)$ is the diffusion coefficient, denoted D_{pp} in the notation used in (8.18) for example. When pitch angle scattering is included the energy changes due to both the betatron effect and to the Fermi mechanism imply such diffusion in momentum space. However, it is not obvious that this is the case, and it is relevant to show that simple models for the betatron effect and for Fermi acceleration do indeed lead to (11.5).

First consider the betatron effect. The changes in the distribution of particles may be described analytically by noting that the changes in f result from two terms, one due to the magnetic changes and the other due to pitch angle scattering. The changes due to changes in B follow from $d(p_\perp^2/B)dt = 0$ giving an expression for $dp_\perp/dt$ in terms of dB/dt, and one then identifies the rate of change of f as $-\nabla\cdot_{\mathbf{p}}(d\mathbf{p}/dt\, f)$, with the divergence, $\nabla\cdot_{\mathbf{p}}$, in momentum space evaluated in cylindrical polar coordinates. Including the pitch angle scattering as in (9.1) then gives

$$\frac{df}{dt} = -\frac{1}{p_\perp}\frac{\partial}{\partial p_\perp}\left(p_\perp^2\,\frac{\dot{B}}{2B}\,f\right) + \frac{1}{\sin\alpha}\frac{\partial}{\partial\alpha}\left(\sin\alpha\, D_{\alpha\alpha}\,\frac{\partial f}{\partial\alpha}\right), \tag{11.6}$$

with $\dot{B} = dB/dt$. Converting fully to spherical polar coordinates involves inserting

$$-\frac{1}{p_\perp}\frac{\partial}{\partial p_\perp}\left(p_\perp^2\,\frac{\dot{B}}{2B}\,f\right) = -\frac{1}{p^2}\frac{\partial}{\partial p}\left(p^3\sin^2\alpha\,\frac{\dot{B}}{2B}\,f\right) - \frac{1}{\sin\alpha}\frac{\partial}{\partial\alpha}\left(\sin^2\alpha\cos\alpha\,\frac{\dot{B}}{2B}\,f\right) \tag{11.7}$$

into (11.6). The scattering term in (11.6) is too cumbersome for our immediate purposes, and is approximated here by

$$\frac{1}{\sin\alpha}\frac{\partial}{\partial\alpha}\left(\sin\alpha\, D_{\alpha\alpha}\,\frac{\partial f}{\partial\alpha}\right) \to \nu_s(f - f_0), \tag{11.8}$$

where ν_s is a scattering frequency, and where f_0 is the average (over pitch angle) distribution function.

A solution of (11.6) with (11.8) may be found by introducing two time scales: a fast time scale, $1/\nu_s$, on which scattering keeps the distribution nearly

isotropic, and a slow time, $B/\dot{B}$, on which the magnetic field changes. On the fast time scale (11.6) with (11.8) may be approximated by

$$0 = -\sin^2\alpha\,\frac{\dot{B}}{2B}\,p\frac{\partial f_0}{\partial p} - \nu_s\,f_1, \tag{11.9}$$

and on the slow time scale by

$$\frac{d\langle f\rangle}{dt} = \left\langle -\frac{1}{p^2}\frac{\partial}{\partial p}\left(p^3\sin^2\alpha\,\frac{\dot{B}}{2B}\,f_1\right)\right\rangle, \tag{11.10}$$

where the average is over the pitch angle distribution. Together (11.9) and (11.10) imply an equation of the form (11.5) with

$$D(p) = \frac{2}{15\nu_s}\left(\frac{\dot{B}}{B}\right)^2 p^2. \tag{11.11}$$

Apart from the numerical coefficient, when depends on the detailed assumptions, (11.11) in (11.5) is the standard result for magnetic pumping.

To treat Fermi acceleration it is convenient to use a Fokker-Planck approach, which is used in another context in lecture 3. The Fokker-Planck equation for energy changes is

$$\frac{dN(\varepsilon)}{dt} = -\frac{\partial}{\partial\varepsilon}\left[\left\langle\frac{d\varepsilon}{dt}\right\rangle N(\varepsilon)\right] + \frac{1}{2}\frac{\partial^2}{\partial\varepsilon^2}\left[\left\langle\frac{d\varepsilon^2}{dt}\right\rangle N(\varepsilon)\right], \tag{11.12}$$

where the quantities in angular brackets are Fokker-Planck coefficients. A simple model that enables one to calculate the Fokker-Planck coefficients is when the magnetic field between the clouds is uniform. Let the momentum, $\mathbf{p}$, of a particle be described by p, α and ϕ, and the velocity, $\mathbf{u}$, of cloud by u, θ, ϕ'. Then one has

$$\left\langle\frac{d\varepsilon}{dt}\right\rangle = \langle\nu\Delta\varepsilon\rangle, \qquad \left\langle\frac{d\varepsilon^2}{dt}\right\rangle = \langle\nu(\Delta\varepsilon)^2\rangle, \tag{11.13}$$

with $\Delta\varepsilon$ given by (11.3) and ν by (11.4). In (11.3) one omits the subscript 1 and writes $\mathbf{p}\cdot\mathbf{u} = pu(\cos\alpha\cos\theta + \sin\alpha\sin\theta\cos(\phi-\phi')$, and in (11.4) one writes $|\mathbf{v}-\mathbf{u}| = |v\cos\alpha + u\cos\theta|$. The averages in (11.13) are then interpreted in terms of averages over the angles:

$$\langle\ \rangle \to \frac{1}{(4\pi)^2}\int_0^{2\pi} d\phi\int_0^{2\pi} d\phi'\int_{-1}^{+1} d\alpha\int_{-1}^{+1} d\theta. \tag{11.14}$$

The Fokker-Planck coefficients then reduce to

$$\left\langle\frac{d\varepsilon}{dt}\right\rangle = \frac{4u^2}{3Lc^2}\,v\varepsilon, \qquad \left\langle\frac{d\varepsilon^2}{dt}\right\rangle = \frac{2u^2}{3Lc^4}\,v^3\varepsilon^2, \tag{11.15}$$

On substituting the forms (11.15) into (11.12), one finds that the result may be rewritten in the form (11.5) with

$$D(p) = \frac{u^2 p^2}{3Lv}, \tag{11.16}$$

which is the standard result for Fermi acceleration.

11.3 Stochastic acceleration due to MHD turbulence

Acceleration can occur due to MHD turbulence in two different ways, one of which is a form of stochastic acceleration. The other, discussed briefly at the end of this section, is due to gyroresonant absorption.

Stochastic acceleration by MHD turbulence is due to a combination of two effects: damping at harmonic number $s = 0$, plus resonant scattering tending to maintain the distribution isotropic. According to (8.18) with (8.19), the effect on the particles of a resonant interaction at $s = 0$ is of the form

$$\frac{df(\mathbf{p})}{dt} = \frac{\partial}{\partial p_\parallel}\left[D_{\parallel\parallel}(\mathbf{p})\frac{\partial f(\mathbf{p})}{\partial p_\parallel}\right],$$
$$D_{\parallel\parallel}(\mathbf{p}) = \int \frac{d^3\mathbf{k}}{(2\pi)^3}\, w_M(\mathbf{p},\mathbf{k},0)\,\hbar^2 k_\parallel^2\, N_M(\mathbf{k}). \tag{11.17}$$

The probability (8.12) for $s = 0$ simplifies considerably when one (a) makes the small gyroradius approximation, and (b) inserts the wave properties (9.8) and (9.9) for Alfvén waves. The first point to note is that the probability is identically zero for Alfvén waves, due to the polarization vector (9.8) being along the x-axis, and the quantity $\mathbf{V}(\mathbf{k},\mathbf{p};s)$ having zero component along the x-axis for $s = 0$. Hence this form of damping applies only to magnetoacoustic waves. For such waves (8.12), with $J_0'(z) \approx z$ for $z \ll 1$, and with the wave properties (9.9), reduces to

$$w_M(\mathbf{p},\mathbf{k},0) \approx \frac{\pi v_A v_\perp^2 p_\perp^2}{2\hbar|v_\parallel|(B^2/2\mu_0)}\left(1 - \frac{v_A^2}{v_\parallel^2}\right)\delta(\cos\theta - v_A/v_\parallel). \tag{11.18}$$

Then for an isotropic distribution of waves, one finds

$$D_{\parallel\parallel}(\mathbf{p}) = \frac{\pi\bar{\omega}}{4}\,\frac{W_M}{B^2/2\mu_0}\,\frac{p^2 v_A}{v}\,\frac{\sin^4\alpha}{|\cos^3\alpha|}\left(1 - \frac{v_A^2}{v^2\cos^2\alpha}\right), \tag{11.19}$$

where W_M is the energy density in the waves, and $\bar{\omega}$ is their mean frequency. The diffusion (11.17) causes the particle distribution to become increasingly

anisotropic, tending to $\partial f/\partial p_{\|} = 0$. This tends to suppress the absorption of the waves, and hence to limit the acceleration.

In the presence of efficient scattering the distribution remains nearly isotropic. On rewriting (11.17) in terms of spherical polar coordinates, cf. (8.18), then averaging over pitch angle assuming that the distribution is nearly isotropic, one obtains an equation of the form (11.5) with $D(p) = \langle D_{pp} \rangle$, which reduces to

$$D(p) = \frac{\pi\bar{\omega}}{4} \frac{W_M}{B^2/2\mu_0} \frac{p^2 v_A}{v} \left\langle \frac{\sin^4\alpha}{|\cos\alpha|} \left(1 - \frac{v_A^2}{v^2\cos^2\alpha}\right) \right\rangle . \qquad (11.20)$$

This has the same dependence on momentum as in (11.16) for $v \gg v_A$. This shows that damping of magnetoacoustic turbulence at $s = 0$ in the presence of efficient resonant scattering is essentially a form of Fermi acceleration.

It follows that stochastic acceleration by MHD is equivalent to damping at $s = 0$ of magnetoacoustic turbulence. This should be distinguished from gyroresonant damping, which involves damping of MHD turbulence due to resonances at high harmonics $|s| \gg 1$. Such damping leads to a diffusion equation of the form (11.16), but with $D(p)$ different from (11.20) both in magnitude and in the functional dependence on p.

11.4 The spectrum of magnetic turbulence

In applying the theory of stochastic acceleration to astrophysical phenomena an obvious problem is the source of the MHD turbulence. Discussion of this aspect of the problem separates naturally into two parts: the primary generation of the MHD turbulence and the formation of the turbulent spectrum. The ultimate source of the energy can be either magnetic energy or kinetic energy. Stored magnetic energy is released explosively in a solar flare, and this must generate MHD turbulence. However, there are relatively few applications where the important source of MHD turbulence is associated with explosive release of magnetic energy. Kinetic energy can produce turbulence, as is familiar from the flow of water in a pipe or of air across an airfoil. The generation of the turbulence can be due to the *Kelvin–Helmholtz* instability, which is associated with generation of turbulence at the interface of two media in relative motion. A familiar example is the generation of water waves by wind blowing over the surface of the water. Plasma flows are common features of astrophysical plasmas, for example, in stellar winds, in jets, in accretion columns, in rotating accretion discs, in collapsing interstellar clouds and in stellar convection zones. When there is an appropriate gradient in the flow velocity, due to a boundary layer or otherwise, kinetic energy in the flow can be partly converted into turbulent energy through the Kelvin–Helmholtz instability.

Generation of turbulence by the Kelvin–Helmholtz instability leads to turbulence with relatively long wavelengths (small k), typical of the size of the boundary layer or the scale of the gradient in the flow velocity. The formation

of the spectrum of MHD turbulence is thought to involve nonlinear wave-wave interactions. The relevant process is referred to as a *cascade of magnetic energy*. The cascade transfers turbulent energy from the long wavelengths, where it can be generated, to short wavelengths, where it is dissipated. In laboratory applications, the dissipation process is usually due to thermalization through viscosity or resistivity. In astrophysical plasmas the dissipation process can be due to stochastic acceleration of fast particles, rather than heating of the plasma. In fact, in a cascade effectively all the energy injected at the largest scale length should end up in the energetic particles if the dominant dissipation mechanism (the sink at short wavelengths) is stochastic acceleration of fast particles. The dominant dissipation mechanism determines the dissipation scale length at which the cascade terminates. The idea is that if one considers all dissipation mechanisms separately, then one can determine for each mechanism the scale length at which it would be an effective sink, so that one estimates a scale length at which the cascade would terminate for each specific dissipation mechanism. The dominant dissipation mechanism is the one that causes the cascade to terminate at the longest scale length; dissipation mechanisms that would cases the cascade to terminate at shorter wavelengths are not relevant because the turbulent energy does not reach the shorter wavelengths. It follows that if the dominant dissipation mechanism is due to Stochastic acceleration of fast particles then virtually all the power going into the turbulent cascade at the longest scale lengths ultimately ends up in the fast particles.

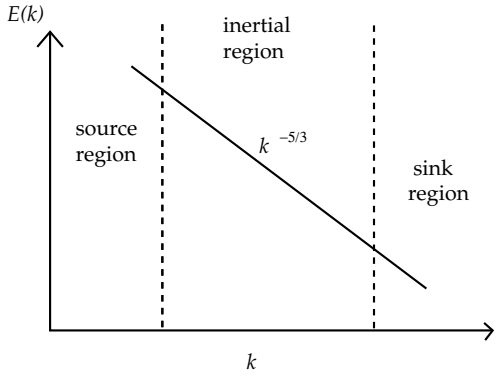


Figure 11.3: The energy per unit wave number in magnetic turbulence is plotted schematically as a function of wave number, showing the Kolmogorov spectrum in the inertial range.

The region of wave numbers between the source at small k and the sink at large k is called the *inertial range*. A simple theory for the shape of the spectrum in this range was proposed by Kolmogorov in 1941 for unmagnetized fluids; the Kolmogorov spectrum explains the typical form of observed fluid

turbulence. Evidence from the solar wind suggests that magnetic turbulence also has essentially the spectrum predicted by Kolmogorov. Kolmogorov's argument is based on dimensional analysis. Let ξ be the power dissipated per unit mass due to the dissipation process. Then, Kolmogorov argued, the turbulent velocity, v_k, at any wave number, k, in the inertial range can depend only on k and ξ. The only quantity with the dimensions of velocity that can be formed from these quantities is $v_k \propto (\xi/k)^{1/3}$. The energy per unit wave number in the spectrum is $E(k) \propto v_k^2/k$, which for a given ξ implies the Kolmogorov spectrum

$$E(k) \propto k^{-5/3}. \tag{11.21}$$

Stochastic acceleration, as described by (11.20), is proportional to $\bar{\omega} W_M \propto \int dk\, kE(k)$ which with (11.21) gives $\bar{\omega} W_M \propto k^{1/3}$, so that the largest k-values dominate. Hence, stochastic acceleration can act as a large-k sink for MHD turbulence with a Kolomgorov-like spectrum.

As remarked above, only the magnetoacoustic component is damped by Fermi-type acceleration. Even if at some k there were pure Alfvénic turbulence, in a cascade a roughly equal mixture of the two modes should result at large k. (To my knowledge, the mode structure has yet to be treated properly in mathematical treatments of the cascade of magnetic turbulence.) If this is the case then, because the magnetoacoustic component dissipates rapidly, ultimately all the turbulent energy is transferred to the energetic particles. Alternatively, both modes can damp through resonant acceleration if this is the relevant dissipation mechanism. Which of Fermi-type or resonant acceleration (or any other dissipation process) is the more important is determined by the one which would be effective at the longest wavelength. That is, if one determines the k at which each dissipation process would be an effective sink terminating the cascade, then the dominant one is that which corresponds to a sink at the smallest k.

11.5 Applications of stochastic acceleration

With all this background, the treatment of stochastic acceleration in specific applications is usually almost trivial. Let us consider two examples, one for a galactic jets, and the other for prompt ions in solar flares.

In a galactic jet, plasma flows out from a central source, perhaps at relativistic speeds, through the intergalactic medium. Such jets can be observed directly through the synchrotron radiation they emit. The synchrotron half lifetimes, cf. lecture 18, are so short that the relativistic electrons which emit the synchrotron radiation need to be accelerated in the jet itself. Two possible acceleration mechanisms are acceleration by shocks and stochastic acceleration by MHD turbulence. One observational feature of some jets is that they are limb brightened, due to the emission being concentrated near the edge of the jet. This suggests that the acceleration may occur preferentially near the edge

of the jet. This could be explained in terms of shocks or MHD turbulence generated due to the interaction in a boundary layer between the outflowing jet plasma and the intergalactic medium. In a model based on stochastic acceleration, one assumes that the Kelvin–Helmholtz instability generates turbulence at a wave number $k_0 \approx 1/R_0$, where R_0 is the radius of the jet. A turbulent cascade forms a Kolmogorov spectrum, with the sink due to stochastic acceleration of relativistic electrons. The wave number, k_1, at which the turbulent cascade ends is determined by the requirement that the power input into the turbulence at k_0 balance the power input into the relativistic electrons at k_1 due to stochastic acceleration. Such a model for acceleration in astrophysical jets seems viable, as does the alternative model based on diffusive acceleration at shocks.

The other possible application of stochastic acceleration mentioned here is to prompt ions in solar flares. In large flares, within about a second of the onset of the flare (as seen at other wavelengths) gamma-ray lines can appear. These are due to ions accelerated to tens of MeV causing nuclear reactions that involve gamma-ray production as these ions hit the denser regions of the solar atmosphere. As in the case of astrophysical jets, the two most plausible acceleration mechanisms for these ions are shock acceleration and stochastic acceleration. However, in this case it is questionable whether shocks can form on this short time scale; at least the shocks that are subsequently observed (through type II solar radio bursts) cannot form so rapidly. The important requirement of a model for stochastic acceleration in this case is that the acceleration time be short enough. Using (11.5) the acceleration rate $\langle \dot{\varepsilon} \rangle$ may be estimated by writing power supplied to the fast particles in the form

$$2\pi \int dpp^2 \, \langle \dot{\varepsilon} \rangle f(p) = 2\pi \int dpp^2 \, \varepsilon \, \frac{1}{p^2} \frac{\partial}{\partial p} \left[p^2 D(p) \frac{\partial f(p)}{\partial p} \right], \tag{11.22}$$

which after two partial integrations gives ($d\varepsilon/dp = v$)

$$\langle \dot{\varepsilon} \rangle = \frac{1}{p^2} \frac{\partial}{\partial p} \left[p^2 D(p) v \right]. \tag{11.23}$$

On defining the acceleration time, t_{a}, by writing

$$\langle \dot{\varepsilon} \rangle = \frac{\varepsilon}{t_{\mathrm{a}}}, \tag{11.24}$$

(11.20) and (11.23) imply $1/t_{\mathrm{a}} \propto \bar{\omega}(W_M/B^2/2\mu_0)(v_A/v)$ for nonrelativistic particles. Assuming that the turbulent energy density is close to the ambient energy density ($W_M \approx B^2/2\mu_0$), one finds that the acceleration time is longer than the typical wave period $2\pi/\bar{\omega}$ by a factor of order v/v_A, which is between 10 and 100 here. Alternatively, one has $t_{\mathrm{a}} \approx \bar{\lambda}/v$, where $\bar{\lambda}$ is the typical wavelength of the turbulence; that is, the acceleration time can be as short as the particle propagation time over one wavelength of the turbulence. Hence, this theory seems

viable provided that the MHD turbulence has high enough frequency (short enough wavelengths). Other evidence suggests scale lengths in solar flares at least as short as 100 km, and turbulence on this scale length might be expected and would be adequate provided it is strong enough (recall that $W_M \approx B^2/2\mu_0$ is assumed in the foregoing discussion).

These rough estimates can be made somewhat more quantitative. However, the theory of particle acceleration is at best semiquantitative, and it would be misleading to be overly quantitative in discussing particular applications. The strongest evidence in favor of specific acceleration mechanisms is usually qualitative. One example is the predicted form of the particle spectrum. For nonrelativistic ions from the Sun there is close agreement between a predicted spectrum, based on stochastic acceleration, which involves a Bessel function, and the observed spectrum. The derivation of this spectrum is as follows.

Consider acceleration in a volume V. Particles are assumed to be injected into V with a momentum $p = p_0$ and to escapes from V with a loss rate ν_L. The diffusion equation in momentum space is

$$\frac{\partial f}{\partial t} = \frac{1}{p^2}\frac{\partial}{\partial p}\left(\nu_A \frac{cp^4}{4v}\frac{\partial f}{\partial p}\right) + \frac{Q\delta(p-p_0)}{4\pi p_0^2} - \nu_L f, \tag{11.25}$$

where Q is the number of particles injected per unit time and per unit volume. Steady state solutions of (11.25) may be obtained in the nonrelativistic and in the ultrarelativistic limits. The Bessel function solution is the steady state solution in the nonrelativistic limit. It is

$$f = \frac{2Q}{\pi p p_0 mc\nu_A}\begin{cases} I_2(z)K_2(z_0), & \text{for } p < p_0, \\ I_2(z_0)K_2(z), & \text{for } p > p_0, \end{cases} \tag{11.26}$$

with

$$z = 4\left(\frac{p}{mc}\frac{\nu_L}{\nu_A}\right)^{1/2}, \tag{11.27}$$

with z_0 given by replacing p by p_0 in (11.27). The analogous solution in the ultrarelativistic limit is

$$f = \frac{Q}{4\pi p_0^2 \nu_A (9/16 + \nu_L/\nu_A)^{1/2}}\left(\frac{p}{p_0}\right)^{-3/2 \pm 2(9/16+\nu_L/\nu_A)^{1/2}} \tag{11.28}$$

with the + sign applying for $p < p_0$ and the − sign applying for $p > p_0$.

The agreement between theory and observation lends support for the validity of a model in which the acceleration is stochastic and reaches a balance between injection and escape. Note that the form of the spectrum for $p \gg p_0$ is insensitive to the form of the injection spectrum at $p \approx p_0$, but that the shape of the escaping spectrum is sensitive to any assumed momentum dependence of escape rate.

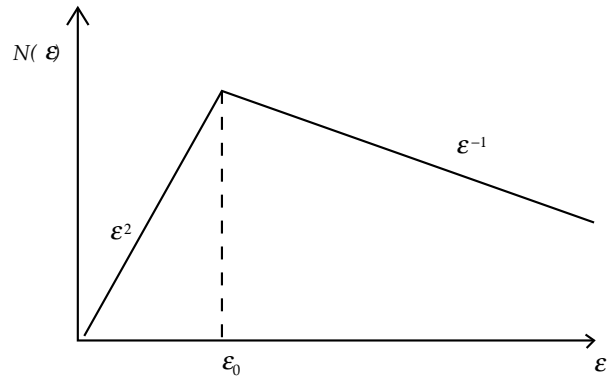


Figure 11.4: The asymptotic energy spectrum (log–log plot) that results from a constant injection at $\varepsilon = \varepsilon_0$ is illustrated schematically.

A final remark concerns the evolution of the spectrum under the influence of acceleration. If one omits the loss terms in (11.25) and sets $\nu_L = 0$ in (11.28), it follows that the solutions (11.28) for relativistic particles are power laws $f(p) \propto p^{-b}$ with $b = 0$ and $b = 3$, which correspond to $N(\varepsilon) \propto \varepsilon^{-a}$ with $a = -2$ and $a = 1$. Thus an initial δ-function distribution at $p = p_0$, or $\varepsilon = \varepsilon_0$, evolves towards a spectrum of the form illustrated in Figure 11.4. That is, the power laws $a = -2$ and $a = 1$ apply below and above the injection energy, respectively. A more complicated injection spectrum may be treated by regarding the solution for a δ-function injection as a Greens function; that is, the resulting spectrum is a convolution of the initial spectrum and the spectrum illustrated in Figure 11.4. It follows that acceleration cannot lead to a spectrum flatter than $a = 1$.

Lecture 12

Diffusive shock acceleration

Diffusive shock acceleration is the favored acceleration mechanism for galactic cosmic rays and for the relativistic electrons in synchrotron sources. It is known to be effective for ions at shocks in the heliosphere.

12.1 Qualitative introduction to diffusive shock acceleration

The acceleration mechanism pointed out by Fermi in 1949 is second order, in the sense that first order, in u/c, changes in the energy are of opposite signs and so cancel each other. In a subsequent paper in 1954, Fermi pointed out that the acceleration can be much more efficient if it is first order, and that first order changes need not always cancel. A simple example is when two magnetized clouds approach each other. A particle bouncing back and forth between them is always reflected head on and so the first order changes in energy is always positive.

It was not until 1977–8 that it was pointed out, independently by four different sets of authors, that first order Fermi acceleration occurs naturally at any shock front, provided only that the fast particles are scattered on either side of the shock, as illustrated in Figure 12.1. To understand this the only important point to note is that the fluid velocity changes across the shock. Indeed the ratio of the velocities is determined by the strength of the shock according to (7.16). Imagine a particle in the upstream plasma about to cross the shock and enter the downstream plasma. The scattering centers, that is, the MHD waves, in the downstream plasma are convected along with the plasma, and so this particle sees these scattering centers as approaching it head on. Once the particle enters the downstream plasma and is scattered, its energy is increased due to the reflection being head on. In effect the scattering centers act like particles with infinite mass, so that the reflection is anologous to that from a moving

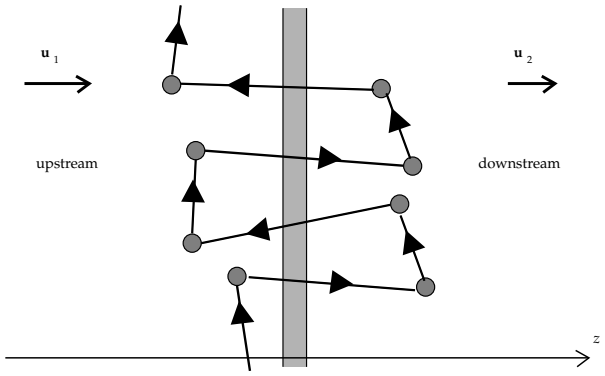


Figure 12.1: Diffusive shock acceleration is illustrated: the shaded vertical region is the shock, the circular blobs denote idealized scattering centers, and the solid line with arrows denotes the path of an idealized fast particle. The coordinate z and the velocities $\mathbf{u}_1$ and $\mathbf{u}_2$ introduced in (12.2) are shown for the case of a parallel shock.

wall. Now consider a particle in the downstream plasma about to cross the shock and enter the upstream plasma. As viewed from the downstream plasma, the upstream plasma is flowing toward the particle, as illustrated in figure 12.2. Again the scattering centers (now in the upstream plasma) are also approaching the particle, and again, when the particle crosses the shock and is scattered, the scattering is head on and so causes an increase in the particle energy. Note that, as viewed by a particle at rest in the upstream plasma, both the shock and the plasma behind it are approaching, with the shock approaching faster than the fluid behind it, and for a particle at rest in the downstream plasma, the shock is receding into the approaching upstream plasma.

Provided that particles cross the shock many times, this mechanism allows efficient acceleration. Moreover, one expects a particle to cross the shock many times provided that it is efficiently scattered on each side of the shock. This is because scattering causes spatial diffusion, so that the motion of a typical scattered particle has a random component that gives it a high probability of returning to the shock many times. Collectively, particles diffuse away from the shock. Particles wandering upstream reach a steady state with those wandering back to the shock, and the upstream particle distribution falls off with distance from the shock, approaching zero far from the shock. Particles wandering downstream can reach arbitrarily far downstream and never return, so that the downstream distribution approaches a constant arbitrarily far away from the shock.

A further ingredient in the theory concerns the generation of the waves required to scatter the upstream particles. Far upstream, that is, a long way

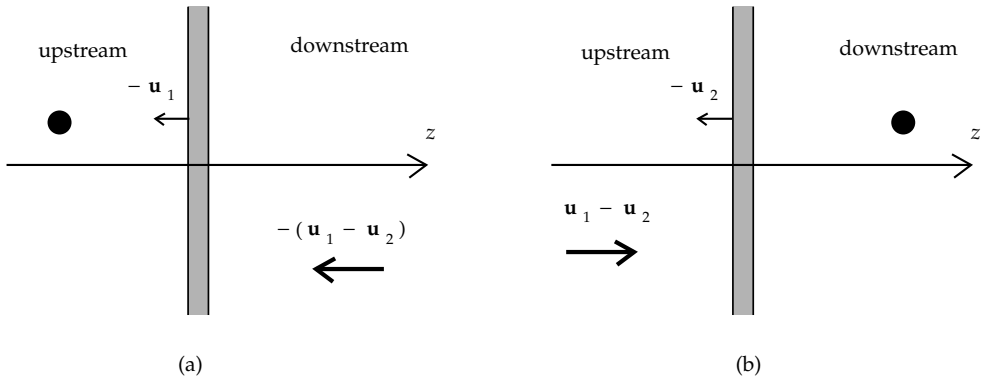


Figure 12.2: (a) A particle, denoted by the black circle, at rest in the upstream plasma sees the shock approaching it and the downstream plasma behind the shock approaching at a slower speed; the velocities indicated are obtained by subtracting the velocity $\mathbf{u}_1$ from the velocities in figure 12.1. (b) As for (a) but for a particle at rest in the downstream plasma; the velocities indicated are obtained by subtracting the velocity $\mathbf{u}_2$ from the velocities in figure 12.1.

ahead of the shock, there are assumed to be no fast particles (or at least many fewer than there are downstream behind the shock), and so there must be a spatial gradient in the number density of fast particles ahead of the shock. As shown in lecture 9, cf. (9.20), a spatial gradient can lead to the growth of waves. This provides a natural mechanism for the generation of the upstream waves to scatter the upstream particles. It is assumed that there are waves in the downstream plasma so that the particles are also scattered there. There is no particularly convincing argument for the generation of the downstream waves, but it is plausible that they are present. For example, the upstream waves convect across the shock into the downstream plasma, that is, the shock overtakes the waves in the upstream plasma. (The waves are propagating away from the shock at the Alfvén speed in the upstream plasma, but the shock is approaching at faster than the Alfvén speed in the upstream plasma, and so overtakes the waves.)

The treatment of diffusive shock acceleration given below is a single-particle theory. That is, the collective effect of the fast particles on the shock is neglected. However, it is thought that the acceleration is so efficient that this assumption is questionable. Once the pressure associated with the cosmic rays accelerated at the front becomes comparable with the gas pressure, the structure of the shock is modified by the cosmic rays. This aspect of the problem has been considered, and it has been found that the stresses can be transferred from the downstream to the upstream plasma primarily through the cosmic rays, with it being possible for there to be no discontinuity in the density of the thermal gas.

Qualititatively, the main point is that the acceleration is so efficient that it is likely to be limited only by such nonlinear effects.

12.2 Diffusive acceleration at a shock front

Consider a distribution of particles $f(p,z)$ that is averaged over pitch angle and is a function of distance z from a shock front in a frame in which the shock is at rest. It is assumed that there is a scattering process that causes the particles to diffuse in the z direction with diffusion coefficient $\kappa(z)$. In the simplest model scattering is assumed to occur only upstream ($z<0$), with $\kappa(z)=0$ for $z>0$; however, for the present let us formulate the problem without making this assumption. The particles are also assumed to be streaming with the streaming speed u. Then one has

$$\frac{df(p,z)}{dt} = \frac{\partial}{\partial z}\left(\kappa(z)\frac{\partial f(p,z)}{\partial z}\right) + Q(p,z) - f_{\rm esc}(p),$$
$$\frac{df(p,z)}{dt} = \frac{\partial f(p,z)}{\partial t} + u\frac{\partial f(p,z)}{\partial z} + \dot{p}\frac{\partial f(p,z)}{\partial p}, \quad \dot{p} = -\frac{1}{3}p\frac{\partial u}{\partial z}, \qquad (12.1)$$

where $Q(p,z)$ is a source term, and where the sink term $f_{\rm esc}(p)$ takes account of escape of particles from the region of the shock front. The term involving a partial derivative with respect to p determines the energy changes, and so is the term that describes the acceleration. In the simplest case, one assumes that the speed u changes abruptly across the front:

$$u = \begin{cases} u_1 & \text{for } z<0 \quad \text{(upstream)}, \\ u_2 & \text{for } z>0 \quad \text{(downstream)}, \end{cases} \qquad \frac{\partial u}{\partial z} = (u_1-u_2)\,\delta(z). \qquad (12.2)$$

Let us seek a stationary solution of (12.1) in which both the source and the sink term are neglected. The acceleration term is zero for $z\neq 0$, and is taken into account only when matching solutions for $z>0$ and for $z<0$. The equation then reduces to $u\,\partial f/\partial z = (\partial/\partial z)(\kappa\,\partial f/\partial z)$ and a general solution is

$$f(p,z) = A + B\exp\left[u\int dz\,\frac{1}{\kappa(z)}\right], \qquad (12.3)$$

with u is a constant on either side of the shock. For $z<0$ one has $u/\kappa(z)\neq 0$, and the solution (12.3) diverges at $z\to-\infty$ unless one has $B=0$ for $z<0$. Hence, writing $f_\pm(p)=\lim_{z\to\pm\infty} f(p,z)$, one has

$$f(p,z) = \begin{cases} f_-(p) + [f_+(p)-f_-(p)]\exp\left[u_1\int_0^\infty dz\,\frac{1}{\kappa(z)}\right] & \text{for } z<0, \\ f_+(p) & \text{for } z>0. \end{cases} \qquad (12.4)$$

On matching the discontinuity in the derivative of this solution with the discontinuity due to the acceleration term, one finds

$$-u_1[f_+(p) - f_-(p)] = \tfrac{1}{3}(u_1 - u_2)\, p \frac{\partial f_+(p)}{\partial p}, \tag{12.5}$$

which integrates to

$$f_+(p) = bp^{-b} \int_0^p dp'\, p'^{(b-1)} f_-(p'), \quad b = \frac{3u_1}{u_2 - u_1}, \tag{12.6}$$

which determines the downstream solution in terms of the upstream solution. That is, if the shock propagates into a region where the electron distribution is $f_-(p)$, then after the shock has passed the distribution is $f_+(p)$. The distribution $f_-(p)$ acts as an injection spectrum in this case.

If the injection spectrum is monoenergetic, $f_-(p) \propto \delta(p - p_0)$ say, then (12.6) implies that the downstream spectrum is a power law at $p > p_0$ of the form $f_+(p) \propto p^{-b}$. According to (12.6), the power law index, b, is determined by the ratio u_1/u_2, which is equal to the inverse of the density ratio, n_1/n_2, which determines the strength of the shock. In terms of the mach number $M = u_1/v_{A1}$, where v_{A1} is the Alfvén speed upstream of the shock, and the ratio Γ of specific heats, one finds

$$b = \frac{3r}{r-1} = \frac{3(\Gamma+1)M^2}{2(M^2-1)}, \quad r = \frac{(\Gamma+1)M^2}{2 + (\Gamma-1)M^2}, \tag{12.7}$$

where r is the compression factor across the shock, cf. (7.16); the result (12.7) follows from (7.20). Then for $\Gamma = 5/3$ and $M^2 \gg 1$, the maximum strength of the shock is $r = 4$ and the minimum value of the spectral index is $b = 4$. This corresponds to an energy spectrum $N(\varepsilon) \propto p^2 f(p) \propto \varepsilon^{-2}$, with $\varepsilon = pc$ in the ultrarelativistic limit. The observed cosmic ray spectrum corresponds to $b = 4.6$, which is consistent with diffusive shock acceleration by relatively strong shocks. (Recall that one has $b = 4$ for arbitrarily strong shocks.)

12.3 Alternative treatment of the acceleration

An alternative treatment of this acceleration is as follows. Consider a cycle in which a particle with momentum p crosses the shock from downstream to upstream and back again. Let $\Delta(p)$ be the change in p in a cycle. Let $P(p)$ be the probability per cycle of the particle escaping into the downstream region. Assuming that the distribution remains approximately isotropic, let $4\pi p^2 f(p)$ be the number of particles between p and $p + dp$. After one cycle one has

$$\begin{aligned} p \to p' = p + \Delta(p), \quad dp \to dp' = (1 + d\Delta(p)/dp)dp', \\ p^2 f(p)dp \to p'^2 f(p')dp' = [1 - P(p)]p^2 f(p)dp. \end{aligned} \tag{12.8}$$

Then one finds

$$\frac{\partial}{\partial p}\left[\Delta(p)\,p^2 f(p)\right] = -P(p)\,p^2 f(p). \tag{12.9}$$

Integrating (12.9) gives

$$f(p) = f(p_0)\,\frac{p_0^2\Delta(p_0)}{p^2\Delta(p)}\,\exp\left[-\int_{p_0}^{p} dp\,\frac{P(p')}{\Delta(p')}\right], \tag{12.10}$$

where $f(p_0)$ is a constant of integration. If the distribution function downstream is determined by advection ('advection' means a change caused by motion of a fluid in a system with some gradient), then it is given by $f_+(p) = P(p)f(p)$. Combining the constants into a single factor, A, determined by the boundary conditions, the distribution function is

$$f_+(p) = \frac{AP(p)}{p^2\Delta(p)}\,\exp\left[-\int_{p_0}^{p} dp\,\frac{P(p')}{\Delta(p')}\right]. \tag{12.11}$$

In applying (12.11) to a shock, the flux into the downstream region is identified as $F_+ = u_2 4\pi p^2 f_+(p)$. The velocity ratio is related to the ratio of the number densities by $u_2/u_1 = n_1/n_2$, with u_1 the shock velocity and $r = n_2/n_1$ given by (12.7). The flux of particles with speed v back into the upstream region is identified as $F_b = v\pi p^2 f_+(p)$, where a factor $v/4$ arises from averaging $v\cos\alpha$ over the backward hemisphere. Then one has $P(p) = F_+/F_b$. The theory implies

$$\Delta(p) = \frac{4(r-1)}{3r}\,\frac{u_1}{v}\,p, \quad P(p) = \frac{4u_1}{vr}. \tag{12.12}$$

Then integrating (12.10) gives

$$f_+(p) = f_0\,\left(\frac{p}{p_0}\right)^{-b}, \tag{12.13}$$

with b given by (12.7), and where f_0 is a constant. For a monoenergetic distribution $f_-(p) = n_1\delta(p-p_0)/4\pi p_0^2$, one finds $f_0 = 3rn_1/(r-1)4\pi p_0^3$.

In this model the cycle time is $t_c = (\lambda_1/u_1 + \lambda_2/u_2) = (\lambda_1 + r\lambda_2)/u_1$, where $\lambda_{1,2}$ are the scattering mean free paths in the upstream and downstream regions, respectively. For isotropic scattering the scattering mean free path, λ, is related to the spatial diffusion coefficient, κ, by $\kappa = \lambda v/3$, where v is the speed of the particle. The mean free path may be written as $\lambda = \eta r_g$, where r_g is the gyroradius, and it is usually assumed that η is a constant of order unity. The rate of momentum gain is given by

$$\frac{dp}{dt} = \frac{p}{t_a} - \left(\frac{dp}{dt}\right)_{\rm loss}, \quad t_a = \frac{u_1^2}{c\bar{\lambda}}, \quad \bar{\lambda} = \frac{3r(\lambda_1 + r\lambda_2)}{4(r-1)}, \tag{12.14}$$

where t_a is an acceleration time, $\bar{\lambda}$ is a mean scattering free path, and where a loss term is included.

12.4 Acceleration at multiple shocks

One specific application of diffusive shock acceleration is to shocks in the wind of an early-type star. Such winds contain multiple shocks, and the acceleration in this case in assumed to involve a sequence of identical shocks. At each shock a new distribution of particles is injected and accelerated, and the particles injected at earlier shocks are accelerated further.

The following specific assumptions in the model:

(a) At each shock particles are assumed to be injected with a spectrum $\phi_0(p)$. The injection spectrum $\phi_0(p)$ is assumed to be the same each shock, and to arise from the acceleration (by processes not discussed explicitly) of thermal particles into particles with $v > v_A$ so that they experience diffusive shock acceleration. That is, there is assumed to be some mechanism (called a preacceleration mechanism) that supplies a distribution $\phi_0(p)$ at each shock.

(b) The spectrum downstream of the first shock is the spectrum, $f_1(p)$, resulting from acceleration of the injection spectrum, $\phi_0(p)$, at the shock, cf. (12.6):

$$f_1(p) = bp^{-b} \int_0^p dq\, q^{b-1} \phi_0(q), \qquad b = \frac{3r}{r-1}. \tag{12.15}$$

(c) All particles accelerated at the first shock are accelerated again when the second shock arrives, so that the spectrum downstream of the second shock includes the particles injected and accelerated at the first shock and reaccelerated at the second shock, plus the particles injected and accelerated at the second shock.

(d) Between each shock a decompression occurs. According to Liouville's theorem f is a constant along the trajectory of a particle in phase space. Let the compression ratio at the shock be written $r = R^{-3}$, so that during decompression p changes to p/R. Then Liouville's theorem implies that the distribution function $f_1'(p)$ after decompression is given by $f_1'(p) = f_1(p/R)$. Hence the injection spectrum into the second shock is not $\phi_0(p) + f_1(p)$, as given by (12.15), but $\phi_0(p) + f_1'(p)$, with

$$f_1'(p) = b(p/R)^{-b} \int_0^{p/R} dq\, q^{b-1} \phi_0(q). \tag{12.16}$$

Solving for r in terms of b gives $r = b/(b-3)$.

(e) The spectrum accelerated at the second shock is decompressed before the third shock arrives, giving a spectrum $f_2'(p)$, just before the third shock arrives, and so on for subsequent shocks.

(f) The number of shocks is assumed to be arbitrarily large.

(g) The injection spectrum at each shocks is assumed to be $\phi_0(p) = K\delta(p-p_0)$.

The particle spectrum downstream of the second shock after decompression, $f_2'(p)$, is obtained by replacing $\phi_0(q)$ in (12.16) by $\phi_0(q) + f_1'(q)$. This takes into account the identical injection spectrum at each shock, and no losses as particles propagate between shocks. The spectrum injected into the nth shock in the injection spectrum, $\phi_0(p)$, plus the particles injected at the n'th shock for all $n' < n$. The contribution from the injection at the first shock to the spectrum downstream of the ith shock (after decompression downstream of the preceding shock) is found to be

$$f_i'(p, p_0) = \frac{Kb^i}{p_0} \left(\frac{p}{R^i p_0}\right)^{-b} \frac{(\ln\, p/R^i p_0)^{i-1}}{(i-1)!}. \tag{12.17}$$

The total spectrum downstream of the nth shock (after decompression) is

$$f_n'(p) = \sum_{i=1}^{n} f_i'(p, p_0). \tag{12.18}$$

Summing over an infinite number of shocks involves summing an exponential series, and this gives

$$f_\infty(p) \propto p^{-3}. \tag{12.19}$$

The spectrum $f(p) \propto p^{-3}$, or $N(\varepsilon) \propto \varepsilon^{-1}$, is that toward which diffusive shock acceleration tends on the upper energy side of the injection, cf. Figure 11.4.

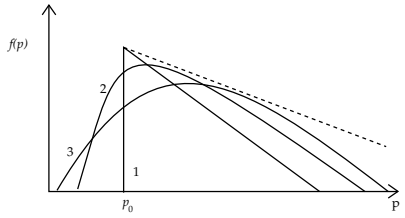


Figure 12.3: The distribution function $f_i'(p, p_0)$ given by (12.17) is indicated schematically for i = 1–3. The dashed line indicates the asymptotic solution, $f_\infty(p)$ given by (12.18).

To derive (12.19) one first writes $f_\infty(p) = p^{-b}S(p)$ and shows that $S(p)$ satisfies the differential equation $pdS(p)/dp = (b-3)R^{b-3}S(p/R)$, then one notes that a power law is a solution of this equation and that the only power law index that is acceptable gives $S(p) \propto p^{b-3}$.

This example of multiple acceleration by shocks is interesting in that it contradicts a suggestion that is sometimes made: that the hardest spectrum that can be produced by diffusive shock acceleration is a power law with $b = 4$ $a = 2$, corresponding to the strongest shock, that is, $r = 4$ in (12.7). An explanation for this is that although the spectrum resulting from a δ-function injection at $p = p_0$ at one shock is a power law with $a \geq 2$ for $p > p_0$, when such a power law spectrum is subject to reacceleration at the next shock, it produces a spectrum that is zero at $p = p_0$, rises to a peak, and then falls off toward the power law $f(p) \propto p^{-b}$, cf. Figure 11.4. Unlike stochastic acceleration, for which a δ-function injection produces a spectrum as illustrated in Figure 11.4, diffusive shock acceleration leads to a spectrum which is identically zero at $p < p_0$. This may be attributed to the acceleration being systematic rather than diffusive. In diffusive shock acceleration, a δ-function injection at $p = p_0$ produces a power law $\propto p^{-b}$ at $p \geq p_0$; however any injection spectrum that is finite at $p = p_0$ (as opposed to a δ-function, which is infinite at $p = p_0$) and is nonzero only for $p > p_0$ produces a spectrum that is zero at $p = p_0$ and has a maximum at higher p, as for the spectra labeled 2 and 3 in Figure 12.3.

12.5 Applications of diffusive shock acceleration

The most direct evidence for diffusive shock accleration is from observations of fast particles in association with shock waves in the interplanetary medium and at planetary bow shocks, notably the Earth's bow shock. The data all apply to ions. In association with the Earth's bow shock, two types of energetic ion distribution are observed: a beamed distribution with $\langle p_\parallel^2 \rangle \gg \langle p_\perp^2 \rangle$ consisting primarily of protons, and a diffuse distribution with $\langle p_\parallel^2 \rangle \approx \frac{1}{2}\langle p_\perp^2 \rangle$ with ionic composition typical of the solar wind. The beamed distribution is attributed to shock drift acceleration, which is discussed in the next lecture, and the diffuse distribution is attributed to diffusive shock acceleration. An important qualitative point is that the beamed component cannot be the source of the diffuse component, that is, the diffuse component cannot be an isotropized form of the beamed component, because their isotopic abundances are different. Other evidence on acceleration at shocks in the heliosphere comes from solar energetic particle (SEP) events. The ionic composition in different SEP events, and at different phases of a single event, can have two distinct characteristics. Either the ionic composition in SEPs can be characteristic of the hot plasma in the solar corona where flares occur, or it can be characteristic of the cooler interplanetary plasma. This provides strong evidence that diffusive shock acceleration occurs, and that the ions accelerated have an ionic composition characteristic of the

ambient plasma.

There are anomalies in the ionic composition in SEPs that seem to require specific acceleration mechanisms. The most notable of these concerns ^{3}He. The ratio of ^{3}He to ^{4}He in the solar photosphere is 5×10^{-4}, but the ratio in SEPS is highly variable, and in some events it exceeds unity. There are several suggestions as to how such preferential acceleration of ^{3}He occurs, but there is no agreement on which mechanism operates.

The acceleration of most galactic cosmic rays is attributed to diffusive shock acceleration. For many decades it has been widely (but never universally) accepted that galactic cosmic rays are accelerated in association with supernova explosions. The original argument was based on energetics: the power needed to maintain the cosmic rays spectrum. For example, assuming that the energy density in galactic cosmic rays is $1\,\mathrm{eV\,cm^{-3}}$ and that the volume to which they are confined is that of the galactic disk, which is $10^{68}\,\mathrm{cm^3}$, one finds a total energy in cosmic rays of $10^{49}\,\mathrm{J}$. (In the astrophysical literature energies are usually expressed in ergs, $1\,\mathrm{erg} = 10^{-7}\,\mathrm{J}$.) If the lifetime is $10^7\,\mathrm{yr} = 3 \times 10^{14}\,\mathrm{s}$, then the power that needs to be supplied to maintain the cosmic rays is of order $0.3 \times 10^{35}\,\mathrm{W}$. The most energetic events in the galaxy are supernovae, each of which releases about $10^{45}\,\mathrm{J}$ on average, and which occur 3 times per century on average. This implies a power of $10 \times 10^{35}\,\mathrm{W}$. It follows from these rough estimates that a significant fraction of the power released in supernova explosions needs to go into cosmic rays to maintain them.

It is now thought that the cosmic rays are accelerated by shocks generated by supernovas and propagating through the hot component of the interstellar medium. The isotopic abundances of the cosmic rays is not consistent with them coming directly from the supernovas themselves, but implies that they come from the interstellar medium, or perhaps from stellar winds that supply some mildly energetic ions to the interplanetary plasma. The isotopic abundances of cosmic rays raises a theoretical question concerning diffusive shock acceleration: Is a preacceleration mechanism needed for ions? The point is that effective acceleration can occur only for particles that can be effectively scattered, and this requires $v \gg v_A$ for ions. One viewpoint is that preacceleration is not needed and that the diffusive shock acceleration process itself regulates the injection of thermal particles. There is no detailed mechanism for this injection, and the main argument in favor of such regulated injection comes from numerical simulations. (A seemingly plausible injection mechanism is shock drift acceleration, but, as already stated, for ions at the Earth's bow shock the ionic compositions of the beamed and diffusive components are different, and the beamed component cannot act as the source for the diffuse component.) The alternative viewpoint, adopted by those who emphasize the plasma physics processes, is that a specific preacceleration mechanism is required. A difficulty is that such preacceleration mechanisms tend to be strongly dependent on ionic species, and the observational evidence suggests that preacceleration is weakly dependent on ionic species. The problem of preacceleration is much more se-

vere for electrons. Despite this and other specific unsolved problems concerning the details of the mechanism, it is now widely accepted that galactic cosmic rays are accelerated by the diffusive shock mechanism due to shocks generated by supernovas and propagating through the hot component of the interstellar medium.

Synchrotron emission is a characteristic signature of relativistic electrons in magnetic fields. Synchrotron sources include the interstellar medium itself, supernova remnants, radio galaxies, galactic jets, and a variety of other specific types of source. In most of these situations it is thought that the relativistic electrons are accelerated through the diffusive shock mechanism. In some sources, such as supernovas and the knots in some jets, there is supporting evidence for the shock waves, but even in the absence of such evidence, it is widely accepted that relativistic electrons in most synchrotron sources are accelerated by shocks. However, there is a major unsolved problem concerning the preacceleration of the electrons, and a related problem concerning the relative efficiency of electron and ion acceleration. Scattering of electrons is more effective, at least in principle, and the threshold condition for such scattering is $v \gg 43 v_A$. However, detailed calculations suggest that the generation of whistlers by electrons may be swamped by damping of the whistlers by energetic ions. Electrons can be scattered by the Alfvén waves generated by the ions, but only for $\gamma v \gg (43)^2 v_A$. Thus even in the most favorable case, one needs to account for preacceleration to $v \gg 43 v_A$, and even this may not be adequate.

A serious practical uncertainty concerns the relative energy densities in of accelerated electrons and ions. In galactic cosmic rays, the ratio of electrons to ions at a fixed energy is about 0.03. There is no reasonable explanation of this number. Evidence on diffusive shock acceleration in the heliosphere is no assistance in this context: there is no evidence for diffusive shock acceleration of electrons at interplanetary shocks and at bow shocks. This is consistent with theory in the sense that no electron acceleration is expected; this is, in part, because the simple theory cannot explain injection of electrons at a speed in excess of the threshold for them to generate the waves required to scatter them. For the Crab Nebula there is strong evidence that the electron to ion ratio is not as small as 0.03: the expansion of the nebula requires a pressure consistent with that implied by the synchrotron emitting electrons. However, the Crab Nebula appears to be powered by its pulsar, and the pulsar is thought to supply relativistic electrons and positrons to the nebula, so that the question of the electron to ion ratio is not really relevant. In applying shock acceleration to synchrotron sources, this difficulty is usually avoided by including an unknown parameter specifying the ratio of the energy going into energetic electrons and ions is usually included, thus avoiding this difficulty.

Lecture 13

Shock drift acceleration

Shock drift acceleration is due to acceleration by an electric field in a shock front when a particle is reflected from or passes through the front. There is observational evidence for such acceleration at interplanetary shocks, and especially at planetary bow shocks. This acceleration mechanism may also be of significance in some other astrophysical applications, but the arguments for this are not entirely convincing. Acceleration in neutral sheets, in association with magnetic reconnection has some similarities to shock drift acceleration.

13.1 Subluminal and superluminal shocks

A single particle that encounters a shock front may either crosses it or be reflected by it. Consider the case where the particle crosses the shock. Let its momentum be $\mathbf{p}_1$ and $\mathbf{p}_2$ on the two sides of the shock. These momenta may be written in terms of spherical polar coordinates relative to the magnetic fields on either side of the shock front: these are the momenta p_1, p_2, the pitch angles, α_1, α_2, and the gyrophase angles, ϕ_1, ϕ_2. In general, p_1 and p_2 are not equal because the electric field is different on either side of the shock (only the tangential component of $\mathbf{E}$ need be continuous) and so the electric field can do work on the particle as it crosses the front. In general, the pitch angle also changes as a particle crosses the front, that is $\alpha_1 \neq \alpha_2$. When one calculates p_2, α_2 for given p_1, α_1 the result depends on the gyrophase at the shock front. However, for most choices of the parameters one finds that the energy of the particles increases. It is this effect that is called shock drift acceleration.

The treatment of reflection and transmission of particles incident on a shock front is facilitated by combining two steps. One is to make a Lorentz transformation to an appropriate inertial frame. In defining the special inertial frames it is necessary to distinguish between two cases: subluminal and superluminal shocks. The definition depends on the speed at which the intersection between

a particular magnetic field line with the shock front moves in the plane of the shock. The shock is *subluminal* if this speed is less than c, and is *superluminal* if this speed is greater than c. The other step that facilitates the treatment of reflection and transmission is to show that in the relevant frames the quantity $p^2 \sin^2 \alpha / B$ is approximately constant as the particle crosses the front. That is, one has

$$\frac{p_2^2 \sin^2 \alpha_2}{B_2} = \frac{p_1^2 \sin^2 \alpha_1}{B_1}. \tag{13.1}$$

The identification of the result (13.1) arose originally from the results of numerical calculations. Although the result appears to be equivalent to the conservation of the adiabatic invariant $p^2 \sin^2 \alpha / B$, cf. (1.19), the justification of (13.1), from hindsight followings its initial identification, is quite different from the justification for the adiabatic invariant (1.19). In fact, (13.1) applies only in an average sense, where the average is over gyrophase.

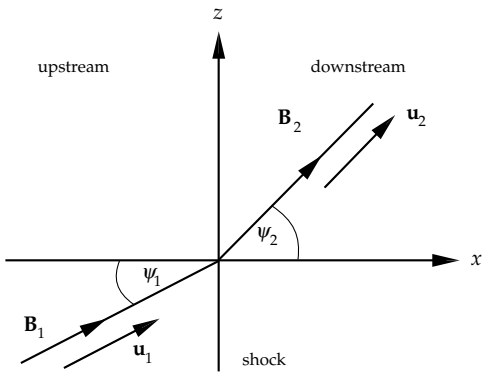


Figure 13.1: In the de Hoffmann–Teller frame the fluid velocities are parallel to the magnetic fields on either side of the shock, cf. Figure 7.3.

First consider a subluminal shock. Then one may make a Lorentz transformation to the frame in which this point of intersection is at rest. The existence of this frame was first pointed out by de Hoffmann and Teller in 1950, and so it is often called the *de Hoffmann–Teller* frame. The shock transition in this frame is illustrated in Figure 13.1. In this frame the fluid velocity is parallel to the field lines on both sides of the shock, because if this were not so the plasma motion perpendicular to the magnetic field would imply a motion of the magnetic field at the shock front, contrary to the assumed properties of the de Hoffmann–Teller frame. Further, because the electric field is determined by the cross product of the fluid velocity and the magnetic field, there is no electric field in this frame. In the absence of an electric field, the energy of a particle is conserved as it crosses the shock (remember: magnetic fields do no work), and so in this frame one has $p_1 = p_2$.

One justification for (13.1) for the case of a subluminal shock may be expressed in terms of Liouville's theorem, which states that both the distribution function and the extension in phase space are constants along the trajectory of a particle in phase space. Conservation of an infinitesimal extension is phase implies

$$d^3\mathbf{p}_1 d^3\mathbf{x}_1 = d^3\mathbf{p}_2 d^3\mathbf{x}_2. \tag{13.2}$$

Let the shock normal be along the x axis, so that the surface areas $dy_1 dz_1 = dy_2 dz_2$ are identically equal. Writing $d^3\mathbf{p}_{1,2} = p_{1,2}^2 dp_{1,2} d\cos\alpha_{1,2} d\phi_{1,2}$, in the de Hoffmann–Teller frame, with $p_1 = p_2 = p$, (13.2) reduces to $d\cos\alpha_1 d\phi_1 dx_1 = d\cos\alpha_2 d\phi_2 dx_2$. The magnetic field is at an angle $\psi_{1,2}$ to the shock normal, cf. Figure 13.1, then the velocity component along the x-axis implies $dx_{1,2} = v\cos\alpha_{1,2}\cos\psi_{1,2} dt$. Finally, the component of the magnetic field along the shock normal is equal on either side of the shock: $B_2\cos\psi_2 = B_1\cos\psi_1$. Hence (13.2) reduces to

$$\frac{\cos\alpha_1 d\cos\alpha_1 d\phi_1}{B_1} = \frac{\cos\alpha_2 d\cos\alpha_2 d\phi_2}{B_2}. \tag{13.3}$$

Now if we assume that the particle distribution is independent of gyrophase, one may integrate of gyrophase: $\int d\phi_{1,2} = 2\pi$. Then (13.3) integrates to $(\cos^2\alpha + \text{constant})/B$ having the same value on either side of the shock, and this implies $\sin^2\alpha_1/B_1 = \sin^2\alpha_2/B_2$ when one evaluates the constant of integration by noting that a particle with zero pitch angle on one side of the shock has zero pitch angle on the other side of the shock, that is, $\alpha_1 = 0 \Rightarrow \alpha_2 = 0$. This establishes (13.1) in the subluminal case.

In the superluminal case one may choose a frame in which the shock is strictly perpendicular. All particles are tied to field lines which convect perpendicularly across the shock, so that no particle can be reflected. The electric field is normal to the shock, and to the magnetic field, so that the parallel momentum is unchanged: $p_{\parallel 1} = p_{\parallel 2}$. The volume elements $d^3\mathbf{x}_1 = r d^3\mathbf{x}_2$ are related by the compression ratio, r of the shock. For a perpendicular shock, the magnetic fields are related by $B_2 = rB_1$, cf. (7.26) for $B_{2n} = B_{1n} = 0$ and $\cos\theta = 0$. Then writing $d^3\mathbf{p}_{1,2} = p_{\perp 1,2} dp_{\perp 1,2} dp_{\parallel 1,2} d\phi_{1,2}$, and assuming that the distribution is independent of gyrophase, (13.2) reduces to $p_{\perp 2} dp_{\perp 2} = r p_{\perp 1} dp_{\perp 1}$, which with $r = B_2/B_1$ integrates to give $p_{\perp 2}^2/B_2 = p_{\perp 1}^2/B_1$, which establishes (13.1) for the superluminal case.

13.2 Reflection and transmission at a shock front

Let quantities in the de Hoffmann–Teller frame be denoted by primes, and make a Lorentz transformation to the laboratory frame, in which the upstream plasma is at rest. In the laboratory frame the fluid velocity $\mathbf{u}_1$ is along the shock normal, cf. Figure 7.3. The relation between the pitch angles in the two frames

may be approximated by assuming that the transformation is nonrelativistic. Then the magnetic field, and hence the angle ψ_1 is the same in the two frames. The velocity of the transformation is such as to transform $\mathbf{u}_1'$, with components $u_1' \sin\psi_1$ along the face of the shock and $u_1' \cos\psi_1$ normal to the shock, into $\mathbf{u}_1$, which is normal to the shock. In the nonrelativistic case this requires a transformation velocity $u_0 = u_1' \sin\psi_1$ with $u_1 = u_1' \cos\psi_1$, implying $u_0 = u_1 \tan\psi_1$. The pitch angles in the de Hoffmann–Teller frame and the rest frame of the upstream plasma are related by

$$p' \cos\alpha' = p\cos\alpha + \gamma m u_0, \qquad p' \sin\alpha' = p \sin\alpha, \tag{13.4}$$

which imply

$$\cos\alpha' = \frac{\cos\alpha + u_1 \tan\psi_1 / v}{\left[1 + \dfrac{2u_1 \tan\psi_1}{v}\,\cos\alpha + \left(\dfrac{u_1 \tan\psi_1}{v}\right)^2\right]^{1/2}}. \tag{13.5}$$

Solving for $\cos\alpha$ gives

$$\cos\alpha = \pm\cos\alpha' \left[1 - \left(\frac{u_1 \tan\psi_1}{v}\right)^2 \sin^2\alpha'\right]^{1/2} - \frac{u_1 \tan\psi_1}{v}\,\sin^2\alpha'. \tag{13.6}$$

For $v > u_1 \tan\psi_1$ only the positive sign is allowed, and for $v < u_1 \tan\psi_1$ both signs are allowed. Positive and negative $\cos\alpha$ are to be interpreted in terms of incidence from the regions 1 and 2, respectively.

The condition for reflection at a shock can now be stated simply: particles with $\sin^2\alpha_1' \geq B_1/B_2$ in the de Hoffmann–Teller frame cannot satisfy (13.1) and so must be reflected rather than transmitted at the shock. Hence the condition for reflection is

$$\cos\alpha' < (1 - B_1/B_2)^{1/2}. \tag{13.7}$$

According to (13.5), for $v > u_1 \tan\psi_1$ all values of $\cos\alpha_1'$ are allowed, for $0 < \cos\alpha_1 < 1$, but for $v < u_1 \tan\psi_1$ only

$$[1 - (v \tan\psi_1/u_1)^2]^{1/2} < \cos\alpha' < 1 \tag{13.8}$$

is allowed. Hence one concludes that reflection is possible only for

$$v > u_1 \tan\psi_1 (B_1/B_2)^{1/2}. \tag{13.9}$$

All particles are transmitted when inequality (13.9) is reversed.

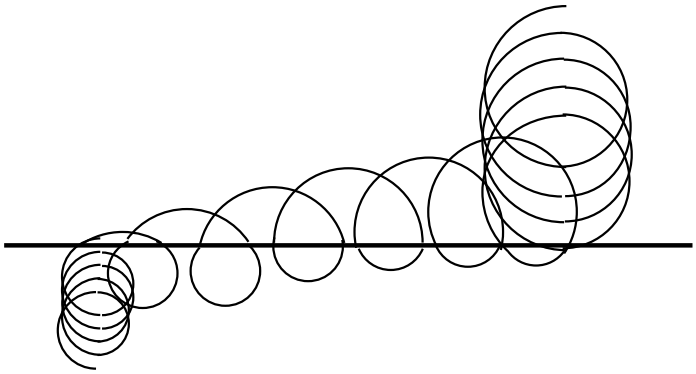


Figure 13.2: The path of an ion is indicated schematically as it crosses a shock discontinuity, represented by the horizontal line. The gyroradius is different on either side of the shock, and this causes the orbit to drift along the shock front.

13.3 Maximum energy change on reflection

Particles tend to gain energy when reflected from or transmitted through a shock. The reason for this is that the particles drift along the shock in the direction of the electric field in the shock (or opposite to the electric field if they are negatively charged). For ions incident on a shock whose thickness is much less than the ion gyroradius, this drift is illustrated in Figure 13.2. For electrons, whose gyroradius is less than the shock thickness the drift can be interpreted in terms of a ∇B drift, cf. (1.10). To see this, consider a coordinate system in which the shock normal is along the x axis, and the magnetic field is in the x-y plane. The relative velocity between the de Hoffmann–Teller frame and the laboratory frame is then a long the y-axis. In the de Hoffmann–Teller frame there is no electric field, and in the laboratory frame the electric field is determined by the vector product of the relative velocity and $\mathbf{B}$, and so is along the z-axis. According to (1.10), the ∇B drift is along $\mathbf{B} \times \nabla B$, and because ∇B is across the shock, that is, along the x-axis, this drift is also along the z-axis.

The energy change on reflection and transmission is most easily determined by noting that there is no change in energy in the de Hoffmann–Teller frame, and then applying the Lorentz transformation to find the change in the laboratory frame. Consider the case of a reflected particle. Reflection is possible only when (13.9) is satisfied, and the interesting case is when $\tan\psi_1$ is large, that is, when the magnetic field in nearly parallel to the shock front. Then the momentum component along the x-axis (the direction of the relative velocity between the two frames) is approximately equal to the parallel velocity. The Lorentz tranformation implies an energy change on reflection equal to $\Delta\varepsilon = u_0 \Delta p'_x$, which therefore implies $\Delta\varepsilon \approx u_0 \Delta p' \cos\alpha'$. On reflection $p' \cos\alpha'$ changes sign, so that

one has $\Delta p' \cos\alpha' = 2p' \cos\alpha'$. Then using (13.4) and (13.5) to reexpress the result in terms of p and α, one finds that the change in energy on reflection is given by

$$(\Delta\varepsilon)_{\rm ref} = 2\varepsilon \left(\frac{u_1 \tan\psi_1 v}{c^2}\right)\left(\cos\alpha + \frac{u_1 \tan\psi_1}{v}\right). \tag{13.10}$$

Shock drift acceleration increases the parallel energy of a reflected particle preferentially, and so leads to a decrease in the pitch angle.

The maximum energy change on reflection is determined by (13.10), subject to the restriction (13.9) for the particle to be reflected. For a nonrelativistic particle this maximum energy change on reflection reduces to

$$\Delta K_{\rm max} = 4K(B_2/B_1 - 1), \tag{13.11}$$

where $K = \varepsilon - mc^2$ is the kinetic energy. The maximum value of $B_2/B_1 = u_1/u_2$ is 4 for a strong shock, so that the maximum change in the kinetic energy is by a factor of 13.

Extending this analysis to include the case of transmitted particles leads to the following maximum changes in the energy for reflected (ref) and transmitted (trans) particles compared with the incident (in) particles in the nonrelativistic limit

$$\text{Reflection } (1 \to 1): \qquad \frac{\varepsilon_{\rm ref}}{\varepsilon_{\rm in}} = 1 + 4\left(\frac{B_2}{B_1} - 1\right), \tag{13.12}$$

$$\text{Transmission } (1 \to 2): \qquad \frac{\varepsilon_{\rm trans}}{\varepsilon_{\rm in}} = 1 + 2\left(\frac{B_2}{B_1} - 1\right), \tag{13.13}$$

$$\text{Transmission } (2 \to 1): \qquad \frac{\varepsilon_{\rm trans}}{\varepsilon_{\rm in}} = 1 + \left(\frac{B_1}{B_2}\right)^{1/2} + \left(1 - \frac{B_1}{B_2}\right)^{1/2}. \tag{13.14}$$

Note that the maximum value of B_2/B_1 across a shock is limited to less than the ratio of the tangential components, which has a maximum value of four.

13.4 The foreshock and interplanetary shocks

The strongest evidence for shock drift acceleration comes from observations of particles associated with shocks in the heliosphere. The shocks that have been studied in detail are the Earth's bow shock and other planetary bow shocks, interplanetary shocks, and bow shocks associated with comets.

The magnetic fields of the Earth and of the giant planets are strong enough to deflect the solar wind around the planet, as discussed in lecture 2, cf. Figure 2.4. As the solar wind is deflected, the plasma flow speed, which is typically around $500\,{\rm km\,s^{-1}}$, must be reduced to a sub-Alfvénic speed near the planet. This occurs at a bow shock on the sunward side of the planet. On crossing

the bow shock, the plasma flows around the planet in a surface layer called the magnetosheath. The inner boundary of the magnetosheath, called the magnetopause, separates the region where the magnetic field lines join to the Sun from the magnetosphere, where the magnetic field lines connect to the planet. The plasma flow velocity in the solar wind is radially outward from the Sun. Ahead of the bow shock the solar wind plasma is unaffected by the bow shock provided that there is no magnetic connection to the bow shock. The *foreshock* region is defined as the region between the field line that is tangent to the bow shock and the bow shock itself. This boundary is identified observationally by a small ($\approx 1\%$) abrupt increase in density. Every point in the foreshock is magnetically connected to the bow shock. The bow shock and the foreshock are illustrated in figure 13.3.

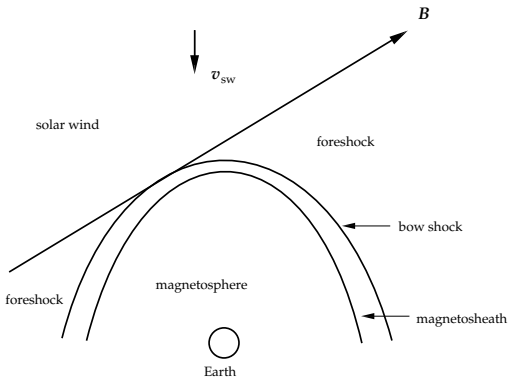


Figure 13.3: The bow shock and the foreshock are illustrated. Only one magnetic field line is shown and this is the field line in the solar wind that is tangent to the bow shock, and which defines the sunward edge of the foreshock.

Particles that encounter a shock experience shock drift acceleration, and a particle reflected from the shock front has access to the foreshock region by propagating along the magnetic field lines. Hence the foreshock is populated by particles accelerated at the shock and escaping upstream ahead of the shock. Consider a representative point in the foreshock. Particles injected at the shock can reach the representative point by propagating along the field lines. However, they reach the representative point only if their velocity, $v_\parallel$, along the field lines is high enough. A particle reaches a point $z = v_\parallel t$ along the field line from the bow shock after a time t, but, due to the motion of the plasma, in this time the particle drifts a distance $u_{\rm SW}t$ from the tangent line that defines the front of the foreshock. Hence particles of higher energy are located in regions of the foreshock closer to the tangent line and further from the bow shock. Electrons tend to have higher speeds than ions, and hence tend to be closer to the tangent line than the ions. Thus there is a spatial separation

between an electron foreshock region and an ion foreshock region, as illustrated in figure 13.4. Observations suggest that the efficiency of conversion of the energy flux in the solar wind into accelerated particles is relatively high, at least 15 percent.

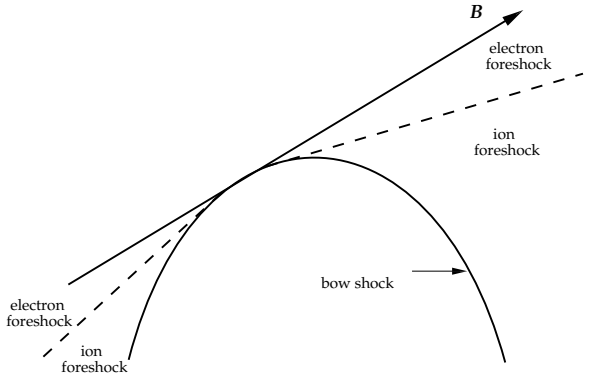


Figure 13.4: The foreshock may be separated into an electron foreshock, nearer to the field line that is tangent to the bow shock, and an ion foreshock, nearer to the bow shock; the dashed line indicates the ill-defined boundary between these two regions, which actually merge continuously into each other.

The particles streaming away from the bow shock in the foreshock region generate waves through streaming instabilities. The waves observed in the electron foreshock include Langmuir waves, whistlers and low-frequency ($\lesssim 2\,\mathrm{kHz}$) acoustic-type waves. In the ion foreshock the dominant waves are higher frequency ($\gtrsim 3\,\mathrm{kHz}$) acoustic-type waves and MHD waves. The Langmuir waves produce radio emission at the fundamental and second harmonic of the plasma frequency.

Acceleration of ions due to the shock drift mechanism also occurs at interplanetary shocks. In addition, some interplanetary shocks produce type II radio emission which can involve beams of electrons propagating away from the shock both ahead and behind it. The only seemingly plausible theory for the acceleration of such electron beams is the shock drift mechanism.

13.5 Shock drift acceleration in the Crab Nebula

A more exotic application of the shock drift mechanism is to acceleration of the highest energy electrons in the Crab Nebula. There is evidence, from the highest energy photons observed from the Crab Nebula, for electrons with energies as high a $10^{16}\,\mathrm{eV}$. (There are probably equal numbers of electrons and positrons,

and for present purpose "electron" means either an electron or a positron.) The synchrotron lifetimes for these electrons are so short that they must be being accelerated continuously.

The site of the electron acceleration is thought to be a shock that forms some ten percent of the distance from the central pulsar to the edge of the nebula. Inside this region there seems to be a deficiency in the synchrotron emission, and just outside this region "wisps" of activity that propagate outward and vary on a time scale of about a month are observed. It is believed that the whole nebula is powered by the pulsar, which generates a relativistic wind of electrons and positrons. The shock is where this wind encounters a pressure in the nebula that is sufficient to stop it, so causing a shock to form.

The suggestion that the highest energy electrons might be generated by shock drift acceleration is acceptable only if the maximum energy that the electrons can reach, according to theory, is as high as the energy inferred from observation. In the discussion here the only point discussed in detail is the maximum possible energy. In principle, a particle can drift completely across the face of a shock and gain an energy equal to the maximum potential drop across the shock. The potential available in a rotating magnetosphere may be estimated using (5.17), which suggests $\Phi_{\max} \approx \Omega_* B_* R_*^2$, cf. comments following (5.19). For the Crab Nebula, taking $B_* = 5 \times 10^5\,\mathrm{T}$, $R_* = 10^4\,\mathrm{m}$ and $\Omega_* = 2 \times 10^2\,\mathrm{s}^{-1}$, this would imply $\Phi_{\max} \approx 5 \times 10^{18}\,\mathrm{V}$, which would be adequate.

However, there is a simple but subtle constraint that must not be violated, and we need to check that the foregoing estimate is not overly optimistic. The maximum potential for a radiating system emitting a power P is limited by a simple circuit restriction. The power radiated can be written as $P = RI^2 = \Phi^2/R$, where R is the resistance and where Ohm's law, $\Phi = IR$ applies. The maximum value of the potential for a given P is determine by the maximum value of the resistance. The resistance cannot exceed the value corresponding to the vacuum: the vacuum resistance (the vacuum wave impedance) is given by

$$R_{\mathrm{vac}} = \mu_0 c = 377\,\Omega. \tag{13.15}$$

The power radiated by the Crab pulsar is $P = 10^{31}\,\mathrm{W}$, and so the maximum potential drop available is $\Phi = 6 \times 10^{16}\,\mathrm{V}$, which is comparable with the estimate made above. However if, as is believed to be the case, most of the power goes into kinetic energy rather than into electromagnetic radiation, then the maximum potential available is correspondingly reduced. Thus, while this exotic application of shock drift acceleration cannot be excluded, it is less favorable than might at first appear.

Lecture 14

Acceleration by parallel electric fields

At the start of lecture 11 eight different acceleration mechanisms are listed. So far only the first three have been discussed. Of the remaining five, two have been mentioned in passing: resonant acceleration by MHD waves has some similarity to stochastic acceleration, and acceleration in neutral sheets during reconnection has some analogy with shock drift acceleration. All five of these mechanisms have one important feature in common. This is that there is an obvious and adequate ultimate source of the energy, which comes from mass motions in a magnetized plasma. Mass motions may drive instabilities that generate MHD turbulence, or they may generate shock waves. In the remaining three acceleration mechanisms, discussed in this lecture, the magnetic field plays no major role, and the source of the particle energy is in electric fields. These acceleration mechanisms tend to be more effective for electrons than for ions.

14.1 Acceleration by Langmuir waves and ion sound waves

Acceleration by Langmuir waves and by ion sound waves has been considered in several astrophysical contexts, mostly connected with radio emission from the solar and stellar coronas. These processes are not discussed in detail here. The following remarks are restricted to introducing the ideas involved, and explaining the problems encountered in applications involving acceleration by these waves.

An *isotropic* plasma is defined to be a plasma (a) with no ambient magnetic field (it is *unmagnetized*), and (b) in which all species of particles have a Maxwellian distribution of velocities. There are two longitudinal wave modes (which are defined to have polarization vector $\mathbf{e} = \kappa = \mathbf{k}/k$) in an isotropic

plasma: Langmuir waves and ion sound waves. Langmuir waves involve only the motion of the electrons, and ion sound waves (also called ion acoustic waves) are associated with motion of the ions.

The dispersion relation for Langmuir waves may be written in the form $\omega = \omega_L(k)$, with

$$\omega_L^2(k) = \omega_p^2 + 3k^2 V_e^2, \tag{14.1}$$

where $\omega_p = (e^2 n_e/\varepsilon_0 m_e)^{1/2}$, is the plasma frequency and $V_e = (\kappa_B T_e/m_e)^{1/2}$ is the thermal speed of electrons. (The ratio of electric to total energy may be approximate by $R_L \approx \frac{1}{2}$.) It turns out that Landau damping is very strong for phase speeds $v_\phi = \omega/k$ close to V_e, and as a consequence Langmuir waves exist as weakly damped propagating waves satisfying (14.1) only for v_ϕ greater than several times V_e. The absorption coefficient for Langmuir waves due to Landau damping by thermal electrons is

$$\gamma_L(k) = \left(\frac{\pi}{2}\right)^{1/2} \frac{\omega_p^4}{k^3 V_e^3}\, e^{-[\omega_L(k)]^2/2k^2V_e^2}. \tag{14.2}$$

One physical interpretation of Landau damping is that it is due to absorption of the waves by the thermal electrons through the inverse of Cerenkov emission. Landau damping exists in the absence of collisions, and is the simplest example of a collisionless damping mechanism.

The dispersion relation for ion sound waves is

$$\omega = \omega_s(k) \approx \frac{k v_s}{(1 + k^2\lambda_{De}^2)^{1/2}}, \tag{14.3}$$

where $v_s = \omega_{pi}\lambda_{De}$ is the *ion sound speed*. In the limit $k\lambda_{De} \ll 1$ the dispersion relation (14.3) reduces to $\omega \approx k v_s$, which is characteristic of a sound wave. The ratio of electric to total energy is

$$R_s(k) \approx \frac{1}{2}\left[\frac{\omega_s(k)}{\omega_{pi}}\right]^2, \tag{14.4}$$

and the absorption coefficient is

$$\gamma_s(k) \approx \left(\frac{\pi}{2}\right)^{1/2} \omega_s(k) \left\{\frac{v_s}{V_e} + \left[\frac{\omega_s(k)}{kV_i}\right]^3 e^{-[\omega_s(k)]^2/2k^2V_i^2}\right\}. \tag{14.5}$$

The two terms inside the curly brackets in (14.5) are due to Landau damping by thermal electrons and by thermal ions, respectively. The damping by thermal ions is strong for $v_s \approx V_i$, and ion sound waves exist as weakly damped waves only for $v_s \gg V_i$, which requires $Z_i T_e \gg T_i$, where $Z_i e$ is the ionic charge.

Acceleration of fast particles by longitudinal waves results from the effect of Landau damping, which requires that the resonance condition (8.1), *viz.*

$$\omega - \mathbf{k}\cdot\mathbf{v} = 0, \tag{14.6}$$

be satisfied. This resonance condition can be satisfied only for particles with speed greater that the phase ω/k of the waves.

Acceleration by longitudinal waves may be described quantitatively using quasilinear theory. The relevant kinetic equation for the particles is, cf. (8.18), (8.19),

$$\frac{df(\mathbf{p})}{d\mathbf{p}} = \frac{\partial}{\partial p_i}\left[D_{ij}(\mathbf{p})\frac{\partial f(\mathbf{p})}{\partial p_i}\right], \tag{14.7}$$

$$D_{ij}(\mathbf{p}) = \int \frac{d^3\mathbf{k}}{(2\pi)^3}\, w_M(\mathbf{p},\mathbf{k})\hbar^2 k_i k_j N_M(\mathbf{k}), \tag{14.8}$$

where the probability is, cf. (8.12),

$$w_M(\mathbf{p},\mathbf{k}) = \frac{q^2 R_M(\mathbf{k})}{\varepsilon_0 \hbar \omega_M(\mathbf{k})}\left|\mathbf{e}_M^*(\mathbf{k})\cdot\mathbf{v}\right|^2 2\pi\delta(\omega_M(\mathbf{k}) - \mathbf{k}\cdot\mathbf{v}). \tag{14.9}$$

The energy transfer between Langmuir waves and fast electrons is very efficient. Indeed this efficiency is a central problem with any model for electron acceleration on a large scale. The coupling between Langmuir waves is so strong that the only effective source for the waves is energetic electrons themselves. A further difficulty is that the waves propagate very slowly (group speed $v_g = 3V_e^2/v_\phi$) so that they cannot transfer energy from one location to another. Hence, one expects acceleration by Langmuir waves to be confined to small spatial and temporal scales. Although this acceleration can be important in detailed aspects of specific process, it is unlikely to be important as the primary mechanism for transferring energy from macroscopic motions into energetic particles.

The problem with ion sound waves is quite different. These waves can be generated efficiently by mass motions through a current instability. The waves have phase speeds that are much less than the thermal speed of electrons, and hence they are damped by all thermal electrons. As a result the damping leads to heating rather than acceleration of the electrons.

14.2 Runaway electrons

The simplest conceivable mechanism for particle acceleration is by an electric field parallel to the magnetic field. Acceleration by a parallel electric field is usually treated in terms of a runaway process. Consider an electron in the presence of an electric field and of a frictional force. The friction is due to collisions with a collision frequency $\nu_e(v) = \nu_e(V_e/v)^3$, where V_e is the thermal speed of electrons. The equation of motion is of the form, cf. (1.25),

$$\frac{d\mathbf{v}}{dt} = -\frac{e\mathbf{E}}{m} - \nu_e\mathbf{v}\left(\frac{V_e}{v}\right)^3, \tag{14.10}$$

It follows that electrons with speed

$$\frac{v}{V_e} > \left(\frac{E_D}{E}\right)^{1/2}, \quad E_D = \frac{mV_e\nu_e}{e}, \tag{14.11}$$

are freely accelerated. These are the *runaway* particles; E_D is the *Dreicer* field. One expects runaway acceleration to result in a spectrum of electrons with energy concentrated around $e\Phi$, where Φ is the total parallel potential drop through which the electrons are accelerated.

Astrophysical plasmas are approximately collisionless for many purposes, and this raises a question concerning the validity of the collisional term in (14.10). It turns out that even in collisionless plasmas a collision-like process develops to give a frictional term of the form in (14.10), but with a different value of ν_e. This is due to a current instability. The simplest form is when the current density $J = en_ev_d$ implies a drift speed, v_d, of the electrons relative to the ions such that v_d exceeds the ion sound speed v_s. Then ion sound waves grow, and these waves scatter the electrons in a way that is analogous to scattering by Coulomb interactions. This scattering results in a form of *anomalous* resistivity or conductivity. The runaway process also occurs in the presence of anomalous resistivity.

There are two serious (but not necessarily independent) difficulties with this mechanism. One concerns how the parallel electric field is set up. In plasmas electrons flow along field lines freely and one expects them to neutralize any parallel electric field rapidly. Another difficulty is more subtle. It concerns the electric current that is necessarily set up when oppositely particles are accelerated in opposite directions by a parallel electric field. From the viewpoint of circuit theory, the inductance of a typical circuit in which such a current flows is relatively large, and inductive effects severely limit the possibility of changing the net current significantly. Specifically, Lenz' law implies that any tendency to change the current results in a charge separation and an associated electric field that opposes the initial electric field. This raises the question of how the parallel electric field could have been set up in the first place. Despite the difficulties, runaway acceleration continues to be invoked in specific astrophysical contexts.

14.3 Acceleration of impulsive solar electrons

In the impulsive phase of solar flares energy is released at a rate up to about 10^{22} W for 10^2–10^3 s. A fraction of at least > 20 percent, goes into electrons with energy around 10 keV. This implies that new electrons are accelerated at a rate $> 10^{36}\,\mathrm{s}^{-1}$, and a total number of precipitating electrons up to $\approx 10^{39}$ during a flare. There is direct evidence for these electrons through the hard X-rays that they emit as they precipitate into the denser regions of the solar

atmosphere, and from type III radio bursts generated by electron beams that escape outward. A simple model for an impulsive flare is outlined in Figure 14.1.

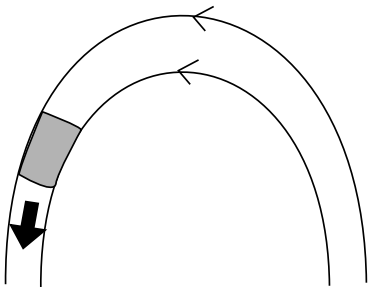


Figure 14.1: Energy is released in a flare kernel, indicated by the shaded region, in a magnetic flux tube. The energy goes into energetic electrons most of which precipitate, indicated by the solid arrow, producing hard X-rays, optical and UV radiation, and heat the ambient plasma so that hot plasma rises into the corona and emits soft X-rays.

In this case it seems that virtually all the electrons in the flare kernel are accelerated. Indeed, some estimates suggest that the number of electrons that precipitates exceeds the number present in the flare kernel initially. This is possible. The current implied by $\dot{N} \approx 10^{36}\,\mathrm{s}^{-1}$ is $I = e\dot{N} \approx 10^{17}\,\mathrm{A}$. Such a current is impossibly large. Suppose it flows in a cylinder of radius r; then the magnetic field at the edge of the cylinder is $B = \mu_0 I/2\pi r$, which for $r \approx 10^6\,\mathrm{m}$ implies $B \approx 10^4\,\mathrm{T}$. Hence such a current cannot be confined by the coronal magnetic field, $B \approx 10^{-2}\,\mathrm{T}$. The electrons can precipitate only if the current associate with them is neutralized by a upward flux of $\approx 10^{36}\,\mathrm{s}^{-1}$ cold electrons. This return current resupplies the electrons lost from the flare kernel region.

There have been several suggestions as to how the 10 keV electrons are accelerated. Runaway and acceleration during magnetic reconnection are two suggestions. Another is that they are heated through anomalous resistivity, and a further possibility involves weak double layers. However, all these mechanisms encounter difficulties and how the electrons are accelerated in the impulsive phase of solar flares remains unclear.

14.4 Electrostatic double layers in the auroral zones

For several decades it was widely accepted that parallel electric field could play no significant role in astrophysical plasmas. However, increasing observational evidence that parallel electric field can be maintained led to a revision of this opinion. The strongest evidence for parallel electric fields is in the auroral zones of the Earth. These field form in localized regions called potential double layers, also called electrostatic shocks. It is possible that double layers are important in other astrophysical contexts, but the discussion here is restricted to the auroral plasma.

Double layers may be classified as *strong* or *weak*. A strong double layer involves a coherent structure with electrons and ions flowing into the double layer and being transmitted or reflected depending on their initial momenta. There are analytical models for strong double layers, but there is no convincing observational evidence for them. In a strong double layer, the potential drop $\phi_{\rm DL}$ across the double layer is such that the potential energy is much greater than the thermal energy of the particles, i.e., $\phi_{\rm DL} \gg m_e V_e^2/e$. In contrast, a weak double layer involves a region of wave turbulence that may be regarded as a region of anomalous resistivity, with a potential drop across it that is usually assumed to be in the range $0.1 m_e V_e^2/e \lesssim \phi_{\rm DL} \lesssim 10 m_e V_e^2/e$. In simple terms the region of wave turbulence may be regarded as providing an anomalous resistance so that the potential drop across the double layer is related to the current flowing through it by Ohm's law. Weak double layers may be classified in terms of the wave mode involved. The two wave modes of most interest are the ion acoustic mode and the ion cyclotron mode.

Figure 14.2: Fig. 14.2
The variation of the relative density ($\delta n/n_0$) and potential (Φ) with height s as observed above the auroral zones from the Viking spacecraft; SW denotes a solitary wave, and WDL denotes a weak double layer.

There is evidence from both observations and simulations that a parallel potential drop tends to break up into a series of weak double layers. That is, if there is an average parallel potential drop over a large distance, then it is found that this breaks up into a series of localized potential drops in weak double layers, with essentially no parallel electric field between the weak double layers. An example from observations in the auroral zones is illustrated in Figure 14.2.

There is detailed evidence on the acceleration of electrons in the Earth's auroral zones from spacecraft that traverse the acceleration region. The accelerated electrons produce the auroral kilometric radiation, through an electron cyclotron maser instability, as mentioned in lecture 16. This radiation may be used to infer gross properties of the accelerated electrons to complement the

detailed data from direct sampling of the electrons. In view of all this observational evidence, one might expect the acceleration of auroral electrons to be well understood. However, this is not the case. There is one school of thought that favors acceleration by weak double layers. The circumstantial evidence for this is strong. However, there is an opposing view: it is argued that the detailed data are not consistent with acceleration by weak double layers. There is no convincing (in my opinion) alternative acceleration mechanism. It is reasonable to expect that further data and further theoretical work on the acceleration of auroral electrons will lead to a detailed understanding of the acceleration mechanism over the next few years.

Lecture 15

Bremsstrahlung

Bremsstrahlung is the radiation emitted by an electron due to its Coulomb interaction with another particle, usually an ion. In the astrophysical literature this is sometimes called free-free emission. Bremsstrahlung in the radio range is the basic thermal emission process from stellar coronas and winds and from HII regions. The inverse of bremsstrahlung, called collisional damping, is the basic absorption process for radio waves in a plasma. Bremsstrahlung is also an important emission process in the X-ray range.

15.1 Treatment based on the Larmor formula

A relatively simple treatment of bremsstrahlung is possible based on the Larmor formula. The Larmor formula gives the power radiated in transverse waves in vacuo by an accelerated, nonrelativistic charge:

$$P(t) = \frac{q^2 |\mathbf{a}(t)|^2}{6\pi\varepsilon_0 c^3}, \tag{15.1}$$

where $\mathbf{a}(t)$ is the acceleration. The acceleration in (5.1) may be rewritten in terms of the force, $\mathbf{F}(t)$, acting on the particle, $\mathbf{a}(t) = \mathbf{F}(t)/m$. Before discussing the use of (15.1), let us note a related formula for the energy radiated per unit frequency in an isotropic dielectric with refractive index $n(\omega)$:

$$U_{\rm rad}(\omega) = \frac{q^2\, n(\omega)\, |\mathbf{F}(\omega)|^2}{6\pi^2\varepsilon_0 m^2 c^3}, \tag{15.2}$$

where $\mathbf{F}(\omega)$ is the temporal Fourier transform of the force $\mathbf{F}(t)$.

Consider an electron moving near a fixed ion with charge Ze situated at the origin, as illustrated in Figure 15.1. The equation of motion of the electron is

$$m_e \ddot{\mathbf{X}}(t) = -\frac{Ze^2 \mathbf{X}(t)}{4\pi\varepsilon_0 |\mathbf{X}(t)|^3}. \tag{15.3}$$

On inserting $\ddot{\mathbf{X}}(t) = \mathbf{a}(t)$ in the Larmor formula (15.1), one obtains

$$P(t) = \frac{2Z^2}{3}\left(\frac{e^2}{4\pi\varepsilon_0}\right)^3 \frac{1}{|\mathbf{X}(t)|^4 m_e^2 c^3}, \tag{15.4}$$

which describes *bremsstrahlung*. Bremsstrahlung in the radio range results from distant encounters, when the hyperbolic orbit in Figure 15.1 may be replaced by a straight line to a first approximation. Classical formulas, such as (15.4), are adequate for this case. However, quantum effects need to be considered for close encounters.

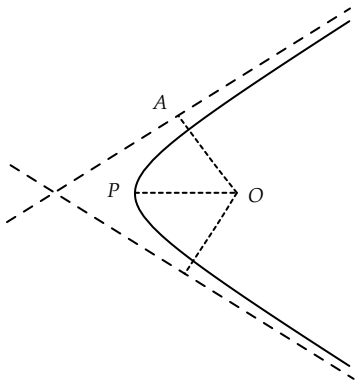


Figure 15.1: A hyperbolic orbit of an electron for an ion at the origin O; $OP = r_{\min}$, $OA = b$, and the dashed lines are asymptotes.

The power radiated is of little interest in a single encounter between an electron and an ion. Each encounter is regarded as an impulsive event, and the quantity of interest is the energy radiated in such an event. The energy radiated in a single encounter is given by integrating (15.4) over all time. For this purpose it is convenient to write $|\mathbf{X}(t)| = r(t)$, where $r(t)$ is the radial distance of the electron from the ion, and to write the integral as one over $r(t)$:

$$E_{\rm rad} = \int_{-\infty}^{\infty} dt\, P(t) = 2\int_{r_{\min}}^{\infty} \frac{dr}{\dot{r}}\,\frac{2Z^2}{3}\left(\frac{e^2}{4\pi\varepsilon_0}\right)^3 \frac{1}{r^4 m_e^2 c^3}, \tag{15.5}$$

where $r_{\min}$ is the distance of closest approach. The integral in (15.5) may be performed exactly. The angular momentum and the energy of the electron are conserved. Let v_0 be the speed of the electron at infinity, and b be the impact parameter; the orbit is illustrated in Figure 15.1. Then these conservation laws become

$$m_e r^2\dot{\theta} = m_e b v_0, \qquad \tfrac{1}{2}m_e(\dot{r}^2 + r^2\dot{\theta}^2) - \frac{Ze^2}{4\pi\varepsilon_0 r} = \tfrac{1}{2}m_e v_0^2, \tag{15.6a,b}$$

respectively. After eliminating $\dot{\theta}$ between (15.6a) and (15.6b), one solves for $\dot{r}$ as a function of r and substitute the result into (15.5). At the distance of closest approach $r = r_{\min}$ one has $\dot{r} = 0$, and (15.6a) and (15.6b) allow one to determine $r_{\min}$ in terms of the other parameters. The integral gives

$$E_{\rm rad} = \frac{2Z^2}{3m_e^2c^3v_0b^3}\left(\frac{e^2}{4\pi\varepsilon_0}\right)^3\left\{3\frac{b_0}{b} + \left(1+3\frac{b_0^2}{b^2}\right)\left[\frac{\pi}{2} + \arctan\left(\frac{b_0}{b}\right)\right]\right\}, \tag{15.7}$$

with $b_0 = (e^2/4\pi\varepsilon_0)(Z^2/m_ev_0^2)$. Distant encounters correspond to $b \gg b_0$ (when one has $b \approx r_{\min}$). In this case (15.7) gives

$$E_{\rm rad} \approx \frac{\pi Z^2}{3m_e^2c^3v_0b^3}\left(\frac{e^2}{4\pi\varepsilon_0}\right)^3. \tag{15.8}$$

The form of the energy spectrum is inferred from the fact that the encounter is impulsive with a characteristic time scale of b/v_0. One expects the energy emitted per unit frequency to be approximately independent of ω for $\omega \ll v_0/b$, to fall off for $\omega \approx v_0/b$, and to be negligible for $\omega \gg v_0/b$. A more detailed calculation using (15.2) gives

$$U_{\rm rad}(\omega) = \frac{8}{3\pi}\frac{n(\omega)Z^2e^6\omega^2}{(4\pi\varepsilon_0)^3m_e^2c^3v^4}\left[K_1^2\left(\frac{b\,\omega}{v}\right) + K_0^2\left(\frac{b\,\omega}{v}\right)\right]. \tag{15.9}$$

The result (15.9) involves the Macdonald functions, which may be approximated for small argument by

$$K_n(z) \approx \begin{cases} \ln(2/\Gamma z) & \text{for } n = 0, \\ 1/z^n & \text{for } n \neq 0. \end{cases} \tag{15.10a}$$

with $\Gamma = 1.7811\ldots$, $\ln\Gamma = 0.5722\ldots$, and for large argument by

$$K_\nu(z) \approx \left(\frac{\pi}{2z}\right)^{1/2} e^{-z}. \tag{15.10b}$$

Using these approximations (15.9) gives

$$U_{\rm rad}(\omega) = \frac{8}{3\pi}\frac{n(\omega)Z^2e^6\omega^2}{(4\pi\varepsilon_0)^3m_e^2c^3v^4}\begin{cases} v^2/b^2\omega^2 & \text{for } b\,\omega/v \ll 1, \\ \pi(v/2b\,\omega)\,e^{-2b\omega/v} & \text{for } b\,\omega/v \gg 1. \end{cases} \tag{15.11}$$

The power emitted by an electron as it experiences distant encounters with many ions along its path is found as follows. Suppose that there are ions of various species, with the ith species having charge Z_ie and number density n_i. The number of encounters that an electron with speed v has per unit time with ions of species i with impact parameters between b and $b + db$ is $2\pi dbb\,n_iv$, cf. Figure 15.2. Hence the power radiated is

$$P(\omega) = \sum_i 2\pi n_iv\int_0^\infty db\,b\,U(\omega), \tag{15.12}$$

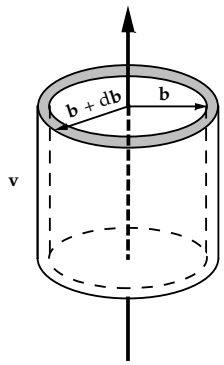


Figure 15.2: Fig. 15.2
An electron has encounters per unit time with impact parameter between b and $b + db$ with all ions in the cylindrical surface shown.

where the sum is over all species of ion. On substituting the expression (15.11) for $U(\omega)$ into (15.12), the integral diverges at $b = 0$. Let us cut the integral off at $b = b_0$. The integral then gives

$$P(\omega) = \sum_i \frac{16}{3} \frac{n(\omega) Z_i^2 n_i e^6}{(4\pi\varepsilon_0)^3 m_e^2 c^3 v} \begin{cases} \ln\left(\frac{2}{\Gamma}\frac{v}{b_0\omega}\right) & \text{for } \omega \ll v/b_0, \\ \frac{\pi}{2} e^{-2b_0\omega/v} & \text{for } \omega \gg v/b_0, \end{cases} \tag{15.13}$$

with $b_0 = (e^2/4\pi\varepsilon_0)(Z^2/m_e v_0^2)$.

15.2 Emission and absorption in a thermal plasma

The result (15.13) applies to bremsstrahlung emission per electron. The power radiated per unit volume by a thermal distribution of electrons is found by integrating over a thermal distribution. Writing $\kappa_B T_e = m_e V_e^2$, the average over a thermal distribution gives

$$\left\langle \frac{1}{v} \right\rangle = \left(\frac{2}{\pi}\right) \frac{1}{V_e}. \tag{15.14}$$

The emissivity (power per unit frequency per unit solid angle) per unit volume is then given by

$$\frac{d\eta(\omega)}{dV} = \frac{n_e \langle P(\omega) \rangle}{4\pi} = \sum_i \frac{4}{3\pi} \frac{n(\omega) Z_i^2 n_i n_e e^6}{(4\pi\varepsilon_0)^3 m_e^2 c^3} \left(\frac{2}{\pi}\right)^{1/2} \frac{1}{V_e} \frac{\pi}{\sqrt{3}} \bar{G}(V_e, \omega). \tag{15.15}$$

where the *Gaunt factor* for a classical thermal plasma may be expressed in terms of the Coulomb logarithm, cf. (1.24):

$$\frac{\pi}{\sqrt{3}}\,\bar{G}(V_e,\omega) = \ln\Lambda - \ln\left(\frac{\Gamma}{2}\frac{\omega}{\omega_p}\right). \tag{15.16}$$

The absorption coefficient for collisional damping may be written in the form

$$\gamma_{\rm c}(\omega) = \frac{\omega_p^2}{\omega^2}\,\nu(\omega), \tag{15.17}$$

with the *collision frequency* identified as

$$\nu(\omega) = \sum_i \frac{1}{3}\left(\frac{2}{\pi}\right)^{1/2} \frac{4\pi Z_i^2 n_i e^4}{(4\pi\varepsilon_0)^2 m_e^2 V_e^3}\,\frac{\pi}{\sqrt{3}}\,\bar{G}(V_e,\omega). \tag{15.18}$$

One may derive (15.18) from (15.16) by appealing to Kirchhoff's law: emission and absorption for a thermal distribution of particles must be in balance for a thermal distribution of radiation. This requires

$$\gamma_{\rm c}(\omega) = \frac{d\eta(\omega)}{dV}\,\frac{(2\pi c)^3}{2n(\omega)\omega^2 m_e V_e^2}, \tag{15.19}$$

with $m_e V_e^2 = \kappa_B T_e$, where T_e is the temperature of the electrons.

15.3 Bremsstrahlung at high frequencies

At high frequencies, when the energy of the emitted photon is comparable with the energy of the electron, a quantum treatment is essential. When the final speed v' of the electron is significantly different from v, conservation of energy implies

$$\tfrac{1}{2}m_e v'^2 = \tfrac{1}{2}m_e v^2 - \hbar\omega. \tag{15.20}$$

A quantum calculation in the Born approximation gives

$$\frac{\pi}{\sqrt{3}}\,G(v,\omega) = \ln\left(\frac{v+v'}{v-v'}\right). \tag{15.21}$$

A more general quantum calculation involves using the exact wavefunctions for an electron in the Coulomb field of an ion. The wave function involves a hypergeometric function. A relevant case is for

$$Z_i\alpha \ll \frac{Z_i e^2}{4\pi\varepsilon_0\hbar v} \ll 1, \tag{15.22}$$

where $\alpha \approx 1/137$ is the fine structure constant. The expression for the Gaunt factor in this case is

$$\frac{\pi}{\sqrt{3}}\, G(v,\omega) = \frac{Z_i e^2}{\varepsilon_0 \hbar v}\, \frac{1}{1-\exp(-Z_i e^2/2\varepsilon_0 \hbar v')}, \tag{15.23}$$

which $Z_i e^2 \ll 2\varepsilon_0 \hbar v'$ for reproduces (15.21) for $v' \ll v$.

An important special case of (15.23) is the limit $v' \to 0$, in which case the denominator is equal to unity. The Gaunt factor approaches a constant value for $v' = 0$, which is the case in which the electron loses all its initial energy to a single photon. The case $\hbar\omega = \frac{1}{2}mv^2$ separates free–free and free–bound transitions. The relatively high efficiency of emission at $\hbar\omega \approx \frac{1}{2}mv^2$ may be attributed to the high density of bound states in the energy range separating bound and free electrons. In practice, X-ray emission from energetic electrons in a plasma is dominated by emission of photons with energy comparable to that of the electrons.

15.4 Optically thin and optically thick sources

In applying bremsstrahlung to the interpretation of radioastronomical sources, it is appropriate to change notation to one more closely related to that used by observers. Radio observations are usually described in terms of angular frequency $\nu = \omega/2\pi$, rather than angular frequency ω. The observational quantity for the level of radiation is the *flux density*, F_ν, which is the power received per unit area and per unit frequency range. The convention unit for F_ν is the (non SI unit) jansky (Jy), with $1\mathrm{Jy} = 10^{-26}\mathrm{W\,m^{-2}Hz^{-1}} = 10^{-23}\mathrm{erg\,s^{-1}\,cm^{-2}Hz^{-1}}$. A more basic quantity, from a theoretical viewpoint, is the *specific intensity* or *brightness*, I_ν, which is the power per unit area per unit frequency range per unit solid angle. Thus one has $I_\nu = F_\nu/\Delta\Omega$, where $\Delta\Omega$ is the solid angle subtended by the source, if it is unresolved by the radio telescope, or the solid angle of the telescope beam for a source that is resolved.

The transfer equation for electromagnetic radiation in a medium with refractive index approximated by unity may be written in the form

$$\frac{dI_\nu}{ds} = \eta_\nu - \kappa_\nu I_\nu, \tag{15.24}$$

where s denotes distance along the ray path, η_ν is the emission coefficient and κ_ν is the absorption coefficient per unit length. *Kirchhoff's law* requires that for a thermal distribution of particles at temperature T these two coefficient be related by

$$\frac{\eta_\nu}{\kappa_\nu} = B_\nu(T) \tag{15.25}$$

where $I_\nu = B_\nu(T)$ corresponds to a thermal distribution of radiation at temperature T. A Planck distribution corresponds to

$$B_\nu(T) = \frac{2h\nu^3}{c^2}\,\frac{1}{e^{h\nu/\kappa_B T}-1} = \begin{cases} \dfrac{2\kappa_B T\nu^2}{c^2} & \text{for } h\nu \ll \kappa_B T \quad \text{(Rayleigh–Jeans law)}, \\ \dfrac{2h\nu^3}{c^2} e^{-h\nu/\kappa_B T} & \text{for } h\nu \gtrsim \kappa_B T \quad \text{(Wien's law)}. \end{cases} \tag{15.26}$$

The specific intensity may be written in terms of the brightness temperature T_b,

$$T_\mathrm{b} = \frac{c^2}{2\nu^2\kappa_B}\, I_\nu, \tag{15.27}$$

which is a function of ν in general.

The transfer equation (15.24) may be integrated through a source. In the simplest case, the source is a uniform slab of thickness ℓ perpendicular to the line of sight. The solution for $I_\nu(s)$ gives

$$I_\nu(\ell) = I_\nu(0)e^{-\tau_\nu} + B_\nu(T)(1-e^{-\tau_\nu}), \qquad T_\mathrm{b}(\ell) = T_\mathrm{b}(0)e^{-\tau_\nu} + T(1-e^{-\tau_\nu}), \tag{15.28}$$

where τ_ν is the *optical depth*:

$$\tau_\nu = \int_0^\ell ds\, \kappa_\nu, \tag{15.29}$$

which is equal to $\kappa_\nu \ell$ in this simple model.

A source is *optically thin* for $\tau_\nu \ll 1$ and *optically thick* for $\tau_\nu \approx 1$. According to (15.28), an optically thin source has $T_\mathrm{b} \ll T$, and an optically thick source has $T_\mathrm{b} \approx T$. For an optically thin source, the intensity or the brightness temperature are proportional to the optical depth.

15.5 Applications of bremsstrahlung

Astrophysical applications of bremsstrahlung include those mentioned at the beginning of this lecture. Modeling of radio sources emitting bremsstrahlung is usually relatively simple. Most sources are optically thick, and then the brightness temperature gives information on the electron temperature in the source. Sources emitting X-rays are usually optically thin, and the X-rays may be due either to thermal particles or to nonthermal particles passing through the plasma.

For thermal bremsstrahlung in the radio range, the emission coefficient may be expressed in terms of the absorption coefficient using (15.25). The absorption

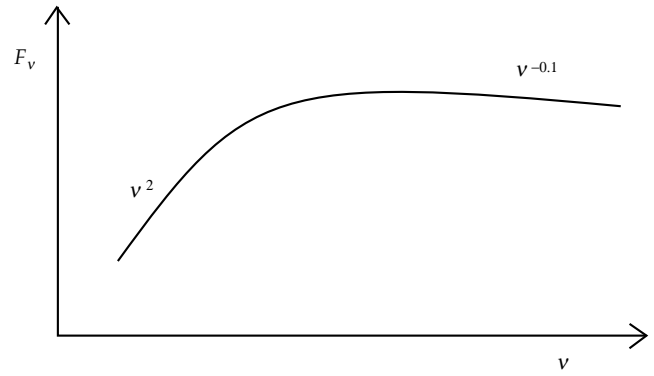


Figure 15.3: Fig. 15.3
An idealized spectrum (log–log plot) for a source emitting bremsstrahlung in the radio range is illustrated.

coefficient is given by $\kappa_\nu = \gamma_{\rm coll}/c$ here. Inserting numerical values in (15.17) with (15.18) gives

$$\kappa_\nu = \frac{\gamma_{\rm c}(\omega)}{c} = 1.0 \left(\frac{n_e}{1\,{\rm cm}^{-3}}\right)^2 \left(\frac{T_e}{1\,{\rm K}}\right)^{-3/2} \left(\frac{\nu}{1\,{\rm Hz}}\right)^{-2.1}, \tag{15.30}$$

where the logarithmic dependence on frequency in (15.16) is approximated by a power law dependence $\propto \omega^{-0.1}$. The spectrum of an idealized source is illustrated in Figure 15.3.

The dependence on the number density of electrons is quite specific, and is used to define the *emission measure*:

$$\mathrm{EM} = \int ds\, n_e^2. \tag{15.31}$$

The optical depth is proportional to the emission measure.

Some numerical values for a specific HII region, the Orion Nebula, are $n_e = 2 \times 10^3\,{\rm cm}^{-3}$, $T_e = 8 \times 10^3\,{\rm K}$, radius of the source, $R = 0.6\,{\rm pc}$, distance to the source, $r = 500\,{\rm pc}$, maximum flux density 400 Jy at 3 GHz.

Another application of bremsstrahlung in the radio range is to the solar corona, which is known to have a temperature of a few times $10^6\,{\rm K}$. Radio emission from the *quiet Sun* is due to optically thick bremsstrahlung. However, the brightness temperature is substantially less, by a factor of about four, than implied by the known temperature of the corona. Why this should be so is not immediately obvious. In the absence of absorption, which is a good approximation here, I_ν and $T_{\rm b}$ are constants along a ray path, and the value at the Earth should faithfully reflect the value in the corona. There is a subtle effect that

seems to account for this inconsistency: the radio waves are generated relatively close to the plasma frequency, and are refracted in random ways so that the apparent radius of the radio Sun is about twice the actual radius of the region from which the radio emission originates at a specific frequency. The apparent brightness temperature is then less than the true brightness temperature by the ratio of the apparent area to the true area of the source, which is indeed a factor of order four.

Bremsstrahlung in the X-ray range due to energetic electrons hitting a dense plasma, is said to be *thick target* if the electrons are all stopped, and *thin target* if most of the electrons pass through the plasma. There is a relatively simple semiquantitative result that is useful in discussing thick-target bremsstrahlung. This relies on the fact that the emission of bremsstrahlung in the X-ray range is dominated by photons with energy $\hbar\omega$ approximately equal to the initial energy of the electrons, cf. the discussion following (15.23) above. It then follows that the power radiated per electrons is of order $P(\omega)\omega$ evaluated at $\omega = \frac{1}{2}m_e v^2/\hbar$. The main loss of energy by an electron is due to its collisional slowing down, which implies a power loss of $\nu(v)\frac{1}{2}m_e v^2$, with $\nu(v)$ given by (1.23). The ratio of these two powers may be interpreted as the probability that an electron emit an X-ray photon with $\hbar\omega = \frac{1}{2}m_e v^2$ before it is stopped. This ratio is proportional to v/c, but is insensitive to other parameters. Emission in the X-ray range corresponds to photon energies, $\varepsilon_{\rm ph}$ tens of keV, and one has $v/c = (\varepsilon_{\rm ph}/10\,{\rm keV})^{1/2}/5$, so that even the dependence on this parameter is relatively weak. One finds a probability of about one part in 10^5 of the electron emitting such a photon. For example, an electron precipitation rate of $\approx 10^{36}\,{\rm s}^{-1}$ for 20 keV electrons in the impulsive phase of a solar flare follows directly from the X-ray data implying that photons of 20 keV are emitted at a rate $\approx 10^{31}\,{\rm s}^{-1}$.

Lecture 16

Gyromagnetic emission

Gyromagnetic emission is the generic name for emission due to a particle spiraling in a magnetic field. Gyromagnetic emission from nonrelativistic electrons is called cyclotron emission; it is concentrated at lines near harmonics $s = 1$, 2, 3 ... of the cyclotron frequency. Gyromagnetic emission from mildly relativistic electrons is called gyrosynchrotron emission, and gyromagnetic emission from highly relativistic electrons is called synchrotron emission.

16.1 Gyromagnetic emission in vacuo

The probability for gyromagnetic emission is written down in a general form in (8.12). Let us consider gyromagnetic emission in a vacuum and derive the total power emitted.

For waves in a vacuum (8.12) gives

$$w(\mathbf{p},\omega,\theta,s) = \frac{\pi q^2 \beta_\perp c^2}{\varepsilon_0 \hbar \omega} \left\{ \left[\frac{z}{x} J_s(sx) \right]^2 + \left[J_s'(sx) \right]^2 \right\} \delta\big[\omega(1 - \beta_\parallel \cos\theta - s\Omega\big],$$
$$z = \frac{\cos\theta - \beta_\parallel}{1 - \beta_\parallel \cos\theta}, \qquad x = \frac{\beta_\perp \sin\theta}{1 - \beta_\parallel \cos\theta}, \tag{16.1}$$

with $\omega = kc$, $\beta_\parallel = v_\parallel/c$, $\beta_\perp = v_\perp/c$, $k_\parallel = k\cos\theta$. In (16.1) a sum over the two transverse states of polarization is performed. In vacuo, only emission at $s \geq 1$ is allowed by the δ-function in (16.1). The emission at each harmonic is completely polarized, with emission at the sth harmonic being elliptical with axial ratio

$$T = -\eta \frac{\cos\theta - \beta_\parallel}{\beta_\perp \sin\theta} \frac{J_s(sx)}{J_s'(sx)}. \tag{16.2}$$

The total power radiated may be evaluated by multiplying the probability

by $\hbar\omega$, summing over $s \geq 1$ and integrating over $d^3\mathbf{k}/(2\pi)^3$,

$$\frac{d^3\mathbf{k}}{(2\pi)^3} = \frac{\omega^2}{(2\pi c)^3}\, d\omega d\cos\theta d\phi, \tag{16.3}$$

The integral over ϕ is trivial, an the ω-integral may be performed using the δ-function, giving

$$P = \sum_{s=1}^{\infty} \int_{-1}^{1} d\cos\theta\, \frac{q^2}{4\pi\varepsilon_0 c}\, \frac{s^2\Omega_0^2\beta_\perp^2}{\gamma^2(1-\beta_\parallel\cos\theta)^3}\, \left\{ \left[\frac{z}{x} J_s(sx)\right]^2 + [J_s'(sx)]^2 \right\}, \tag{16.4}$$

The remaining integral and sum may be preformed explicitly. Moreover they may be performed explicitly in either order, and they may be performed separately for the two terms inside the square brackets in (16.4), denoted $\parallel$ and $\perp$, respectively. Performing the integral first gives

$$P^{\parallel,\perp} = \sum_{s=1} \frac{2q^2 s^2 \Omega_0^2}{4\pi\varepsilon_0 c}\, \beta'^2(1-\beta'^2)\, G_s^{\parallel,\perp},$$

$$\begin{aligned} G_s^{\parallel} &= \frac{1}{\beta'^2}\int_0^{\beta'} dy\, \frac{J_{2s}(2sy)}{y} - \frac{1}{\beta'^3}\int_0^{\beta'} dy\, J_{2s}(2sy), \\ G_s^{\perp} &= \frac{1}{s\beta'} J'_{2s}(2s\beta') - \frac{1}{\beta'^2}\int_0^{\beta'} \frac{dy}{y}\, J_{2s}(2sy) + \frac{1}{\beta'}\int_0^{\beta'} dy\, J_{2s}(2sy). \end{aligned} \tag{16.5}$$

Performing the sum before the integral in (16.4) gives

$$\begin{aligned} P = {} & \frac{q^2\Omega_0^2}{16(4\pi\varepsilon_0)c}\, \beta_\perp^2(1-\beta^2) \\ & \times \int_{-1}^{1} d\cos\theta \left\{ \frac{(\cos\theta-\beta_\parallel)^2[4(1-\beta_\parallel\cos\theta)^2 + \beta_\perp^2\sin^2\theta]}{[(1-\beta_\parallel\cos\theta)^2 - \beta_\perp^2\sin^2\theta]^{7/2}} \right. \\ & \left. + \frac{4(1-\beta_\parallel\cos\theta)^2 + 3\beta_\perp^2\sin^2\theta}{[(1-\beta_\parallel\cos\theta)^2 - \beta_\perp^2\sin^2\theta]^{5/2}} \right\}. \end{aligned} \tag{16.6}$$

Performing the sums in (16.5) give

$$\begin{aligned} P = P^{\parallel} + P^{\perp} &= \frac{2q^2\Omega_0^2 p_\perp^2}{12\pi\varepsilon_0 m^2 c^3} = \frac{2}{3}\, \frac{q^2\Omega_0^2\gamma^2\beta_\perp^2}{4\pi\varepsilon_0 c}, \\ \frac{P^{\parallel} - P^{\perp}}{P^{\parallel} + P^{\perp}} &= -\frac{2+\beta_\perp^2 - 2\beta_\parallel^2}{4(1-\beta_\parallel^2)} = -\frac{2m^2c^2 + 3p_\perp^2}{4(m^2c^2 + p_\perp^2)}, \end{aligned} \tag{16.7}$$

and the expression for the total power emitted P may be derived from (16.6) by performing the integral.

16.2 Cyclotron emission and absorption by thermal electrons

Cyclotron emission and absorption by nonthermal electrons may be treated in an approximate way by noting that x in the argument of the Bessel functions in (16.1) is small. The power series expansion

$$J_\nu(z) = \sum_n \frac{(-)^n}{n!\Gamma(\nu+n+1)} \left(\frac{z}{2}\right)^{2n+\nu}. \tag{16.8}$$

then converges rapidly and only the leading term need be retained. It is sometimes important to take account of the wave properties being different from those of waves in vacuo, and it complicates the algebra only in a modest way to include the details of these properties. Hence, consider the emission of waves in an arbitrary mode M with refractive index n_M and other wave properties given by

$$\mathbf{e}_M = (\frac{L_M\kappa + T_M\mathbf{t} + i\mathbf{a})}{(L_M^2+T_M^2+1)^{1/2}}, \qquad R_M = \frac{L_M^2+T_M^2+1}{2(T_M^2+1)\,\partial(\omega n_M)/\partial\omega}, \tag{16.9}$$

where in a coordinate system with $\mathbf{B}$ along the z-axis and $\mathbf{k}$ in the x-z plane one has

$$\kappa = (\sin\theta, 0, \cos\theta), \quad \mathbf{t} = (\cos\theta, 0, -\sin\theta), \quad \mathbf{a} = (0,1,0). \tag{16.10}$$

A Maxwellian distribution function of the electrons may be written

$$f(p) = \frac{n_e\, e^{-v^2/2V_e^2}}{(2\pi)^{3/2} m_e^3 V_e^3}, \tag{16.12}$$

with $\kappa_B T_e = m_e V_e^2$, written in terms of $\beta_0 = V_e/c$ in the following. The absorption coefficient at the sth harmonic in this case is given by

$$\gamma_M(s,\omega,\theta) = \frac{\pi^{1/2}\omega_p^2 A_M(s,\omega,\theta)}{2\omega n_M\beta_0|\cos\theta|\partial(\omega n_M)/\partial\omega}\, e^{-(\omega-s\Omega_e)^2/2\omega^2 n_M^2\beta_0^2\cos^2\theta}, \tag{16.13}$$

$$A_M(s,\omega,\theta) \approx \frac{(\frac{1}{2}\lambda_M)^{s-1}}{2s!(1+T_M^2)}\left[\frac{\omega}{\Omega_e}(L_M\cos\theta - T_M\sin\theta)\tan\theta + sT_M\sec\theta + s\right]^2, \tag{16.14}$$

with $\lambda_M = (\omega/\Omega_e)^2 n_M^2\beta_0^2\sin^2\theta$. A further approximation is when the polarization of the magnetoionic waves, written $M = \sigma = \pm$ with $\sigma = +1$ for the o mode and $\sigma = -1$ for the x mode are approximately circular:

$$A_\sigma(s,\omega,\theta) \approx \frac{s^2}{4}\frac{(\frac{1}{2}\lambda_M)^{s-1}}{s!}(1-\sigma|\cos\theta|)^2. \tag{16.15}$$

The width (in frequency) of the line at the sth harmonic, $(\Delta\omega)_s$, is determined by the width of the exponential function in (16.13):

$$(\Delta\omega)_s/s\Omega_e = n_M\beta_0|\cos\theta|. \tag{16.16}$$

Thermal electron cyclotron emission occurs in the slowly varying component of solar microwave emission, where is arises from hot plasma near sunspots ($B \approx$ 0.1–0.3 T), and in optical cyclotron emission from white dwarf stars with strong magnetic fields $B \approx 10^3$ T. In these astrophysical applications the emission is optically thick. Gyromagnetic absorption by thermal electrons can also be important in preventing escape of radio emission generated at $\omega < 2\Omega_e$ or $\omega < 3\Omega_e$ from the solar corona. Specifically, suppose that the magnetic field decreases with radial distance r from the center of the Sun as $dB/dr = -B/L_B$; then the optical depth for absorption at the sth harmonic may be estimated from the product of the maximum value of the absorption coefficient (16.13) times the light propagation time over a distance in which $s\Omega_e$ changes by $(\Delta\omega)_s$. This gives

$$\tau_{Ms} = \sqrt{2}n_e\beta_0|\cos\theta|L_B\gamma_{Ms}/c. \tag{16.17}$$

For $n_e = 10^{15}\,\mathrm{m}^{-3}$, $T_e = 3\times10^6$ K, $B = 10^{-2}$ T, $L_B = 10^7$ m, $\theta = 45°$, one finds that τ_{Ms} is very larger for both modes at $s = 2$, and is large for the x-mode at $s = 3$. Hence one expects no radiation to escape through the second harmonic layer $\omega = 2\Omega_e$.

16.3 Cyclotron maser emission

Electron cyclotron emission can be negative leading to electron cyclotron maser emission. The most favorable case for negative absorption is for x mode waves at $\theta \approx \pi/2$ at the fundamental $s = 1$ in a plasma with $\omega_p \ll \Omega_e$. The required source of free energy is a distribution of electrons with $\partial f/\partial p_\perp > 0$ at small $p_\perp$. The discussion of electron cyclotron maser emission here is centered around explaining why this case is the most favorable, and outlining some applications.

The case of nearly perpendicular propagation $\theta \approx \pi/2$ is of special significance because of the form of the resonance condition (8.7), $\omega - s\Omega_e/\gamma - \omega[1 - n_M(\omega,\theta)\beta_\parallel\cos\theta] = 0$. The nonrelativistic approximation involves replacing the Lorentz factor γ by unity. In terms of the resonance ellipse, cf. Figure 8.4, this corresponds to assuming that the ellipse is approximated by a vertical line. In contrast, for $\theta = \pi/2$ the resonance ellipse reduces to a circle centered on the origin of velocity space and with radius $\beta = (1-\omega^2/s^2\Omega_e^2)^{1/2}$. A vertical line is a poor approximation to this circle when its radius is small. The nonrelativistic approximation is invalid for $\theta = \pi/2$ no matter how small the energy of the electron. A "semirelativistic" approximation to the resonance condition is

$$\omega - s\Omega_e(1 - \tfrac{1}{2}\beta_\perp^2 - \tfrac{1}{2}\beta_\parallel^2) - \omega[1 - n_M(\omega,\theta)\beta_\parallel\cos\theta] = 0. \tag{16.18}$$

This corresponds to a resonance ellipse that is a circle with its center displaced from the origin. Three cases are illustrated in Figure 16.1. The resonance curve labeled (a) is a vertical line at $v_\parallel = (\omega - s\Omega_e)/k_\parallel$, which corresponds to an ellipse with its center far to the left so that the line is actually an arc of a very large ellipse. The resonance curve labeled (b) illustrates an intermediate case where the center is closer to the shaded regions. The resonance curve labeled (c) corresponds to the semirelativistic case (16.18). The circle drawn is chosen so that it lies just outside the region where the thermal electrons are located so that thermal gyromagnetic absorption is absent.

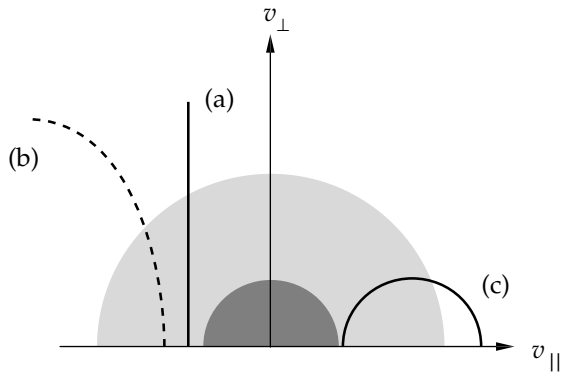


Figure 16.1: Examples of resonance ellipses in the non-relativistic region of velocity space. Thermal electrons occupy the darkly shaded region, and non-thermal electrons occupy the lightly shaded region. Curves (a), (b) and (c) are discussed in the text.

The gyromagnetic absorption coefficient is, cf. (8.17),

$$\gamma_M(\mathbf{k}) = -\sum_{s=-\infty}^{\infty} 2\pi \int_{-\infty}^{\infty} dp_\parallel \int_0^{\infty} dp_\perp p_\perp \, w_M(\mathbf{p}, \mathbf{k}, s) \times\hbar \left(\frac{s\Omega}{v_\perp} \frac{\partial}{\partial p_\perp} + k_\parallel \frac{\partial}{\partial p_\parallel} \right) f(p_\parallel, p_\perp). \tag{16.19}$$

The integrals over $p_\perp$ and $p_\parallel$ are transformed into ones over $\beta_\perp$ and $\beta_\parallel$. One of these integrals is performed over the δ-function, and the remaining integral then has a simple geometric interpretation: it is an integral around the resonance ellipse.

There are two important requirements that need to be satisfied for electron cyclotron maser emission to be a possible source of escaping radiation. One is that there be a distribution of electrons with available free energy in the appropriate form, specifically with $\partial f/\partial p_\perp > 0$ at small $p_\perp$. One such distribution is the loss cone distribution illustrated in Figure 16.2. Loss cones form naturally

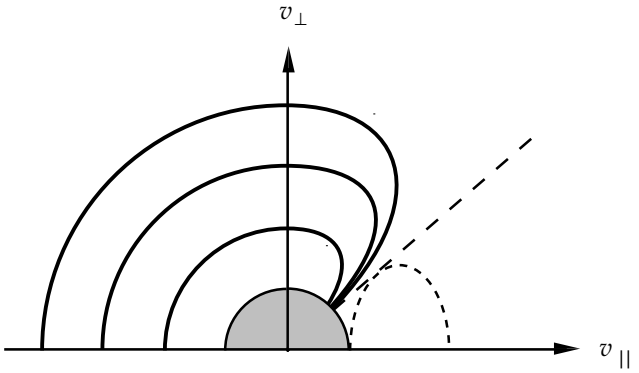


Figure 16.2: A loss-cone distribution with loss-cone angle indicated by the dashed line. The resonance ellipse illustrated favors maser action.

in a magnetic trap where electrons are confined due to the magnetic mirror effect. Thus this first requirement is expected to be satisfied for magnetically trapped electrons under a wide variety of circumstances. The other requirement is that there be waves with the dispersion properties needed (i) to define the most favorable resonant ellipse, as in the case illustrated in Figure 16.2, and (ii) to allow direct escape from the plasma.

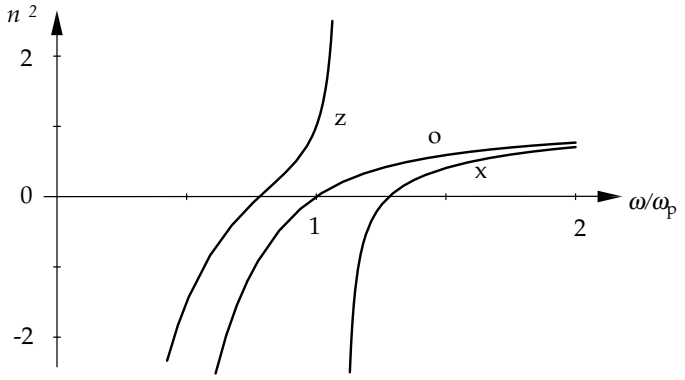


Figure 16.3: The squares of the refractive indices for magnetoionic modes are plotted for frequencies near the plasma frequency. The three modes are labeled as the o-mode, the x-mode and the z-mode.

The most favorable case for emission is at $s = 1$ in the x mode above the cutoff frequency at $\omega = \omega_x$, cf. Figure 16.3. One has $\omega_x \approx \omega_p + \frac{1}{2}\Omega_e$ for $\omega_p \gg \frac{1}{2}\Omega_e$ and $\omega_x \approx \Omega_e + \omega_p^2/\Omega_e$ for $\omega_p \ll \frac{1}{2}\Omega_e$. Hence fundamental emission

at $\omega \approx \Omega_e$ can be in the x mode only in a plasma with $\omega_p \ll \Omega_e$. Moreover, only emission that is above the line center at $\omega = \Omega_e$ by at least ω_p^2/Ω_e can be in the x mode.

This latter requirement restricts the emission to be at an angle θ significantly different from $\pi/2$. To see this, first recall from Figure 16.1 that $\theta - \pi/2$ must be small for the resonance ellipse to be approximated by a small circle rather than by a vertical line in the region of velocity space where the condition $\partial f/\partial p_\perp > 0$ is satisfied, and that it is only when the resonance ellipse closes within this region that effective growth can occur. The resonance condition (16.9) implies $\omega < \Omega_e$ for $\theta = \pi/2$, and it is only for $n_M \beta_\parallel \cos\theta > \frac{1}{2}\beta^2$ that resonance occurs at $\omega > \Omega_e$. The parameters (8.11) of the resonance ellipse imply that for small $|\cos\theta|$, the separation of the center of the resonance ellipse from the origin is proportional to $\cos\theta$, and the semimajor axis of the ellipse is determined by the difference $s\Omega_e - \omega$. Thus one infers that negative absorption occurs for emission at $|\theta - \pi/2|$ small but non-zero.

There is one further restriction on electron cyclotron maser emission that applies when the emission is attributed to a distribution of suprathermal electrons with a density much lower than that of thermal electrons. The resonance ellipse must avoid the region of velocity space occupied by the thermal electrons in order for gyromagnetic absorption by the thermal electrons swamping the maser emission. This point is also illustrated in Figures 16.1 and 16.2.

The evaluation of the absorption coefficient for electron cyclotron maser emission needs to be performed numerically in practice. By making appropriate simplifying assumptions one can estimate the maximum growth rate (the modulus of the maximum negative value of the absorption coefficient). For a loss-cone distribution of suprathermal electrons with mean speed $\langle\beta\rangle$ and loss-cone angle α_0, the maximum growth is estimated to be

$$|\gamma_x| \approx \frac{\pi\omega_p^2}{\Omega_e}\,\frac{n_1}{n_e}\,\frac{1}{\langle\beta\rangle^2 \sin\alpha_0}, \tag{16.20}$$

where n_1 is the number density that would be required to fill the loss cone.

Electron cyclotron maser emission (ECME) is the accepted emission mechanism for auroral kilometric radiation (AKR) from the Earth and analogous emissions from Jupiter and the other giant planets. AKR consists of bursts of cyclotron emission from regions above the auroral zones of the Earth. The electrons that generate AKR are accelerated downward by a potential drop of several kilovolts along the magnetic field lines. Below the acceleration region conservation of the adiabatic invariant $v^2 \sin^2\alpha/B$ implies that $\sin^2\alpha$ increases as B increases leading to the magnetic mirror effect. Electrons with sufficiently large initial pitch angle α reflect and electrons with smaller initial α precipitate into the atmosphere. One source of free energy that results from this is in the reflected electrons: there are no reflected electrons with small $\sin\alpha$ and hence the distribution is of the form illustrated in Figure 16.2. This is the favored source of free energy.

ECME is also the favored mechanism for one particular class of solar radio burst (but not for most solar radio bursts), called spike bursts, and for very bright radio emission from some flare stars. The brightest emission cannot be explained in terms of incoherent emission, such as gyrosynchrotron emission. The reason is that incoherent emission by particles with kinetic energy $\varepsilon - m_e c^2$ cannot exceed about $T_\mathrm{b} = (\varepsilon - m_e c^2)/\kappa_B$, because when absorption limits the emission to this value. Thus any emission observed to have $T_\mathrm{b} \gg 10^{10}\,\mathrm{K}$ must be due to some coherent mechanism. Brightness temperatures up to $10^{15}\,\mathrm{K}$ are observed from specific types of flare stars, and it is this emission that is thought to be due to ECME. Most of the radio emission from flare stars is due to gyrosynchrotron emission.

16.4 Gyrosynchrotron emission

Gyrosynchrotron emission is more difficult to treat analytically than either cyclotron emission or synchrotron emission. The emission is at intermediate harmonic numbers (s = 2–100). The available analytic approximations involve somewhat unfamiliar approximations to the Bessel functions. In practice interpolation formulas based on numerical calculations are often more useful that analytic approximations. Some relevant interpolation formulas are summarized here.

For a thermal distribution of electrons, with number density n_1 and temperature in the range $10^8 \lesssim T \lesssim 10^9\,\mathrm{K}$, relatively simple approximate formulas are available for $10 \lesssim s \lesssim 100$. In cgs units, the emissivity, η_ν, and the absorption coefficient, κ_ν, may be approximated

$$\frac{\kappa_\nu B}{n_1} \approx 50\, T^7 \sin^6\theta\, B^{10}\, \nu^{-10}, \qquad \frac{\eta_\nu B}{n_1} \approx 1.2 \times 10^{-24} T \left(\frac{\nu}{\nu_B}\right)^2 \frac{\kappa_\nu B}{n_1}. \quad (16.21)$$

The spectrum has a peak, which is more pronounced than for bremsstrahlung, cf. Figure 15.3, with the peak frequency at

$$\nu_\mathrm{peak} \approx \begin{cases} 1.4 \left(\dfrac{n_1 L}{B}\right)^{0.1} (\sin\theta)^{0.6}\, T^{0.7}\, B & \text{for } 10^8 < T < 10^9\,\mathrm{K}, \\ 475 \left(\dfrac{n_1 L}{B}\right)^{0.05} (\sin\theta)^{0.6}\, T^{0.5}\, B & \text{for } 10^7 < T < 10^8\,\mathrm{K}, \end{cases} \quad (16.22)$$

where L is the depth of the source along the line of sight. More detailed formulas are available for thermal distributions and for power law distributions (e.g., the review article, G.A. Dulk, "Radio emission from the Sun and stars", *Annual Reviews of Astronomy and Astrophysics*, **23**, 169–224, 1985).

The microwave emission from solar flares, and analogous emission from flare stars are the most notable applications for gyrosynchrotron emission. A radio spectrum are sometimes observed to have a peak at frequencies of a few

gigahertz, and then the parameter (16.22) can be used to place a significant constraint on the model of the source. Theory also implies that the circular polarization should change sense across the peak.

Lecture 17

Synchrotron radiation

Synchrotron emission, which is gyromagnetic emission from ultrarelativistic particles, is the dominant nonthermal emission mechanism in radioastronomy.

17.1 Emission by relativistic particles

Before discussing synchrotron emission in particular, it is appropriate to discuss emission by relativistic particles in general. Important features of the emission by relativistic particles may be determined by the special theory of relativity and are only weakly dependent on the specific emission mechanism involved. Emission by a particle which is highly relativistic in the laboratory frame, K, may be inferred from its emission pattern in its rest frame, K_0. Let the instantaneous velocity, $\mathbf{v}$, of the particle in K be along the z axis. The frame K_0, as viewed from K, is moving along the z axis at velocity $-\mathbf{v}$, as illustrated in Figure 17.1, and the frame K, as viewed from K_0, is moving along the z_0 axis at velocity $\mathbf{v}$.

In K_0 the particle is momentarily at rest, and its emission pattern is dipolar, with dipole axis along the direction of the instantaneous acceleration in K_0. For present purposes the only important point is that the emission pattern in K_0 is not highly anisotropic. Let ω_0, k_0, ψ_0 describe a plane wave in K_0, with ψ_0 the angle between $\mathbf{k}_0$ and the z_0 axis. Applying a Lorentz transformation, the wave parameters ω, k, ψ in K are related to those in K_0 by

$$\begin{aligned} \omega &= \gamma(\omega_0 + k_0 v\cos\psi_0), & \omega_0 &= \gamma(\omega - kv\cos\psi), \\ k\sin\psi &= k_0\sin\psi_0, & k_0\sin\psi_0 &= k\sin\psi, \\ k\cos\psi &= \gamma(k_0\cos\psi_0 + \omega_0 v/c^2), & k_0\cos\psi_0 &= \gamma(k\cos\psi - \omega v/c^2), \end{aligned} \quad (17.1)$$

with $\gamma = (1-\beta^2)^{-1/2}$, $\beta = v/c$. For emission in vacuo one has $k = \omega/c$,

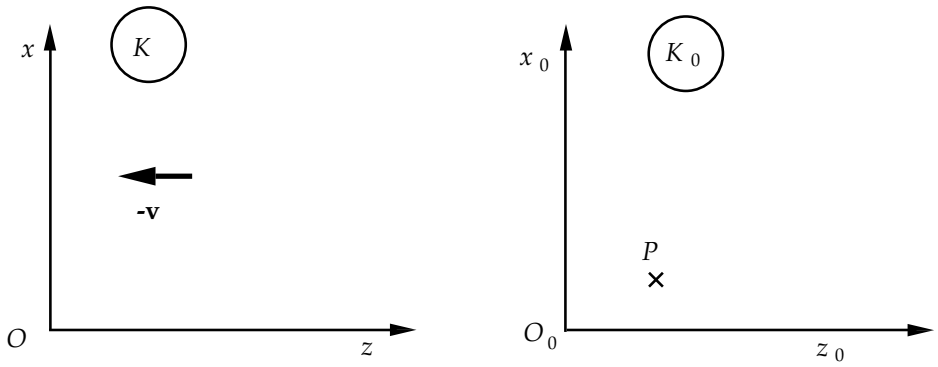


Figure 17.1: The laboratory frame K moves at $-\mathbf{v}$ relative to the rest frame K_0. A particle at the point P at rest in K_0 has velocity $\mathbf{v}$ in K

$k_0 = \omega_0/c$, and then (17.1) implies

$$\frac{\omega}{\omega_0} = \gamma(1 + \beta\cos\psi_0), \quad \cos\psi = \frac{\cos\psi_0 + \beta}{1 + \beta\cos\psi_0}, \quad \sin\psi = \frac{\sin\psi_0}{\gamma(1 + \beta\cos\psi_0)}. \tag{17.2a, b, c}$$

For a highly relativistic particle one assumes $\gamma \gg 1$ and expands in powers of γ^{-1}. Thus one refers to terms of order unity, terms of order γ^{-1} and so on. In particular one has

$$\beta \approx 1 - 1/2\gamma^2, \tag{17.3}$$

so that β differs from unity by a term of order γ^{-2}. Equation (17.2c) implies that, provided $1 + \beta\cos\psi_0$ and $\sin\psi_0$ are of order unity, $\sin\psi$ is of order $1/\gamma$. The proviso is satisfied except for ψ_0 of order γ^{-1}. This leads to the following property: *All angles of emission in K_0, with the exception of a range of order γ^{-1} about the direction $-\mathbf{v}$, transform into a forward cone with half-angle of order γ^{-1} about the direction $\mathbf{v}$ in K.*

Inspection of (17.2a) shows that the frequency ω in K is of order γ times the frequency ω_0 in K_0, except for the small range of ψ_0 of order γ^{-1} around $\cos\psi_0 = -1$. In semiclassical language, it follows that nearly all the photons emitted in K are confined to the forward cone and have an energy of order γ times the energy of the corresponding photon in K_0. Thus: *All but a fraction γ^{-1} of the power emitted in K is confined to a forward cone with half-angle of order γ^{-1} about the direction $\mathbf{v}$.* For gyromagnetic emission this implies that all but a fraction γ^{-1} of the power emitted by a particle with pitch angle α is confined to angles θ satisfying

$$\theta = \alpha + O(\gamma^{-1}). \tag{17.4}$$

That is, all but a fraction γ^{-1} of the power is emitted on the surface of a cone with half-angle $\theta = \alpha$ and thickness of order γ^{-1}.

17.2 Semiquantitative Treatment

The properties of synchrotron emission may be derived in a semiquantitative manner by combining some exact results for gyromagnetic emission and the foregoing arguments concerning emission by relativistic particles.

One exact result from the classical electromagnetic theory is for the power radiated by an accelerated charge in vacuo. For a particle spiraling in a magnetic field this gives

$$P = \frac{2e^2}{12\pi\varepsilon_0 c}\,\Omega_e^2\gamma^2\beta^2\sin^2\alpha, \tag{17.5}$$

with $\Omega_e = eB/m_e$. The emission is due to a resonant interaction, and so must occur at a sequence of harmonics, s, with emission at each harmonic satisfying the resonance condition (12.7), which in the notation used here becomes

$$\omega(1 - n\beta\cos\alpha\cos\theta) - s\Omega_e/\gamma = 0. \tag{17.6}$$

The properties of emission by relativistic particles discussed above imply that the emission occurs only on the cone defined by (17.4). Then, for emission in vacuo ($n = 1$), (17.6) implies

$$\omega = s\Omega_e/\gamma\sin^2\theta. \tag{17.7}$$

However, we need an independent argument to estimate the range of either ω or s over which synchrotron emission predominantly occurs.

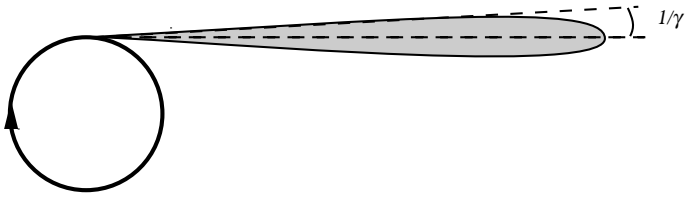


Figure 17.2: The emission pattern of a gyrating particle with $\gamma \gg 1$. The beam sweeps around in a circle as the particle gyrates.

The following argument enables one to estimate the typical frequency in synchrotron emission by considering the temporal distribution of the radiation received by an observer. First note that because of the highly anisotropic angular distribution of synchrotron emission, significant emission from a particle with pitch angle α is seen by an observer only in the narrow range of angles about $\theta = \alpha$. One pulse of radiation is received each revolution of the particle, with this pulse corresponding to the range of angles of order γ^{-1} during which the emission cone of the particle sweeps across the observer, as illustrated in

Figure 17.2. Consider the case $\alpha = \pi/2$, which corresponds to the particle moving in a circle. The particle performs one gyration in a time $2\pi\gamma/\Omega_e$. The time interval in which it is traveling toward the observer is a fraction of order γ^{-1} of this, that is, a time interval $\approx 2\pi/\Omega_e$. Consider the pulse received when a particle with speed βc moves directly toward the observer for a time Δt. Suppose that the particle is traveling in the x direction and that it starts radiating at $t = 0$ when it is at $x = 0$ and stops radiating at time $t = \Delta t$ when it is at $x = \beta c \Delta t$. (This starting and stopping of emission in the direction of the observer simulates the sweeping of the beam illustrated in Figure 17.2 across the direction to the observer.) The length of the wave train received by the observer is then $(1 - \beta)c\Delta t \approx c\Delta t/2\gamma^2$, and the duration of the pulse received per gyration by the observer is $\Delta t_{\rm rec} \approx \pi/\gamma^2\Omega_e$. A pulse of duration Δt contains Fourier components $\omega \lesssim 1/\Delta t$. Hence one expects synchrotron radiation to contain frequencies

$$\omega \lesssim \frac{\gamma^2 \Omega_e}{\pi}. \tag{17.8}$$

The estimate (17.8) applies only to emission by particles with pitch angle $\alpha = \pi/2$. However, it may be used to derive the corresponding result for arbitrary pitch angle by applying a Lorentz transformation. Consider a frame K in which a particle has pitch angle α and Lorentz factor γ, and another frame K' in which the particle has pitch angle $\alpha' = \pi/2$ and Lorentz factor γ'. The frame K' is moving relative to the frame K along the direction of the magnetic field with velocity equal to the parallel velocity of the particle, that is, at $\beta c \cos\alpha$. The Lorentz factor of the transformation between the two frames is therefore $(1 - \beta^2 \cos^2\alpha)^{-1/2} \approx 1/\sin\alpha$. Hence we have

$$\omega \approx \omega'/\sin\alpha, \quad \gamma \approx \gamma'/\sin\alpha. \tag{17.9}$$

Then (17.8), now written $\omega' \lesssim \gamma'^2 \Omega_e/\pi$, implies

$$\omega \lesssim \pi^{-1}\gamma^2 \Omega_e \sin\alpha. \tag{17.10}$$

It follows from (17.7) that the important harmonic numbers s involved in emission at $\theta \approx \alpha$ are

$$s \lesssim \pi^{-1}(\gamma \sin\theta)^3. \tag{17.11}$$

Thus, for $(\gamma \sin\theta)^3 \gg 1$, the emission is dominated by high harmonics and one is justified in treating s as a continuous variable.

17.3 The Emissivity for Synchrotron Emission

A detailed treatment of synchrotron emission may be developed as follows (cf. Melrose & McPhedran 1991, chapter 22 for more details). The initial objective is to evaluate the emissivity in synchrotron emission. The emissivity is the

power emitted per unit frequency and per unit solid angle, and it contains all relevant information on the distribution in frequency and angle.

In view of (17.4), the angular distribution is singular, and so may be described by a δ-function distribution, $\propto \delta(\theta - \alpha)$. Specifically, given the total power radiated per, the power radiated per unit solid angle is found by multiplying by $\delta(\cos\theta - \cos\alpha)/2\pi$.

The power radiated per unit frequency can be determined from the power radiated at the sth harmonic. The total power may be written as a sum over the power at each harmonic, and because very high harmonics are involved, s may be regarded as a continuous variable, and the sum over s may be rewritten as an integral over s. This integral can be rewritten as an integral over ω by appealing to (17.7), and the integrand of this integral is the power per unit frequency. Finally one needs to evaluate the power radiated at the sth harmonic and make the ultrarelativistic approximation to it. After a rather lengthy calculation, one can evaluate the power radiated at the sth harmonic exactly. It involves Bessel functions, and these are approximated in terms of Airy integrals. This calculation may be performed separately for two linear polarizations, denoted here by $\parallel$ and $\perp$, corresponding to the direction of the electric vector in the radiation relative to the projection of the magnetic field on the plane normal to the direction of propagation. (There is no circular polarization to lowest order in the expansion in $1/\gamma$.)

The resulting expression for the emissivity is

$$\eta^{\parallel,\perp}(\omega,\theta) = \frac{\sqrt{3}}{8\pi^2 c}\,\frac{q^2\Omega_e}{4\pi\varepsilon_0}\,\delta(\theta-\alpha)\,F^{\parallel,\perp}(\omega,\theta),$$

$$F^{\parallel,\perp}(\omega,\theta) = \frac{\omega}{\omega_{\rm c}}\int_{\omega/\omega_{\rm c}}^{\infty} dt\,K_{5/3}(t) \mp (\omega/\omega_{\rm c})K_{2/3}(\omega/\omega_{\rm c}),$$

$$\omega_{\rm c} = \tfrac{3}{2}\Omega_e\gamma^2\sin\theta. \tag{17.12}$$

The functions $K_\nu(x)$ are modified Bessel function, also called Macdonald functions, and they appear when making the Airy integral approximation.

The power emitted per unit frequency, found from (17.12) by integrating over solid angle and summing over the two states of polarization, is

$$P(\omega) = \frac{\sqrt{3}}{2\pi c}\,\frac{q^2\Omega_e\sin\theta}{4\pi\varepsilon_0}\,F(\omega/\omega_{\rm c}), \tag{17.13}$$

with

$$F(R) = \tfrac{1}{2}[F^{\parallel}(R) + F^{\perp}(R)] = R\int_R^{\infty} dt\,K_{5/3}(t). \tag{17.14}$$

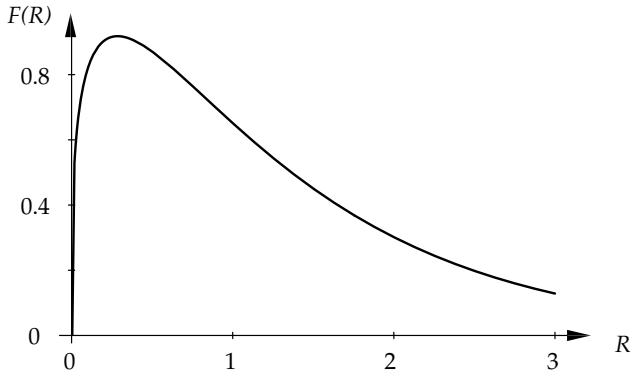


Figure 17.3: The function $F(R)$ defined by (17.19)

Expansions of the function $F(R)$ for small and large arguments are given by

$$F(R) = \begin{cases} \dfrac{4\pi}{\sqrt{3}\Gamma(1/3)} \left(\dfrac{R}{2}\right)^{1/3} \left[1 - \tfrac{1}{2}\Gamma(1/3)\left(\dfrac{R}{2}\right)^{2/3} + \cdots\right], \\ \left(\dfrac{\pi R}{2}\right)^{1/2} e^{-R} \left[1 + \dfrac{55}{72R} + \cdots\right], \end{cases} \tag{17.15}$$

for $R \ll 1$, $R \gg 1$, respectively. In between these limiting cases, there is a maximum at $F(0.29) = 0.92$. The function $F(R)$ is plotted in Figure 17.3. A simple analytic approximation to it is $F(R) \approx 1.8R^{0.3}e^{-R}$.

17.4 Emission by a Power-Law Distribution

Synchrotron emission is the emission mechanism for most radioastronomical sources, including supernova remnants, quasars and radio galaxies. In these sources the frequency spectrum is typically a power of the form $I(\omega) \propto \omega^{-\alpha}$, where α is the *spectral index*. This type of spectrum is interpreted in terms of emission from relativistic electrons with a power-law energy distribution.

The emission due to a distribution $f(p, \alpha) = f(p)\phi(\alpha)$ of electrons is found by integrating the emissivity (17.13) over the distribution of electrons; this gives the emissivity per unit volume, $J(\omega, \theta)$. For a power-law distribution in energy it is convenient to introduce the *energy spectrum* $N(\varepsilon)$, with $\varepsilon = pc$ for relativistic particles, by writing

$$4\pi \int_0^\infty dp\, p^2\, f(p) = \int_0^\infty d\varepsilon\, N(\varepsilon), \quad N(\varepsilon) = \frac{4\pi}{c^3}\, \varepsilon^2 f(\varepsilon/c). \tag{17.16}$$

The energy spectrum is assumed to be a power-law of the form

$$N(\varepsilon) = \begin{cases} K\varepsilon^{-a} & \text{for } \varepsilon_1 \le \varepsilon \le \varepsilon_2, \\ 0 & \text{otherwise.} \end{cases} \tag{17.17}$$

In integrating the emissivity over this energy spectrum, the limits ε_1, ε_2 can often be ignored, but at least one limit is essential in normalizing the number of particles, specifically, one must have $\varepsilon_1 > 0$ for $a > 1$.

The volume emissivity follows by integrating (17.12) over the distribution of particles

$$J^{\parallel,\perp}(\omega,\theta) = \frac{\sqrt{3}}{8\pi^2 c}\,\frac{q^2\Omega_e \sin\theta\phi(\theta)}{4\pi\varepsilon_0}\int_0^\infty d\varepsilon\, N(\varepsilon)\, F^{\parallel,\perp}\left(\omega/\omega_{\rm c}\right). \tag{17.18}$$

The integral over ε is performed by setting $\varepsilon_1 = 0$, $\varepsilon_2 = \infty$, noting that the energy dependence of $F(\omega/\omega_{\rm c})$, as given by (17.14) with (17.15), involves $\omega/\omega_{\rm c} \propto \varepsilon^{-2}$, and using the standard integral

$$\int_0^\infty dx\, x^\mu K_\nu(ax) = 2^{\mu-1}a^{-\mu-1}\,\Gamma\left(\frac{1+\mu+\nu}{2}\right)\Gamma\left(\frac{1+\mu-\nu}{2}\right), \tag{17.19}$$

with

$$\Gamma(1+x) = x\Gamma(x), \quad \Gamma(1-x)\Gamma(x) = \frac{\pi}{\sin(\pi x)}. \tag{17.20}$$

The result is

$$J^{\parallel,\perp}(\omega,\theta) = A(\theta)\, j^{\parallel,\perp}(a)\left(\frac{2\omega}{3\Omega_e \sin\theta}\right)^{-(a-1)/2}, \tag{17.21}$$

with

$$\begin{aligned} A(\theta) &= \frac{K(mc^2)^{-a+1}}{16\pi^2 c}\,\frac{\sqrt{3}q^2\Omega_e\sin\theta\phi(\theta)}{4\pi\varepsilon_0}, \\ j^{\parallel}(a) &= \frac{2/3}{a+1}2^{(a-3)/2}\,\Gamma\left(\frac{3a+7}{12}\right)\Gamma\left(\frac{3a-1}{12}\right), \\ j^{\perp}(a) &= \frac{a+5/3}{2/3}\,j^{\parallel}(a). \end{aligned} \tag{17.22}$$

It follows that the spectral index a of the particle energy spectrum, and the spectral index α of the synchrotron emission are related by

$$\alpha = \tfrac{1}{2}(a-1), \quad a = 2\alpha + 1. \tag{17.23}$$

The degree of linear polarization of the synchrotron emission is given by the difference between the power in the two polarizations divided by their sum.

Choosing to write the degree of polarization as positive if the $\parallel$ component dominates, (17.21) implies

$$r_{\rm l} = -\frac{a+1}{a+7/3} = -\frac{\alpha+1}{\alpha+5/3}. \tag{17.24}$$

The relatively high degree of polarization is a characteristic signature of synchrotron emission. For a value of $a = 3$, which is typical for many sources, (17.24) implies $r_{\rm l} = -3/4$, that is a degree of linear polarization of 75% in the direction orthogonal to the projection of the magnetic field $\mathbf{B}$ in the source on the plane of the sky. This prediction is based on the seemingly unrealistic assumption that the magnetic field in the source is uniform. In practice one expects the orientation of $\mathbf{B}$ to vary along the line of sight and across the source, and this would reduce the degree of polarization observed; the degree of polarization would be zero for a source in which the orientation of the magnetic field is random. Nevertheless many synchrotron sources show relatively high degrees of polarization, for example, $r_{\rm l} \gtrsim 30\%$ and a few have $r_{\rm l} \approx 70\%$.

17.5 Synchrotron Absorption

The absorption process corresponding to synchrotron emission is called synchrotron self-absorption, or more simply, *synchrotron absorption*. A general expression for the absorption coefficient is given by (12.17), which needs to be rewritten in the notation used here. This involves changing variables from p to ε, using (17.16) to introduce $N(\varepsilon)$, and rewriting the probability of emission in terms of the emissivity. In this way the absorption coefficient per unit length for the two states of linear polarization becomes

$$\mu^{\parallel,\perp}(\omega,\theta) = -\frac{(2\pi)^4 c}{\omega^2}\int_{-1}^{1} d\cos\alpha\,\phi(\alpha)\int_0^\infty d\varepsilon\,\varepsilon^2\,\eta^{\parallel,\perp}(\omega,\theta)\,\frac{d}{d\varepsilon}\left[\frac{N(\varepsilon)}{\varepsilon^2}\right]. \tag{17.25}$$

The absorption coefficient (17.25) cannot be negative. This may be shown by partially integrating in (17.25) and using (17.12) and the properties of the Macdonald functions to show $\partial[\varepsilon^2\,\eta^{\parallel,\perp}(\omega,\theta)]/\partial\varepsilon > 0$.

For the power-law distribution (17.17), with the limits ε_1, ε_2 ignored, (17.25) gives

$$\mu^{\parallel,\perp}(\omega,\theta) = \frac{(2\pi)^3 c}{\omega^2}\,\frac{K(mc^2)^{-a}}{16\pi^2 c^2}\,\frac{\sqrt{3}q^2\Omega_e\sin\theta\phi(\theta)}{4\pi\varepsilon_0} \times(a+2)j^{\parallel,\perp}(a+1)\left(\frac{2\omega}{3\Omega_e\sin\theta}\right)^{-a/2}, \tag{17.26}$$

with $j^{\parallel,\perp}(a)$ given by (17.22).

The transfer equation for synchrotron radiation including the effects of the polarization is written in the Mueller calculus, that is, in terms of the Stokes parameters I, Q, U, V. For simplicity the circularly polarized component is ignored (it is of order γ^{-1} compared with the linearly polarized components), and the natural modes of the medium are assumed to be circularly polarized. The transfer equation, including spontaneous emission, is then of the form

$$\frac{d}{ds}\begin{pmatrix} I \\ Q \\ U \end{pmatrix} = \begin{pmatrix} \alpha_I \\ \alpha_Q \\ 0 \end{pmatrix} + \begin{pmatrix} -\mu_I & -\mu_Q & 0 \\ -\mu_Q & -\mu_I & -\rho_V \\ 0 & \rho_V & -\mu_I \end{pmatrix} \begin{pmatrix} I \\ Q \\ U \end{pmatrix}, \tag{17.27}$$

where $\rho_V = \Delta k$ describes the effect of Faraday rotation, with Δk the difference in wave number between the two magnetoionic modes.

The emission coefficients are given by

$$\alpha_I = \frac{1}{c}\big[J^{\parallel}(\omega,\theta) + J^{\perp}(\omega,\theta)\big], \quad \alpha_Q = \frac{1}{c}\big[J^{\parallel}(\omega,\theta) - J^{\perp}(\omega,\theta)\big], \tag{17.28}$$

and the absorption coefficients by

$$\mu_I = \tfrac{1}{2}\big[\mu^{\parallel}(\omega,\theta) + \mu^{\perp}(\omega,\theta)\big], \quad \mu_Q = \tfrac{1}{2}\big[\mu^{\parallel}(\omega,\theta) - \mu^{\perp}(\omega,\theta)\big]. \tag{17.29}$$

In a slab model, one integrates (17.27) through a source of thickness L in general. Consider the simpler cases in which only one of the last two terms in (17.12) is retained. First, suppose that the absorption term is neglected. Then integration of (17.27) gives

$$I = \alpha_I L, \quad Q = \frac{\alpha_Q}{\rho_V}\sin(\rho_V L), \quad U = \frac{\alpha_Q}{\rho_V}\left[1 - \cos(\rho_V L)\right]. \tag{17.30}$$

This result describes the effect of spontaneous emission in a medium in which significant Faraday rotation occurs. It follows from (17.30) that in the limit of large Faraday rotation, that is, for $\rho_V L \gg 1$, Q and U do not exceed α_Q/ρ_V. This implies that the degree of polarization is of order $\alpha_Q/\alpha_I\rho_V L$, which is a factor $1/\rho_V L$ smaller than in the absence of Faraday rotation. This reduction in the degree of polarization due to Faraday rotation is called *Faraday depolarization*.

Next, suppose that Faraday rotation is neglected in (17.27) so that only emission and absorption are included. Integration of (17.27) then gives

$$I \pm Q = \frac{\alpha_I \pm \alpha_Q}{\mu_I \pm \mu_Q}\left(1 - e^{-(\mu_I \pm \mu_Q)L}\right), \quad U = 0. \tag{17.31}$$

The solution of (17.31) for I is plotted as a function of $\omega/\omega_{\rm c}$ for given L and given $N(\varepsilon)$ in Figure 17.4. The intensity has a maximum as a function of frequency. The *optically thin* region, where absorption is unimportant, corresponds to frequencies above the maximum, where one has $I \propto \omega^{-(a-1)/2}$, as in (17.21).

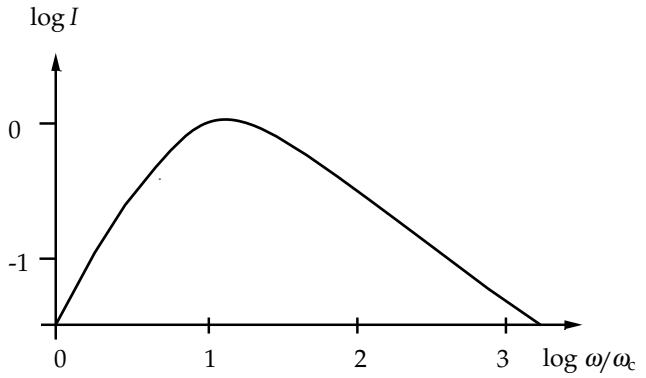


Figure 17.4: Spectrum for a slab model of a synchrotron source with $a = 3$. The asymptotic limits are $I \propto \omega^{5/2}$ and $I \propto \omega^{-(a-1)/2}$.

The *optically thick* region, where absorption is important, corresponds to frequencies below the maximum, where one has $I \propto \omega^{5/2}$. The spectrum for an optically thick source may be interpreted in terms of a thermal-like spectrum $I \propto \omega^2 \Theta$ with the temperature Θ replaced by the energy ε of the particle, and with this energy rewritten as $\varepsilon \propto \omega^{1/2}$ which follows from $\omega \approx \omega_c$, cf. (17.15) and Figure 17.3.

The degree of linear polarization $r_l = Q/I$ reduces to a simple form for $(\mu_I \pm \mu_Q)L \ll 1$ and for $(\mu_I \pm \mu_Q)L \gg 1$. The former limit corresponds to an optically thin source with degree of polarization given by (17.24). The latter case corresponds to an optically thick source with the degree of polarization then given by

$$r_l = \frac{Q}{I} = \frac{3}{6a+13}, \quad \text{for} \quad \mu_I L \gg 1. \tag{17.32}$$

The polarization implied by (17.32) for $a = 3$ is about 10%. Comparison of (17.24) and (17.32) shows that as a source becomes optically thick the plane of polarization flips through 90° and decreases in magnitude.

17.6 The Razin Effect

Synchrotron emission in a plasma with plasma frequency ω_p, is suppressed at $\omega \lesssim \gamma\omega_p$, which is the *Razin effect*. This effect causes a turnover in the spectrum that is somewhat similar to synchrotron absorption, cf. Figure 17.4. However, the causes of the two types of turnover are quite different.

A treatment of synchrotron emission including the Razin effect involves repeating the derivation given in section 15.3 retaining the refractive index $n \neq 1$. The refractive index appears in such a way that most of its effects correspond

to replacing β in vacuo by $n\beta$ in a plasma. This is the case in the resonance condition and in the arguments of the Bessel functions. Suppose that if one writes $n\beta = \beta_{\rm eff}$ and $\gamma_{\rm eff} = 1/(1-\beta_{\rm eff}^2)^{1/2}$, then for $n = (1-\omega_p^2/\omega^2)^{1/2}$, $\gamma \gg 1$ implies $\gamma_{\rm eff} \gg 1$ only for $\omega \gg \gamma\omega_p$. This suggests that emission at $\omega \lesssim \gamma\omega_p$ is more analogous to emission by a nonrelativistic particle than to emission by a relativistic particle. Thus one expects a relativistic particle in a plasma to emit much less effectively at $\omega \lesssim \gamma\omega_p$ than a corresponding particle in vacuo.

An alternative argument based on a Lorentz transformation between the frames K and K_0 introduced in section 1, cf. Figure 17.1, also suggests that emission in a plasma with refractive index $n = (1-\omega_p^2/\omega^2)^{1/2}$ must be suppressed relative to emission in vacuo for $\omega \lesssim \gamma\omega_p$. First, note that $\omega^2 - k^2c^2$ is an invariant under Lorentz transformations, and is equal to ω_p^2. In K_0 one has $\omega_0^2 - k_0^2c^2 = \omega_p^2$, so that the refractive index in K_0 is $n_0 = (1-\omega_p^2/\omega_0^2)^{1/2}$. It follows that emission at $\omega_0 < \omega_p$ in K_0 cannot occur. Consequently, the corresponding emission in K also cannot occur. The first of (17.2) with $n_0 = 0$ implies that the frequency $\omega_0 = \omega_p$ in K_0 corresponds to $\omega = \gamma\omega_p$ in K. This leads one to expect that emission at $\omega < \gamma\omega_p$ is not possible in K, and while this is not strictly correct, emission at $\omega < \gamma\omega_p$ is strongly suppressed.

A semiquantitative treatment treatment of the Razin effect involves introducing the effective Lorentz factor, as defined above, in (17.15) by replacing R by

$$R = 2s/3\gamma_{\rm eff}^3 \sin^3\theta = \omega/\omega_{\rm c}, \quad \gamma_{\rm eff} = \gamma(1+\gamma^2\omega_p^2/\omega^2)^{-1/2}, \tag{17.33}$$

where the latter formula results from expanding $\gamma_{\rm eff} = 1/(1-\beta_{\rm eff}^2)^{1/2}$ for $\gamma^2 \gg 1$, $\omega^2/\omega_p^2 \gg 1$. The presence of the plasma is unimportant for $\omega \gg \gamma\omega_p$ when one has $\gamma_{\rm eff} \approx \gamma$. An important point in the following argument is that the emissivity is peaked around $R \approx 0.3$, that is, at $\omega \approx 0.3\omega_{\rm c}$, and it decreases strongly ($\propto e^{-R}$) with $R \gg 1$. For $\omega \ll \gamma\omega_p$ inspection of (17.33) shows that $R \propto 1/\omega^3$ increases with decreasing frequency. Thus the effect of the plasma is to cause R to have a maximum value as a function of ω at $\omega \approx \gamma\omega_p$. Provided that this maximum occurs at $\omega \gg \omega_{\rm c}$ the plasma has little effect on the synchrotron emission. However, if the maximum occurs at $\omega \ll \omega_{\rm c}$ (with $\omega_{\rm c}$ determined by ignoring the plasma), then the emission is confined to the exponentially weak region. The frequency at which this suppression effect sets in is found by solving $\omega/\omega_{\rm c} = 1$ for the value of γ at which the turnover occurs, and substituting this into $\omega = \gamma\omega_p$. This gives the *Razin Tsytovich frequency*

$$\omega_{\rm RT} = 2\omega_p^2/3\Omega_e \sin\theta. \tag{17.34}$$

Below this turnover the spectrum rises must more steeply than for the self-absorbed case illustrated in Figure 17.4.

The Razin effect is of interest from a formal viewpoint. There is no known synchrotron source for which the effect is recognized as important.

Lecture 18

Plasma emission

Solar radio bursts are due to an emission mechanism that has properties that are quite different from the emission mechanisms that operate in other radio sources. The emission is at the fundamental or second harmonic of the plasma frequency, and it is polarized in the sense of the o-mode, and not the x-mode as for gyromagnetic emission. The brightness temperatures can be high ($T_\mathrm{b} > 10^{10}\,\mathrm{K}$) so that the emission process must be coherent. The relevant emission processes are referred to as plasma emission.

18.1 Qualitative discussion of plasma emission

Plasma emission is any indirect emission process in which (a) the exciting agency generates plasma turbulence which cannot escape directly from the plasma, and (b) this turbulence leads to escaping radiation through some secondary process. Plasma emission occurs in solar radio bursts, of which there are several types with the most familiar being type III bursts due to beams of fast electrons.

The first detailed theory for plasma emission from the solar corona was for type III bursts by Ginzburg and Zheleznyakov in 1958. At the time the theory was proposed it was accepted that type III bursts involves emission at the fundamental ($\omega = \omega_p$) and second harmonic ($\omega = 2\omega_p$), and that the emission is excited by a stream of electrons. A variant of their theory is outlined schematically in Figure 19.1. It consists of three stages: 1) generation of Langmuir turbulence through a streaming instability, 2) production of fundamental plasma emission by scattering of Langmuir waves into transverse waves by plasma particles, and 3) production of second harmonic emission through coalescence of two Langmuir waves to form a transverse wave. Since the theory was originally proposed, the details of each of these stages has been updated several times as ideas on the underlying plasma theory developed over the past three decades. There is now a vast literature on the theory of solar radio bursts, and there is

a wide variety of detailed ideas on the specific processes that are important.

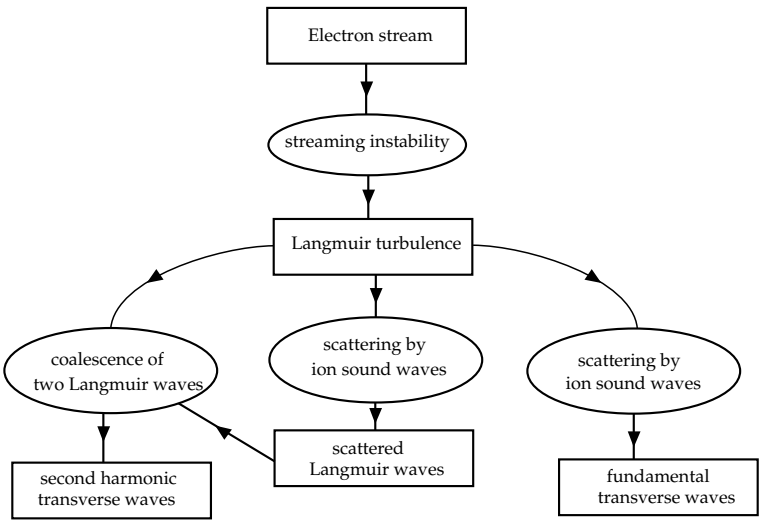


Figure 18.1: Flow diagram for a variant of the theory Ginzburg and Zheleznyakov for the generation of plasma emission. In other variants the processes indicated that involve ion sound waves are replaced by other nonlinear plasma processes.

In the variant of the theory of plasma emission illustrated schematically in Figure 19.1 the scattering is attributed to ion sound waves, also called ion acoustic waves. These scattering processes may be described using weak turbulence theory. Qualitatively, the relevant processes involve three-wave interactions in which two waves beat to generate a third wave. Let the three initial waves be described by frequencies ω_1, ω_2, ω_3 and wave vectors $\mathbf{k}_1$, $\mathbf{k}_2$, $\mathbf{k}_3$. Then the specific beat process $1 + 2 \rightarrow 3$, in which waves 1 and 2 beat to form wave 3, satisfies the Manley-Rowe conditions

$$\omega_1 + \omega_2 = \omega_3, \quad \mathbf{k}_1 + \mathbf{k}_2 = \mathbf{k}_3. \tag{19.1}$$

One process that can lead to fundamental plasma emission is the beat $L + S \rightarrow T$, where L refers to a Langmuir wave, S to an ion sound wave and T to a transverse wave. The process corresponding to scattering of Langmuir waves is the beat $L + S \rightarrow L'$, where L' refers to the scattered Langmuir wave. These are wave coalescence processes. The inverse of a coalescence is a decay process. The decay processes $L \rightarrow T + S$ and $L \rightarrow L' + S$ are qualitatively similar to the corresponding coalescence processes, and lead to a slight downshift, rather than a slight upshift, in frequency. Second harmonic emission results from the coalescence process $L + L' \rightarrow T$.

When the effect of a magnetic field is included there is a richer variety of possible coalesce and decay processes, including, for example, upper hybrid or z-mode waves in place of the L-waves, lower hybrid or electron acoustic waves in place of the S-waves, and o-mode or x-mode waves in place of the T-waves. An observation which suggests that the low-frequency waves may be lower hybrid waves concerns a specific class of type III bursts which exhibit fine structure, which are now referred to as stria bursts. The stria have a characteristic splitting that has a simple explanation in terms of a lower hybrid waves.

Each wave must satisfy the relevant dispersion relations. The dispersion relation for an arbitrary wave mode labeled M may be written in the form $\omega = \omega_M(\mathbf{k})$. For Langmuir waves, ion sound waves and transverse waves are, respectively,

$$\omega_L(\mathbf{k}) \approx \omega_p + 3k^2V_e^2/2\omega_p, \quad \omega_S(\mathbf{k}) \approx kv_s, \quad \omega_T(\mathbf{k}) = (\omega_p^2 + k^2c^2)^{1/2}, \quad (19.2)$$

where$V_e = (T_e/m_e)^{1/2}$ is the thermal speed of electrons and $v_s \approx V_e/43$ is the ion sound speed. In practice the dispersion relations can place severe restrictions on when specific three-wave interactions are allowed. How the ion sound waves are generated is not clear, but there is evidence that low-frequency turbulence (which may be ion sound waves or lower hybrid waves) on an appropriate form is normally present in the interplanetary medium.

There are other possible versions of the plasma emission processes. Of particular interest from a formal viewpoint is the possibility that strong-turbulence effects might play a role. This possibility is discussed separately below.

18.2 Weak turbulence theory

In weak turbulence theory an expansion of the basic equations is made in powers of the amplitude of the waves and the waves are treated in the random phase approximation. Consider waves with wavevector $\mathbf{k}$ in a wave mode M with dispersion relation $\omega = \omega_M(\mathbf{k})$. A distribution of such waves may be regarded as a collection of wave quanta with energy $\hbar\omega$ and momentum $\hbar\mathbf{k}$ and occupation number $N_M(\mathbf{k})$. The occupation number is a dimensionless quantity that corresponds to an energy $T_M(\mathbf{k})d^3\mathbf{k}/(2\pi)^3$ in the range $d^3\mathbf{k}/(2\pi)^3$, where

$$T_M(\mathbf{k}) = \hbar\omega_M(\mathbf{k})N_M(\mathbf{k}) \quad (19.3)$$

is the effective temperature of the waves.). In a time-independent medium the transfer of radiation has the simple form, in the absence of emission or absorption, $N_M(\mathbf{k})$ = constant or $T_M(\mathbf{k})$ = constant along a ray, corresponding to conservation of wave quanta. In principle, measurement of the brightness temperature of radiation at the Earth provides direct information on the effective temperature of the radiation as it leaves the source, provided that the radiation is not scattered or absorbed significantly along the ray path from the source

to the observer. Kinetic equations describe the sources and sinks for the wave quanta.

The kinetic equation for the wave in the mode M due to a three-wave process $L + S \leftrightarrow M$,with $M = T$ or L', is

$$\frac{dN_M(\mathbf{k})}{dt} = \int \frac{d^3\mathbf{k}'}{(2\pi)^3}\frac{d^3\mathbf{k}''}{(2\pi)^3}\, u^{\pm}_{MLS}(\mathbf{k},\mathbf{k}',\mathbf{k}'')\left\{N_L(\mathbf{k}')N_S(\mathbf{k}'') - N_M(\mathbf{k})\big[[N_S(\mathbf{k}'') \pm N_L(\mathbf{k}')\big]\right\}. \tag{19.4}$$

The actual form of the probabilities $u^{\pm}_{MLS}$ is not important in the following discussion: the two probabilities differ only in the sign in the beat conditions $\omega_L \pm \omega_S = \omega_M$, $\mathbf{k}_L \pm \mathbf{k}_S = \mathbf{k}_M$, which appear as the arguments of δ-functions in the probability.

A quantitative treatment of plasma emission based on the foregoing equations is cumbersome even when simplifying assumptions are made. However, there is one limiting case that may be treated simply and which may be understood in terms of simple physical arguments. Moreover, there is supporting evidence that this case may be relevant in practice. This is the limit in which the three wave interaction saturates. Imagine the processes $L \pm S \rightarrow M$ with $N_S \gg N_L$ building up the level of waves in the mode M. If initially one has $N_M \ll N_L$ then the factors inside the braces in (19.4) that involves the occupation numbers may be approximated by $N_L N_S$. As N_M increases this factor in (19.4) decreases. The three-wave interaction is said to saturate when this factor reaches zero and the interaction ceases to caused N_M to increase. Saturation occurs for

$$N_M = N_L N_S/(N_S \pm N_L) \approx N_L. \tag{19.5}$$

In particular, for fundamental emission due to scattering by ion sound waves with $N_S \gg N_L$, the process saturates at $N_T = N_L$, and as the frequencies of the fundamental transverse and Langmuir waves are nearly equal, this saturation level corresponds to $T_T = T_L$. Saturation of the two stages in second harmonic emission leads to its saturation at the same level, at least to within a factor of two. Thus despite the highly nonthermal nature of these processes, this saturation level is what one might expect on a thermodynamic-type argument. Before discussing saturation of the plasma emission processes further, it is appropriate to consider the generation of the Langmuir turbulence and the saturation of the beam instability.

18.3 Growth of the Langmuir waves

The growth of Langmuir waves is attributed to an instability driven by a beam of electrons. In principle the instability may be either kinetic or reactive. For type III electron streams the kinetic version appears to be appropriate, but this is not necessarily the case for electron beams from the Earth's bow shock.

The kinetic instability is amenable to a relatively simple treatment in a one-dimensional model, where 'one-dimensional' refers to a three-dimensional theory in which the only Langmuir waves considered have **k** along the direction of the beam and the particle distribution function is integrated over the two momentum component orthogonal to the beam direction.

The beam instability saturates when the distribution function becomes sufficiently distorted in the sense that the growth rate is reduced substantially from its initial value. Unlike the saturation of a three-wave process, as discussed above, there is no well-defined saturation level. Semi-quantitatively, saturation occurs when the energy density in the Langmuir waves becomes a significant fraction of the initial energy density in the electron beam. A rough estimate of the effective temperature of the Langmuir waves at saturation may be made as follows. For simplicity assume that the saturation level of the Langmuir waves is approximately constant, $W(v_\phi) = W_0$ say, for waves in a range $v_b - \Delta v < v_\phi < v_b$ and zero otherwise, where v_b is the beam speed. If the energy density of these waves is a fraction ζ of the initial energy density in the beam, then we have

$$W_0 = \zeta n_1 m v_b^2 / 2\Delta v. \tag{19.11}$$

Further assume that the Langmuir waves are confined to a range $\Delta\Omega$ of solid angle. The range of solid angles $\Delta\Omega$ can be estimated only in terms of a three-dimensional theory for the instability. One simple estimate is to write $\Delta\Omega = \pi(\Delta\theta)^2$,with $\Delta\theta = \Delta v/v_b$. Then the effective temperature T_L and W_0 are related by, for $\Delta v \ll v_b$,

$$T_L = \frac{(2\pi)^3 v_b^4}{\omega_p^3 \Delta\Omega} W_0 = \frac{\zeta (2\pi)^2 n_1 m v_b^8}{\omega_p^3 (\Delta v)^3}. \tag{19.12}$$

18.4 Saturation model for plasma emission

There are two saturation models for aspects of plasma emission that fit with observational data, despite the fact that the saturation models seem unrealistically simple. One such model is the one-dimensional quasilinear model for the evolution of the electron beam.The measured distribution functions for electron beams in the solar corona are consistent with the predictions of one-dimensional quasilinear theory in which the Langmuir waves generated by the faster electrons are subsequently reabsorbed by the slower electrons. Several assumptions in this model are in conflict with the observations, notably that the Langmuir waves are generated uniformly over the cross-section of the beam, whereas they are observed to be very inhomogeneously distributed in clumps. It can be shown that quasilinear evolution due to interaction with a random distribution of clumps of Langmuir turbulence has the same form as for a uniform distribution of Langmuir turbulence, so that the assumption of a uniform distribution of Langmuir waves is unnecessarily restrictive. Nevertheless the success of such

a simple model is surprising in view of the many and various processes thought to affect the evolution of the Langmuir turbulence.

Another saturation model is for the radio emission. As explained above, if the weak turbulence processes saturate due to processes involving ion sound (or other appropriate low frequency) turbulence, then all secondary wave distributions tend to saturate at the same effective temperature T_L, at least to within a factor of two. The estimate (19.12) is a rough one, and should be used only for an order of magnitude estimate of the saturation level of the Langmuir waves. However, the important qualitative point is that if the plasma emission processes saturate, then the observed brightness temperature should be equal to T_L. Thus in a saturation model, an estimate such as (19.12) implies a prediction on the functional relation between the properties of the beam of electrons and the brightness of the observed plasma emission. Such predictions can be tested given appropriate statistical data on the electron beams and their radio emission.

It may be concluded that saturation models for plasma emission, despite the simple and almost naïve assumptions made, provide a reasonable basis for a semiquantitative treatment of plasma emission.

18.5 Strong turbulence and plasma emission

For present purposes a strong turbulence effects is defined as any processes that cannot be treated using weak turbulence theory. Strong turbulence concepts that have been applied to plasma emission include parametric instabilities, modulational instabilities and Langmuir collapse. Specific strong turbulence effects are relevant only when the energy density of the Langmuir turbulence exceeds an appropriate threshold value.

Qualitatively, in the case of a three-wave interaction (a parametric instability) the important difference between weak and strong turbulence effects is as follows. In weak turbulence theory a three-wave processes causes the turbulence to evolve only when two excited wave distributions beat together to generate the third wave distribution. In strong turbulence theory one excited wave distribution acts as a pump that generates the other two wave distributions. A modulational instability is similar to a parametric instability in that only one wave distribution (the pump) is required but, from a kinematic viewpoint, the evolution is regarded as a four-wave interaction rather than as a three-wave interaction. One such instability (the oscillating two stream instability) was proposed and explored in connection with the evolution of the Langmuir turbulence in type III streams.

One observation that supports the suggestion that such effects play a significant role in plasma emission is that low-frequency turbulence, with wavevectors in the range expected in a parametric instability, is sometimes closely correlated with the Langmuir turbulence. Such turbulence has been seen both in type III

events and in association with plasma emission from planetary bow shocks.

Langmuir collapse is related to self-focusing of laser light in nonlinear optics: a uniform distribution of turbulence breaks up into localized regions in which the energy density in the turbulence increases to very high values on a very short timescale. Radiation during Langmuir collapse is usually treated as a weak turbulence process that has no significant effect on the collapse. The role that such strong turbulence effects may play remains unclear because of the paucity of definitive observational tests for specific theories.

Lecture 19

Pulsar electrodynamics

Pulsars are rotating, magnetized neutron stars which emit radio waves in beams that are detected as pulses as the beam sweeps across the Earth. Despite considerable progress in understanding pulsars, some aspects the electrodynamics of the neutron star magnetosphere and the details of the emission mechanism of the radio waves are inadequately understood.

19.1 Goldreich–Julian density

Consider a magnetized neutron star, of radius R_*, rotating with angular velocity $\mathbf{\Omega}$. Inside the star, which is assumed to be a good conductor, there is a *corotation electric field*, $\mathbf{E} = -(\mathbf{\Omega} \times \mathbf{x}) \times \mathbf{B}$, which is such as to cause an electric drift, $\mathbf{E} \times \mathbf{B}/B^2$, that is equal to the rotation velocity $\mathbf{\Omega} \times \mathbf{x}$.

As a first attempt to understand the magnetosphere of a pulsar, consider let us assume that the star is surrounded by a vacuum and explore the consequences of this assumption. Then there is a discontinuity in $\mathbf{E}$ at the surface of the star, implying a surface charge on the star. Outside the star there is no corotation field in vacuo, but due to the charge on the surface of the star, there is an electric field. This electric field has a component, $E_\parallel$, along the magnetic field lines. Assuming a magnetic dipole at the center of the star, the surface charge has a quadrupolar distribution and the component of the exterior electric field parallel to the magnetic field is of magnitude

$$E_\parallel = R_* \Omega B_* (R_*/r)^4 \cos^3\theta, \tag{19.1}$$

where B_* is the polar magnetic field. Taking the values $\Omega = 10\,\mathrm{s}^{-1}$, $B_* = 10^8\,\mathrm{T}$ and $R_* = 10\,\mathrm{km} = 10^4\,\mathrm{m}$, (19.1) implies $E_\parallel \approx 10^{13}\,\mathrm{V\,m}^{-1}$. The scale height of the pulsar atmosphere is a few millimeters. The electric field (19.1) across a scale height implies a large potential difference In fact, this potential difference is large enough to rip particles off the surface and accelerate them to relativistic

energies. This accelerate should continue until there is enough charge around the pulsar to shield out $E_\parallel$. Hence, one expects the star to have a magnetosphere populated with particles. This plasma tends to shield out $E_\parallel$ and to set up the corotation electric field in the magnetosphere. Thus it is inconsistent to assume a vacuum model because the vacuum field itself implies that vacuum conditions cannot be maintained.

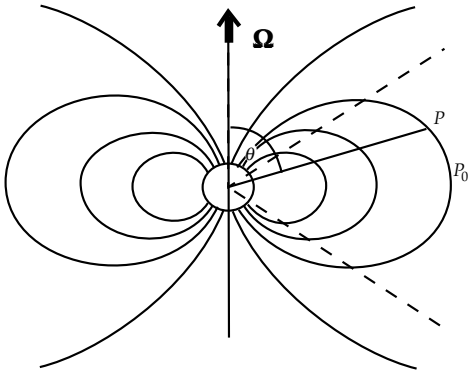


Figure 19.1: On a dipolar magnetic field line, the point P satisfies $r = r_0 \sin^2\theta$, where the point P_0 is in the equatorial plane at $r = r_0$. For an aligned rotator, the locus of points corresponding to $\mathbf{\Omega}\cdot\mathbf{B} = 0$ that separates regions of opposite sign of the charge density correspond to $|\cos\theta| = 1/3$, indicated by the dashed lines.

The simplest alternative assumption is that there are sufficient charged particles to set up the corotation electric field. Let us assume this to be the case and explore the consequences of this assumption. Assuming complete screening of $E_\parallel$ requires that the corotation field be set up everywhere in the magnetosphere. This requires a charge density whose value is implied by the divergence of this field:

$$\rho = \varepsilon_0 \nabla\cdot\mathbf{E} = \varepsilon_0\left(-2\mathbf{\Omega}\cdot\mathbf{B} + \mathbf{\Omega}\times\mathbf{x}\cdot\nabla\times\mathbf{B}\right). \tag{19.2}$$

The final term in (19.2) is negligible close to the star, and is ignored here. The number density determined by setting $\rho = en_e$ in (19.2) is called the *Goldreich–Julian* density:

$$n_{\rm GJ} = -\frac{2\varepsilon_0\mathbf{\Omega}\cdot\mathbf{B}}{e}. \tag{19.3}$$

Consider the case of a dipolar magnetic field, with rotation axis coaligned with the magnetic axis. Then (19.3) implies that $n_{\rm GJ}$ has the same sign above the northern and southern polar caps, and has the opposite sign above the equator. The sign of $n_{\rm GJ}$ reverses where $\mathbf{B}$ is perpendicular to the rotation axis, as illustrated in Figure 19.1.

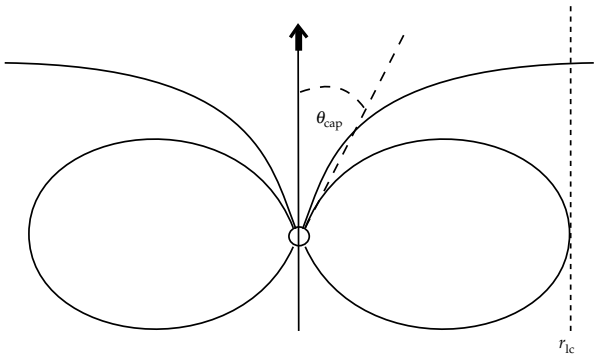


Figure 19.2: The polar cap is defined by the set of field lines that extend to beyond the light cylinder.

The region of corotation cannot extend beyond a surface, called the *light cylinder*, at which the corotation speed would equal the speed of light, as illustrated in Figure 19.2. In a model in which the rotation and magnetic axes are aligned, this corresponds to a radius (in the equatorial plane)

$$r_{\rm lc} = c/\Omega. \tag{19.4}$$

The *polar cap* may be defined by the locus on the surface of the star of the footpoints of the last closed field lines. All field lines from within the polar cap extend beyond the light cylinder and these cannot be corotating because that would violate the special theory of relativity beyond the light cylinder.

In spherical polar coordinates, let the polar cap define a polar cap angle, $\theta_{\rm cap}$. Then for small $\theta_{\rm cap}$, the radius of the polar cap is $R^* \sin\theta_{\rm cap}$. The value of $\theta_{\rm cap}$ may be estimated using the equation (3.22) for a dipolar field line and the fact that a field line that is at at radius R^* angle $\theta_{\rm cap}$ is at radius $r_{\rm lc}$ at $\theta = \pi/2$. This gives

$$R_* = r_{\rm lc} \sin^2\theta_{\rm cap}, \qquad \theta_{\rm cap} = \arcsin(R_*/r_{\rm lc})^{1/2} \approx (\Omega_* R_*/c)^{1/2}. \tag{19.5}$$

For a slowly rotating pulsar, $r_{\rm lc}$ is of order 10^8 m, which is about $10^4\, R_*$, implying $\theta_{\rm cap} \approx 10^{-2} \approx 0.5^\circ$. For the Crab pulsar, rotating with a period of 33 ms implying $\Omega = 2 \times 10^2\,{\rm s}^{-1}$, the light cylinder is of order a hundred stellar radii, and the polar cap angle is of order 5°. For millisecond pulsars, the light cylinder can be as small as about ten stellar radii, and the polar cap angle of order 15°.

Particles on field lines in the polar cap are accelerated outward by the parallel electric field, and apparently must escape from the pulsar. Particles of the same sign escape from both polar caps. Particles of the opposite sign are trapped in corotating regions nearer the equatorial plane, and so do not escape.

The outflow of charge leads to various inconsistencies in the corotation model unless there is an independent source of charge somewhere above the stellar surface. One inconsistency occurs along the field lines for which the sign of $n_{\rm GJ}$ reverses, cf. Figure 19.1. Along an open field line the reversal in the sign of $n_{\rm GJ}$ cannot be occur without there being an additional source of charge above the stellar surface. Another inconsistency arises when the flow along the open field lines becomes relativistic, so that the current density is $J = n_{\rm GJ}ec$. Conservation of charge then requires $J \propto B$ which expresses the requirement that the number of charges crossing unit area at $v \sim c$ be constant as they follow along field lines whose number density per unit area proportional to B. This requires $n \propto B$. However, corotation requires $n = n_{\rm GJ}$, and (19.3) implies $n_{\rm GJ} \propto B\cos\theta$. These two variations are not compatible. This implies that the corotation field cannot be maintained unless there is some source of charge above the surface.

In existing pulsar models these difficulties are overcome by assuming that pairs are created in *gaps* in the magnetosphere. In a gap it is required that $E_\parallel$ be large enough to cause one component of the pair to be accelerated toward the star while the other escapes. This provides a net source of charge in the magnetosphere.

Nevertheless problems remain with this model, and it is not clear how these are to be resolved. The remaining problems include the following: (1) The escape of charges from both polar caps seems to imply a net current and hence a net loss of charge. This is unacceptable because the star would charge up rapidly, and set up a Coulomb field until this is strong enough to prevent further loss of charge. There must be a return current, but how this is set up remains a problem with most models. (2) One expects there to be a difference between neutron stars with different signs of $\mathbf{\Omega}\cdot\mathbf{m}$, where $\mathbf{m}$ is the magnetic moment. Using more colorful language, Ruderman has been remarked that pulsars should come in two sexes, one corresponding to electrons being ripped off the polar cap, and the other corresponding to ions being ripped of the polar cap. However, there is no evidence for two such distinct classes of pulsars. (3) There is a problem in accounting simultaneously for the energy loss and angular momentum loss. As explained further below, together these require that the slowing down torque be applied near the light cylinder, presumably associated with the acceleration of a pulsar wind, but the detailed nature of this torque is unclear. (In principle some information about the torque can be inferred from the 'braking index', which is an observable parameter, but little progress has been made in doing so.)

19.2 Vacuum Gaps

Although there is no completely consistent model for a pulsar magnetosphere, it is possible to identify several ingredients that are likely to be incorporated in an acceptable model. One of these is the development of vacuum gaps which

discharge through the production of electron-positron pairs.

The first detailed model that incorporates a vacuum gap was proposed by Ruderman and Sutherland in 1975. Their model starts with an assumption that is now considered unacceptable: that the binding energy of ions on the surface of a neutron star is so strong that they cannot be ripped off by the available potential. Let us accept this assumption for the present. An implication is that, provided the sign of $\mathbf{\Omega} \cdot \mathbf{B}$ is negative so that (19.2) requires a positive charge density over the polar cap, the strong binding of the ions implies that it is not possible to provide the required charge from the stellar surface. It follows that there is a large electric field above the polar cap, which is essentially the vacuum field.

In this vacuum gap, a stray gamma ray can decay into an electron-positron pair. The positron is accelerated upward and the electron downward in this case. These primary particles can themselves emit gamma rays through a variety of processes, and these gamma rays may also decay into pairs. As a result of this secondary production of pairs, the region above the polar cap becomes sufficiently well populated by the pair plasma to screen the vacuum electric field. The discharge leading to this screening is referred to as *sparking*. The pair plasma then flows out along the field lines in the polar cap. This pair plasma contains both a primary positron beam, with a typical particle energy $\varepsilon \approx e\Delta\Phi$, where $\Delta\Phi$ is potential drop across the polar cap region, plus lower energy, secondary pairs. This model is illustrated schematically in Figure 19.3.

In this model the existence of a vacuum gap effectively implies that the field lines inside the polar cap are not rotating at all, while those outside the polar cap are rotating somewhat slower than the star. Although this is unrealistic in detail, it is thought to provide a reasonable estimate of the expected potential drop that becomes available. The potential for a rotating magnetic field is $\Phi = \mathbf{\Omega} \times \mathbf{x} \cdot \mathbf{A}$, and for a dipolar magnetic field one has $\mathbf{A} = (\mu_0/4\pi)\mathbf{m} \times \mathbf{x}/r^3$, with $r = |\mathbf{x}|$, and where $\mathbf{m}$ is the magnetic dipole moment. The difference in this potential between the pole ($\theta = 0$) and the edge of the polar cap ($\theta = \theta_{\rm cap}$) is

$$\Delta\Phi = \frac{\mu_0}{4\pi} \frac{\Omega_* |\mathbf{m}| \sin^2 \theta_{\rm cap}}{R_*} = \frac{\Omega_*^2 R_*^3 B_*}{2c}, \tag{19.6}$$

where $B_* = (\mu_0/4\pi)|\mathbf{m}|R_*^3$ is the magnetic field at the (magnetic) equator of the star. With $\Omega_* = 10\,\mathrm{s}^{-1}$, $R_* = 2 \times 10^4\,\mathrm{m}$ and $B_* = 10^8\,\mathrm{T}$ (19.6) gives $\Delta\Phi \approx 10^{14}\,\mathrm{V}$. The potential is larger for young pulsars, notably the Crab and Vela pulsars, which are more rapidly rotating.

As already remarked, one of the essential assumptions in the Ruderman and Sutherland model is now thought to be invalid. There are several alternative models based on the opposite assumption, that is, that charges can flow freely away from the surface of the star. The models differ in detail, but all involve gaps with large $E_\parallel$ in which pair creation occurs. In a model proposed by Arons and coworkers, so-called 'slot gaps' occur along field lines where $\cos\theta$ decreases.

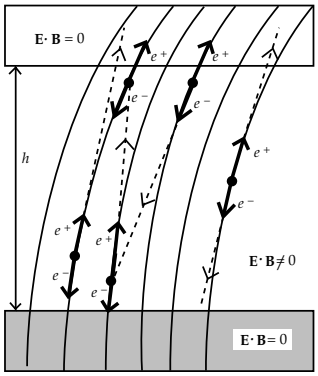


Figure 19.3: The vacuum gap of height h is illustrated. The parallel electric field in the region between the stellar surface (shaded region) and the overlying magnetosphere accelerates electrons (e^-) downward and positrons (e^+) upward. Curvature photons, indicated by dashed lines, generate secondary pairs.

This occurs because current conservation implies $n \propto B$ and this is inconsistent with $n = n_{\rm GJ}$, with $n_{\rm GJ} \propto B\cos\theta$. Along such field lines, in the absence of another source of free charges, $E_\parallel$ must build up toward the vacuum value. The $E_\parallel$ that develops in these gaps is large, but not as large as in the Ruderman and Sutherland model.

The foregoing models are examples of 'polar gap' models, in the sense that the regions with large $E_\parallel$ occur above the surface of the star in the polar cap regions. There are also 'outer gap' models, where the gap occurs near points where the field lines are orthogonal to the rotation axis, where $n_{\rm GJ}$ vanishes according to (19.3). Charged particles from the surface of the star cannot supply a charge density that smoothly changes sign along such field lines, and hence the corotation electric field cannot be set up along them. It follows that vacuum conditions, with $E_\parallel \neq 0$, must develop. Such outer gaps can be a source of pairs well away (in both height and angle) from the polar caps.

19.3 The Pair Plasma

Pair creation is an essential ingredient in models for pulsars, and also in models for AGN (lecture 21). A number of different processes can lead to pair formation. These include (i) spontaneous breakdown of the vacuum in the presence of a sufficiently strong electric field ($E \sim E_{\rm cr} = m_e^2c^3/e\hbar$), which condition is never satisfied in pulsar magnetospheres, (ii) photon-photon interactions, $\gamma + \gamma \rightarrow e^- + e^+$, (iii) decay of a photon in the field of a nucleus (Ze), $\gamma + Ze \rightarrow e^- + e^+ + Ze$, (iv) decay of a photon in a superstrong magnetic field, $\gamma + B \rightarrow e^- +$

$e^+ + B$, plus other more complicated processes. In a pulsar magnetosphere the most important pair-creation process is the decay of a photon in a superstrong magnetic field. (In pulsars in which ions flow freely from the surface, process (iii) can also be important: when the ions become highly relativistic the threshold condition, that the energy of the photon in the center of momentum frame exceed $2m_ec^2$, can be satisfied for thermal photons from the stellar surface.)

The process $\gamma + B \to e^- + e^+ + B$ is possible only for a gamma ray with a high enough energy, and then is significant only if the magnetic field is sufficiently strong. Obviously, one photon can decay into a pair only if the energy of the photon, $\varepsilon_{\rm ph} = \hbar\omega$, exceeds the rest energy, $2m_ec^2$, of the pair. In fact, only the perpendicular (to $\mathbf{B}$) energy of the photon is relevant, and the actual threshold condition is $\varepsilon_{\rm ph} \sin\psi \geq 2m_ec^2 \approx 1\,\mathrm{MeV}$, where ψ is the angle between $\mathbf{k}$ and $\mathbf{B}$. The decay mechanism $\gamma + B \to e^- + e^+ + B$ is effective only in a *superstrong* magnetic field, that is, one which is a significant fraction of the so called *critical magnetic field*

$$B_c = m_e^2c^2/e\hbar = 4.4 \times 10^9\,\mathrm{T}. \tag{19.7}$$

The field B_c corresponds to the cyclotron energy, $\hbar\Omega_e = \hbar eB/m_e$, being equal to the rest energy, m_ec^2. There are several processes that are strictly forbidden in the absence of a superstrong magnetic field and that are allowed in its presence. These include the decay of a gamma ray into a pair and the splitting of one photon into two photons.

The probability per unit time of decay of a gamma ray with frequency $\omega \gg m_ec^2/\hbar$ into a pair due to propagation at an angle ψ to the magnetic field lines is

$$\alpha_\perp \approx \frac{\alpha}{\omega}\,\frac{m_e^2c^4}{\hbar^2}\,\frac{B\sin\psi}{B_c}\,T(\chi), \qquad T(\chi) \approx 0.46\,\exp\left(\frac{-4}{3\chi}\right),$$
$$\chi = \frac{\hbar\omega}{2m_ec^2}\,\frac{B\sin\psi}{B_c}, \tag{19.8}$$

with $\alpha = e^2/4\pi\varepsilon_0\hbar c \approx 1/137$ the fine structure constant. The result (19.8) applies for $\chi \ll 1$. For relevant parameters this decay occurs quickly except for ψ very close to zero.

A further important feature of the pair plasma, at least in the inner part of a pulsar magnetosphere, is that it is one dimensional, that is, from a classical viewpoint, the electrons and positrons have no spiraling motion about the field lines. To see why this is so, first note that the motion of an electron perpendicular to the magnetic field is circular motion classically, and circular motion is simple harmonic motion. Treating an electron quantum mechanically, it has energy eigenvalues (ignoring its spin)

$$\varepsilon = \varepsilon_n(p_\parallel) = (m^2c^4 + p_\parallel^2c^2 + 2n|q|B\hbar c^2)^{1/2}, \tag{19.9}$$

where $p_\parallel$ is a continuous quantum number and $n = 0, 1, 2, \ldots$ is the simple harmonic quantum number corresponding to circular motion. When a particle

emits a cyclotron photon it jumps from state n to state $n-1$. The rate per unit time that such transitions occur for $p_\parallel = 0$ is

$$R_{n\to n-1} = \frac{4}{3}\alpha \frac{m_e c^2}{\hbar}\left(\frac{B}{B_c}\right)^2. \tag{19.10}$$

where r_0 is the classical radius of the electron, and α is the fine structure constant. For $p_\parallel \neq 0$ the decay rate can be obtained by a Lorentz transformation, and it is given by $R_{n\to n-1}/\gamma$. Electrons in a high energy state jump down to lower states, predominantly in a stepwise process in which n decreases by unity in each transition. In the absence of other effects, the electron eventually reaches its ground state. The lifetime for the decay to the ground state is of order the inverse of the transition rate for the slowest transition, which is the last transition from the first excited state to the ground state. This gives a lifetime of order $3/\gamma B^2$ s, where B is in tesla, and where the factor $1/\gamma$ is included because (19.10) applies in the rest frame of the center of gyration. For the magnetic field near a pulsar one has $B \approx 0.1 B_c = 4 \times 10^8$ T, and the lifetime for the decay is then extremely short $\approx 2 \times 10^{-17}/\gamma$ s. One expects all electrons and positrons to be in their ground state, corresponding to one-dimensional motion.

Curvature emission occurs as follows. The primary particles experience an acceleration of magnitude $c^2/R_{\rm c}$, where $R_{\rm c}$ is the radius of curvature, toward the center of curvature. As a result of this acceleration the particles emit *curvature* radiation. This is somewhat analogous to synchrotron emission in that both may be regarded as emission by a relativistic particle moving in an arc of a circle. The characteristic frequency of emission is $(c/R_{\rm c})\gamma^3$. ($R_{\rm c}$ is replaced by the gyroradius, $\gamma c/\Omega_e$, in the synchrotron case.) Applied to a pulsar, one has $R_{\rm c} \approx 3R_*$ near the surface of the star, and hence $(c/R_{\rm c})\gamma^3 \approx 10^4 \gamma^3\,{\rm s}^{-1}$. Expressing the frequency ω as a photon energy $\varepsilon_{\rm ph} = \hbar\omega$ one has $\varepsilon_{\rm ph} \approx 10^{-11}\gamma^3$ eV.

The highly relativistic primary particles emit photons along dipolar magnetic field lines which are curved, so that initially the photons have $\sin\psi \lesssim 1/\gamma$. Even for a gamma ray with $\varepsilon_{\rm ph} \gg 2m_e c^2$ cannot decay into a pair until $\sin\psi$ has increased sufficiently for the condition $\varepsilon_{\rm ph}\sin\psi > 2m_e c^2$ to be satisfied. A systematic increase occurs because the photon travels in (approximately) a straight line and so deviates increasingly away from the curved field line. In this way, one initial primary particle, with $\gamma \approx 10^6$ emitting photons with $\varepsilon_{\rm ph} \gg 2m_e c^2$ produces a large number of secondary pairs. The predicted properties of this secondary pair plasma are somewhat model dependent. However, a multiplicity (number of secondaries per primary) $\sim 10^4$ and a mean Lorentz factor $\gamma \sim 10^2$ are typical estimates.

It is accepted that the radio emission is generated by the relativistic pairs, although the detailed mechanism is not understood. It follows that neutron stars that do not generate pairs cannot be radio pulsars. As the rotation slows down, $E_\parallel$, as given by (19.1), and the available potential drop, $\Delta\Phi$, decrease, and eventually become inadequate to sustain pair production. Thus a pulsar

dies as it ages and slows down. The so-called *death line* depends on the model, with different functional dependences in different models, due primarily to the different values of $\Delta\Phi$. For example, in the Ruderman and Sutherland model, one has $P_{\rm max} \propto B^{8/13}$, in the model of Arons one has $P_{\rm max} \propto B^{8/17}$, and in the model of Beskin, Gurevich and Istomin one has $P_{\rm max} \propto B^{8/15}$. By way of illustration, Ruderman and Sutherland (1975) derived the maximum period by requiring that the optical depth for absorption of a gamma ray to produce a pair is greater than unity. this leads to the condition

$$\frac{\hbar\omega}{2m_ec^2}\,\frac{B\sin\theta}{B_c} \gtrsim \frac{1}{15} \tag{19.11}$$

where $\omega \sim (c/R_c)\gamma^3$, with $\gamma \sim e\Delta\Phi/m_ec^2$, is the typical frequency of curvature photons emitted by the primary particles.

19.4 The Loss of Energy and Angular Momentum

It is now widely accepted that pulsars lose energy through a relativistic wind. There are strong arguments for this. However, before discussing the wind, consider the case of a magnetized neutron star rotating in a vacuum.

In a vacuum a rotating magnet losses energy and angular momentum through the emission of electromagnetic radiation. The radiation is magnetic dipole radiation, and the power radiated is given by a standard formula of electromagnetic theory:

$$\mathcal{P} = \frac{\mu_0|\ddot{\mathbf{m}}|^2}{6\pi c^3}, \tag{19.12}$$

with $\ddot{\mathbf{m}} = \mathbf{\Omega} \times (\mathbf{\Omega} \times \mathbf{m})$.

A simple model that reproduces (19.12) approximately is as follows. Inside the light cylinder one is in the inductive zone and the magnetic field may be approximated by a dipolar magnetic field, which falls off as $1/r^3$, so that the field at the light cylinder, $B_{\rm lc}$, is given in terms of the magnetic field at the equator of the star, $B_{\rm eq} = (\mu_0/4\pi)|\mathbf{m}|/R_*^3$, where $\mathbf{m}$ is the magnetic moment, by $B_{\rm lc} = B_{\rm eq}(R_*/r_{\rm rl})^3 = (\mu_0/4\pi)|\mathbf{m}|/r_{\rm rl}^3$. Beyond the light cylinder, one is in the radiative zone, corresponding to an outgoing spherical wave. This requires that the field, $B_{\rm wave}$, falls off as $1/r$, so that the power radiated, given by the outward directed Poynting vector integrated over a sphere of radius r be independent of r. This corresponds to $B_{\rm wave}E_{\rm wave}/\mu_0 = B_{\rm wave}^2 c/\mu_0$ times $4\pi r^2$ being independent of r. Evaluating the power radiated at the light cylinder gives

$$\mathcal{P} = \frac{B_{\rm lc}^2}{\mu_0}\, c4\pi r_{\rm rl}^2 = \frac{\mu_0|\mathbf{m}|^2\Omega^4}{4\pi c^3}, \tag{19.13}$$

which reproduces (19.12) apart from a factor of order unity.

The energy loss may also be equated to the rate of decrease of the rotational energy, $\frac{1}{2}I\Omega^2$, where $I \sim M_* R_*^2$ is the moment of inertia of the star:

$$\mathcal{P} = -I\Omega\dot{\Omega} = \frac{I(2\pi)^2\dot{P}}{P^3}, \tag{19.14}$$

where $P = 2\pi/\Omega$ is the period of the pulsar. Equating (19.13) and (19.14) allows one to estimate the surface magnetic field on the neutron star from the two observable quantities P and $\dot{P}$. The magnetic field at the poles is estimated using the formula

$$B_{\rm pole} = 3.2 \times 10^{15}\,(P\dot{P})^{1/2}\,{\rm T}. \tag{19.15}$$

The slowing down also involves a loss of angular momentum. The angular momentum, $I\Omega$, of the star changes at the rate $-I\dot{\Omega}$. Comparing this with (19.14) shows that the ratio of the rates of energy and angular momentum loss is Ω. Consider the emission of one photon with energy $\hbar\Omega$, where the fact that a rotating magnetic dipole radiates at $\omega = \Omega$ is taken into account. If this is radiated in a direction perpendicular to the rotation axis at a 'lever-arm' distance r from the star then it carries off angular momentum $\hbar\Omega r/c$. It follows that the effective lever-arm distance is the light cylinder distance, $R_{\rm lc} = c/\Omega$. Although the vacuum model is not correct, this argument on the lever-arm distance remains valid: the torque that provides both the energy loss and the angular momentum loss must be applied close the light cylinder.

Another observable quantity is the second derivative of the rotation frequency, $\ddot{\Omega}$, which is used to defined the *braking index*, n, by writing

$$\dot{\Omega} = -K\Omega^n, \tag{19.16}$$

and noting that one has

$$n = \frac{\ddot{\Omega}\Omega}{\dot{\Omega}^2}. \tag{19.17}$$

Comparison of (19.12) and (19.14) shows that one has $n = 3$ for magnetic dipole radiation. Observed values are typically smaller than $n = 3$ for single, isolated pulsars. (A very recent estimate for the Vela pulsar suggests $n \approx 1.4$.)

One can integrate (19.16) to find Ω as a function of t and the initial value, Ω_0, at $t = 0$. For $\Omega \ll \Omega_0$, this gives the age of the pulsar as

$$t_{\rm age} = -\frac{\Omega}{(n-1)\dot{\Omega}} = \frac{P}{(n-1)\dot{P}}. \tag{19.18}$$

It is conventional to define the age using (19.18) with $n = 3$, that is, $t_{\rm age} = P/2\dot{P}$.

The vacuum model cannot be correct. In particular it predicts emission of monochromatic electromagnetic radiation at $\omega = \Omega$, which is well below the plasma frequency in the magnetosphere and in the surrounding ISM, and

hence which cannot propagate. Observations preclude high-frequency radiation carrying off the energy and angular angular momentum, even for the five (radio) pulsars from which high-frequency emission is detected. The angular momentum carried off by the observed radiation is far too small to account for the spin down. Thus the emission of electromagnetic radiation is excluded as a slowing down mechanism. The only remaining possibilities are a kinetic energy flux or a flux of Alfvén waves. Both have been suggested in the literature, but no mechanism has been proposed for generating the required flux of Alfvén waves. By exclusion one is left with a flux of kinetic energy, that is, a wind. However, there is no adequate theory that allows one to interpret these observed values of the braking index.

There is indirect evidence for a wind from the structure of the Crab Nebula. At about one tenth the radius of the nebula there are wisps which are now interpreted in terms of acceleration of synchrotron-emitting particles at a termination shock of the wind. There is also evidence for winds from other pulsars, through the bow shocks that they generate in the ISM.

Although it is widely believed that pulsars do lose their angular momentum through a wind consisting of relativistic electrons and positrons, there is no detailed theory as to how the angular momentum is transferred to this wind. The resolution of this may be connected with another problem mentioned above, namely that of current closure. The argument given above implies that the torque required to accelerate the wind, and to transfer angular momentum to it, must be applied, at least in some average sense, at about the light cylinder. The only relevant torque is a $\mathbf{J} \times \mathbf{B}$ force, which requires a cross-field current flow. Thus, it appears that the currents flowing from the polar caps must close by flowing across field lines near the light cylinder, presumably connecting with a return current path to the surface of the neutron star near the edge of the polar cap region.

Although some considerable progress has been made in understanding the electrodynamics of pulsar magnetosphere, a view held widely (but not universally) by the experts is that there are serious weaknesses in our present understanding.

Additional Reading

Beskin, V.S., Gurevich, A.V., and Istomin, Ya.N., *Physics of the Pulsar Magnetosphere*, Cambridge University Press (1993)

Manchester, R.N., and Taylor, J.H., *Pulsars*, W.H. Freeman, New York, (1977)

Michel, F.C., *Theory of neutron star magnetospheres*, University of Chicago Press (1991)

Lecture 20

Pulsar radiation

The details of the pulsar radio emission mechanism are poorly understood. It must be a coherent mechanism, in the sense that a maser or laser emits coherently. There are several other known classes of coherently emitting radio-astronomical sources. These are OH and other molecular line masers, plasma emission from the solar corona, and electron cyclotron maser emission (ECME) from the planetary magnetospheres and from flare stars. Pulsar radio emission does not appear to be closely related to any of these, although perhaps it may be loosely related to plasma emission.

20.1 Pulsar radio emission

Several general observational criteria (brightness temperature, beaming and linear polarization, microstructures) are incorporated into nearly all suggested pulsar emission mechanisms.

The brightness temperature of pulsar radio emission is very much higher than for any other radio-astronomical sources. Its maximum value, estimated to be in the range $T_\mathrm{b} \sim 10^{25}$–$10^{30}\,\mathrm{K}$, is at least ten orders of magnitude greater than the next brightest sources, which are flare stars. The very high T_b cannot be explained in terms of thermal emission, which cannot have a T_b higher than the temperature of the emitting particles, and it cannot be incoherent emission, which cannot have $k_B T_\mathrm{b}$ higher than the typical energy of the emitting particles. Hence, very bright emission ($T_\mathrm{b} \gg 10^{10}\,\mathrm{K}$) requires a coherent emission mechanism.

In all pulsar models, the pulsed nature of the emission is interpreted in terms of beamed radiation from the pulsar sweeping across the line of sight of the observer as the neutron star rotates. In polar cap models, the beaming is explained in terms of the relativistic beaming effect for relativistic particles propagating along the magnetic field lines, which implies that the radiation is

confined to a cone with half-angle $\sim 1/\gamma$ about the direction of the magnetic field. This beaming is a feature of emission by relativistic particles and applies to all emission mechanisms that involve relativistic particles in one-dimensional motion along $\mathbf{B}$.

The most obvious features of the polarization of pulsars are the high degree of linear polarization, and the systematic sweep of the plane of polarization through a pulse. In young pulsars the polarization is typically 100% linear. Older (slower) pulsars tend to have a more complicated polarization structure, sometimes with a significant circular component. The linear polarization is well explained in terms of the 'single-vector model' in which the plane of polarization is determined by the magnetic field in the source. The projection of the magnetic field on the sky rotates as the beam sweeps across the observer.

There is indirect evidence for a frequency to radius mapping, specifically, evidence that higher frequency emission comes from lower heights. Such a mapping favors emission from the polar cap regions well inside the light cylinder. In particular, the separation between multiple peaks in the average pulse profile is observed to decrease with increasing frequency. This is interpreted in terms of emission at higher frequency coming from nearer to the star, where the cone defined by the open field lines is narrower.

There are two subclasses of pulsars that potentially provide complementary information on pulsar emission mechanisms. These are radio pulsars that also radiate at high-frequency and millisecond pulsars.

Most pulsars have no detectable high-frequency emission, that is, X-ray and gamma-ray emission. However, strong high-frequency radiation is observed from five radio pulsars: PSR 0531+21, PSR 0833-45, PSR 1509-58, PSR 1706-44 and PSR 0540-69 in the Large Magellanic Cloud. The main part of the radiation of all these pulsars falls in the X-ray and γ-ray ranges. The γ-ray pulsar Geminga is probably also a radio pulsar, but at a level too low to detect. The high frequency emission is not discussed in detail here, but it provides an important constraint on pulsar models. This is because some (although not all) pulsars must be capable of emitting a substantial fraction of their power at high frequencies.

Ordinary pulsars, with periods of a few tenth to a few seconds, are assumed to be formed in supernova explosions. The radio emission from millisecond pulsars, which have periods from just over a millisecond to several tens of milliseconds, is remarkably similar to that from most fast, young pulsars. Millisecond pulsars are thought to start life as ordinary pulsars, evolve until they slow down and die, then to reappear as pulsars after having been spun up due to mass transfer from a companion star. From the pulse profile alone one could not distinguish between a millisecond pulsar and an ordinary pulsar. This suggests that the emission mechanism is essentially the same for the two classes of pulsars. Whatever this mechanism is, it must be capable of accounting for similar emission from slow pulsars and from millisecond pulsars. The magnetic field, which is typically $B \sim 10^8\,\mathrm{T}$ in ordinary pulsars, is $B \sim 10^4\,\mathrm{T}$ in millisecond pulsars.

20.2 Coherent emission mechanisms

The bright radio luminosity of pulsars implies that the emission mechanism must be coherent. Three different types of coherent emission are possible in principle:

1. An antenna mechanism involves emission by bunches of particles with negligible velocity dispersion. The coherence is attributed to the particles radiating in phase with each other. Then N particles radiate a power N^2 times the power radiated by an individual particle alone.

2. A reactive instability involves an intrinsically growing, phase-coherent wave whose growth rate exceeds the intrinsic bandwidth of the growing waves. The intrinsic bandwidth is due to the velocity dispersion of the particles, which is required to be very narrow so that all the particles remain in phase with the growing wave.

3. A maser mechanism involves negative absorption. Maser (or laser) emission is familiar in bound-state systems (atoms or molecules), in which the emission between two levels produces a narrow line. Negative absorption between two levels results from the higher-energy level being overpopulated relative to the lower-energy level, called an inverted population. In a plasma, the particle distribution needs to have some feature that corresponds in a meaningful sense to an inverted energy population. In the case of a one-dimensional distribution of relativistic particles, this condition is $\partial f/\partial\gamma > 0$ over some range of γ. Maser emission applies in the random phase approximation, so that (unlike the other two types of coherent emission) the phase of the growing waves is irrelevant.

In all cases the back reaction to the coherent emission tends to reduce the feature in the distribution of particles that is causing the coherent emission. The radiation reaction to an antenna mechanism tends to disperse the bunch, the back reaction to a reactive instability tends to increase the velocity dispersion, and the back reaction to a maser mechanism is the so-called quasilinear relaxation that tends to reduce the inverted energy population. The time scale on which the back reaction occurs is determined roughly by the time required for the energy density in the radiation to become comparable to the energy density in the radiating particles. This is fastest for an antenna mechanism and slowest for a maser mechanism.

There is a strong argument for considering only maser mechanisms in astrophysical applications. In order for coherent emission to occur some process is required to set up the feature in the distribution function that causes the coherent emission. The time scale on which this feature must be set up needs to be at least as short as the time scale on which the back reaction would destroy this feature. For a maser, the process that sets up the effective inverted population

is referred to as a pump. It is convenient to use the word 'pump' to describe the process that sets up the conditions for the other types of coherent emission, that is, for the creation of a bunch with negligible velocity dispersion required to allow an antenna mechanism to operate, or for the creation of an appropriate distribution with a very narrow velocity dispersion required to allow a reactive instability to develop. The requirements on the pump are most demanding for an antenna mechanism (extreme localization in coordinate space and velocity space) and are very demanding for a reactive instability (extreme localization in velocity space). The requirements on the pump are least demanding for a maser mechanism, and hence the simplest explanation of any coherent emission is in terms of a maser mechanism. In fact, there is no well-established case of a coherent emission in astrophysics that is not due to a maser mechanism.

The two forms of coherent emission that are well established are plasma emission and ECME. As a modified form of plasma emission may well be relevant to pulsar radio emission, the theory for plasma emission is outlined briefly below. In ECME, an anisotropic distribution of electrons with an excess of perpendicular momentum over parallel momentum cause radiation at close to the electron cyclotron frequency to grow. For this radiation to escape one requires a plasma with $\Omega_e > \omega_p$.

20.3 Specific emission mechanisms

In the literature you will find that following mechanisms have been proposed for pulsar radio emission.

Emission by bunches: An early theory for the radio emission is coherent curvature emission by bunches. The basic idea is the familiar one that N particles in a volume less than a cubic wavelength radiate like a macrocharge $Q = Ne$, and because the power radiated is proportional to Q^2, the power is N^2 times the power from an individual particle.

There are seemingly insurmountable difficulties with this theory. For example, when one takes into account the highly anisotropic nature of curvature emission by relativistic particles, the bunch really needs to be a pancake with its normal within an angle $1/\gamma$ of the direction of the magnetic field. An obvious difficulty is to identify a mechanism that allows such an exotic bunch to form; none of the suggested mechanisms works. Moreover, even if such a bunch did form, the bending of the field lines would cause its normal to deviate to more than $1/\gamma$ away from the direction of the magnetic field in a very short time. For these and other reasons, coherent curvature emission by bunches is unacceptable.

Relativistic plasma emission: Perhaps the most widely favored mechanism may be regarded as a relativistic version of plasma emission. Plasma emission may be defined as a multi-stage process in which an instability first transfers energy from an exciting agency (typically a beam of particles) into waves that

cannot propagate out of the plasma. In subsequent stages the energy in these waves is converted into escaping radiation (in another wave mode) through nonlinear or other processes in the plasma.

Maser curvature emission: In the simplest case, the absorption coefficient corresponding to curvature emission (like that corresponding to synchrotron emission) cannot be negative, so that maser emission cannot occur. However, when other effects are included, maser curvature emission can occur due to at least two effects: the curvature drift across field lines and a twist of the field lines, corresponding to curved field lines that are not confined to a plane.

Cyclotron instability: An instability that involves electrons (or positrons) having a cyclotron transition through the anomalous cyclotron resonance $s = -1$ in (12.7). This leads to emission near the cyclotron frequency. The waves that grow in this instability have refractive index > 1 and so cannot escape directly. To produce escaping radiation these waves need to be converted into the high-frequency modes through a plasma-emission process. For this process to produce radiation in the radio range, the instability must occur at large distances from the star, where B is sufficiently small.

Free electron maser emission: A form of linear acceleration emission in which the relativistic particles are accelerated by an oscillating electric field can lead to maser emission. The characteristic frequency of the emission is $\omega \sim \omega_0 \gamma^2$, where ω_0 is the greater of the typical frequency of the oscillating electric field, or the typical wavenumber times c. This mechanism can operate in two regimes. One corresponds to a form of relativistic plasma emission, in which the energy in the emitted waves comes primarily from the energy in the oscillating electric field. The other corresponds to a form of free electron maser emission, in which the oscillating field acts as a wiggler field.

20.4 Relativistic plasma emission

Relativistic plasma emission may be interpreted as an exotic form of an emission process that is known to operate in the solar corona.

Solar radio bursts of types I, II and III are due to plasma emission. The first detailed theory for plasma emission was proposed for type III bursts by Ginzburg and Zheleznyakov in 1958. At the time it was accepted that type III bursts involve emission at the fundamental ($\omega = \omega_p$) and second harmonic ($\omega = 2\omega_p$) of the plasma frequency, and that the emission is excited by a stream of electrons. A variant of their theory is outlined schematically in Figure 20.1. It consists of three stages: 1) generation of Langmuir turbulence through a streaming instability, 2) production of fundamental plasma emission by scattering of Langmuir waves into transverse waves by plasma particles, and 3) production of second harmonic emission through coalescence of two Langmuir waves to form a transverse wave. Since the theory was originally proposed, the details of each of these stages has been updated several times as ideas on the underlying plasma theory

have developed.

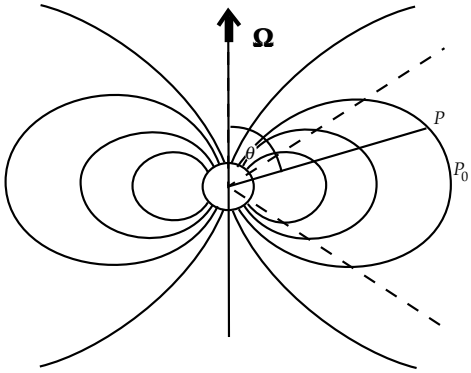


Figure 20.1: Flow diagram for a variant of the theory Ginzburg and Zheleznyakov for the generation of plasma emission. In other variants the processes indicated that involve ion sound waves are replaced by other nonlinear plasma processes.

In the variant of the theory of plasma emission illustrated schematically in Figure 20.1 the scattering is attributed to ion sound waves, also called ion acoustic waves. The relevant processes involve three-wave interactions in which two waves beat to generate a third wave. Let the three initial waves be described by frequencies ω_1, ω_2, ω_3 and wave vectors $\mathbf{k}_1$, $\mathbf{k}_2$, $\mathbf{k}_3$. Then the specific beat process $1 + 2 \rightarrow 3$, in which waves 1 and 2 beat to form wave 3, satisfies the three-wave beat conditions

$$\omega_1 + \omega_2 = \omega_3, \quad \mathbf{k}_1 + \mathbf{k}_2 = \mathbf{k}_3. \tag{20.1}$$

One process that can lead to fundamental plasma emission is the beat $L + S \rightarrow T$, where L refers to a Langmuir wave, S to an ion sound wave and T to a transverse wave. The process corresponding to scattering of Langmuir waves is the beat $L + S \rightarrow L'$, where L' refers to the scattered Langmuir wave. These are wave coalescence processes. The inverse of a coalescence is a decay process. The decay processes $L \rightarrow T + S$ and $L \rightarrow L' + S$ are qualitatively similar to the corresponding coalescence processes, and lead to a slight downshift, rather than a slight upshift, in frequency. Second harmonic emission results from the coalescence process $L + L' \rightarrow T$.

Each wave must satisfy the relevant dispersion relations. The dispersion relation for an arbitrary wave mode labeled M may be written in the form $\omega = \omega_M(\mathbf{k})$. For Langmuir waves, ion sound waves and transverse waves these are, respectively,

$$\omega_L(\mathbf{k}) \approx \omega_p + 3k^2V_e^2/2\omega_p, \quad \omega_S(\mathbf{k}) \approx kv_s, \quad \omega_T(\mathbf{k}) = (\omega_p^2 + k^2c^2)^{1/2}, \tag{20.2}$$

where $V_e = (k_B T_e/m_e)^{1/2}$ is the thermal speed of the electrons and $v_s \approx V_e/43$ is the ion sound speed. In practice the dispersion relations can place severe restrictions on when specific three-wave interactions are allowed. How the ion sound waves are generated is not clear, but there is evidence that low-frequency turbulence (which may be ion sound waves or lower hybrid waves) or an appropriate form is normally present in the interplanetary medium.

Relativistic plasma emission from pulsars involves essentially the same ideas. An instability is assumed to generate waves in a wave mode that cannot escape, and nonlinear processes convert these waves into waves in a mode that can escape. However, all the details (properties of the wave modes, the instability and the nonlinear processes) for pulsars are different from plasma emission in the solar corona.

The streaming instability, in the case of a one-dimensional ultrarelativistic distribution, requires a distribution with $\partial f/\partial\gamma > 0$ in the inertial frame in which the bulk motion of the plasma is zero. This could be due to the high energy beam of positrons moving through the pair plasma, to a relative motion of the electrons and positrons in the pair plasma, or to other less obvious types of relative motion. However, the growth rates for most of these instabilities are too small. It is estimated that the pair plasma leaves the magnetosphere before it has given up significant energy. A larger growth rate can result, for example, if the generation of the pair plasma fluctuates in time, producing a sequence of beams, with the faster particles in a following beam overtaking the slower particles in a preceding beam.

A pump is needed to maintain the condition $\partial f/\partial\gamma > 0$ for wave growth. In a pair cascade, one expects there to be few electrons or positrons at low energies, and for their number to rise to a peak at some characteristic energy. One then has $\partial f/\partial\gamma > 0$ below this peak and $\partial f/\partial\gamma < 0$ above it. The back reaction to the streaming instability reduces the gradient $\partial f/\partial\gamma > 0$. In the absence of a pump to reestablish the peak, the instability would be confined to the near vicinity of the region where the pairs are produced. Such a model is very restrictive, both on location of the emission region and on the energy available for the maser emission, which is given by the energy in the pairs below the peak multiplied by a presumably small efficiency factor for the relativistic plasma emission. A model in which the production of pairs fluctuates or is appropriately spatially structured can provide a pump that continues to operate away from the region where the pairs are produced.

20.5 Maser curvature emission

Curvature emission is like synchrotron radiation in that, in the simplest case, the absorption coefficient cannot be negative. It can be negative when the curvature drift is included.

The curvature drift speed is

$$v_{\rm cd} = \frac{v^2\gamma}{\Omega_e R_{\rm c}}, \tag{20.3}$$

where v is the parallel velocity of the particle and $R_{\rm c}$ is the radius of curvature of the field line. The direction of the curvature drift (opposite for opposite signs of the charge) is such that the Lorentz force $q\mathbf{v}_{\rm cd} \times \mathbf{B}$ provides the centripetal acceleration, toward the center of curvature of the field line, needed to make the particle follow the curved path. In the following discussion it is more convenient to introduce a drift angle,

$$\theta_{\rm cd} = v_{\rm cd}/v, \tag{20.4}$$

which is the angle between the momentum of the particle and $\mathbf{B}$. The drift angle is proportional to the particle energy, and when the particle energy changes as a result of emission, the drift angle also changes. This dependence is the essential ingredient in curvature-drift induced maser emission in that it allows negative absorption to occur.

The total emissivity (summed over the two polarizations) in the presence of curvature drift is

$$\eta(\omega,\theta,\gamma) = \frac{q^2}{4\pi\varepsilon_0}\,\frac{\omega^2 R_{\rm c}}{6\pi^3 c^2}\left[(\theta-\theta_{\rm cd})^2\left(\xi^{-1}{\rm K}_{1/3}(y)\right)^2 + \left(\xi^{-2}{\rm K}_{2/3}(y)\right)^2\right]. \tag{20.5}$$

with $\xi = [2(1-N) + N(\gamma^{-2} + (\theta-\theta_{\rm cd})^2)]^{-1/2}$, $y = \omega/3n^{1/2}\omega_R\xi^3$, where N is the refractive index, and with $\omega_R = v/R_{\rm c}$. The absorption coefficient per unit time for a one-dimensional distribution function is

$$\Gamma(\omega,\theta) = -\frac{(2\pi c)^3 n_0}{2\omega^2 mc^2}\int d\gamma\,\frac{df(\gamma)}{d\gamma}\,\eta(\omega,\theta,\gamma), \tag{20.6}$$

where n_0 is the particle number density, and where $f(\gamma)$ is normalized to unity. From (20.6) the conditions for negative absorption are

$$\frac{df(\gamma)}{d\gamma} > 0, \qquad \frac{d\eta}{d\gamma} < 0. \tag{20.7}$$

The first of (20.7) follows from (20.6) and the fact that $\eta(\omega,\theta,\gamma)$ is non-negative, and the second follows by partially integrating (20.6) and the fact that $f(\gamma)$ is non-negative. The original proof that maser action is not possible is based on the inequality $d\bar{\eta}/d\gamma > 0$, where $\bar{\eta}$ is the emissivity for $\theta_{\rm cd} = 0$. For $\theta_{\rm cd} \neq 0$ one has

$$\frac{d\eta}{d\gamma} = \frac{\partial\eta}{\partial\gamma} - \frac{\theta_{\rm cd}}{\gamma}\frac{\partial\eta}{\partial\theta}. \tag{20.8}$$

The term involving the θ-derivative allows negative absorption. A graphical description of this is provided in Figure 20.2.

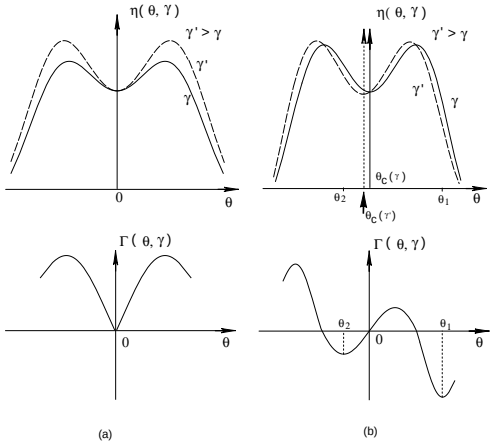


Figure 20.2: Plots of the emissivity (upper) and absorption coefficients (lower), (a) in the absence of curvature drift, and (b) when curvature drift is taken into account. Maser emission is possible only where the absorption coefficient is negative.

Maser curvature emission can also occur due to a twist of the field lines, corresponding to curved field lines that are not confined to a plane. Unlike curvature-drift induced maser emission, this mechanism is not sensitive to the magnitude of the magnetic field B, and electrons and positrons contribute with the same sign. Although this version of curvature maser emission seems more favorable than the curvature-drift induced mechanism, it does require a highly non-dipolar magnetic structure at the surface of the star for the growth rate to be large enough to be of relevance.

Any form of maser curvature emission requires $df/d\gamma > 0$, cf. equation (20.7), and a pump is required to continually regenerate such a distribution. The requirement $df/d\gamma > 0$ is the same as for a relativistic streaming instability, and hence any process that operates as a pump for relativistic plasma emission may also act as a pump for maser curvature emission.

Opinion remains divided on which (if any) of these various emission processes actually operates in pulsars.

Lecture 21

Active Galactic Nuclei

Active galactic nuclei (AGN) include quasars, BL Lacertae objects and the nuclei of Seyfert galaxies. These objects exhibit activity over a wide range of frequencies, from $\lesssim 100\,\mathrm{MHz}$ radio waves to $\gtrsim 100\,\mathrm{MeV}$ gamma rays, and over a wide scale of distance scales, from radio lobes of order several Mpc, to the scales implied by the light travel time corresponding to the shortest time scale on which the X rays vary, $\sim 100\,\mathrm{s}$. They have luminosities ranging up to $\sim 10^{40}\,\mathrm{W}$, with the major energy output in the most powerful objects being in the UV to X-ray range. The most widely favored interpretation is that AGN are powered by accretion onto a supermassive black hole.

Three aspects of AGN are discussed in this lecture. The first aspect concerns the UV, X rays and gamma rays that come from a central region filled with very hot plasma. The second aspect concerns the high-frequency radio emission from some AGN, which is interpreted in terms of a core-jet model. An important ingredient in this model is relativistic beaming, which provides an explanation for so-called superluminal motions. The third aspect is the formation and propagation of jets.

21.1 Very hot plasmas in AGN

Matter falling into a black hole can be seen only while it is outside the gravitational (Schwarzschild) radius,

$$R_S = 2GM/c^2 \approx 3 \times 10^3 (M/M_\odot)\,\mathrm{m}, \tag{21.1}$$

where G is Newton's constant, M is the mass of the black hole and $M_\odot$ is the solar mass. The mean gravitational energy released when matter is accreted is thought to be of order 10% of the rest energy. For each hydrogen atom this corresponds to an energy $0.1 m_p c^2 \approx 100 m_e c^2$ released. One can define a temperature for 'virialized matter', where the virial theorem is used in the

sense that the mean kinetic energy (defining the temperature) is equal to half the potential energy (defined by the gravitational potential on the surface from which the radiation originates). Then, at a radius r from the black hole, this corresponds to $T(r) \approx 0.1 m_p c^2 (R_S/r) \approx 10^{12} (R_S/r)\,\mathrm{K}$. If one defines a very hot plasma as one with $T \sim m_e c^2 = 0.5 \times 10^{10}\,\mathrm{K}$, then the plasma present within $r \sim 10^3 R_S$ of an accreting black hole is very hot. The physics of such plasmas is dominated by the interactions between photons and pairs.

The generic term for interaction between photons and free electrons is Compton scattering. (Strictly, conventional usage is that 'Compton' scattering is the scattering of electrons by photons, and 'Thomson' scattering is the classical theory for the scattering of photons by electrons.) As an example of the role of Compton scattering, consider the effect of the radiation pressure on infalling matter. Specifically, consider a central object mass M accreting at a rate $\dot{M}$ producing a luminosity, L. If the central object is of radius R, the release of gravitational energy GM/R per unit accreted mass implies a luminosity $L = \eta G\dot{M}M/R$, where η is the efficiency of conversion of the energy released into escaping radiation. The escaping radiation exerts a radiation pressure on matter around the accreting object. The radiation pressure may be attributed to Compton scattering of electrons by the escaping radiation. The cross section is the Thomson cross section $\sigma_T = 8\pi r_0^2/3$, where $r_0 = e^2/4\pi\varepsilon_0 = 3 \times 10^{-15}\,\mathrm{m}$ is the classical radius of the electron. A radial outward flow of momentum $L/4\pi r^2 c$ in the radiation then provides a force $L\sigma_T/4\pi r^2 c$ on an electron. The electrons and ions are tied together by the electric field, so that this may be interpreted as the net force on the mean mass per electron. This outward force is opposed by the inward force of gravity, which is GM/r^2 per unit mass. The *Eddington luminosity*, L_{Edd}, is defined as the luminosity at which these two opposing forces balance. Assuming a hydrogen plasma, one finds

$$L_{\mathrm{Edd}} = \frac{4\pi G M m_p c}{\sigma_T} = \frac{2\pi m_p c^3 R_S}{\sigma_T} \approx 10^{31} \left(\frac{M}{M_\odot}\right) \mathrm{W}. \tag{21.2}$$

It appears that AGN have a luminosity of order their Eddington luminosity, implying that Compton scattering must be an important process in such sources.

The optical depth for Thomson scattering depends on the electron number density, n_e: the *Thomson optical depth* is defined by

$$\tau_T = \sigma_T \int ds\, n_e, \tag{21.3}$$

where the integral is along the path of the photon. Thomson scattering is important for the escaping photons for $\tau_T \gtrsim 1$. For $\tau_T \gg 1$ photons are scattered many times before escaping, implying that they diffuse out of the scattering region.

Suppose that the Thomson optical depth is greater than unity and that the energy density in the photons is much greater than the energy density in the

bulk of the electrons (ignoring any distribution of high-energy electrons). Then Compton scattering has several important consequences: it causes the hard photons to be downscattered (due to the quantum recoil) to lower frequencies, soft photons to be upscattered (due to the Doppler effect) to higher frequencies, an energy transfer between the photons and electrons that tends to determine the electron temperature.

Qualitatively one can understand these two energy transfer processes as follows. For an electron initially at rest, any recoil motion of the electron implies that the photon loses energy to the electron, and hence this must cause a downshift in the frequency of the photon and to an energy transfer from photons to electrons. As a result of the Doppler effect, photons gain energy in head on collisions with electrons and lose energy in overtaking collisions with electrons. The average energy change per collision is the same. However, there is a higher probability of a head-on collision, and hence on average the photons gain energy from the electrons.

Ignoring relativistic effects and for $\hbar\omega \ll m_e c^2$, the average frequency *downshift* of hard photons due to the recoil is

$$\left\langle \frac{\Delta\omega}{\omega} \right\rangle_{\rm recoil} = -\frac{\hbar\omega}{m_e c^2}. \tag{21.4}$$

The frequency change due to the Doppler effect gives $\Delta\omega/\omega \sim v/c$. The relative probabilities of head on over overtaking scatterings imply that the net average upshift is $\Delta\omega/\omega \sim (v/c)^2$. Averaging over a Maxwellian distribution gives the average frequency *upshift* of soft photons due to the thermal Doppler effect

$$\left\langle \frac{\Delta\omega}{\omega} \right\rangle_{\rm Doppler} = \frac{4T}{m_e c^2}. \tag{21.5}$$

Both these effects are included in the Kompaneets equation, which applies to an isotropic distribution of photons with occupation number $N(\omega)$ scattered by thermal electrons:

$$\frac{dN(\omega)}{dt} = \frac{\sigma_T n_e}{m_e c} \left(\frac{1}{\omega^2} \frac{d}{d\omega} \omega^4 \right) \left\{ \frac{T}{\hbar} \frac{dN(\omega)}{d\omega} + N(\omega) + [N(\omega)]^2 \right\}. \tag{21.6}$$

The first term inside the curly brackets is the only one that involves the temperature of the electrons, and it describes the effect of the Doppler shift. The middle term describes the effect of the quantum recoil. The final term describes the effect of induced scattering.

The production of pairs can occur due to several different processes, with photon-photon interactions being the most important in many cases. Just above the threshold for the process $\gamma + \gamma \rightarrow e^- + e^+$, the cross section is of order the Thomson cross section σ_T. The effective optical depth for pair production is of order $n_{\rm ph}(m_e c^2/\hbar)\sigma_T R$, where R is the radius of the source and $n_{\rm ph}(m_e c^2/\hbar)$

is the number density of photons with energy of order the threshold, corresponding to $\hbar\omega \approx 0.5\,\mathrm{MeV}$. In terms of the luminosity at this photon energy, in a spherical source one has $n_{\rm ph}(m_e c^2/\hbar) \approx L_{1\,\mathrm{MeV}}/4\pi R^2 m_e c^3$. Hence for $n_{\rm ph}(m_e c^2/\hbar)\sigma_T c \gtrsim 1$ the production of pairs by this process must be important. It is conventional to express this condition in terms of a *compactness parameter*, defined by

$$\ell_\alpha = \frac{L_\alpha \sigma_T}{m_e c^3 R} = 2\pi \left(\frac{m_p}{m_e}\right)\left(\frac{L_\alpha}{L_{\rm Edd}}\right)\left(\frac{R_S}{R}\right), \tag{21.7}$$

for any any photon component labeled α. For $\approx 1\,\mathrm{MeV}$ gamma rays, one concludes that pair production must be important for $\ell_{1\,\mathrm{MeV}} \gtrsim 10$.

The compactness parameter may be estimated from observation for a source that shows temporal variations on a time Δt. Then one estimates $R = c\Delta t$, and inserts the observed values for Δt and for the luminosity near 1 MeV. The evidence suggests that pair production is important in accretion models for AGN.

21.2 Synchrotron-self-Compton model

Synchrotron emission is the primary source of soft (i.e. radio to mm wavelength) photons in an AGN. Inverse Compton scattering of these photons can contribute to the high frequency emission, through the so-called synchrotron-self-Compton model. Compton scattering also limits the brightness temperature of synchrotron sources.

In a model for the core of an AGN it is assumed that both a magnetic field and photons and present simultaneously. The photons include those generated by synchrotron emission. The power radiated per electron may be written

$$P = \frac{4}{3}\,\sigma_T c \gamma^2 (U_{\rm mag} + U_{\rm rad}), \tag{21.8}$$

where $U_{\rm mag} = B^2/2\mu_0$ and $U_{\rm rad}$ are the energy densities in the magnetic field and in the radiation field, respectively. The term involving $U_{\rm mag}$ describes synchrotron emission and that involving $U_{\rm rad}$ describes inverse Compton emission. The typical frequencies radiated are

$$\omega \approx \begin{cases} 0.5\Omega_e \gamma^2 \sin\theta & \text{with } \Omega_e = eB/m_e, \\ \frac{4}{3}\,\bar{\omega}\gamma^2 & \text{with } \hbar\bar{\omega} = U_{\rm rad}/n_{\rm ph}, \end{cases} \tag{21.9}$$

where θ is the angle of emission relative to the magnetic field, and where $n_{\rm ph}$ is the number density of photons.

The loss of energy by the relativistic electrons leads to their cooling on a time scale

$$t_{\rm cool} = \frac{\gamma m_e c^2}{P} = \frac{3 m_e c^2}{4\sigma_T c \gamma (U_{\rm mag} + U_{\rm rad})} \tag{21.10}$$

For synchrotron emission, $t_{\rm cool}$ corresponds to the synchrotron half life time

$$t_{1/2} = \frac{5 \times 10^8}{\gamma} \left(\frac{B \sin^2 \theta}{1\,{\rm G}} \right)^{-2} \ {\rm s}. \tag{21.11}$$

Very bright synchrotron emission is limited by self absorption to (setting Boltzmann's constant equal to unity)

$$\frac{T_{\rm max}}{m_e c^2} \sim \left(\frac{\omega_{\rm max}}{\Omega_e} \right)^{1/2}, \qquad I_{\rm max} = \frac{2 \omega_{\rm max}^2 T_{\rm max}}{(2\pi)^3 c^2}, \tag{21.12}$$

where $I_{\rm max}$ is the maximum value of the specific intensity. If both $T_{\rm max}$ and $\omega_{\rm max}$ are observed, then (21.12) enables one to estimate the magnetic field in the source.

Inverse Compton scattering of photons generated by synchrotron emission can lead, in principle, to an *inverse Compton catastrophe*. Consider $U_{\rm rad}$ in (21.10): it consists of the photons generated by synchrotron emission, the photons generated by inverse Compton scattering of the synchrotron photons, the photons generated by inverse Compton scattering of these scattered photons, and so on. Provided that one ignores the suppression of Compton scattering at high photon energies (the Klein-Nishina cross section falls below the Thomson cross section), this sequence of terms corresponds to a geometric series, which diverges when the ratio of successive terms approaches unity. As a result, one expects inverse Compton losses to become catastrophic, and to limit the brightness temperature of any synchrotron source. Writing $U_{\rm rad} \sim \omega^3 T_{\rm b}/(2\pi c)^3$, and assuming $\omega \sim \Omega_e \gamma^2$, $T_{\rm b} \sim \gamma m_e c^2$, then $U_{\rm rad} = U_{\rm mag}$ implies

$$\omega \left(\frac{T_{\rm b}}{m_e c^2} \right)^5 \sim \frac{c}{r_0}, \tag{21.13}$$

which implies a maximum $T_{\rm b} \sim 10^{12}\,{\rm K}$ for a typical frequency around a few gigaherz. No synchrotron spectrum should have a higher brightness temperature than implied by (21.13). Although some sources have $T_{\rm b}$ apparently greater than this limit, it appears that these can be explained in terms of Doppler boosting, or in terms of scintillations, as discussed below.

The radio emission from some AGN is interpreted in terms of a core-jet model, as illustrated in Figure 21.1. Suppose that the core is optically thick with a spectrum of relativistic particles of the form $N(\gamma) = K \gamma^{-2}$ for $\gamma_{\rm min} < \gamma < \gamma_{\rm max}$. The jet is assumed to move with constant speed within a cone. Particle acceleration is assumed to take place continuously, so that $\gamma_{\rm min}$, $\gamma_{\rm max}$ remain constant. Then conservation of particles implies $K \propto r^{-2}$. Assuming that the azimuthal magnetic field dominates, it scales as $B \propto r^{-1}$.

The synchrotron emission from an optically thin and an optically thick portion of the jet then scale as

$$S(\omega, r) \propto K B^{1.5} \omega^{-0.5} r^3 \propto (\omega r)^{-0.5}, \qquad S(\omega, r) \propto B^{-0.5} \omega^{2.5} r^2 \propto (\omega r)^{2.5}, \tag{21.14}$$

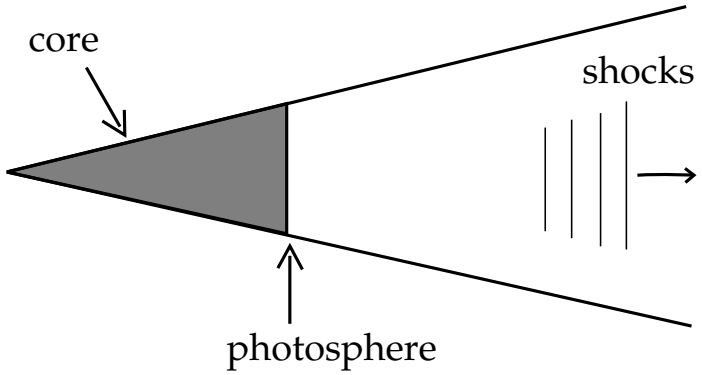


Figure 21.1: A simple model of a nonrelativistic core-jet source. The core is the unresolved optically thick inner region, shown shaded. The resolved optically thin jet contains shocks that propagate outward accelerating relativistic electrons.

respectively. Both relations (21.14) depend only on the combination ωr. It follows that if the source is unresolved at one frequency, then it is unresolved at all frequencies for which the model is valid.

The maximum emission occurs approximately where the two estimates in (21.14) are equal. It follows that one has

$$S_{\max}(\omega) = \text{const}, \qquad \omega_{\max} \propto r^{-1}. \tag{21.15}$$

The predicted spectrum is flat ($\alpha = 0$), as illustrated in Figure 21.2. This model is roughly in accord with the observations, except that it is nonrelativistic.

21.3 Relativistic beaming effects

Compact sources vary on various time scales $t_{\rm var}$ from less than a day to many years. For variation on time scales less than about a year, the implied angular size, $\sim ct_{\rm var}/D$ where D is the distance to the source, is so small that it implies a brightness temperature in excess of the maximum, allowed for synchrotron emission, $T_{\rm b} \sim 10^{12}\,{\rm K}$. There appear to be only three ways of overcoming this difficulty: appeal to a relativistic beaming effect, assume that the variations are not intrinsic to the source (but are due to scintillations), or abandon the synchrotron hypothesis.

A simple model for relativistic expansion is illustrated in Figure 21.3. There are two sources, one of which remains at rest relative to the observer, and the other moves toward the observer at speed βc at an angle θ to the line of sight. If the two sources coincide at $t = 0$, then after time t the moving one is displaced a

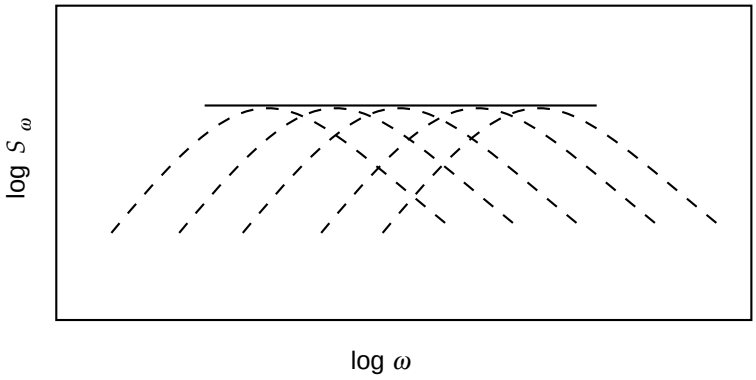


Figure 21.2: A flat integrated spectrum, shown by the solid line, results from the model discussed in the text in which the emission from each point in the source has the self-absorbed shape indicated by the dashed curves.

distance $\Delta x = ct\beta \sin\theta$. The observer measures a time delay $\Delta t = t(1-\beta\cos\theta)$ that is shorter than t due to the motion of the source. Hence the transverse expansion speed is estimated to be

$$\beta_{\rm obs} = \frac{\beta\sin\theta}{1-\beta\cos\theta}. \tag{21.16}$$

The maximum expansion speed is for $\cos\theta = \beta$, when one has $\beta_{\rm obs} = \gamma\beta$, which is superluminal for $\beta > 1/\sqrt{2}$.

Note that this model does not require that the actual motions be as illustrated. The stationary source could be the surface of the self-absorbed region of a jet, which surface can remain stationary despite the motion of the jet, and the moving source could be a shock or other disturbance propagating outward along the jet.

The emission and absorption coefficients from a moving source are modified by Lorentz transformation factors from the corresponding coefficients for a stationary source. Consider two frames: the observer's rest frame, $\mathcal{K}$, in which quantities are unprimed, and the rest frame of the source, $\mathcal{K}'$, in which quantities are primed. The frequency is the time-component of a 4-vector, and so transforms according to

$$\omega = \mathcal{D}\omega', \qquad \mathcal{D} = [\gamma(1-\beta\cos\theta)]^{-1}, \tag{21.17}$$

where $\mathcal{D}$ is referred to as the Doppler factor.

The specific intensity may be written in the form

$$I(\omega) = \frac{2\hbar\omega^3 N(\omega)}{c^2}, \tag{21.18}$$

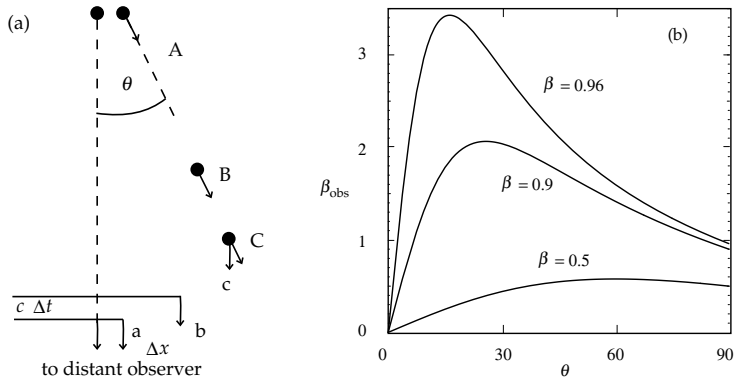


Figure 21.3: (a) An idealized model for superluminal motion: emission from the moving source at A and B, denoted by a and b, appears to move a distance Δx in a time Δt, so that its apparent transverse speed is $\Delta x/\Delta t$. (b) Plot of the function (21.16) showing the apparent transverse speed.

where the photon occupation number $N(\omega)$ is an invariant. Hence the transformed specific intensity involves a factor $\mathcal{D}^3$. Similarly, the brightness temperature may be written in the form $T_{\rm b}(\omega) = \hbar\omega N(\omega)$, so that it transforms with a factor $\mathcal{D}$. Thus one has

$$I(\omega) = \mathcal{D}^3 I'(\omega'), \qquad T_{\rm b}(\omega) = \mathcal{D} T'_{\rm b}(\omega'). \tag{21.19}$$

Now consider the equation of radiative transfer,

$$\frac{d}{dx} I(\omega) = J(\omega) - \mu(\omega) I(\omega), \tag{21.20}$$

where $J(\omega)$ is the emission coefficient and $\mu(\omega)$ is the absorption coefficient, and where the derivative is along the ray path. An element of path length transforms with a factor $\mathcal{D}$, and hence one deduces

$$dx = \mathcal{D} dx', \qquad \mu(\omega) = \mathcal{D}^{-1}\mu'(\omega'), \qquad J(\omega) = \mathcal{D}^2 J'(\omega'). \tag{21.21}$$

These formulas allow one to evaluate the specific intensity in the source frame and then to transform to the observer's frame. To take the redshift of the source into account one then includes an additional factor in accord with

$$\mathcal{D} \to (1+z)^{-1}\mathcal{D}. \tag{21.22}$$

For an optically thin source the flux density is given by integrating the specific intensity over solid angle. This gives

$$S(\omega) = \int d^2\mathbf{\Omega}\, I(\omega) = \frac{1}{D^2}\int d^2A dl\, J(\omega) = \frac{\mathcal{D}^{2+\alpha}}{D^2}\int d^3V\, J'(\omega'), \tag{21.23}$$

where D is the distance to the source, α is the spectral index, dl is an element of length along the line of sight, d^2A is an element of area, and where $d^3V = d^2A dl$ is an element of volume.

One may use (21.23) to estimate the relative intensity of two identical jets propagating in opposite directions along an axis at an angle θ to the line of sight. One finds

$$\frac{S_1}{S_2} = \left(\frac{1+\beta\cos\theta}{1-\beta\cos\theta}\right)^{2+\alpha}. \tag{21.24}$$

For $\gamma \gtrsim \theta^{-1} \gtrsim 1$, this may be approximated by $S_1/S_2 \sim (2/\theta)^{4+2\alpha}$, which can be a very large factor, e.g., $\sim 10^3$ for $\theta \sim 30^\circ$ and $\alpha \sim 0.5$. This provides a natural explanation of why relativistic jets should appear one sided. (However, there is evidence that some one-sided jets are nonrelativistic, and hence that the one-sidedness is intrinsic in these cases.)

Another implication of Doppler boosting is on the interpretation of source counts. Suppose all sources are identical except that they are viewed at different angles. If all viewing angles are equally probable, then the fraction of all sources viewed at $< \theta$ is $\frac{1}{2}(1-\cos\theta)$. Smaller θ implies higher flux density according to (21.23). Assuming $\theta \lesssim \gamma^{-1} \ll 1$, this implies that the number of sources with S greater than a specific value would vary according to $N(>S) \propto S^{-1/(2+\alpha)}$. Although the actual source counts do not follow this relation, it is clear that the effect of Doppler boosting and the viewing angle needs to be taken into account in interpreting the source counts.

The explanation in terms of scintillations applies to sources for which $T_{\rm b}$ is estimated using an argument based on the time scale for variations. A brightness temperature may be estimated using the observed flux density and the variability time scale, $t_{\rm var}$. This is

$$T_{\rm var}(\omega) = \frac{4\pi^2 S(\omega) D^2}{2\omega^2 c^2 t_{\rm var}^2}. \tag{21.25}$$

Suppose that the proper time scale for variations is R/c, with R the source size, and further suppose that the time transforms as the inverse of the frequency, cf. (21.17). Then one has $t_{\rm var} \sim R/c\mathcal{D}$, and using the relations (21.19) one finds

$$T_{\rm var} \sim \mathcal{D}^3 T'_{\rm var}, \tag{21.26}$$

where T' is the brightness temperature in the source frame. Note that (21.26) implies a different dependence than (21.19) because $T_{\rm var}$ is an inferred maximum brightness temperature, which depends on several observational parameters, and not the measured maximum brightness temperature. Specifically, $T_{\rm var}$ depends on the variation time scale $t_{\rm var}$, which is used to infer the size of the radiating region. Thus, the interpretation in terms of scintillations is that the observed variations are not directly related to the source, and are dependent on the properties of the intervening medium. Hence, the argument that small $t_{\rm var}$ implies high $T_{\rm b}$ is not necessarily valid.

21.4 Extragalactic jets

The existence of the extended radio sources (the lobes) in extragalactic radio sources has been known since the early days of radio astronomy. It was subsequently recognized that the lobes must be powered continuously, and it is now accepted that the energy is channelled along highly collimated jets. Extended jets are conveniently divided into two types: sometimes called Fanaroff-Riley jets of class 1 and class 2. Class 1 are edge-brightened jets with their radio emission concentrated at their extremities; they have a high power, $L > 10^{35}$ W. Class 2 are edge-darkened jets with a low power, $L < 10^{35}$ W.

VLBI observations established the existence of compact sources that may also contain jets. Compact sources have their emission originating from less than 1 kpc from the central engine, and their radio spectra are relatively flat, $\alpha \sim 0$–0.5. It is now believed that the extended and compact sources are essentially the same phenomenon viewed from different perspectives. Compact radio sources have a variety of morphological types. *Core-jet* sources comprise a flat spectrum ($\alpha \sim 0$) core and a steeper spectrum one-sided jet. Features in these jets are often seen to be moving away from the core with apparent superluminal speeds.

A large class of objects are referred to collectively as *blazars*. These include BL Lac objects and optically violent variable (OVV) sources. These objects exhibit rapid variability and high optical polarization. Most compact objects exhibit blazar characteristics at some level. The favored interpretation is that blazars and AGN are related objects, with apparent differences due in part to the viewing angle. For example, in such so-called "unified" schemes, BL Lac objects are interpreted in terms of a viewing angle directly along the axis of an oncoming relativistic jet.

The simplest model for a jet is for a one-dimensional and nonrelativistic jet that is pressure confined by an external ideal gas. However, there is evidence that jets are not confined by external gas pressure. The pressure, estimated from the observed synchrotron radiation, does not conform to the predicted form for a nonrelativistic jet, and the magnitude of the estimated pressure can be implausibly large. As a consequence, confinement by a hoop stress in a magnetic field wound into a spiral about the jet has been proposed. Such a model involves a current flowing out along the body of the jet. The axial current produces an azimuthal magnetic field that has a value $B_\phi \propto r^{-1}$ for $r > r_{\rm jet}$, where $r_{\rm jet}$ is the radius of the jet. The return current is assumed to flow on a sheath (the surface of a "cocoon"), with $r_{\rm sheath} \gg r_{\rm jet}$. The pressure $B^2_{\rm sheath}/2\mu_0$ is assumed to balance the ambient pressure. This allows the pressure confining the jet to be larger than the ambient pressure by the factor $(r_{\rm sheath}/r_{\rm jet})^2$. This model encounters difficulties in terms of the current path and closure by flow across field lines.

The MHD processes involved in the formation of jets in AGN are thought to be closely related to the processes involved in the formation of galactic jets, such

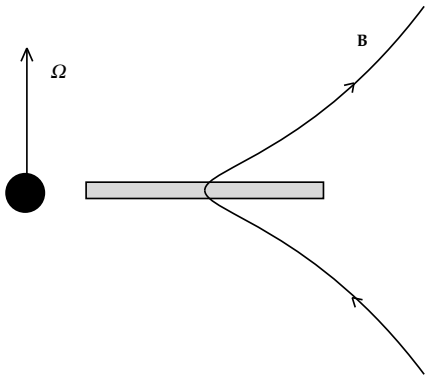


Figure 21.4: A schematic magnetic field lines threading an accretion disk is illustrated. The angle that the magnetic field line leaves the disk determines whether cold plasma is flung outward due to the component of the centrifugal force along the field line exceeding the component of the gravitational force.

as SS 433 and the bipolar outflows from young stellar objects (YSOs). With few exceptions, theories of MHD jets have been developed with applications to both galactic and extragalactic jets in mind. These theories are related to theories for magnetized winds. There is no clear formal distinction between a jet and a wind. One qualitative distinction is that in jets the outflow is predominantly bipolar, being collimated approximately along the rotation axis of the central object, and in a wind the outflow is approximately isotropic, or perhaps concentrated toward the equatorial plane.

Hydromagnetic models for jets can be classified in various ways. A "disk jet" is one that originates from a disk and not from a centrally condensed object. A "centrifugally-driven" jet implies that the driving force for the outflow is the centrifugal force on a corotating blob of gas in the disk, rather than a driving force involving some combination of magnetic stress and gas pressure.

A particularly attractive feature of centrifugally-driven jet models is that they link the formation of the jet with the angular momentum transfer in the disk. A small fraction of the accreting mass is thrown off with a lever arm of order the Alfvén radius. This small fraction of mass carries off the excess angular momentum and allows the accretion to proceed. An essential ingredient is that there be an ordered magnetic field perpendicular to the disk, and that this be anchored in the disk, as illustrated in Figure 21.1. Thus, a centrifugally-driven jet model automatically incorporates a model for angular momentum transport in the accretion disk, and should provide a relation between the accretion rate and the properties of the jet.

Plasma Astrophysics

Exercise Set 1

1.1 Calculate the gyroradius of a particle under the following conditions.

(a) A 2 keV electron with pitch angle $\alpha = \pi/2$ at the Earth's magnetic equator at $L = 3$. Assume that the Earth's magnetic field is dipolar with the magnetic field at the pole $B = 0.3\,\mathrm{G}$. Express your answer in centimeters.

(b) A 10 keV electron with pitch angle $\alpha = \pi/4$ at in a solar flux tube with $B = 100\,\mathrm{G}$. Express your answer in centimeters.

(c) A1 MeV proton with pitch angle $\alpha = \pi/2$ in the solar wind with $B = 5\,\mathrm{nT}$. Express your answer in kilometers.

(d) A 10^{18} eV ion with pitch angle $\alpha = \pi/2$ in the interstellar magnetic field with $B = 3\,\mu\mathrm{G}$. Express your answer in parsecs.

1.2 The flux J_L of light elements, Li, Be, B, denoted collectively by L, and the flux J_M of medium elements, C, N, O, denoted collectively by M, in cosmic rays increases with distance, s, traversed due to spallation. Let the distance traversed be parametrized by x, defined by $dx = \eta ds$, where η is the mass density in the interstellar medium. The fluxes eveolve according to

$$\frac{dJ_M(x)}{dx} = -\frac{J_M(x)}{x_M}, \qquad \frac{dJ_L(x)}{dx} = -\frac{J_L(x)}{x_L} + \frac{P_{ML} J_M(x)}{x_M},$$

where $x_M = 6.7\,\mathrm{g\,cm^{-2}}$, $x_L = 10\,\mathrm{g\,cm^{-2}}$, and $P_{ML} = 0.35$ are constants.

(a) Show that the solution for $J_M(0) = \text{constant}$, $J_L(0) = 0$ implies

$$\frac{J_L(x)}{J_M(x)} = \frac{P_{ML} x_L}{x_l - x_M} \left[\exp\left(\frac{x}{x_M} - \frac{x}{x_L}\right) - 1\right].$$

(b) Show that the observed abundances, $J_L(x)/J_M(x) = 0.25$, imply $x = 4.3\,\mathrm{g\,cm^{-2}}$.

(c) Assuming that the cosmic rays traverse a hydrogen plasma with uniform number density $n = 1\,\mathrm{cm^{-3}}$, estimate the distance traveled by an average cosmic ray.

1.3 A magnetic field has cartesian components

$$\mathbf{B} = (\alpha y^2, \beta x, 0),$$

where α and β are constants.

(a) Find the equations for the field lines.

(b) Calculate the current density $\mathbf{J}$.

(c) Calculate the force density $\mathbf{J} \times \mathbf{B}$.

1.4 The magnetic potential for a dipole, $\mathbf{m}$, in cylindrical coordinates is

$$\phi_M = \frac{mz}{(\varpi^2 + z^2)^{3/2}}.$$

(a) Evaluate $\mathbf{B} = -\mathrm{grad}\,\phi_M$ in cylindrical coordinates.

(b) Evaluate $\mathrm{curl}\,\mathbf{B}$ in cylindrical coordinates, and show that it is identically zero.

Remark: Relevant formulas are given in the appendix to the lecture notes.

2.1 The electric field in a rigidly rotating magnetosphere ($\mathbf{\Omega}$ = constant) is given by

$$\mathbf{E} = -(\mathbf{\Omega} \times \mathbf{x}) \times \mathbf{B}. \qquad (E2.1)$$

(a) Show that the charge density is given by

$$\rho = -\varepsilon_0(2\mathbf{\Omega} \cdot \mathbf{B} - \mathbf{\Omega} \times \mathbf{x} \cdot \operatorname{curl} \mathbf{B}). \qquad (E2.2)$$

(b) Estimate ρ/e for the magnetosphere of the Earth at a distance one Earth radius above the equator, and compare this number with the typical number density $\approx 10^3\,\mathrm{cm}^{-3}$ in the plasmasphere. The terrestrial magnetic field is to be assumed dipolar with a polar magnetic field $B_p = 0.3\,\mathrm{G}$.

(c) A rotating charged magnetosphere has a current density $\mathbf{J} = \rho\mathbf{\Omega} \times \mathbf{x}$, and this current generates a nonpotential component of magnetic field. Estimate the way the z-component of this nonpotential field varies with radial distance ϖ (and the constants Ω, B_p) in the equatorial plane of an aligned, rigid rotator.

(d) Suggest an astrophysical situation where this nonpotential magnetic field might cause a significant distortion to the net field, and use semiquantitative estimates to justify (or otherwise) your suggestion.

2.2 In very low density plasmas the displacement current cannot be neglected when treating MHD waves.

(a) Show that if the displacement current is included, the dispersion equation is of the form (6.15) with (6.16) replaced by

$$\Gamma_{ij}(\omega, \mathbf{k}) = \omega^2 \left[\delta_{ij} + \frac{v_A^2}{c^2}(\delta_{ij} - b_i b_j)\right] - k^2 c_s^2 \kappa_i \kappa_j - k^2 v_A^2 [\kappa_i \kappa_j - \cos\theta(\kappa_i b_j + b_i \kappa_j) + \cos^2\theta\, \delta_{ij}]. \qquad (E2.3)$$

(b) Hence show that the dispersion relation for Alfvén waves, $\omega = \omega_A(\mathbf{k})$, has the form

$$\omega_A(\mathbf{k}) = \frac{k v_A |\cos\theta|}{(1 + v_A^2/c^2)^{1/2}}. \qquad (E2.4)$$

(c) As the plasma density approaches zero, $n \to 0$, the Alfvén speed approaches infinity, $v_A \to \infty$. Show that the dispersion relation for Alfvén mode approaches $\omega = kc|\cos\theta|$ in this limit, and find the corresponding limiting value of the dispersion relation for the fast mode (assuming $c_s \ll c$). [One expects the wave properties to approach those of electromagnetic waves in vacuo in this limit, but the dispersion relation for Alfvén mode approaches $\omega = kc|\cos\theta|$ and not $\omega = kc$. Can you suggest how this dilemma might be resolved?]

2.3 Consider shock waves in an unmagnetized, ideal gas with $\Gamma = 5/3$.

(a) Plot the compression ratio, $r = \eta_2/\eta_1$, as a function of the Mach number, M.

(b) Find the Mach number at which the compression ratio has the value $r = 2$.

(c) Find the pressure ratio P_2/P_1 and the temperature ratio T_2/T_1 corresponding to $r = 2$.

(d) Find the increase in the entropy ($S = C_V \ln(P/\eta^\Gamma) + \text{constant}$) under the same conditions.

2.4 The equation (7.23) that determines the relation between the compression ratio and the Mach number for a discontinuity in a magnetized plasma is derived from the jump conditions.

(a) Fill in the steps in the derivation (7.23), viz.

$$\left(\frac{aM_A^2}{r} - \beta\right)\left(\frac{M_A^2}{r} - \cos^2\theta\right)^2 - \frac{M_A^2}{r}\sin^2\theta\left[\frac{M_A^2}{r}\left(\frac{a}{r} - \frac{1-r}{2}\right) - a\cos^2\theta\right] = 0. \tag{E2.5}$$

(b) Show that the solutions of ($E2.5$) for M_A^2 in the limit of a small amplitude jump, that is in the limit $r \to 1$, reduce to the dispersion equations for the MHD waves, with M_A^2 interpreted as $\omega^2/k^2v_A^2$ and with $\beta = c_s^2/v_A^2$.

3.1 The resonance ellipse is usually drawn in velocity space. The following exercise involves drawing the corresponding curve in momentum space.

(a) Show that the Doppler condition $\omega - s\Omega - k_\parallel v_\parallel = 0$ implies

$$(m^2c^2 + p_\perp^2 + p_\parallel^2)\omega^2 = (s\Omega_0 m + k_\parallel p_\parallel)^2 c^2. \qquad (E3.1)$$

(b) On plotting (E3.1) in $p_\perp$-$p_\parallel$ space, the resulting curve is a conic section. Identify the parameters of the conic section.

(c) Show that the relevant conic section for resonant scatttering by Alfvén waves ($\omega = k_\parallel v_A$) of highly relativistic particles is an hyperbola for $\cos\alpha \ll v_A/c$.

(d) Sketch the hyperbola in part (c) for (i) $s = 1$, and (ii) $s = -1$

3.2 Fill in the details of the derivations of equations (9.13) and (9.14). Specifically,

(a) show that on inserting (9.8) into (9.11) and thence into (8.17), after summing over $s = \pm 1$, the result reduces to (9.13), and

(b) show that making similar approximations in (8.19), with $Q = Q' = \alpha$, the result reduces to (9.14).

3.3 It is found that the bulk of the energetic particles released from the Sun in a solar flare arrive at the Earth later than would be predicted by propagation directly along the Archimedean spiral field line that joins the source to the Earth. It is assumed that the particles are scattered by waves and so diffuse along the field lines reaching a distance z at time t related by $z^2 = \kappa t$, where κ is the spatial diffusion coefficient.

(a) Assuming that the distance from the Sun to the Earth along the spiral path is 1.5 AU, estimate κ in the case where particles are observed to arrive one day after a flare.

(b) Consider the particular case of protons with an energy of 30 MeV that are scattered by Alfvén waves. Assuming that the magnetic field is $B = 8$ nT, estimate the required wavenumber of the Alfvén waves.

(c) Using (9.14) and (9.19), make an order of magnitude estimate of the energy density in resonant Alfvén waves required to cause the scattering in part (a) for the particles in part (b).

3.4 In the preceding problem, the magnetic field decreases with distance approximately as $B \propto r^{-2}$ inside the orbit of the Earth.

(a) Estimate the pitch angle that a particle injected with $\alpha = \pi/2$ at the Sun would have at the orbit of the Earth if there were no scattering.

(b) Estimate the pitch angle diffusion coefficient, $D_{\alpha\alpha}$, from the value of κ derived for the particles in part (b) of the preceding problem. (You cannot derive the functional dependence on α without further information.)

(c) Assuming that the particles diffuse through an angle $\Delta\alpha$ in a time t given by $(\Delta\alpha)^2 = D_{\alpha\alpha}t$, estimate the expected degree of anisotropy, δ, of the particles by equating by equating δ to the ratio of the rate at which the effect in part (a) tends to make the particles anisotropic to the rate at which the effect in part (b) tends to make them isotropic.

4.1 The derivation of the Kolmogoroff spectrum is based on dimensional analysis. The spectrum is $S(k) \propto k^{-5/3}$, where $S(k)$ is the energy density per unit wavenumber k in the turbulent spectrum. An important assumption is that the energy flux F through a spherical shell in k-space is independent of k; physically this corresponds to assuming that larger eddies (smaller k) break up into smaller eddies (larger k) conserving energy in the turbulence. Specifically, F is set equal to the power input at the largest eddies, which is equal to the power dissipated by viscocity at the smallest eddies.

(a) With M, L and T denoting mass, length and time, respectively, write down the dimensions of F, $S(k)$, k and the mass density η.

(b) The dimensional argument is that the energy flux F can depend only only the other three variables, and hence that one has

$$F \propto \eta^{\alpha}[S(k)]^{\beta}k^{\gamma}, \qquad (E4.1)$$

with α, β, γ determined by the requirement that the dimensions of both sides of $(E4.1)$ be the same.

(c) Show that the argument in part (b) implies the Kolmogoroff spectrum.

4.2 The observation of gamma-ray lines in large solar flares implies that ions with energy $\gtrsim 40\,\mathrm{MeV/nucleon}$ start bombarding the solar surface soon after the onset of a flare. The gamma-ray emission apears within about one second of the onset of the flare implying that the ions are accelerated within this time scale. Consider a theory for this acceleration in which stochastic acceleration due to MHD turbulence is assumed.

(a) Using the result (11.20), viz.

$$D(p) = \frac{\pi\bar{\omega}}{4}\,\frac{W_M}{B^2/2\mu_0}\,\frac{p^2 v_A}{v}\left\langle \frac{\sin^4\alpha}{|\cos\alpha|}\left(1 - \frac{v_A^2}{v^2\cos^2\alpha}\right)\right\rangle. \qquad \text{(E4.2)}$$

for the diffusion coefficient due to magnetoacoustic waves, assume $v^2 \gg v_A^2$ and estimate the magnitude of $D(p)$ for $\alpha = \pi/4$, $B = 300\,\mathrm{G}$, $n_e = 10^{11}\,\mathrm{cm}^{-3}$.

(b) Using (11.22) and (11.23), viz.

$$\langle\dot{\varepsilon}\rangle = \frac{1}{p^2}\frac{\partial}{\partial p}\left[p^2 D(p) v\right], \quad t_{\mathrm{a}} = \frac{\varepsilon}{\langle\dot{\varepsilon}\rangle}, \qquad (E4.3)$$

estimate the acceleration time t_{a} under the same conditions used in part (a).

(c) Estimate the required ratio $W_M/(B^2/2\mu_0)$ of the energy density in the waves to the energy density in the background magnetic field for the inferred acceleration of ions to occur.

4.3 Fill in the details of the solution of (12.1) to derive the result (12.6), viz.

$$f_+(p) = bp^{-b}\int_0^p \mathrm{d}p'\, p'^{(b-1)} f_-(p'), \quad b = \frac{3u_1}{u_2 - u_1}, \qquad (E4.4)$$

for the spectrum, $f_+(p)$, of particles downstream of a shock in terms of the spectrum, $f_-(p)$, upstream of the shock.

4.4 Consider diffusive shock acceleration of electrons in a synchrotron source. Assume a strong shock, $r = 4$, and a scattering mean free path $\lambda = \frac{1}{3}r_g$. Estimate the maximum energy to which the electrons can be accelerated in the presence of synchrotron losses.

(a) Specifically, use (12.14), viz.

$$\frac{\mathrm{d}p}{\mathrm{d}t} = \frac{p}{t_\mathrm{a}} - \left(\frac{\mathrm{d}p}{\mathrm{d}t}\right)_\mathrm{loss}, \quad t_\mathrm{a} = \frac{u_1^2}{c\bar{\lambda}}, \quad \bar{\lambda} = \frac{3r(\lambda_1 + r\lambda_2)}{4(r-1)}, \tag{E4.5}$$

to estimate the value of the energy at which the acceleration balances the loss term.

(b) Apply the result in part (a) to synchrotron losses as given by (18.14), viz.

$$\dot{\varepsilon} = -b(\alpha)\varepsilon^2, \qquad b(\alpha) = \frac{2e^2\Omega_e^2 \sin^2\alpha}{3(4\pi\varepsilon_0)m_e^2c^5}, \tag{E4.6}$$

for $\alpha = \pi/4$.

(c) Assuming that this mechanism limits the energy of cosmic ray electrons in the galaxy, make estimates that you consider plausible for the parameters involved, and estimate the maximum energy to which you would expect electrons to be acclerated in the interstellar medium by shocks from supernovas.

Plasma Astrophysics Exercise Set 2

2.1 In Figure 4.3b an idealized magnetic field corresponding to an X-type neutral point is illustrated schematically. An idealized mathematical model for such a field is $\mathbf{B} = ay\hat{\mathbf{x}} + bx\hat{\mathbf{y}}$, where a and b are constants

1. Calculate the current density implies by this field.
2. Determine the Maxwell stress for this field, separating it into an isotropic pressure force and a tension force.

2.2 In very low density plasmas the displacement current cannot be neglected when treating MHD waves. When the displacement current is included, the dispersion equation is of the form (5.15) with (5.16) replaced by

$$\Gamma_{ij}(\omega, \mathbf{k}) = \omega^2 \left[\delta_{ij} + \frac{v_A^2}{c^2}(\delta_{ij} - b_i b_j)\right] - k^2 c_s^2 \kappa_i \kappa_j \\ -k^2 v_A^2 [\kappa_i \kappa_j - \cos\theta(\kappa_i b_j + b_i \kappa_j) + \cos^2\theta\, \delta_{ij}].$$

1. Write Γ_{ij} as a matrix, choosing axes such that $\mathbf{b}$ is along the z-axis, and κ is in the x-z plane.
2. Hence (or otherwise) show that the dispersion relation for Alfvén waves, $\omega = \omega_A(\mathbf{k})$, is $\omega_A(\mathbf{k}) = kv_A|\cos\theta|/(1 + v_A^2/c^2)^{1/2}$.
3. As the plasma density approaches zero, $\eta \to 0$, the Alfvén speed approaches infinity, $v_A \to \infty$. Show that the dispersion relation for Alfvén mode approaches $\omega = kc|\cos\theta|$ in this limit.
4. One expects the wave properties to approach those of electromagnetic waves in vacuo in the limit $v_A \to \infty$, but the dispersion relation for Alfvén mode approaches $\omega = kc|\cos\theta|$ and not $\omega = kc$. Can you suggest how this dilemma might be resolved?

2.3 The angular momentum loss rate due to magnetic braking of a star is $\ell = \Omega r_A^2$ per unit mass.

1. Show that if r_A and the mass loss rate, $\dot{M}$, are constant, then Ω decreases $\propto \exp(-t/\tau)$ with $\tau = I/\dot{M}r_A^2$.
2. Estimate τ for the Sun ($M_\odot = 2 \times 10^{30}\,\mathrm{kg}$, $R_\odot = 7 \times 10^8\,\mathrm{m}$). You may assume the Sun to be a sphere of uniform density rotating once every 28 days, that the solar wind is spherical, composed of hydrogen and has steady speed of $500\,\mathrm{km\,s^{-1}}$ and a number density $n_e = 10\,\mathrm{cm^{-3}}$ at the orbit of the Earth, and that $r_A = 20\,R_\odot$.