Radiation I

Jan Trulsen

The Institute of Theoretical Astrophysics
University of Oslo

Winter 2011

Foreword

The following compendium constitutes the curriculum for the course AST 3210, Radiation I as taught at the Institute of Theoretical Astrophysics at the University of Oslo for the spring semester 2009. The compendium is in a preliminary form, figures and the selection of problems in particular are still incomplete. Suggestions for improvement of the text and reports of misprints are most welcomed.

The course aims at providing quantitative training in basic physics concepts constituting the foundation for any advanced study of astrophysical phenomena. The course starts with a brief review of Maxwell equations and electromagnetic waves, then proceeds to a discussion of electromagnetic radiation emitted by accelerated charged particles. Next, an introduction to quantum mechanics at an intermediate level is given to lay the foundation for an understanding of atomic and molecular spectra. An introduction to basic statistical concepts, statistical physics and thermodynamics including the Saha equation is included, mostly for reference purposes. Finally, the course includes an introduction to the topic of radiation transport. The latter topic deals with the question of how radiation is generated in, interacts with, and is transported through material media. It has been necessary to limit the major part of the discussion of the latter topic to the local thermal equilibrium (LTE) approximation.

For reference purposes the compendium also contains an introduction to fluid mechanics including magnetohydrodynamics (MHD) and dimensional analysis, in addition to an appendix outlining useful vector calculus results.

The different topics covered, are discussed as topics of physics. The selection of topics, examples and problems have been made on the basis of their astrophysical relevance. The more detailed discussion of astrophysical phenomena is deferred to subsequent and more specialized courses in astrophysics.

The course is best suited for students who have completed at least one year training in both mathematics and in physics. She is expected to have working knowledge of linear algebra, including the eigenvalue problem, calculus including differential equations and the Gauss and Stoke integral theorems. Likewise he is expected to be acquainted with basic properties of Maxwell's equations, Newtonian mechanics and also to have passed a first courses in quantum mechanics and statistical physics.

In the compendium emphasis is put on the derivation of import results from the basic equations, sometimes based on simplified models. The focus of the course is, never the less, on applying these results for the solution of relevant astrophysical questions. This is reflected in the fact that the course makes use of open book exams.

The compendium is supplied with a number of problems to be solved, some simple, others more demanding. Most of these problems can be and should be solved analytically. Others will require the use of computers for numerical or graphical reasons. The student is therefore expected to be familiar with a suitable programming language such as MATLAB, Python,

IDL or others.

SI units are used throughout the compendium. This modern unit system is by now totally dominating the science literature. The student should, however, be warned that one may still find astrophysical texts, particularly within the field of radiation transport, clinging to obsolete unit systems.

No explicit reference to literature is made in the compendium. Instead the student is directed to standard textbooks for further reading. Much additional information can of course be found on the web, for instance on Wikipedia, and in the professional literature. The student is encouraged to familiarize herself and make regular use of these resources. Below is collection of textbooks that have been consulted during the writing of the compendium:

- J. D. Jackson: Classical Electrodynamics, John Wiley, ISBN 0-471-30932-X
- D. J. Griffiths: Introduction to Quantum Mechanics, Pearson Prentice Hall, ISBN 0-13-191175-9
- C. Kittel & H. Kroemer: Thermal Physics, Freeman, ISBN 0-77167-1088-9
- B. H. Bransden & C. J. Joachain: Physics of atoms and molecules, Longman Scientific and Technical, ISBN 0-582-44401-2
- D. Mihalas & B. Weibel-Mihalas: Foundations of Radiation Hydrodynamics, Dover, ISBN 0-486-40925-2

Oslo, Winter 2011 Jan Trulsen

Contents

List of Symbols

a_B	Bohr radius
c	speed of light (in vacuum)
e	elementary charge
$\hat{\boldsymbol{e}}$	polarization vector
g	Landé factor
g	multiplicity
$\boldsymbol{g}$	gravitaional acceleration
h	Planck's constant
$\hbar$	$\hbar = h/2\pi$
j	angular momentum quantum number
$\boldsymbol{j}$	electric current density
k	wave number
$\boldsymbol{k}$	wave vector
k_σ	spring constant
ℓ	quantum number
m	particle mass
$\boldsymbol{m}_E$	electric dipole moment
m_j	azimuthal angular momentum quantum number
m_ℓ	quantum number
$\boldsymbol{m}_L$	orbital magnetic dipole moment
m_s	azimuthal spin quantum number
m_{H}	mass of hydrogen atom
$\boldsymbol{m}_S$	spin magnetic dipole moment
n	principal quantum number
n_Q	quantum consentration
p	degree of polarization
$\boldsymbol{p}$	momentum
p_F	Fermi momentum
$\boldsymbol{r}$	position vector
s	spin quantum number
t	time
u	energy density
u	specific internal energy
u_B	magnetic field energy density
u_E	electric field energy density
v	vibrational quantum number
$\boldsymbol{v}$	flow velocity

v_{ph}	phase speed
$\boldsymbol{v}_{gr}$	group velocity
w_{fi}	transition rate
$\boldsymbol{A}$	directed surface area
B	rotational constant
$\boldsymbol{B}$	magnetic field (magnetic flux density)
B_ν	Planck's radiation function
C_s	sound speed
$\mathcal{D}$	viscous dissipation rate
D_e	dissociation energy
$\boldsymbol{E}$	electric field
$\boldsymbol{F}$	force
F	Helmholtz free energy
$\boldsymbol{\mathcal{F}}_c$	Coriolis force per unit volume
$\boldsymbol{\mathcal{F}}_{em}$	Lorentz force per unit volume
$\boldsymbol{\mathcal{F}}_\eta$	viscous force per unit volume
G	gravitational constant
I	electric current
I	Stokes parameter
I_e	moment of inertia
H	Hamiltonian
J	total angular quantum number
$\boldsymbol{J}$	total angular momentum
J_z	azimuthal total angular momentum
L	total orbital angular quantum number
L_ν^μ	associate Laguerre polynomial
$\boldsymbol{L}$	orbital angular momentum
L_z	azimuthal component of orbital angular momentum
$\mathcal{M}$	Mach number
M_J	total azimuthal angular quantum number
M_L	total azimuthal orbital angular quantum number
M_N	azimuthal nuclear orbital angular quantum number
M_S	total azimuthal spin angular quantum number
N	number of particles
N	nuclear orbital angular quantum number
N_z	azimuthal component of nuclear orbital angular momentum
$\boldsymbol{N}$	nuclear orbital angular momentum
$\mathcal{N}$	refractive index
$\mathcal{N}_r$	ray refractive index
P	pressure
$\boldsymbol{P}$	Poynting's vector
$\mathcal{P}$	spectral power density
P_ℓ^m	associate Legendre polynomial
Q	electric charge
Q	Stokes parameter
R	radius of curvature
R_e	equilibrium inter–nuclear distance

$R_{n\ell}$	Laguerre function
R_H	Rydberg constant
S	total spin angular quantum number
$\boldsymbol{S}$	spin angular momentum
S_z	azimuthal component of spin angular momentum
T	temperature
U	potential energy
U	Stokes parameter
V	Stokes parameter
V	volume
V_A	Alfvén speed
W	energy
Z	charge number
Z	partition function
$\mathcal{Z}$	partition function
α	fine structure constant
δ	phase shift
ϵ_0	permitivity of vacuum
η	coefficient of viscosity
κ	Boltzmann constant
λ	coefficient of heat conduction
λ	wavelength
μ	chemical potential
μ	mean molecular weight
μ	reduced mass
μ_0	permeability of vacuum
ν	frequency
ν_0	vibrational frequency
ϕ	phase
φ	azimuthal angle
ρ_e	electric charge density
ρ_m	mass density
σ	repetence, $\sigma = 1/\lambda$
σ	electric conductivity
σ	fundamental entropy
σ	Stefan–Boltzmann constant
τ	fundamental temperature
θ	polar angle
ω	angular frequency
ω_L	angular Larmor frequency
$\boldsymbol{\omega}$	vorticity
Γ	circulation
Φ	magnetic flux
$\boldsymbol{\Phi}$	rate of strain increase
Φ_c	centripetal potential
Φ_g	gravitational potential
Ψ	velocity potential

Ψ	wave function
$\boldsymbol{\Omega}$	angular velocity

Chapter 1

Electromagnetic Waves

Any study of distant astronomical objects is totally dependent on the information that may be gathered on these objects. Apart from sending spacecrafts to perform actual *in situ* measurements, four different information channels are available to us. We may collect information through the electromagnetic radiation reaching us from these distant objects. We may study cosmic rays, material particles moving with velocities close to the speed of light and originating in different parts of the universe. We may study the flux of neutrinos and anti-neutrinos, that is, near mass-less particles moving with the speed of light and created by nuclear reactions in the interior of stars. Finally, we may try to infer information from the presence of gravitational waves.

Gravitational waves have been predicted since the time of birth of the theory of general relativity. The time history of the Hulse-Taylor binary pulsar provides indirect evidence for their existence, but we will have to wait well into this century before gravitational waves possibly become an important information channel of astrophysical phenomena.

Neutrinos are produced through several nuclear reactions, for instance, the reaction

$$^1\mathrm{H} + {}^1\mathrm{H} \rightarrow {}^2\mathrm{H} + e^+ + \nu,$$

which plays a central role in the energy production in the interior of any star. The neutrino flux is difficult to detect because these particles hardly interact with matter, in fact, the neutrino flux passes through the solid Earth almost unobstructed. One successful detection scheme has been to make use of the reaction

$$^{37}\mathrm{Cl} + \nu \rightarrow {}^{37}\mathrm{Ar} + e^-$$

in large chlorine-filled tanks located in deep underground pits, well shielded from the perturbing influence of cosmic rays. An outstanding problem in astrophysics for many years was that the neutrino flux detected in this way was a factor 2 - 3 less than theoretically expected. We now have an explanation of this puzzle.

Cosmic rays have been observed over many years through the tracks left in photographic emulsions placed behind leaden shields of varying thicknesses. Cosmic rays are produced by several astrophysical processes. Supernova remnants are one of the more important sources. Cosmic ray observations provide important constraints for any understanding of processes responsible for the acceleration of these high-energy particles.

The by far most important information channel on astronomical objects is, however, still the electromagnetic radiation with frequencies ranging from radio-waves through the infrared,

visible and ultraviolet parts of the spectrum to energetic X-rays and γ-rays. Technological developments have over recent years significantly improved our ability to observe this spectrum both from ground-based observatories, but also from space-born platforms without the perturbing effects of our atmosphere. It is therefore only natural that our discussion of fundamental topics in astrophysics starts with a short review of basic parts of electromagnetic theory.

1.1 Electric and Magnetic Fields

The Maxwell equations constitute the basis for any discussion of electromagnetic phenomena. The reader is assumed to be acquainted with the physical content of these equations already. We shall therefore only briefly review some of their properties. In the simplest version these equations take the form

$$\nabla \cdot \boldsymbol{E} = \frac{\rho}{\epsilon_0} \tag{1.1}$$

$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \tag{1.2}$$

$$\nabla \cdot \boldsymbol{B} = 0 \tag{1.3}$$

$$\nabla \times \boldsymbol{B} = \mu_0 \left(\boldsymbol{j} + \epsilon_0 \frac{\partial \boldsymbol{E}}{\partial t} \right). \tag{1.4}$$

Here $\boldsymbol{E}$ is the electric field intensity and $\boldsymbol{B}$ the magnetic flux density. For convenience, $\boldsymbol{E}$ and $\boldsymbol{B}$ will be referred to as the electric and magnetic fields in the following. The electric charge and current densities ρ and $\boldsymbol{j}$ as well as the fields $\boldsymbol{E}$ and $\boldsymbol{B}$ are all functions of space $\boldsymbol{r}$ and time t. The constants ϵ_0 and μ_0 are the *permittivity* and *permeability* of vacuum. In SI-units, the latter coefficient is defined as $\mu_0 \equiv 4\pi \cdot 10^{-7}$ H/m while $\epsilon_0 = 1/(\mu_0 c^2)$ where c is the *speed of light* in vacuum, $c = 2.997925 \cdot 10^8$ m/s. Basic properties of the divergence and curl operators, div $= \nabla\cdot$ and curl $= \nabla\times$, acting on the electric and magnetic fields are reviewed in appendix A. The explicit forms of the divergence and curl operators depend on the choice of coordinate system. For Cartesian, cylindrical and spherical coordinates, as well as for an arbitrary orthonormal curvilinear coordinate system, these forms are listed in appendix A.

Maxwell equations in the form (1.1) - (1.4) are sometimes referred to as the vacuum version of the Maxwell equations. This means that *all* sources of charge and current densities are included in ρ and $\boldsymbol{j}$. In material media, it is often convenient to restrict ρ and $\boldsymbol{j}$ to the free charge and conduction current densities. Polarization charge and current densities and the magnetization current density are then included through the permittivity and permeability of the medium. We shall return to this aspect in a later chapter.

The Maxwell equations express the fact that the charge and current densities ρ and $\boldsymbol{j}$ are the sources for the electric and magnetic fields $\boldsymbol{E}$ and $\boldsymbol{B}$. *Gauss law* for electric fields (1.1) states that electric field lines originate in electric charges. This is seen by making use of the definition (A.3) of the divergence operator as applied to the electric field $\boldsymbol{E}$,

$$\nabla \cdot \boldsymbol{E} \equiv \lim_{V \to 0} \frac{1}{V} \oint_{\mathcal{A}} \mathrm{d}^2 \boldsymbol{\mathcal{A}} \cdot \boldsymbol{E}. \tag{1.5}$$

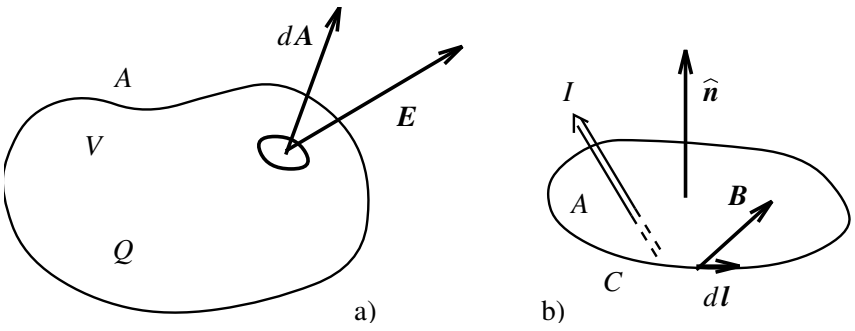

Figure 1.1: Illustrating Gauss and Ampère's laws

where $d^2\boldsymbol{\mathcal{A}}$ is an outward pointing differential element of the closed surface A bounding the infinitesimal volume V. The geometry is illustrated in figure 1.1a. According to (1.1), this means that the total *flux* of the electric field, $\oint_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}} \cdot \boldsymbol{E}$, out of any volume element V, is equal to the total amount of electric charge $Q = \int_V \rho \, \mathrm{d}^3\boldsymbol{r}$ inside that volume, except for the factor of proportionality ϵ_0^{-1},

$$\oint_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}} \cdot \boldsymbol{E} = \frac{Q}{\epsilon_0}.$$

The corresponding *Gauss law* for magnetic fields (1.3), expresses the fact that magnetic charges do not exist. This means that the magnetic field has a *solenoidal* character: a magnetic field line never ends, or equivalently, the net magnetic flux $\oint_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}} \cdot \boldsymbol{B}$ leaving any volume element V bounded by the surface $\mathcal{A}$ vanishes,

$$\oint_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}} \cdot \boldsymbol{B} = 0.$$

For slowly time-varying phenomena the last term of the *Maxwell law* (1.4) may often be neglected, $|\epsilon_0 \partial \boldsymbol{E}/\partial t| \ll |\boldsymbol{j}|$. The simplified version of the Maxwell law,

$$\nabla \times \boldsymbol{B} = \mu_0 \boldsymbol{j}, \tag{1.6}$$

is referred to as the *Ampère law*. The Ampère law may be given a simple geometric interpretation by making use of the definition (A.4) of the curl operator as applied to the magnetic field $\boldsymbol{B}$. The component of the curl of the vector field $\boldsymbol{B}$ along a direction $\hat{\boldsymbol{n}}$ is given by

$$\hat{\boldsymbol{n}} \cdot \nabla \times \boldsymbol{B} \equiv \lim_{\mathcal{A} \to 0} \frac{1}{\mathcal{A}} \oint_{\mathcal{C}} \mathrm{d}\boldsymbol{\ell} \cdot \boldsymbol{B}. \tag{1.7}$$

where the line integral on the right hand side is taken along the closed contour $\mathcal{C}$ bounding the open surface $\boldsymbol{\mathcal{A}} = \mathcal{A}\hat{\boldsymbol{n}}$. The positive direction of the contour $\mathcal{C}$ and the direction $\hat{\boldsymbol{n}}$ of the surface $\boldsymbol{\mathcal{A}}$ are related through the well-known right hand rule as indicated in figure 1.1b. Thus, the Ampère law states that the *circulation* $\oint_{\mathcal{C}} \mathrm{d}\boldsymbol{\ell} \cdot \boldsymbol{B}$ of the magnetic field around the perimeter $\mathcal{C}$ of any open surface $\mathcal{A}$, equals the total electric current $I = \int_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}} \cdot \boldsymbol{j}$ passing through that surface, except for the constant of proportionality μ_0,

$$\oint_{\mathcal{C}} \mathrm{d}\boldsymbol{\ell} \cdot \boldsymbol{B} = \mu_0 I.$$

For slowly time-varying (low-frequency) phenomena, the *Faraday law* (1.2) represents the main coupling between electric and magnetic fields. A time varying magnetic flux $\int_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}} \cdot \boldsymbol{B}$ inside a given electric circuit $\mathcal{C}$ bounding an area $\mathcal{A}$ results in an induced *electromotive force*[1] $\oint_{\mathcal{C}} \mathrm{d}\boldsymbol{\ell} \cdot \boldsymbol{E}$,

$$\oint_{\mathcal{C}} \mathrm{d}\boldsymbol{\ell} \cdot \boldsymbol{E} = -\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}} \cdot \boldsymbol{B}.$$

This in turn may drive electric currents which couple back to the magnetic field. This class of phenomena includes for instance the dynamics of stellar atmospheres, the generation of stellar and large scale inter-stellar magnetic fields. We shall return to this class of phenomena in later chapters.

In the high-frequency limit, the *displacement current*, the $\epsilon_0 \partial \boldsymbol{E}/\partial t$-term of the Maxwell law (1.4), contributes importantly to the coupling between the electric and magnetic fields. This term accounts for the existence of electromagnetic waves and therefore for how electric and magnetic fields may decouple from their proper sources ρ and $\boldsymbol{j}$ and propagate away. Before we turn to a review of basic properties of these waves, we shall need another important property of electromagnetic fields.

Quiz 1.1 : Write the explicit form of Maxwell equations (1.1) - (1.4) in Cartesian and spherical coordinates.

Quiz 1.2 : Make use of the Gauss and Stoke integral theorems, (A.40) and (A.41), to recast Maxwell equations (1.1) - (1.4) into the equivalent integral form

$$\oint_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}} \cdot \boldsymbol{E} = \frac{Q}{\epsilon_0} \tag{1.8}$$

$$\oint_{\mathcal{C}} \mathrm{d}\boldsymbol{\ell} \cdot \boldsymbol{E} = -\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}} \cdot \boldsymbol{B} \tag{1.9}$$

$$\oint_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}} \cdot \boldsymbol{B} = 0 \tag{1.10}$$

$$\oint_{\mathcal{C}} \mathrm{d}\boldsymbol{\ell} \cdot \boldsymbol{B} = \mu_0 \left(I + \epsilon_0 \frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}} \cdot \boldsymbol{E} \right). \tag{1.11}$$

Give the proper definitions of all quantities involved.

Quiz 1.3 : Show that the equation

$$\partial\rho/\partial t + \nabla \cdot \boldsymbol{j} = 0 \tag{1.12}$$

follows from Maxwell's equations. What is the corresponding integral form of this equation? Give a physical interpretation of the terms appearing in the equation.

Quiz 1.4 : Make use of the Maxwell equations in integral form to
a) show that the electric field from a point charge Q at the origin is given by

$$\boldsymbol{E}(\boldsymbol{r}) = \frac{Q}{4\pi\epsilon_0\, r^2}\, \hat{\boldsymbol{r}}$$

[1] The traditional notation is rather unfortunate, the electromotive force is no force but an induced electric voltage!

b) show that the magnetic field from a constant current I flowing along the positive z-axis is given by

$$\boldsymbol{B}(\boldsymbol{r}) = \frac{\mu_0\, I}{2\pi\, \rho}\, \hat{\boldsymbol{\phi}}$$

c) find the current I flowing in a plane circular circuit of radius a and resistance R during the time interval T in which a uniform magnetic field perpendicular to the plane of the circuit increases linearly from 0 to $\boldsymbol{B}$. Determine the direction of the current.
Interprete the symbols used in these results!

1.2 The Poynting Theorem

To create electric or magnetic fields, energy is required. This energy may be recovered when the fields are destroyed. Therefore, the fields represent "potential" energy. The electromagnetic fields are also able to transport this "potential" energy from one location to another. Explicit expressions for the energy density and the energy flux associated with the electromagnetic fields are readily derived from Maxwell equations. Scalar multiplications of (1.2) and (1.4) with $\boldsymbol{B}/\mu_0$ and $-\boldsymbol{E}/\mu_0$ and a subsequent addition of terms lead to the result

$$\nabla \cdot \frac{1}{\mu_0}(\boldsymbol{E} \times \boldsymbol{B}) = -\frac{\partial}{\partial t}(\frac{\epsilon_0}{2}\boldsymbol{E}^2 + \frac{1}{2\mu_0}\boldsymbol{B}^2) - \boldsymbol{j} \cdot \boldsymbol{E}. \tag{1.13}$$

In deriving the left hand term we made use of the vector identity (A.37). Equation (1.13) is known as the *Poynting theorem* and allows for the following interpretation. The last term $-\boldsymbol{j} \cdot \boldsymbol{E}$ is recognized as the work performed by the current source $\boldsymbol{j}$ on the fields per unit volume and time. With the geometric interpretation of the divergence operator in mind, the left hand term must represent the outflow of energy per unit volume and time, and therefore that the *Poynting vector*,

$$\boldsymbol{P} = \frac{1}{\mu_0}(\boldsymbol{E} \times \boldsymbol{B}), \tag{1.14}$$

expresses the energy flux carried by the fields. If the work done by the source $\boldsymbol{j}$ on the field $\boldsymbol{E}$ does not balance the energy outflow, the energy density in the fields must change with time. Thus, we interpret

$$u = u_E + u_B = \frac{\epsilon_0}{2}\boldsymbol{E}^2 + \frac{1}{2\mu_0}\boldsymbol{B}^2 \tag{1.15}$$

as the sum of electric and magnetic field energy densities.

Quiz 1.5 : Show for any volume V with boundary surface $\mathcal{A}$ that

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_V u\, \mathrm{d}^3\boldsymbol{r} + \oint_{\mathcal{A}} \boldsymbol{P} \cdot \mathrm{d}^2\boldsymbol{\mathcal{A}} = -\int_V \boldsymbol{j} \cdot \boldsymbol{E}\, \mathrm{d}^3\boldsymbol{r} \tag{1.16}$$

where $\boldsymbol{P}$ and u are defined by (1.14) and (1.15). Interpret the physical meaning of each term.

1.3 Electromagnetic Waves in Vacuum

The Maxwell equations may be written as a vector equation in, for instance, the electric field alone. Taking the curl of (1.2) and substituting the expression for $\nabla \times \boldsymbol{B}$ from (1.4) leads to the *wave equation* for the electric field in vacuum

$$\nabla^2 \boldsymbol{E} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\boldsymbol{E} = 0. \tag{1.17}$$

Physically allowable electric fields in vacuum must satisfy (1.17) together with the divergence condition $\nabla \cdot \boldsymbol{E} = 0$. For instance, any field of the form

$$\boldsymbol{E} = \hat{\boldsymbol{x}} f(z \pm ct)$$

where f is an arbitrary, twice differentiable function will be allowed.

In the following we shall often be satisfied with a simpler class of solutions. Direct substitution will show that the electromagnetic fields

$$\begin{aligned} \boldsymbol{E}(\boldsymbol{r}, t) &= \boldsymbol{E}_0 \cos(\boldsymbol{k} \cdot \boldsymbol{r} - \omega t) \\ \boldsymbol{B}(\boldsymbol{r}, t) &= \boldsymbol{B}_0 \cos(\boldsymbol{k} \cdot \boldsymbol{r} - \omega t) \end{aligned} \tag{1.18}$$

constitute a solution of Maxwell equations (1.1) -(1.4) in a source-free region of space, $\rho = \boldsymbol{j} = 0$, provided the conditions

$$\omega = kc \tag{1.19}$$

and

$$\boldsymbol{k} \times \boldsymbol{E}_0 = \omega \boldsymbol{B}_0 \quad \text{and} \quad \boldsymbol{k} \cdot \boldsymbol{E}_0 = 0 \tag{1.20}$$

are satisfied. Here $\boldsymbol{k}$ and ω are the *wave vector* and *angular frequency* of the wave, and $\boldsymbol{E}_0$ and $\boldsymbol{B}_0$ are the amplitude vectors of the electric and magnetic fields. The *dispersion relation* (1.19) shows that a definite relationship exists between the frequency of the wave and the *wave number* $k = |\boldsymbol{k}|$. The condition (1.20) means that the three vectors $\boldsymbol{E}$, $\boldsymbol{B}$ and $\boldsymbol{k}$ constitute a right-handed orthogonal set of vectors.

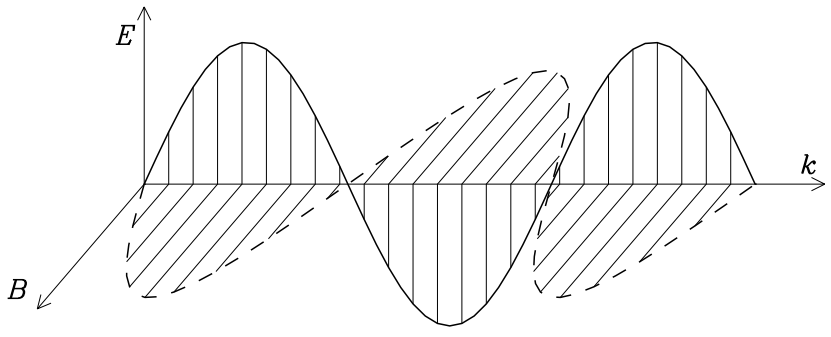

Figure 1.2: Field distribution in a plane wave

In figure 1.2 the distribution of electric and magnetic fields in the harmonic wave (1.18) along the propagation direction $\hat{\boldsymbol{k}}$ at a fixed time t is given. The argument

$$\phi = \boldsymbol{k} \cdot \boldsymbol{r} - \omega t \tag{1.21}$$

in (1.18) is the *phase* of the wave. The wave is called a *plane wave* because the phase at any given time t is constant in the planes $\boldsymbol{k}\cdot\boldsymbol{r} = \text{constant}$.[2] To remain at a constant phase of the wave, an observer must move with the *phase velocity* $\boldsymbol{v}_{ph} = v_{ph}\hat{\boldsymbol{k}}$ in the direction of the wave vector $\boldsymbol{k}$ and magnitude given by the *phase speed*

$$v_{ph} = \omega/k. \tag{1.22}$$

With the dispersion relation (1.19), the phase speed is equal to the speed of light c in vacuum.

The energy density and energy flux in the wave, as given by (1.15) and (1.14), oscillate in space and time. Physically more meaningful quantities are often found by forming *time averages* $\overline{f}$ of the relevant quantities $f(t)$,

$$\overline{f} \equiv \frac{1}{T}\int_0^T f(t)\,\mathrm{d}t,$$

where T represents the period of the time varying quantity. For the time averaged energy density of the plane wave (1.18) we find

$$\overline{u} = \left(\frac{\epsilon_0}{2}\boldsymbol{E}_0^2 + \frac{1}{2\mu_0}\boldsymbol{B}_0^2\right)\overline{\cos^2\phi} = \frac{\epsilon_0}{4}\boldsymbol{E}_0^2 + \frac{1}{4\mu_0}\boldsymbol{B}_0^2 = \frac{\epsilon_0}{2}\boldsymbol{E}_0^2. \tag{1.23}$$

In (1.23) we made use of the conditions (1.19) and (1.20). We note that the electric and the magnetic fields in the wave contribute equally to the time averaged energy density in the wave. The time averaged energy flux in the wave is

$$\overline{\boldsymbol{P}} = \frac{1}{\mu_0}\boldsymbol{E}_0 \times \boldsymbol{B}_0\overline{\cos^2\phi} = \frac{\epsilon_0}{2}\boldsymbol{E}_0^2 c\,\hat{\boldsymbol{k}} = \overline{u}c\,\hat{\boldsymbol{k}}. \tag{1.24}$$

This means that the electromagnetic wave transports its average energy density $\overline{u}$ with the speed of light c in the direction of the wave vector $\boldsymbol{k}$, that is, the phase velocity and the energy propagation velocity of the wave are identical.

The above conclusions are valid for electromagnetic waves in vacuum. For electromagnetic waves in material media these conclusions will be modified. The phase and energy propagation velocities, the latter represented by the *group velocity*

$$\boldsymbol{v}_g = \frac{\partial\omega(\boldsymbol{k})}{\partial\boldsymbol{k}}, \tag{1.25}$$

will generally differ not only in magnitude, but sometimes also in direction.

1.3.1 Complex field notation

Physically acceptable electric and magnetic fields are both real quantities. However, since the Maxwell equations are linear equations in $\boldsymbol{E}$ and $\boldsymbol{B}$ with real coefficients it is often mathematically convenient to introduce complex notation for these fields. In this notation the wave solution (1.18) is written

$$\begin{aligned}\boldsymbol{E}(\boldsymbol{r},t) &= \boldsymbol{E}_0\exp(\imath\boldsymbol{k}\cdot\boldsymbol{r} - \imath\omega t)\\ \boldsymbol{B}(\boldsymbol{r},t) &= \boldsymbol{B}_0\exp(\imath\boldsymbol{k}\cdot\boldsymbol{r} - \imath\omega t),\end{aligned} \tag{1.26}$$

[2] We may also find wave solutions with constant phase surfaces in the form of cylinders or spheres. These waves are correspondingly called *cylindrical* or *spherical waves*.

where $\iota = \sqrt{-1}$ is the imaginary unit and $\boldsymbol{E}_0$, $\boldsymbol{B}_0$, $\boldsymbol{k}$ and ω have to satisfy conditions (1.19) and (1.20). Due to the de Moivre identity

$$\exp(\iota\phi) = \cos\phi + \iota \sin\phi, \tag{1.27}$$

the physically relevant fields are recovered by taking the real part (or alternatively the imaginary part) of the corresponding complex fields.

A convenient feature of the complex notation – and in fact the main reasons for its introduction – is that taking derivatives with respect to $\boldsymbol{r}$ and t of the complex exponential function in (1.26) is seen to be equivalent to simple algebraic operations,

$$\nabla \to \iota\boldsymbol{k} \quad \text{and} \quad \frac{\partial}{\partial t} \to -\iota\omega \tag{1.28}$$

that is, any ∇-operator may be replaced with the vector $\iota\boldsymbol{k}$, any time derivative $\partial/\partial t$ with the scalar factor $-\iota\omega$. We shall make repeated use of complex fields in the following.

The electric field amplitude vector $\boldsymbol{E}_0$ in the complex notation (1.26) represents not only the (real) amplitude E_0 of the wave but also the (complex) polarization vector $\hat{\boldsymbol{\epsilon}}$ of the wave,

$$\boldsymbol{E}_0 = E_0\hat{\boldsymbol{\epsilon}}. \tag{1.29}$$

The polarization vector $\hat{\boldsymbol{\epsilon}}$ has unit length, $\hat{\boldsymbol{\epsilon}} \cdot \hat{\boldsymbol{\epsilon}}^* = 1$. We shall return to a discussion of the physical importance of the polarization vector in section 1.5.

Useful as the complex field notation may be, it still has its limitations. The complex fields *cannot* be used directly when calculating non-linear field quantities. It is necessary first to find the real (physical) part of the fields before substitution into the non-linear expression. One notable exception to this rule, however, exists. *Time averaged quadratic field quantities* may be calculated as one half the real part of the product of the two complex fields with one of the field factors complex conjugated. Thus, the time averaged energy density and energy flux in the electromagnetic wave are calculated in terms of the complex fields $\boldsymbol{E}$ and $\boldsymbol{B}$ as

$$\overline{u} = \frac{1}{2}\,\mathrm{Re}\left(\frac{\epsilon_0}{2}\boldsymbol{E}\cdot\boldsymbol{E}^* + \frac{1}{2\mu_0}\boldsymbol{B}\cdot\boldsymbol{B}^*\right) = \frac{\epsilon_0}{4}\,|\,\boldsymbol{E}_0\,|^2 + \frac{1}{4\mu_0}\,|\,\boldsymbol{B}_0\,|^2 = \frac{\epsilon_0}{2}\,|\,\boldsymbol{E}_0\,|^2 \tag{1.30}$$

and

$$\overline{\boldsymbol{P}} = \frac{1}{2}\,\mathrm{Re}\left(\frac{1}{\mu_0}\boldsymbol{E}\times\boldsymbol{B}^*\right) = \frac{\epsilon_0}{2}\,|\,\boldsymbol{E}_0\,|^2\, c\,\hat{\boldsymbol{k}}. \tag{1.31}$$

These results are seen to agree with our previous results (1.23) and (1.24).

Quiz 1.6 : Define wavelength λ, period T and frequency ν in terms of wave number k and angular frequency ω. Rewrite the dispersion relation (1.19) in terms of λ and ν.

Quiz 1.7 : Convince yourself of the correctness of the *superposition theorem*: A sum of two independent solutions of the Maxwell equations is also a valid solution. Explain the mathematical reasons for this result.

Quiz 1.8 : Verify that

$$\overline{\cos^2(\boldsymbol{k}\cdot\boldsymbol{r} - \omega t)} = \frac{\omega}{2\pi}\int_0^{2\pi/\omega} \cos^2(\boldsymbol{k}\cdot\boldsymbol{r} - \omega t)\,\mathrm{d}t = \frac{1}{2}$$

and therefore that (1.23) and (1.24) are correct.

Quiz 1.9 : Show that

$$\begin{aligned} \boldsymbol{E}(\boldsymbol{r},t) &= c\,\hat{\boldsymbol{\theta}}\cos(kr-\omega t)/r \\ \boldsymbol{B}(\boldsymbol{r},t) &= \hat{\boldsymbol{\varphi}}\cos(kr-\omega t)/r \end{aligned} \tag{1.32}$$

is an asymptotic solution of (1.1) - (1.4) for large r, provided $\rho = \boldsymbol{j} = 0$ and the dispersion relation (1.19) is satisfied. Here $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\varphi}}$ together with $\hat{\boldsymbol{r}}$ are unit vectors in a spherical coordinate system. How would you classify this solution?

[*Hint*: Asymptotic means that the expressions (1.32) satisfy (1.1) - (1.4) to any given degree of relative accuracy for large enough values of r.]

Quiz 1.10 : Two plane waves with identical amplitude vectors, but with different wave numbers $k = k_0 \pm \Delta k$, $|\,\Delta k\,| \ll k_0$, both propagate along $\hat{\boldsymbol{k}}_0$. Make use of identities for trigonometric functions to show that the resulting wave field consists of a rapidly varying harmonic part with phase function $\phi_0 = \boldsymbol{k}_0 \cdot \boldsymbol{r} - \omega_0 t$ multiplied with a slowly varying envelope

$$\cos(\Delta\boldsymbol{k}\cdot\boldsymbol{r} - \Delta\omega\, t).$$

Verify that the envelope (and therefore that also the energy carried by the wave) moves with the group velocity

$$\boldsymbol{v}_g = \frac{\partial\omega(\boldsymbol{k})}{\partial\boldsymbol{k}}\,|_{\boldsymbol{k}=\boldsymbol{k}_0}\,.$$

Quiz 1.11 : Write the spherical wave solution (1.32) in complex notation.

Quiz 1.12 : Prove that

$$\overline{f_{\rm phys}(t)g_{\rm phys}(t)} = \frac{1}{2}\,{\rm Re}\left(f_{\rm cmpl}(t)g^*_{\rm cmpl}(t)\right)$$

where $f_{\rm cmpl}(t) \sim \exp(-\iota\omega t)$ and $f_{\rm phys}(t) = {\rm Re}\,(f_{\rm cmpl}(t))$ and similarly for g. What is the corresponding value for $\overline{f_{\rm cmpl}(t)g_{\rm cmpl}(t)}$?

1.4 The Electromagnetic Spectrum

Electromagnetic waves of astrophysical interest extend from radio waves through the infrared, visible and ultraviolet spectral ranges to X-rays and hard γ-rays. The frequencies and wavelengths corresponding to these classifications are illustrated in figure 1.3.

The Earths ionosphere represents a limiting factor at low frequencies. Only waves with frequencies exceeding a minimum value depending on the maximum value of the electron density in the ionosphere and the angle of incidence are able to penetrate. This minimum frequency is normally in the 5 - 10 MHz range. Thus, the so-called kilo-metric radiation, $\lambda \sim 1$ km, originating in the Jovian magnetosphere was first found when satellites made observations outside the ionosphere possible. Similarly, atmospheric gases are responsible for absorption bands at different frequencies. H_2O-vapor and CO_2 give rise to absorption bands in the infrared range. The ozone-layer is an effective absorbing agent in the ultraviolet.

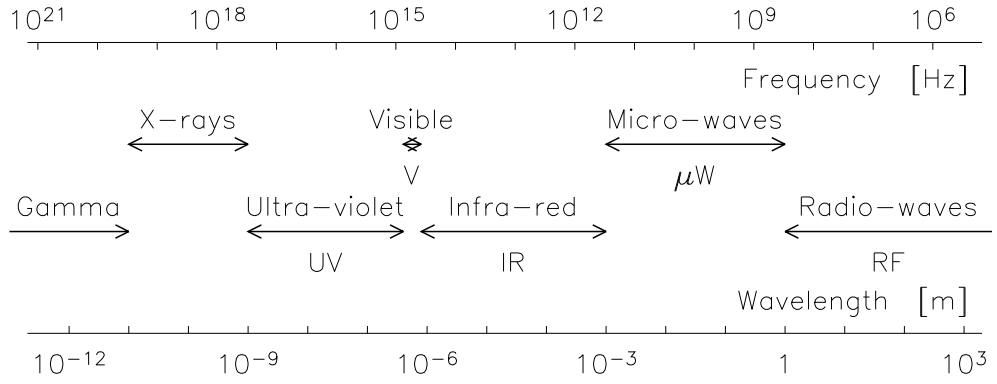

Figure 1.3: The electromagnetic spectrum

1.5 Wave Polarization

At a given position $\boldsymbol{r}$, the electric field vector of (1.18) oscillates along the fixed $\boldsymbol{E}_0$-direction as a function of time. The wave is therefore said to be *linearly* or *plane polarized*. In the complex notation (1.26) this case corresponds to a real polarization vector, for instance, $\hat{\boldsymbol{\epsilon}} = \hat{\boldsymbol{x}}$ for $\hat{\boldsymbol{k}} = \hat{\boldsymbol{z}}$. This is, however, not the most general case. If the polarization vector is chosen as $\hat{\boldsymbol{\epsilon}} = (\hat{\boldsymbol{x}} \pm \imath\hat{\boldsymbol{y}})/\sqrt{2}$ for $\hat{\boldsymbol{k}} = \hat{\boldsymbol{z}}$, it is easily seen that the tip of the (physical) electric field vector at a given position will perform a circular motion as a function of time.

In the more general plane wave solution the (physical) electric field can be written in the form

$$\boldsymbol{E}(\boldsymbol{r}, t) = E_0(\hat{\boldsymbol{p}} \cos\beta \cos\phi + \hat{\boldsymbol{q}} \sin\beta \sin\phi) \tag{1.33}$$

where β is a constant, $\beta \in [-\pi/4, \pi/4]$, where the constant unit vectors $\hat{\boldsymbol{p}}$ and $\hat{\boldsymbol{q}}$ satisfy the requirement

$$\hat{\boldsymbol{p}} \times \hat{\boldsymbol{q}} = \hat{\boldsymbol{k}},$$

and where the phase function ϕ is given by (1.21). With β in the specified range we have $\cos\beta \geq |\sin\beta|$. In this case the tip of the electric field vector will at any given position trace an ellipse as a function of time with semi-major and semi-minor axis, $a = E_0 \cos\beta$ and $b = E_0 |\sin\beta|$ oriented along $\hat{\boldsymbol{p}}$ and $\hat{\boldsymbol{q}}$, respectively. The geometry is illustrated in figure 1.4.

The *ellipticity* of the ellipse is given by $\tan\beta$. With $\tan\beta = 0, \pm\infty$ the ellipse reduces to a line and the linearly polarized wave is recovered. For $|\tan\beta| = 1$ the ellipse reduces to a circle. The wave is then *circularly polarized*. For any other value of β the wave is *elliptically polarized*. If $\beta < 0$ the angle χ in figure 1.4 increases with time. The tip of the electric field vector therefore traces the ellipse (circle) in a clockwise direction looking *along* the wave vector $\boldsymbol{k}$, or alternatively, with the right-hand thumb pointing along $\boldsymbol{k}$, in the direction of the other fingers. The wave is then said to be *right-handed* elliptically (circularly) polarized. With the opposite sign for β, $\beta > 0$, the wave is *left-handed* polarized.[3]

[3]Be aware that in the literature left/right handed polarization is sometimes defined as looking along $-\boldsymbol{k}$, that is, looking toward the infalling radiation! Note also that the connection between the sign of β and right- or left-handed polarization assumes the phase function to be defined as $\phi = \boldsymbol{k} \cdot \boldsymbol{r} - \omega t$. Different definitions of ϕ are in use in the litterature.

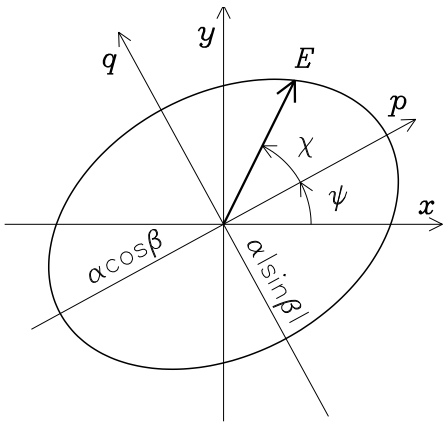

Figure 1.4: Polarization ellipse

The polarization of an electromagnetic wave is often conveniently described in terms of the *Stoke parameters*. With the z-axis along the wave vector $\boldsymbol{k}$ and for an arbitrarily chosen set of orthogonal x- and y-axis, these parameters are defined by

$$\begin{aligned} I &= 2\,\overline{(E_x^2(t) + E_y^2(t))} \\ Q &= 2\,\overline{(E_x^2(t) - E_y^2(t))} \\ U &= 4\,\overline{E_x(t)E_y(t)} \\ V &= 4\,\overline{E_x(t)E_y(t - \pi/2\omega)}. \end{aligned} \tag{1.34}$$

The over-bar in (1.34) again represents time average. In the last expression a time delay corresponding to a phase shift of $\pi/2$ is introduced in the y-component before the averaging is performed. With the electric field given in terms of (1.33), the Stoke parameters reduce to

$$\begin{aligned} I &= E_0^2 \\ Q &= E_0^2 \cos 2\beta \cos 2\psi \\ U &= E_0^2 \cos 2\beta \sin 2\psi \\ V &= E_0^2 \sin 2\beta \end{aligned} \tag{1.35}$$

and therefore also

$$I^2 = Q^2 + U^2 + V^2. \tag{1.36}$$

The Stoke parameters all have the same physical dimension. For an electromagnetic wave in vacuum, I is proportional to the energy density or the energy flux associated with the wave. We notice that the ratio of the two principal axis of the polarization ellipse, the ellipticity $\tan\beta$, may be determined from the ratio of V and I,

$$\sin 2\beta = \frac{V}{I},$$

while the orientation ψ of the major axis of the polarization ellipse with respect to the chosen x- and y-axis follows from

$$\tan 2\psi = \frac{U}{Q}.$$

Quiz 1.13 : Verify that (1.26) with $\hat{\boldsymbol{k}} = \hat{\boldsymbol{z}}$ and $\hat{\boldsymbol{\epsilon}} = (\hat{\boldsymbol{x}} \pm \imath\hat{\boldsymbol{y}})/\sqrt{2}$ represents circularly polarized waves.

Quiz 1.14 : From the Stoke parameters, how do you determine if the arriving wave is left- or right-handed polarized?

Quiz 1.15 : The Stoke parameters I, Q, U and V refer to a particular choice of orientation (x, y) of the measuring instrument. We may indicate this by marking the Stoke parameters with this particular orientation, $I_{x,y}$ and so on. Show that if the measuring instrument is rotated an angle $\pi/4$ to the new orientation (ξ, η) then

$$I_{\xi,\eta} = I_{x,y}, \quad V_{\xi,\eta} = V_{x,y}, \quad Q_{\xi,\eta} = U_{x,y} \quad \text{and} \quad U_{\xi,\eta} = -Q_{x,y}.$$

Quiz 1.16 : Find the corresponding magnetic field $\boldsymbol{B}$ that will make (1.33) part of a solution of Maxwell equations (1.1) - (1.4) in vacuum. What relation must exist between ω and $\boldsymbol{k}$? What is the amplitude of the magnetic field in the wave? Are you able to classify this wave as plane, cylindrical or spherical? How would you characterize the polarization of the wave? Discuss the following statement: "The superposition of two plane polarized plane waves with identical wave vectors $\boldsymbol{k}$ generally results in an elliptically polarized plane wave".

Quiz 1.17 : With the z-axis along $\boldsymbol{k}$ and given orthogonal x- and y-axis (see figure 1.4) show that the E_x and E_y components of the electric field (1.33) can be written

$$E_x = E_{0x}\cos(\phi + \delta_x)$$
$$E_y = E_{0y}\cos(\phi + \delta_y).$$

Show that the amplitudes E_{0x} and E_{0y} and the relative phase shift $\delta = \delta_y - \delta_x$ may be expressed uniquely in terms of β and the angle ψ as

$$E_{0x}^2 = E_0^2(\cos^2\beta\cos^2\psi + \sin^2\beta\sin^2\psi)$$
$$E_{0y}^2 = E_0^2(\cos^2\beta\sin^2\psi + \sin^2\beta\cos^2\psi)$$
$$\tan\delta = -\tan 2\beta/\sin 2\psi.$$

What are the Stoke parameters I, Q, U and V in terms of E_{0x}, E_{0y} and δ?

1.6 Non-Monochromatic Waves

The solutions of Maxwell equations studied in the previous sections were all examples of monochromatic waves. To produce a strictly monochromatic wave, rather stringent conditions have to be met by the wave source. In particular, the source will have to maintain a strictly harmonic character over (ideally) infinite times. The two best known examples of sources approximately satisfying this condition are the non-modulated electronic oscillator and the laser. In this section we will consider three different aspects related to the non-monochromatic character of electromagnetic radiation: finite wave coherence, partial polarization and power spectra.

1.6.1 Wave coherence

A monochromatic wave may be characterized as having infinite phase memory, the wave phase ϕ varying strictly linearly with time. Naturally occurring electromagnetic radiation have finite phase memory. After a *coherence time* τ, the wave tends to forget its previous phase. As an example we will here visualize the wave as consisting of a superposition of finite time segments, each of duration τ during which the electric and magnetic fields vary according to (1.18), but where the phase experiences a random jump between each successive segment. This is illustrated in figure 1.5. The segment length is the *coherence length* $\ell = c\tau$.

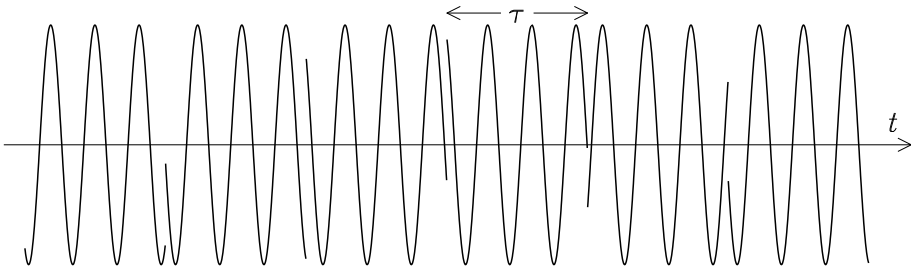

Figure 1.5: Wave with finite coherence time

At the same time any harmonically oscillating time segment of duration τ, for instance,

$$E(t) = \begin{cases} \cos(\omega_0 t) & -\frac{\tau}{2} \le t < \frac{\tau}{2} \\ \\ 0 & \text{otherwise,} \end{cases} \tag{1.37}$$

may be represented as a superposition of harmonic oscillations of different frequencies. The classical theory of *Fourier transforms* tells that any square integrable function $f(t)$, that is a function satisfying $\int_{\infty}^{\infty} |\, f(t)\,|^2 \, \mathrm{d}t < \infty$, may be written in the form

$$f(t) = \int_{-\infty}^{\infty} \tilde{f}(\omega) \exp(-\iota\omega t) \, \mathrm{d}\omega \tag{1.38}$$

where the complex amplitude $\tilde{f}(\omega)$ is given by

$$\tilde{f}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) \exp(\iota\omega t) \, \mathrm{d}t. \tag{1.39}$$

The two relations (1.38) and (1.39) form a *Fourier transform pair*. A summary of some important properties of the Fourier transform[4] is given in table 1.1. Fourier transforms are also valid for generalized functions, a class of functions that are not square integrable. Two important examples are (1.44) and (1.45).

[4]Different definitions of the Fourier transform are found in the literature. One common alternative to the transform pair (1.40)-(1.41) is

$$\tilde{f}(\nu) = \int_{-\infty}^{\infty} f(t) \exp(2\pi\iota\nu t) \, \mathrm{d}t \quad \text{and therefore} \quad f(t) = \int_{-\infty}^{\infty} \tilde{f}(\nu) \exp(-2\pi\iota\nu t) \, \mathrm{d}\nu.$$

With this definition the factors 2π in the Fourier transform of the convolution product and in the Parseval theorem in (1.46) and (1.47) will be absent.

For reference purposes we also note that if $f(t)$ is any periodic function, $f(t+T) = f(t)$, then it may be expanded in a *Fourier series*, consisting of harmonically varying terms in the fundamental angular frequency $2\pi/T$ and its harmonics. A summary of important properties of Fourier series is given in table 1.2.

The Fourier transform and its inverse are defined as

$$\tilde{f}(\omega) = \mathcal{F}[f(t)] \equiv \frac{1}{2\pi}\int_{-\infty}^{\infty} f(t)\exp(\iota\omega t)\,\mathrm{d}t \tag{1.40}$$

$$f(t) = \mathcal{F}^{-1}[\tilde{f}(\omega)] = \int_{-\infty}^{\infty} \tilde{f}(\omega)\exp(-\iota\omega t)\,\mathrm{d}\omega. \tag{1.41}$$

If $f(t)$ is real, then $\tilde{f}(-\omega) = \tilde{f}^*(\omega)$.

The following relations hold

$$\mathcal{F}[\frac{\mathrm{d}}{\mathrm{d}t}f(t)] = -\iota\omega\tilde{f}(\omega) \tag{1.42}$$

$$\mathcal{F}[f(at+b)] = \frac{1}{|a|}\exp(-\iota\frac{\omega}{a}b)\tilde{f}\left(\frac{\omega}{a}\right). \tag{1.43}$$

Some generalized Fourier transform pairs are

$$\mathcal{F}[1] = \delta(\omega) \tag{1.44}$$

$$\mathcal{F}[\mathrm{sign}(t)] = -(\iota\pi\omega)^{-1}. \tag{1.45}$$

For the convolution product $h(t) = \int f(\tau)g(t-\tau)\,\mathrm{d}\tau \equiv f(t)\otimes g(t)$

$$\tilde{h}(\omega) = 2\pi\tilde{f}(\omega)\tilde{g}(\omega). \tag{1.46}$$

Parseval theorem: If $f(t)$ and $g(t)$ are real, $\tilde{f} = \mathcal{F}[f]$ and $\tilde{g} = \mathcal{F}[g]$, then

$$\int_{-\infty}^{\infty} f(t)g(t)\,\mathrm{d}t = 2\pi\int_{-\infty}^{\infty}\tilde{f}^*(\omega)\tilde{g}(\omega)\,\mathrm{d}\omega = 4\pi\int_0^{\infty}\mathrm{Re}(\tilde{f}^*(\omega)\tilde{g}(\omega))\,\mathrm{d}\omega. \tag{1.47}$$

Table 1.1: The Fourier transform

After this intermezzo on mathematical methods let us return to our discussion on wave coherence. Substituting the expression (1.37) for $E(t)$ with

$$\cos\omega_0 t = \frac{1}{2}(\exp(\iota\omega_0 t) + \exp(-\iota\omega_0 t))$$

into (1.39) and performing the integral we find

$$\tilde{E}(\omega) = \frac{\tau}{4\pi}\left(\mathrm{sinc}\,(\omega+\omega_0)\frac{\tau}{2} + \mathrm{sinc}\,(\omega-\omega_0)\frac{\tau}{2}\right) \tag{1.52}$$

Let $f(t)$ be a periodic function, $f(t+T) = f(t)$.
The function $f(t)$ may be expanded in the trigonometric series

$$f(t) = \sum_{n=-\infty}^{\infty} \tilde{f}_n \exp(-\iota \frac{2\pi n}{T} t) \tag{1.48}$$

with

$$\tilde{f}_n = \frac{1}{T} \int_{-T/2}^{T/2} f(t) \exp(\iota \frac{2\pi n}{T} t)\, \mathrm{d}t \tag{1.49}$$

If $f(t)$ is a real function, then $\tilde{f}_{-n} = \tilde{f}_n^*$.
If $g(t) = f(t+b)$ then

$$\tilde{g}_n = \exp(-\iota \frac{2\pi n}{T} b)\, \tilde{f}_n. \tag{1.50}$$

Parseval theorem: For real periodic functions $f(t)$ and $g(t)$, both of period T and with $\overline{f}\overline{g} = 0$, then

$$\frac{1}{T} \int_{-T/2}^{T/2} f(t) g(t)\, \mathrm{d}t = \sum_{n=-\infty}^{\infty} \tilde{f}_n^* \tilde{g}_n = \int \mathrm{d}\omega \sum_{n=1}^{\infty} 2\,\mathrm{Re}(\tilde{f}_n^* \tilde{g}_n)\, \delta(\omega - \frac{2\pi n}{T}). \tag{1.51}$$

Table 1.2: The Fourier series

where the sinc-function is defined as

$$\mathrm{sinc}\, x \equiv \frac{\sin x}{x}. \tag{1.53}$$

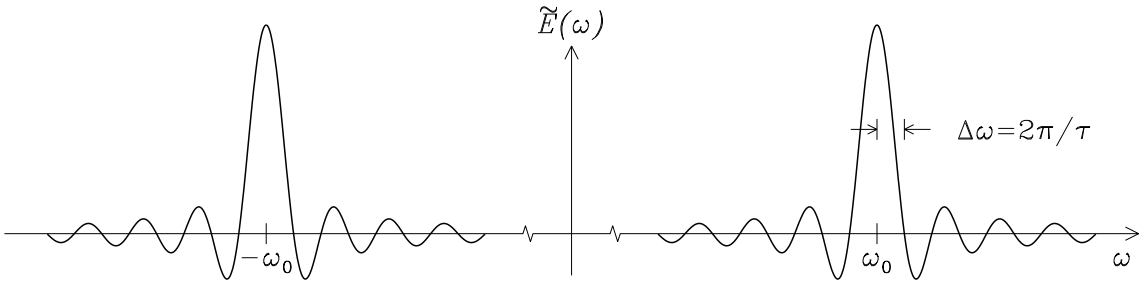

Figure 1.6: Amplitude distribution for a finite harmonic segment

The result (1.52) is plotted in figure 1.6 for a case where $\omega_0 \tau \gg 1$, that is, where there is negligible "cross-talk" between the two sinc-functions. The two wings of the plot have typical half-widths (from line center to first zero-crossing)

$$\Delta\omega = \frac{2\pi}{\tau}. \tag{1.54}$$

According to (1.38) the finite harmonic segment may be considered made up from a superposition of (an infinite number of) harmonic oscillations with frequencies within an interval

with typical half-width $\Delta\omega$ varying inversely with the duration τ.

Formula (1.52) only takes into account the contribution to the complex amplitude $\tilde{E}(\omega)$ from the time segment $-\tau/2 \leq t < \tau/2$. Adjacent time segments will give similar contributions. In fact, if a constant phase δ is added to the argument of the cos-function in (1.37), the two terms in (1.52) get additional complex conjugate phase factors $\exp(\pm\imath\delta)$ and such that $\tilde{E}(-\omega) = \tilde{E}^*(\omega)$ in accordance with table 1.1. The contribution from the different time segments will contain the identical sinc functions multiplied with random phase factors of the form $\exp(\pm\imath\delta)$. The frequency dependence of the amplitude function $\tilde{E}(\omega)$ will, therefore, be identical for a single time segment and a sequence of such segments. We may conclude that any wave with finite coherence time τ can be considered as a non-monochromatic wave, with a characteristic angular frequency half-width $\Delta\omega$ given by (1.54).

For a simple estimation of the coherence time τ of a given radiation field, Young's two-slit interference experiment may be used (see quiz 1.21). The experiment is illustrated schematically in figure 1.7. A plane wave, illustrated by selected phase fronts, is falling perpendicularly on a screen with two parallel slits, each of width $2b$ and a distance d apart. The resulting interference pattern is focused by a lens L on a second screen S one focal length f away. For an ideal mono-chromatic incident wave a series of interference intensity maxima will be observed on the screen. The intensity of the n.th interference fringe, at a location on the screen corresponding to an optical path difference $n\lambda$ for individual parallel rays emerging from the two slits, $\sin\theta = n\lambda/d$, will be determined by the diffraction pattern for each slit, that is, $I \sim [\sin(kb\sin\theta)/kb\sin\theta]^2$ (dash-dotted curve in the figure). For an incident wave with finite coherence time τ, however, the internal coherence between the rays emanating from the two slits will be lost as soon as $n\lambda \approx c\tau$. The effect of constructive interference will then be reduced and the intensity of the interference fringes thus fall below the diffraction limit. A count of the number of visible fringes is thus a direct measure of the time that the radiation field "remembers" its own phase, that is, the coherence time τ of the radiation field.

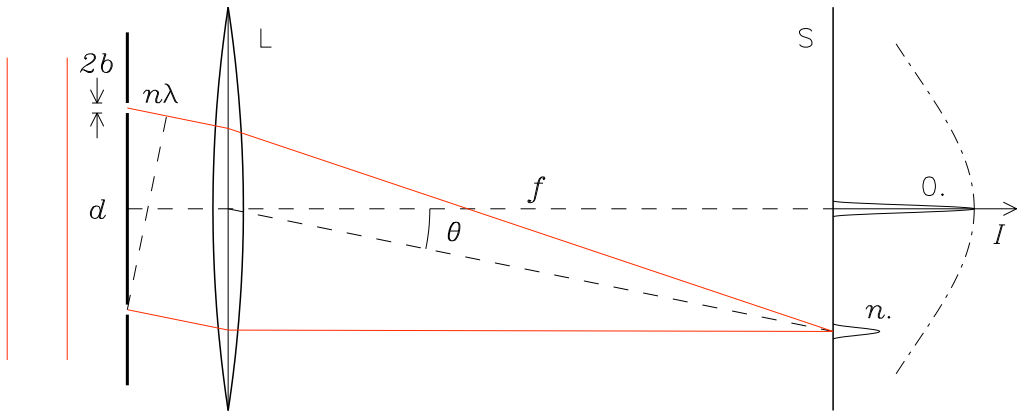

Figure 1.7: Young's two slit interference experiment

1.6.2 Partial polarization

Apart from the finite phase memory discussed above, also the polarization state of the wave may change as a function of time. For an idealized model we again consider a wave consisting

of a sequence of harmonic segments as depicted in figure 1.5. This time we allow both phase and polarization state to change from one segment to the next. The situation is conveniently analyzed in terms of the Stoke parameters (1.35). The time averaging in (1.35) will now have to be extended over many coherence times over which the parameters β and ψ take different values for adjacent segments. For simplicity we assume E_0 to remain constant. The result is then

$$\begin{aligned} I &= E_0^2 \\ Q &= E_0^2\ \overline{\cos 2\beta\ \cos 2\psi} \\ U &= E_0^2\ \overline{\cos 2\beta\ \sin 2\psi} \\ V &= E_0^2\ \overline{\sin 2\beta}\,, \end{aligned} \tag{1.55}$$

where the time averaging is taken over a time interval $N\tau$ with N large. Making use of the Schwarz inequality for integrals,

$$\left(\int_a^b f(t)g(t)\,\mathrm{d}t\right)^2 \le \int_a^b f^2(t)\,\mathrm{d}t \int_a^b g^2(t)\,\mathrm{d}t, \tag{1.56}$$

it will now be apparent that for instance

$$U^2 = E_0^4\ \overline{\cos 2\beta\ \sin 2\psi}^2 \le E_0^4\ \overline{\cos^2 2\beta}\ \overline{\sin^2 2\psi}\,,$$

and therefore that

$$I^2 \ge Q^2 + U^2 + V^2. \tag{1.57}$$

The *degree of polarization* p, defined by

$$p \equiv \frac{\sqrt{Q^2+U^2+V^2}}{I}, \tag{1.58}$$

describes the degree to which the wave remembers its own polarization. With $p = 1$ the wave is *completely polarized*, with $p = 0$ the wave has *random polarization*. In the first case the polarization is constant between successive segments. In the second case the wave chooses its polarization in each segment completely at random.

Two waves with finite coherence times, propagating in the same direction with electric fields $\boldsymbol{E}^{(1)}$ and $\boldsymbol{E}^{(2)}$ are said to be *incoherent* if the time averaged product of any two components of the two fields vanishes,

$$\overline{E_x^{(1)}(t)E_y^{(2)}(t)} = 0, \cdots .$$

The Stoke parameters for incoherent waves are additive. That is, if we define the Stoke parameters for the two waves $(I^{(1)}, Q^{(1)}, U^{(1)}, V^{(1)})$ and $(I^{(2)}, Q^{(2)}, U^{(2)}, V^{(2)})$, then the corresponding parameters for the resultant wave is

$$(I, Q, U, V) = (I^{(1)}, Q^{(1)}, U^{(1)}, V^{(1)}) + (I^{(2)}, Q^{(2)}, U^{(2)}, V^{(2)}). \tag{1.59}$$

This means that any *partially polarized* wave, $0 < p < 1$, may be considered as a superposition of two incoherent waves, one completely polarized and one with random polarization,

$$I^{(1)} = \sqrt{Q^2+U^2+V^2}, \quad Q^{(1)} = Q, \quad U^{(1)} = U, \quad V^{(1)} = V$$

and

$$I^{(2)} = I - I^{(1)}, \quad Q^{(2)} = U^{(2)} = V^{(2)} = 0.$$

The degree of polarization of the radiation arriving from the universe is important for the interpretation of the physical conditions responsible for the emitted wave. Our knowledge of stellar, interstellar and galactic magnetic fields and conditions in interstellar dust clouds depends to a large extent on this type of information.

1.6.3 Spatial-temporal Fourier representation

The Fourier formalism summarized in table 1.1 is trivially extended to functions varying in both space and time. Thus, any square integrable spatial function $f(\boldsymbol{r})$ may be expressed in terms of its spatial Fourier amplitude $\tilde{f}(\boldsymbol{k})$ as

$$f(\boldsymbol{r}) = \int \tilde{f}(\boldsymbol{k}) \exp(\imath \boldsymbol{k} \cdot \boldsymbol{r}) \, \mathrm{d}^3\boldsymbol{k} \tag{1.60}$$

$$\tilde{f}(\boldsymbol{k}) = \frac{1}{(2\pi)^3} \int f(\boldsymbol{r}) \exp(-\imath \boldsymbol{k} \cdot \boldsymbol{r}) \, \mathrm{d}^3\boldsymbol{r}. \tag{1.61}$$

The integrals over $\boldsymbol{r}$ and $\boldsymbol{k}$ here extend over the entire $\boldsymbol{r}$ and $\boldsymbol{k}$ space. Note that we for later convenience chose the opposite sign convention in (1.60)-(1.61) compared to (1.40)-(1.41). Substituting expressions for the electromagnetic fields $\boldsymbol{E}$ and $\boldsymbol{B}$ and the corresponding source densities $\boldsymbol{j}$ and ρ in terms of their spatial-temporal Fourier amplitudes, for instance,

$$\boldsymbol{E}(\boldsymbol{r}, t) = \iint \tilde{\boldsymbol{E}}(\boldsymbol{k}, \omega) \exp(\imath(\boldsymbol{k} \cdot \boldsymbol{r} - \omega t)) \, \mathrm{d}^3\boldsymbol{k} \, \mathrm{d}\omega, \tag{1.62}$$

in the Maxwell equations leads to

$$\iint \begin{bmatrix} \imath \boldsymbol{k} \cdot \tilde{\boldsymbol{E}} - \tilde{\rho}/\epsilon_0 \\ \imath \boldsymbol{k} \times \tilde{\boldsymbol{E}} - \imath\omega \tilde{\boldsymbol{B}} \\ \imath \boldsymbol{k} \cdot \tilde{\boldsymbol{B}} \\ \imath \boldsymbol{k} \times \tilde{\boldsymbol{B}} + \imath\omega\epsilon_0\mu_0 \tilde{\boldsymbol{E}} - \mu_0 \tilde{\boldsymbol{j}} \end{bmatrix} \exp(\imath(\boldsymbol{k} \cdot \boldsymbol{r} - \omega t)) \, \mathrm{d}^3\boldsymbol{k} \, \mathrm{d}\omega = 0. \tag{1.63}$$

For this set of equations to be satisfied identically for any allowed $\boldsymbol{r}$ and t, the integrands (in the square bracket) must all vanish. For electromagnetic waves in vacuum, $\tilde{\boldsymbol{j}} = \tilde{\rho} = 0$, this leads to a set of algebraic equations identical to those previously studied for plane waves in the complex notation (see section 1.3.1), and requires $\tilde{\boldsymbol{E}}$ and $\tilde{\boldsymbol{B}}$ to vanish identically, *except* for combinations of $\boldsymbol{k}$ and ω satisfying the dispersion relation for electromagnetic waves in vacuum, $\omega = kc$.

This result means that any (non-monochromatic) wave field may be constructed as a superposition of plane, harmonic waves. For instance, for any plane wave propagating in the z-direction the Fourier amplitude of the electric field will be of the form

$$\tilde{\boldsymbol{E}}(\boldsymbol{k}, \omega) = \tilde{\boldsymbol{E}}(\omega) \delta(k_x) \delta(k_y) \delta(k_z - \omega/c), \quad \tilde{\boldsymbol{E}}(\omega) \cdot \hat{\boldsymbol{z}} = 0, \tag{1.64}$$

and therefore

$$\boldsymbol{E}(\boldsymbol{r}, t) = \int \tilde{\boldsymbol{E}}(\omega) \exp(\imath \frac{\omega}{c}(z - ct)) \, \mathrm{d}\omega \tag{1.65}$$

since the Dirac δ-functions allow for immediate integration over the wave vector $\boldsymbol{k}$. With

$$\tilde{\boldsymbol{B}}(\boldsymbol{k}, \omega) = \frac{\boldsymbol{k}}{\omega} \times \tilde{\boldsymbol{E}}(\boldsymbol{k}, \omega), \tag{1.66}$$

the corresponding magnetic field is

$$\boldsymbol{B}(\boldsymbol{r},t) = \frac{1}{c}\hat{\boldsymbol{z}} \times \int \tilde{\boldsymbol{E}}(\omega) \exp(\iota \frac{\omega}{c}(z-ct))\,\mathrm{d}\omega = \frac{1}{c}\hat{\boldsymbol{z}} \times \boldsymbol{E}(\boldsymbol{r},t). \tag{1.67}$$

It should be born in mind that the details of a computed amplitude spectrum $\tilde{\boldsymbol{E}}(\omega)$ for a given location $\boldsymbol{r}$ depends not only on the properties of the physical wave field $\boldsymbol{E}(t)$ present, but also on the duration T of the observational time series available. The observational amplitude spectrum is a combination of properties inherent in the physical signal present and a "windowing" effect.

1.6.4 Power spectrum

From the Parseval theorem (1.47) given in table 1.1, the average value of the product of two real functions $f(t)$ and $g(t)$ over the time interval $(-T/2, T/2)$ can be expressed as

$$\overline{f(t)g(t)} = \frac{1}{T}\int_{-T/2}^{T/2} f(t)g(t)\,\mathrm{d}t = \frac{4\pi}{T}\int_0^\infty \mathrm{Re}(\tilde{f}_T^*(\omega)\tilde{g}_T(\omega))\,\mathrm{d}\omega \tag{1.68}$$

where $\tilde{f}_T$ and $\tilde{g}_T$ are the Fourier transforms of f and g over the given time interval. The theorem (1.68) may be used to express the average value of the power density carried by an electromagnetic wave, in terms of the Fourier transforms of the electric and magnetic fields in the wave, $\tilde{\boldsymbol{E}}_T(\boldsymbol{r},\omega)$ and $\tilde{\boldsymbol{B}}_T(\boldsymbol{r},\omega)$. Thus, the average value of the Poynting vector (1.14) at position $\boldsymbol{r}$ may be written

$$\overline{\boldsymbol{P}}(\boldsymbol{r}) \equiv \int_0^\infty \boldsymbol{\mathcal{P}}_\omega(\boldsymbol{r},\omega)\,\mathrm{d}\omega \equiv \int_0^\infty \boldsymbol{\mathcal{P}}_\nu(\boldsymbol{r},\nu)\,\mathrm{d}\nu, \tag{1.69}$$

with *spectral power density*

$$\boldsymbol{\mathcal{P}}_\omega(\boldsymbol{r},\omega) = \frac{4\pi}{T}\frac{1}{\mu_0}\,\mathrm{Re}\left(\tilde{\boldsymbol{E}}_T^*(\boldsymbol{r},\omega) \times \tilde{\boldsymbol{B}}_T(\boldsymbol{r},\omega)\right) \tag{1.70}$$

and the corresponding expression for $\boldsymbol{\mathcal{P}}_\nu(\boldsymbol{r},\nu)$. In the following we will be interchanging freely between spectral quantities in terms of angular frequency, $[\omega]$ = rad/s, and frequency, $[\nu]$ = Hz. A simple relation exists between these quantities. With $\mathrm{d}\omega = 2\pi\mathrm{d}\nu$ it thus follows that $\boldsymbol{\mathcal{P}}_\nu(\boldsymbol{r},\nu) = 2\pi\boldsymbol{\mathcal{P}}_\omega(\boldsymbol{r},\omega = 2\pi\nu)$. $\hat{\boldsymbol{n}} \cdot \boldsymbol{\mathcal{P}}_\nu(\boldsymbol{r},\nu)$ represents the energy transported by the electromagnetic field per unit time interval per unit frequency interval through a unit area located at position $\boldsymbol{r}$ and oriented in the direction $\hat{\boldsymbol{n}}$, that is, $[\boldsymbol{\mathcal{P}}_\nu] = \mathrm{W\ m^{-2}Hz^{-1}}$.

From (1.70) we see that each frequency component in the radiation field contributes to the total power density only at that frequency. The contribution $\boldsymbol{\mathcal{P}}_\nu(\boldsymbol{r},\nu)\,d\nu$ to the power density from an infinitesimal frequency interval $(\nu, \nu + d\nu)$ only depends on the amplitudes of the electric and magnetic fields in that frequency interval. We also note that with the spectral power density in the form (1.69) only "positive" frequencies contribute to the total power density.

As an example let us return to the case of a general plane wave propagating in vacuum along the z-axis with electric and magnetic fields as described by (1.65) and (1.66). Substitution into (1.70) leads to the result

$$\boldsymbol{\mathcal{P}}_\omega = \frac{4\pi}{T}c\epsilon_0 \mid \tilde{\boldsymbol{E}}_T(\omega) \mid^2 \hat{\boldsymbol{z}}, \tag{1.71}$$

independent of position $\boldsymbol{r}$. We remember that the length T of the observational time window enters the result not only as an explicit factor, but also through the actual form of the amplitude spectrum $\tilde{\boldsymbol{E}}_T(\omega)$.

Quiz 1.18 : What is the width $\Delta\lambda$ of the wavelength range corresponding to (1.54)?

Quiz 1.19 : Two photons (in vacuum) have energies W and $W + \Delta W$. It is often said that corresponding relative differences in wavelengths, angular frequencies and energies of the two photons are given by the relation

$$\frac{\Delta\lambda}{\lambda} \approx \frac{\Delta\omega}{\omega} \approx \frac{\Delta W}{W}.$$

Is this correct? What are the errors in $\Delta\lambda/\lambda$ and $\Delta\omega/\omega$ made when using this relation for a given value of $\Delta W/W$? If $\Delta\lambda/\lambda$ is required with an accuracy $\epsilon < 10^{-4}$, what is the maximum value of $\Delta W/W$ allowed without taking correctional terms into account?

Quiz 1.20 : For a harmonic oscillation with a Gaussian shaped envelop of characteristic width $\Delta t = \tau$, that is,

$$f(t) = \exp(-\iota\omega_0 t)\, \exp\left(-\frac{t^2}{2\tau^2}\right) \quad -\infty < t < \infty,$$

show that the corresponding width of the amplitude spectrum $\tilde{f}(\omega)$ is $\Delta\omega = \tau^{-1}$.

Quiz 1.21 : Light from a distant source with finite coherence time τ falls perpendicularly on a screen containing two parallel slits of width b and a distance d apart. The light is subsequently focused by a lens on another screen. A series of interference maxima (fringes) of increasing order n will be seen. What is the maximum order n that you expect to see? Which of the quantities τ, b and d will be involved? Estimate the reduction factor for the interference fringe of order n due to the finite coherence time.

Quiz 1.22 : A plane-polarized, monochromatic plane wave propagating along the z-axis with electric field $E_x(z,t) = E_0 \cos(k_0 z - \omega_0 t)$ is sampled over a time interval T. What is the observed power spectrum $\hat{\boldsymbol{z}} \cdot \boldsymbol{\mathcal{P}}_\omega$? What is the full width at half height $\Delta\omega_{FWHH}$ of the power spectrum?

Quiz 1.23 : For a plane wave with an electric field given as a series of harmonic segments as discussed in section 1.6.1, what is the corresponding spectral power density $\mathcal{P}_\omega$? For simplicity, choose a time interval for the Fourier transform of length $T = N\tau$ where N is an integer number. Compare with the corresponding result for one harmonic segment of length $T = N\tau$.

Quiz 1.24 : Two identical waves as described in quiz 1.22 but with slightly different frequencies, ω_1 and ω_2 with $|\,\omega_2 - \omega_1\,|\,/\omega_1 \approx 0.01$, falling on an ideal detector are sampled over a time interval T. What is the power spectrum of the combined wave? Will your result be different whether the power spectrum is calculated by first adding the electric field of the two individual waves or by first calculating the power spectrum of the individual waves and then adding? What time interval T is needed in order to identify the two frequencies ω_1 and ω_2 in the combined spectrum?

1.7 Specific Intensity of Radiation

The energy transported through a given area $\mathrm{d}^2\boldsymbol{\mathcal{A}} = \mathrm{d}^2\mathcal{A}\,\hat{\boldsymbol{n}}$ generally has contributions from individual electromagnetic waves propagating in different directions $\hat{\boldsymbol{s}}$. It will be convenient to work with quantities that explicitly reveal this directional dependence. We thus define

$$\hat{\boldsymbol{n}} \cdot \boldsymbol{\mathcal{P}}_\omega(\boldsymbol{r}, \omega) \equiv \int \mathcal{I}_\omega(\boldsymbol{r}, \hat{\boldsymbol{s}}, \omega)\, \hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{s}}\, \mathrm{d}^2\Omega \tag{1.72}$$

where $\mathrm{d}^2\Omega$ represents the differential solid angle element centered on $\hat{\boldsymbol{s}}$ as illustrated in figure 1.8. $\mathcal{I}_\omega$ or the corresponding $\mathcal{I}_\nu$ is the *specific intensity* of the radiation field. $\mathcal{I}_\nu$ represents the energy transported per unit time through a unit area oriented in direction $\hat{\boldsymbol{n}}$ per unit frequency interval and per unit solid angle centered on the direction $\hat{\boldsymbol{s}}$, that is $[\mathcal{I}_\nu] = \mathrm{W\ m^{-2}\ Hz^{-1}\ sr^{-1}}$.

If the specific intensity is independent of the direction $\hat{\boldsymbol{s}}$, the radiation field is said to be *isotropic*. If the intensity is independent of position $\boldsymbol{r}$ the radiation field is *homogeneous*.

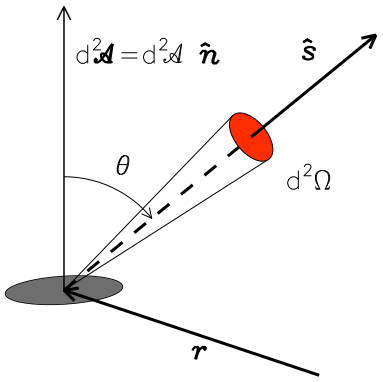

Figure 1.8: Geometry for specific intensity

From the specific intensity several useful quantities may be derived. The *mean intensity* is defined as the intensity averaged over all directions,

$$\mathcal{J}_\omega \equiv \frac{1}{4\pi}\int \mathcal{I}_\omega\, \mathrm{d}^2\Omega = \frac{1}{4\pi}\int_0^{2\pi}\int_{-1}^{1} \mathcal{I}_\omega\, \mathrm{d}\cos\theta\, \mathrm{d}\varphi. \tag{1.73}$$

For an isotropic radiation field we have $\mathcal{J}_\omega = \mathcal{I}_\omega$.

The *net energy flux* $\mathcal{F}_\omega$ is defined as

$$\mathcal{F}_\omega(\boldsymbol{r}, \hat{\boldsymbol{n}}) \equiv \int \mathcal{I}_\omega \cos\theta\, \mathrm{d}^2\Omega = \int_0^{2\pi}\int_{-1}^{1} \mathcal{I}_\omega\, \cos\theta\, \mathrm{d}\cos\theta\, \mathrm{d}\varphi. \tag{1.74}$$

$\mathcal{F}_\omega(\boldsymbol{r}, \hat{\boldsymbol{n}})$ represents the *net* flow of energy per unit frequency interval and per unit time through a unit area with normal $\hat{\boldsymbol{n}}$ located at $\boldsymbol{r}$. Flow of energy in direction $\hat{\boldsymbol{n}}$ contributes positively, flow of energy in the opposite direction contributes negatively. It is often convenient to write the net flow of energy as the difference between the *unidirectional energy flux* in each direction

$$\mathcal{F}_\omega = \mathcal{F}_\omega^+ - \mathcal{F}_\omega^- = \int_0^{2\pi}\int_0^{1} \mathcal{I}_\omega\, \cos\theta\, \mathrm{d}\cos\theta\, \mathrm{d}\varphi + \int_0^{2\pi}\int_{-1}^{0} \mathcal{I}_\omega\, \cos\theta\, \mathrm{d}\cos\theta\, \mathrm{d}\varphi. \tag{1.75}$$

$\mathcal{F}_\omega^+$ represents the net energy flux per unit area in the $+\hat{\boldsymbol{n}}$-direction per unit specified frequency interval, $\mathcal{F}_\omega^-$ the corresponding net flux in the opposite direction. For an isotropic radiation field we have $\mathcal{F}_\omega^+ = \mathcal{F}_\omega^- = \pi\mathcal{I}_\omega$ and therefore $\mathcal{F}_\omega = 0$.

All quantities discussed above are spectral quantities measured per unit angular frequency ω or per unit frequency ν interval. From the spectral quantities the corresponding total quantities are derived by integrating over (angular) freguency, for instance the total energy flux

$$F^+ = \int \mathrm{d}\omega \mathcal{F}_\omega^+ = \int \mathrm{d}\nu \mathcal{F}_\nu^+.$$

We have so far expressed quantities like spectral power density or specific intensity in terms of the electromagnetic wave fields $\boldsymbol{E}$ and $\boldsymbol{B}$. According to the correspondence principle it is often advantageous to express these quantities in terms of the alternative corpuscular view, that is, that the energy transported by the radiation is carried by photons of the proper frequency. To this end the *photon probability density* $f_{\boldsymbol{r}\hat{\boldsymbol{s}}\,\omega}$ or alternatively $f_{\boldsymbol{r}\hat{\boldsymbol{s}}\,\nu}$ are needed. $f_{\boldsymbol{r}\hat{\boldsymbol{s}}\,\nu}$ represents the number of photons per unit volume, per unit frequency interval and traveling per unit solid angle in the specified direction $\hat{\boldsymbol{s}}$. Each photon with energy $\hbar\omega = h\nu$ and traveling (in vacuum) with the speed of light c in direction $\hat{\boldsymbol{s}}$ contributes to the specific intensity $\mathcal{I}_\omega$ in that direction. We note that $[f_{\boldsymbol{r}\omega\hat{\boldsymbol{s}}}] = \mathrm{m}^{-3}\ \mathrm{Hz}^{-1}\ \mathrm{sr}^{-1}$. In terms of these quantities the specific intensity is given as

$$\mathcal{I}_\omega \equiv \hbar\omega\, c\, f_{\boldsymbol{r}\hat{\boldsymbol{s}}\,\omega} \quad \text{or} \quad \mathcal{I}_\nu \equiv h\nu\, c\, f_{\boldsymbol{r}\hat{\boldsymbol{s}}\,\nu}. \tag{1.76}$$

We end this initial introduction to the specific intensity with a simple, yet illustrative example. Consider the radiation field in vacuum from an extended source, a star with a projected area $\mathcal{A}$ as indicated in figure 1.9. We assume that at a distance r_1 the radiation from the star fills the solid angle $\Delta\Omega_1 = \mathcal{A}/r_1^2$ uniformly. The energy flux through a radially directed unit area is then given by

$$\mathcal{F}_\omega(r_1) = \mathcal{I}_\omega(r_1)\,\Delta\Omega_1,$$

where $\mathcal{I}_\omega(r_1)$ is the specific intensity at radius r_1. For this result we did assume $\Delta\Omega \ll 1$ and therefore that we could assume $\cos\theta \approx 1$. The total energy flux through a spherical shell of radius r_1 is $4\pi r_1^2 \mathcal{F}_\omega(r_1) = 4\pi\mathcal{A}\,\mathcal{I}_\omega(r_1)$. The corresponding flux through a spherical shell of radius r_2, where the specific intensity is $\mathcal{I}_\omega(r_2)$, is given by $4\pi r_2^2 \mathcal{F}_\omega(r_2) = 4\pi\mathcal{A}\,\mathcal{I}_\omega(r_2)$. If no energy is created or lost between radii r_1 and r_2, these quantities must be equal. This means that we have arrived at the conclusion that the specific intensity is independent of the distance from the star,

$$\mathcal{I}_\omega(r_2) = \mathcal{I}_\omega(r_1). \tag{1.77}$$

At first sight this result may come as a surprise: the specific intensity of radiation from the Sun is the same at the distance of the pluton Pluto as it is at the planet Earth!

Quiz 1.25 : Comment on the claim: "The specific intensity resulting from a radiating point source is singular".

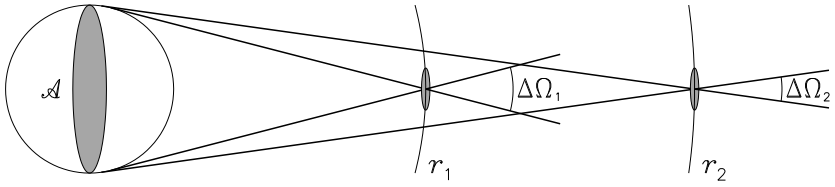

Figure 1.9: Intensity along rays

Quiz 1.26 : Two identical bolometers are directed toward the Sun but placed at the distances of the Earth and Pluto. Will they measure identical energy deposition by radiation from the Sun? Discuss. How is a constant specific intensity with distance consistent with the expectation of a r^{-2} dependent energy flux? [An ideal bolometer measures the total energy deposition from infalling radiation within the opening angle of the instrument.]

1.8 Interaction of Waves and Matter

We have so far considered properties of electromagnetic waves in vacuum. In a medium additional effects will be important. The electric field $\boldsymbol{E}$ of an electromagnetic wave present will induce electric currents in the medium. These currents will act back to determine the properties of the wave. The simple dispersion relation $\omega = kc$ valid for electromagnetic waves in vacuum will have to be replaced. The phase and group velocities of the wave will generally take different values, possibly even pointing in different directions. Waves may suffer refractive effects in spatially varying media, they may suffer absorption, and the medium may act to generate additional wave energy. The latter two effects constitute the main subjects for the subsequent chapters. Here we first consider refractive effects.

1.8.1 Non-magnetized Plasma

The common astrophysical medium is an ionized gas, or plasma, with free electron density $n_e(\boldsymbol{r})$. The electric field $\boldsymbol{E}$ of an electromagnetic wave will set these electrons into oscillating motion and thus generate an electric current density $\boldsymbol{j}_p$ in the medium. Ions will also contribute to the total induced currents. But because of their much larger mass their effects will be much smaller. For a simplest model we assume the ions to be immobile. The electrons will be assumed mobile but cold. The latter condition means that we are neglecting effects from thermal motions of the electrons. As long at the thermal speed of the electrons is much less than the phase speed of the waves, this is normally a useful approximation.

For a quantitative analysis we assume the electromagnetic wave to be harmonically varying in space and time with angular frequency and wave vector ω and $\boldsymbol{k}$,

$$\boldsymbol{E} = \boldsymbol{E}_0 \exp(\imath \boldsymbol{k} \cdot \boldsymbol{r} - \imath\omega t).$$

This electric field will give rise to an oscillating velocity $\boldsymbol{v}$ for electrons at $\boldsymbol{r}$ given by

$$-\imath\omega m\boldsymbol{v} = -e\boldsymbol{E}. \tag{1.78}$$

Here effects due to the Lorentz force from the magnetic field $\boldsymbol{B}$ associated with the wave were neglected. We also assumed no external constant magnetic field $\boldsymbol{B}_0$ to be present. With n_e electrons per unit volume the resulting induced electric current density is

$$\boldsymbol{j}_p = -en_e\boldsymbol{v} = \iota\frac{\omega_p^2}{\omega}\epsilon_0\boldsymbol{E} = -\iota\omega\epsilon_0(K-1)\boldsymbol{E} \tag{1.79}$$

where

$$K = 1 - \frac{\omega_p^2}{\omega^2} \tag{1.80}$$

is the *dielectric coefficient* of the medium. The *plasma frequency* ω_p is defined by

$$\omega_p^2 \equiv \frac{n_e e^2}{m\epsilon_0}. \tag{1.81}$$

We notice that $\boldsymbol{j}_p$ is parallel, but 90° out of phase, with $\boldsymbol{E}$. This means that the wave will not suffer damping due to Ohmic heating of the medium (see quiz 1.28). The induced current $\boldsymbol{j}_p$ will be accompanied by the charge density $\rho_p = \boldsymbol{k}\cdot\boldsymbol{j}_p/\omega$, which when substituted into Gauss law (1.1) gives rise to the condition

$$K\boldsymbol{k}\cdot\boldsymbol{E} = 0. \tag{1.82}$$

For $\omega \neq \omega_p$ this requires $\boldsymbol{k}\cdot\boldsymbol{E} = 0$, that is, the wave will be a transverse wave.

Substitution of the induced current density (1.79) into (1.2) leads to the wave equation

$$\mathcal{N}^2\hat{\boldsymbol{k}}\times(\hat{\boldsymbol{k}}\times\boldsymbol{E}) + K\boldsymbol{E} = 0 \tag{1.83}$$

where

$$\mathcal{N}^2 \equiv \frac{c^2k^2}{\omega^2} \tag{1.84}$$

defines the *refractive index* $\mathcal{N}$ of the medium. The wave equation (1.83) allows for non-trivial transverse solutions (solutions with $\boldsymbol{k}\cdot\boldsymbol{E} = 0$ and $\boldsymbol{E} \neq 0$!) only if $\boldsymbol{k}$ and ω together satisfy the dispersion relation

$$\mathcal{N}^2 = K \quad \text{or equivalently} \quad \omega^2 = \omega_p^2 + c^2k^2. \tag{1.85}$$

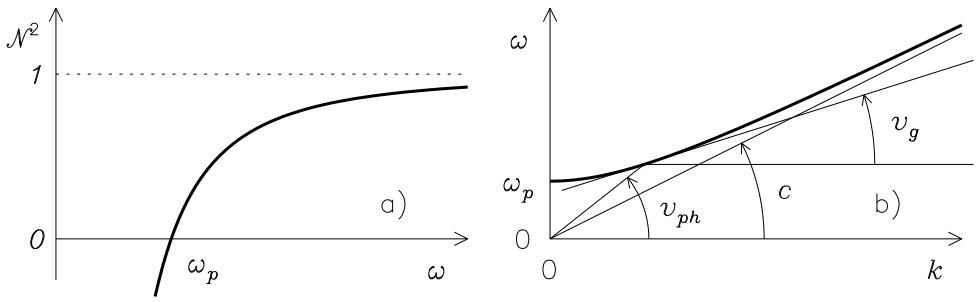

Figure 1.10: Refractive index of the non-magnetized plasma

The properties of transverse electromagnetic waves in a plasma differ significantly from the corresponding property of electromagnetic waves in vacuum. The waves are now *dispersive*,

that is, the phase velocity of the wave varies with frequency ω (or wave number k). For frequencies $\omega < \omega_p$ the refractive index N takes imaginary values. This means that the electric field of the wave will be decaying exponentially with distance. Such a wave is called an *evanescent wave*. Propagating electromagnetic waves only exist for frequencies $\omega > \omega_p$ where refractive index $\mathcal{N}$ is real and less than unity. This means that the phase velocity magnitude $v_{ph} = \omega/k$ of the wave exceeds the velocity of light in vacuum. The magnitude of the group velocity $\boldsymbol{v}_g = \partial\omega/\partial\boldsymbol{k}$ is, however, less than c and tends to zero as the frequency approaches the plasma frequency. The result is plotted in figure 1.10. The quantities v_{ph}, c and v_g are graphically represented by the tangent of the indicated angles.

In the simple cold plasma model described above the propagating electromagnetic wave suffer no damping. In reality even low-frequency electromagnetic waves will be damped due to the action of elastic collisions. The oscillating electrons will sometimes collide with ions or neutral atoms and suffer deflections. No net current perpendicular to the wave electric field $\boldsymbol{E}$ will be produced from these random deflections. The net effect on an "average" oscillating electron can be described by an extra collision (or friction) term in (1.78),

$$-\iota\omega m\boldsymbol{v} + m\nu_{coll}\boldsymbol{v} = -e\boldsymbol{E}, \tag{1.86}$$

where ν_{coll} is the effective *collision frequency*. The extra term means that dielectric coefficient of the medium takes complex values,

$$K = 1 - \frac{\omega_p^2}{\omega^2 + \iota\nu_{coll}\omega} \tag{1.87}$$

and thus that $\boldsymbol{j}_p$ is no longer 90° out of phase with $\boldsymbol{E}$. This means that the wave will be damped due to Ohmic heating (see quiz 1.28).

1.8.2 Magnetized Plasma

The result of the previous subsection was an isotropic dispersion relation, $\omega = \omega(k)$. That is, the frequency depends on the magnitude of the wave vector, but is independent of its direction. This means that the group velocity $\boldsymbol{v}_g = \partial\omega/\partial\boldsymbol{k}$ and the wave vector $\boldsymbol{k}$ are parallel. This result will be different if a static magnetic field $\boldsymbol{B}_0 = B_0\hat{\boldsymbol{z}}$ is permeating the ionized gas.

With the magnetic field present the oscillating velocity of the electrons (1.78) must be modified by including the Lorentz force from the static magnetic field $\boldsymbol{B}_0$

$$-\iota\omega m\boldsymbol{v} = -e(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}_0). \tag{1.88}$$

The induced current density is now given by

$$\boldsymbol{j}_p = -en_e\boldsymbol{v} = -\iota\omega\epsilon_0\left(\mathbf{K} - \mathbf{1}\right)\cdot\boldsymbol{E} \tag{1.89}$$

where the *dielectric tensor* $\mathbf{K}$ is a tensor of rank 2 (a 3×3 matrix for any given choice of coordinate system) with elements

$$\mathrm{K}_{xx} = \mathrm{K}_{yy} = S = 1 + \frac{\alpha^2}{1-\beta^2} \tag{1.90}$$

$$\mathrm{K}_{xy} = -\mathrm{K}_{yx} = \iota D = \frac{\iota\alpha^2\beta}{1-\beta^2} \tag{1.91}$$

$$\mathrm{K}_{zz} = P = 1 - \alpha^2 \tag{1.92}$$

$$\mathrm{K}_{xz} = \mathrm{K}_{yz} = \mathrm{K}_{zx} = \mathrm{K}_{zy} = 0 \tag{1.93}$$

with

$$\alpha \equiv \omega_p/\omega, \quad \beta \equiv \omega_b/\omega, \quad \text{and} \quad \omega_b \equiv \frac{eB_0}{m}. \tag{1.94}$$

The induced current $\boldsymbol{j}_p$ is no longer parallel with $\boldsymbol{E}$. The dielectric tensor is Hermitean ($K^*_{ij} = K_{ji}$), the wave will therefore still not suffer damping due to Ohmic heating of the medium (see quiz 1.29). Gauss law (1.1) leads to the condition

$$\boldsymbol{k} \cdot \mathbf{K} \cdot \boldsymbol{E} = 0, \tag{1.95}$$

and we can no longer assume the wave to be transverse ($\boldsymbol{k} \cdot \boldsymbol{E} = 0$).

The wave equation this time reads

$$\mathcal{N}^2 \hat{\boldsymbol{k}} \times (\hat{\boldsymbol{k}} \times \boldsymbol{E}) + \mathbf{K} \cdot \boldsymbol{E} = 0. \tag{1.96}$$

Without loss of generality we may choose $\hat{\boldsymbol{k}}$ to lie in the xz-plane, $\hat{\boldsymbol{k}} = (\sin\theta, 0, \cos\theta)$. The linear, homogeneous set of equations (1.96) may then be written in matrix form

$$\begin{pmatrix} -\mathcal{N}^2\cos^2\theta + S & \iota D & \mathcal{N}^2 \sin\theta\cos\theta \\ -\iota D & -\mathcal{N}^2 + S & 0 \\ \mathcal{N}^2\sin\theta\cos\theta & 0 & -\mathcal{N}^2\sin^2\theta + P \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} = 0. \tag{1.97}$$

Non-trivial solutions for $\boldsymbol{E}$ ($\boldsymbol{E} \neq 0$) will only exist if this set of equations is linearly dependent, that is, if the system determinant vanishes,

$$\det \begin{vmatrix} -\mathcal{N}^2\cos^2\theta + S & \iota D & \mathcal{N}^2 \sin\theta\cos\theta \\ -\iota D & -\mathcal{N}^2 + S & 0 \\ \mathcal{N}^2\sin\theta\cos\theta & 0 & -\mathcal{N}^2\sin^2\theta + P \end{vmatrix} = A\mathcal{N}^4 - B\mathcal{N}^2 + C = 0 \tag{1.98}$$

with

$$\begin{aligned} A &= S\sin^2\theta + P\cos^2\theta \\ B &= (S^2 - D^2)\sin^2\theta + SP(1 + \cos^2\theta) \\ C &= (S^2 - D^2)P. \end{aligned} \tag{1.99}$$

Equation (1.98) is the dispersion relation for electromagnetic waves in our cold, magnetized plasma. It is a quadratic polynomial equation with two solutions for $\mathcal{N}^2$ as functions of ω and θ, solutions that may be written in the form $\omega = \omega(k, \theta)$. These solutions correspond to different wave modes with wave polarization $\hat{\boldsymbol{\epsilon}}$ determined by substituting the proper value for $\mathcal{N}^2$ for a given value of θ in (1.96) and solving for $\boldsymbol{E}$. In figure 1.11 the different solutions for $\mathcal{N}^2$ are plotted as functions of ω for $\theta = 0°$ (green) and $\theta = 90°$ (red) for two cases: a) $\omega_p > \omega_b$ and b) $\omega_p < \omega_b$. Symbols o, x, z and w stand for *ordinary*, *extraordinary*, z and *whistler* modes. The o and x modes are of special importance in radio astronomy. These are the only modes that may escape from an astrophysical radio source. Propagation away from the source towards decreasing ω_p and ω_b is equivalent to moving to the right in figure 1.11. The o and x modes meet no obstacle in this respect, whereas the z and whistler modes will encounter a *resonance* ($\mathcal{N}^2 \to \infty$) and be absorbed. Contrary, when propagating towards increasing ω_b and ω_p the o and x modes will eventually meet a *cutoff* ($\mathcal{N}^2 \to 0$) where they will be refracted or reflected.

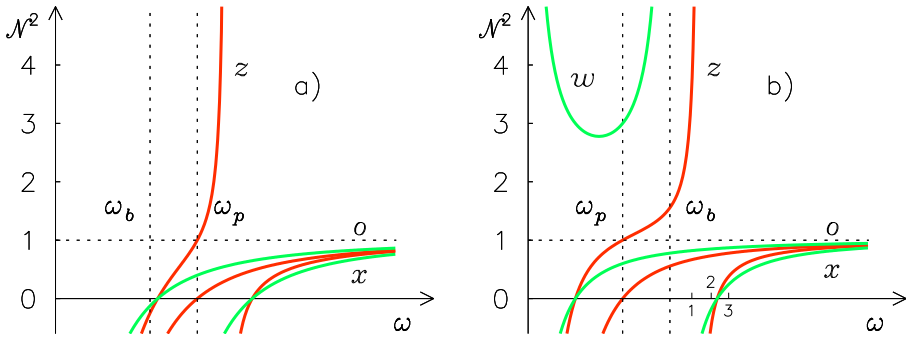

Figure 1.11: Refractive index for magnetized plasma

With a dispersion relation for the magnetized plasma of the general form $\omega = \omega(k, \theta)$ the corresponding group velocity $\boldsymbol{v}_g$ is no longer parallel with the wave vector $\boldsymbol{k}$, in fact,

$$\boldsymbol{v}_g = \frac{\partial \omega}{\partial k}|_\theta \, \hat{\boldsymbol{k}} + \frac{1}{k}\frac{\partial \omega}{\partial \theta}|_k \, \hat{\boldsymbol{\theta}} = \frac{\partial \omega}{\partial k}|_\theta \left(\hat{\boldsymbol{k}} - \frac{\partial \ln \mathcal{N}}{\partial \theta}|_\omega \, \hat{\boldsymbol{\theta}} \right). \tag{1.100}$$

and therefore also

$$v_g = \left| \frac{\partial \omega}{\partial k}|_\theta \right| \sqrt{1 + \left(\frac{\partial \ln \mathcal{N}}{\partial \theta}|_\omega \right)^2} \tag{1.101}$$

For this result we made use of the mathematical identity (1.103) (see quiz 1.33). The group velocity $\boldsymbol{v}_g$ makes an angle ξ with the magnetic field $\boldsymbol{B}_0$ where

$$\cos \xi = \frac{\cos \theta + \frac{\partial \ln \mathcal{N}}{\partial \theta}|_\omega \, \sin \theta}{\sqrt{1 + \left(\frac{\partial \ln \mathcal{N}}{\partial \theta}|_\omega \right)^2}}. \tag{1.102}$$

We note that $\boldsymbol{v}_g$, $\boldsymbol{k}$ and $\boldsymbol{B}_0$ are coplanar.

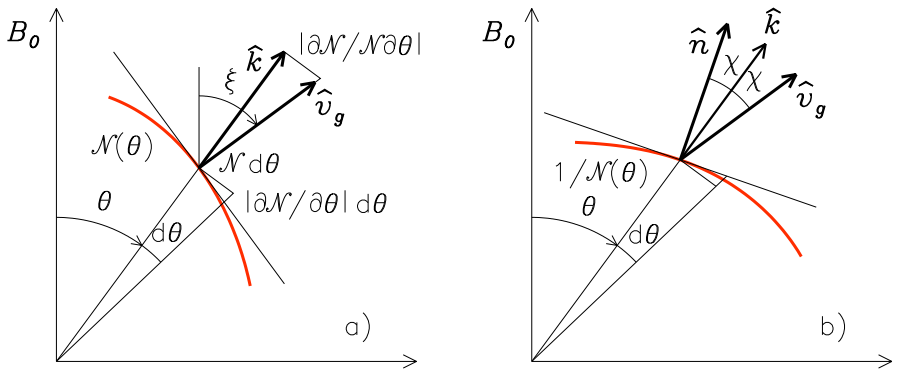

Figure 1.12: Refractive index a) and phase velocity b) surfaces with graphical construction of the group velocity direction for an axisymmetric medium

The situation is illustrated in figure 1.12. In a) the red curve represents the function $\mathcal{N}(\theta) = ck(\theta)/\omega$ for a given ω. This curve will when rotated around the $\boldsymbol{B}_0$ axis generate

what is known as the *refractive index surface* for the chosen wave mode. From the graphical construction one readily notes that the group velocity will always be perpendicular to the refractive index surface. An alternative construction, where the refractive index surface is replaced by the *phase velocity surface* (inverse refractive index surface) $1/\mathcal{N}(\theta)$ for a given ω, is shown in b). The normal $\hat{\boldsymbol{n}}$ to this surface is parallel with

$$\hat{\boldsymbol{k}} - \frac{\partial \ln 1/\mathcal{N}}{\partial \theta}|_\omega \, \hat{\boldsymbol{\theta}} = \hat{\boldsymbol{k}} + \frac{\partial \ln \mathcal{N}}{\partial \theta}|_\omega \, \hat{\boldsymbol{\theta}}$$

while the group velocity $\boldsymbol{v}_g$ is still given by (1.100). Thus, the wave vector $\boldsymbol{k}$ bisects the angle between the phase velocity surface normal $\hat{\boldsymbol{n}}$ and the group velocity $\boldsymbol{v}_g$.

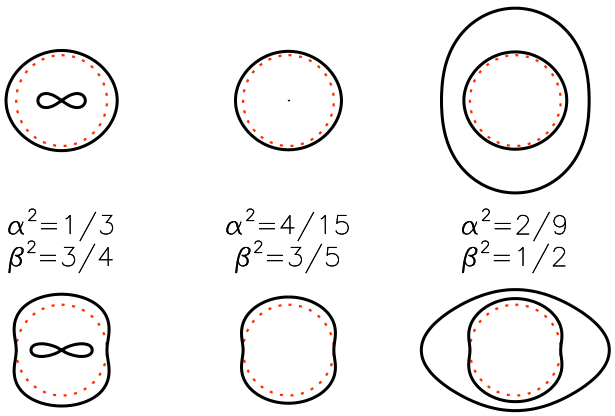

Figure 1.13: Phase velocity and inverse ray refractive index surfaces for three frequencies $\omega > \omega_b > \omega_p$

In the upper panel of figure 1.13 cuts through the phase velocity surface corresponding to the situation in figure 1.11b ($\omega_b > \omega_p$) for the three frequencies denoted 1, 2 and 3 in figure 1.11b are given. The magnetic field is parallel with the vertical axis. The dotted circles (red) represents an electromagnetic wave in vacuum. At the lower frequency (1) the ordinary and the z modes exist, the latter one only for angles θ between the magnetic field $\boldsymbol{B}_0$ and the wave vector $\boldsymbol{k}$ exceeding a minimum value. For the middle frequency (2) only the ordinary mode is allowed. At the higher frequency (3) the ordinary and the extraordinary modes both propagate. With still increasing frequencies the phase velocity surfaces for both modes approach that of electromagnetic waves in vacuum. In the lower panel of figure 1.13 corresponding cuts through the inverse ray refractive index are shown (see section 1.9).

In an isotropic plasma electromagnetic waves are transverse and they may take any polarization, much like electromagnetic waves in vacuum. In the magnetized plasma the situation is different. Each mode takes a particular polarization as required by the wave equation, and the waves are not generally transverse. For a short summary we restrict our discussion to $\omega > (\omega_p, \omega_b)$. For $\theta = 0°$ the ordinary and extraordinary modes are both transverse and circularly polarized, but with opposite handedness. For $\theta = 90°$ the ordinary mode is transverse and plane-polarized with the electric field parallel with $\boldsymbol{B}_0$. The corresponding

extraordinary mode, however, is non-transverse with the non-vanishing components of the electric field satisfying $E_x/E_y = -\iota D/S$.

Quiz 1.27 : Calculate the group velocity v_g resulting from the dispersion relation (1.85) and show that $v_{ph}v_g = c^2$ for electromagnetic waves propagating in an ionized non-magnetized plasma.

Quiz 1.28 : Argue that the Ohmic heating rate per unit volume of the propagating electromagnetic wave is given by the expression $\frac{1}{2}\,\mathrm{Re}\,\overline{\boldsymbol{j}_p^* \cdot \boldsymbol{E}}$. Show that the heating rate vanishes for the isotropic cold plasma in the absence of collisions. What is the corresponding result when collisions are included (see equation (1.86))?

Quiz 1.29 : With $\boldsymbol{j}_p$ as given by (1.89) for the the anisotropic cold plasma, show that the Ohmic heating rate per unit volume vanishes whatever the choice of the electric field $\boldsymbol{E}$. [Hint: Make use of the Hermitean character of the dielectric tensor $\mathbf{K}$.]

Quiz 1.30 : Show that cutoff and resonance in a magnetized plasma occur where the conditions $(S^2 - D^2)P = 0$ and $\tan^2\theta = -P/S$ are satisfied, respectively. Determine the corresponding values of ω for given ω_p and ω_b.

Quiz 1.31 : For o and x modes propagating parallel with the magnetic field ($\theta = 0$) in a plasma determine the corresponding wave polarizations $\hat{\boldsymbol{\epsilon}}$. Argue that a plane polarized wave can be formed from a superposition of o and x modes. What happens to the polarization plane as the wave propagates along the magnetic field. Why is the effect called Faraday rotation? [*Hint:* Note that the refractive index and therefore also the wave number k for the two modes will be different.]

Quiz 1.32 : Determine the wave polarizations $\hat{\boldsymbol{\epsilon}}$ for o and x modes propagating perpendicular to the magnetic field in a plasma.

Quiz 1.33 : For three variables x, y and z, functionally connected, $z = z(x, y)$, show that

$$\frac{\partial z}{\partial x}|_y \, \frac{\partial x}{\partial y}|_z \, \frac{\partial y}{\partial z}|_x = -1. \tag{1.103}$$

[*Hint:* Note that $\mathrm{d}z = \frac{\partial z}{\partial x}|_y \, \mathrm{d}x + \frac{\partial z}{\partial y}|_x \, \mathrm{d}y$, write the corresponding expression for $\mathrm{d}y$ and then eliminate $\mathrm{d}y$.]

Quiz 1.34 : Electromagnetic waves will be modified also when propagating in a neutral gas. An empirical formula for the refractive index in (dry) air at standard pressure ($P = 760$ mm-Hg) and standard temperature ($T = 0°$C) for wavelength λ [μm] is tabulated as

$$\mathcal{N} = 1 + 2.876 \times 10^{-4} + 1.629 \times 10^{-6}/\lambda^2.$$

Two nearby spectral lines have wavelengths $\lambda_0 = 5000$ Å and $\lambda_0 + \Delta\lambda_0 = 5001$ Å in vacuum. Express corresponding values of λ and $\Delta\lambda/\lambda$ when measured in a standard atmosphere in terms of λ_0 and $\Delta\lambda_0$.

1.9 The Ray Equations

We now turn to the more quantitative discussion of refractive effects for the transport of radiation in media, for the time being neglecting absorption and emission processes. We limit our discussion to radiation transport in a time stationary medium. The medium will be allowed to be inhomogeneous, but the length scales for spatial variations in the medium will be assumed to be long compared to the wave lengths of interest.

Electromagnetic waves in such a medium are governed by a local dispersion relation

$$\omega = \omega(\boldsymbol{r}, \boldsymbol{k}). \tag{1.104}$$

As the wave propagates through the time stationary medium, the wave frequency ω remains constant, while the wave vector $\boldsymbol{k}$ changes with position $\boldsymbol{r}$. The corresponding changes in wave vector and position are found by forming the total time derivative of the dispersion relation (1.104)

$$0 = \frac{\mathrm{d}\omega}{\mathrm{d}t} = \frac{\partial\omega}{\partial\boldsymbol{k}}\frac{\mathrm{d}\boldsymbol{k}}{\mathrm{d}t} + \frac{\partial\omega}{\partial\boldsymbol{r}}\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t},$$

that is,

$$\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t} = \frac{\partial\omega}{\partial\boldsymbol{k}} \tag{1.105}$$

$$\frac{\mathrm{d}\boldsymbol{k}}{\mathrm{d}t} = -\frac{\partial\omega}{\partial\boldsymbol{r}}. \tag{1.106}$$

We must here interpret $\boldsymbol{r}$ and $\boldsymbol{k}$ as the instantaneous central location and wave vector of a propagating wave packet. In particular, we have identified the propagation velocity of the wave packet with the group velocity $\boldsymbol{v}_g = \partial\omega/\partial\boldsymbol{k}$, that is, the velocity at which energy is transported in the wave (see quiz 1.10). The result (1.105)-(1.106) is referred to as the *ray equations* and forms the basis of *geometrical optics*.[5]

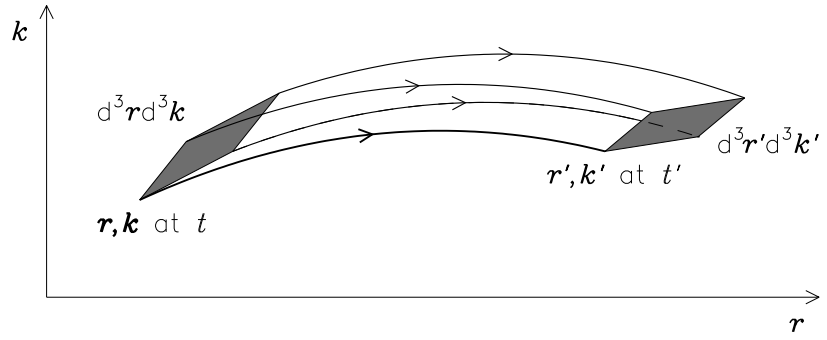

Figure 1.14: Incompressible flow through the ray equations

A remarkable property of the ray equations is that they represent an incompressible flow in the six dimensional phase space $(\boldsymbol{r}, \boldsymbol{k})$. That is, the divergence of the six dimensional phase space velocity $(\mathrm{d}\boldsymbol{r}/\mathrm{d}t, \mathrm{d}\boldsymbol{k}/\mathrm{d}t) = (\partial\omega/\partial\boldsymbol{k}, -\partial\omega/\partial\boldsymbol{r})$ vanishes identically,

$$\frac{\partial}{\partial\boldsymbol{r}} \cdot \frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t} + \frac{\partial}{\partial\boldsymbol{k}} \cdot \frac{\mathrm{d}\boldsymbol{k}}{\mathrm{d}t} \equiv 0. \tag{1.107}$$

[5]Note that with momentum $\boldsymbol{p} = \hbar\boldsymbol{k}$ and Hamiltoniam $\mathcal{H} = \hbar\omega$ the ray equations are identical in form with the Hamilton equations (3.5)-(3.6).

Physically, this result means that any 6-dimensional volume element $\mathrm{d}^3\boldsymbol{r}\,\mathrm{d}^3\boldsymbol{k}$ in $(\boldsymbol{r},\boldsymbol{k})$-space remains constant as transformed by the ray equations. Mathematically this may be reformulated in the following way: any infinitesimal six dimensional volume element $\mathrm{d}^3\boldsymbol{r}\,\mathrm{d}^3\boldsymbol{k}$ centered at $(\boldsymbol{r},\boldsymbol{k})$ at time t transforms into the volume element $\mathrm{d}^3\boldsymbol{r}'\,\mathrm{d}^3\boldsymbol{k}'$ centered at $(\boldsymbol{r}',\boldsymbol{k}')$ at time t' and such that

$$\mathrm{d}^3\boldsymbol{r}'\,\mathrm{d}^3\boldsymbol{k}' \equiv J\,\mathrm{d}^3\boldsymbol{r}\,\mathrm{d}^3\boldsymbol{k} = \mathrm{d}^3\boldsymbol{r}\,\mathrm{d}^3\boldsymbol{k} \tag{1.108}$$

since the Jacobian

$$J \equiv \frac{\partial(\boldsymbol{r}',\boldsymbol{k}')}{\partial(\boldsymbol{r},\boldsymbol{k})} = 1, \tag{1.109}$$

due to the property (1.107) (see quiz 1.35). This incompressibility property is illustrated schematically in figure 1.14. The volume of the shaded 6-dimensional $(\boldsymbol{r},\boldsymbol{k})$-element remains constant as it is transformed in time by the ray equations.

To study an important effect of the constraint (1.108), let us first consider an *isotropic* medium, that is, a medium with a dispersion relation for a chosen wave mode of the form,

$$\omega = \omega(\boldsymbol{r},\boldsymbol{k}) = \omega(\boldsymbol{r},k). \tag{1.110}$$

This is in accordance with the properties of the non-magnetized plasma of section 1.8.1. The angular wave frequency ω only depends on the magnitude k of the wave vector $\boldsymbol{k}$ and the group velocity $\boldsymbol{v}_g$ and the wave vector $\boldsymbol{k}$ are parallel. Let us consider waves with wave vectors in $\mathrm{d}^3\boldsymbol{k} = k^2\mathrm{d}k\mathrm{d}^2\Omega$ centered on $\boldsymbol{k}$ with $\mathrm{d}^2\Omega = \mathrm{d}\cos\theta\mathrm{d}\phi$. The polar angle θ is given relative to an arbitrary chosen z-axis. It is convenient to express the $\boldsymbol{k}$-space volume element in terms of the corresponding angular frequency ω, that is,

$$\mathrm{d}^3\boldsymbol{k} = k^2\,\mathrm{d}k\,\mathrm{d}^2\Omega = \frac{\omega^2}{c^2}\mathcal{N}^2\frac{\partial k}{\partial\omega}\,\mathrm{d}\omega\,\mathrm{d}^2\Omega = \frac{\omega^2}{c^2}\frac{\mathcal{N}^2}{v_g}\,\mathrm{d}\omega\,\mathrm{d}^2\Omega. \tag{1.111}$$

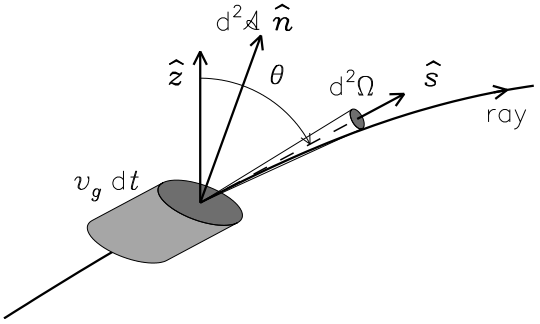

Figure 1.15: Energy transport in an isotropic medium

The energy carried by waves that passes through an arbitrary surface element $\mathrm{d}^2\boldsymbol{\mathcal{A}}$ at position $\boldsymbol{r}$ with wave vectors within the chosen $\mathrm{d}^3\boldsymbol{k}$ element centered at $\boldsymbol{k}$ during the time interval $\mathrm{d}t$ will be located within the volume element

$$\mathrm{d}^3\boldsymbol{r} = \mathrm{d}^2\boldsymbol{\mathcal{A}}\cdot\boldsymbol{v}_g\,\mathrm{d}t. \tag{1.112}$$

The geometry is illustrated in figure 1.15. The curved line represents a ray of radiation with local direction of propagation $\hat{\boldsymbol{s}} = \hat{\boldsymbol{v}}_g$ which again coincides with the direction of the wave vector $\boldsymbol{k}$. But, according to (1.108) this means that

$$\mathrm{d}^3\boldsymbol{r}\,\mathrm{d}^3\boldsymbol{k} = \frac{\omega^2}{c^2}\,\mathcal{N}^2\,\mathrm{d}^2\boldsymbol{\mathcal{A}}\cdot\hat{\boldsymbol{s}}\,\mathrm{d}t\,\mathrm{d}\omega\,\mathrm{d}^2\Omega = \text{constant along the ray.} \tag{1.113}$$

If the radiation travels unscattered and without energy gain or loss through the background medium, then the amount of energy that is transported along a ray in direction $\hat{\boldsymbol{s}}$ through the given surface element $\mathrm{d}^2\boldsymbol{\mathcal{A}}$ at position $\boldsymbol{r}$, within the solid angle element $\mathrm{d}^2\Omega$ centered on $\hat{\boldsymbol{s}}$, in the frequency interval $\mathrm{d}\omega$ during the time interval $\mathrm{d}t$, must also remain constant as transformed along the ray, that is,

$$\mathcal{I}_\omega\,\mathrm{d}^2\boldsymbol{\mathcal{A}}\cdot\hat{\boldsymbol{s}}\,\mathrm{d}t\,\mathrm{d}\omega\,\mathrm{d}^2\Omega = \text{constant along the ray.} \tag{1.114}$$

If we now compare (1.113) and (1.114) it then follows that also the ratio

$$\frac{\mathcal{I}_\omega}{\mathcal{N}^2} = \text{constant along the ray.} \tag{1.115}$$

For anisotropic media the above result needs to be generalized. We limit our discussion to an axisymmetric anisotropic medium with a dispersion relation of the form

$$\omega = \omega(\boldsymbol{r},\boldsymbol{k}) = \omega(\boldsymbol{r},k,\theta). \tag{1.116}$$

This is in accordance with the properties of the magnetized plasma of section 1.8.2. The main difference is that the direction of the ray no longer coincides with the wave vector direction. This means that we must distinguish a solid angle $\mathrm{d}^2\Omega_k = \mathrm{d}\cos\theta\mathrm{d}\phi$ centered on $\hat{\boldsymbol{k}}$ from the corresponding solid angle $\mathrm{d}^2\Omega = \mathrm{d}\cos\xi\mathrm{d}\phi$ centered on $\hat{\boldsymbol{s}}$. The polar angles θ and ξ are here both measured relative to the magnetic field $\boldsymbol{B}_0$. The geometry is illustrated in figure 1.16.The vectors $\boldsymbol{B}_0$, $\hat{\boldsymbol{s}}$ and $\hat{\boldsymbol{k}}$ are coplanar. We may now make use of (1.101) and (1.102) to find

$$\mathrm{d}^3\boldsymbol{k} = k^2\,\mathrm{d}k\,\mathrm{d}^2\Omega_k = \frac{\omega^2}{c^2}\mathcal{N}^2\left|\frac{\partial k}{\partial\omega}\,|_\theta\right|\left|\frac{\partial\cos\theta}{\partial\cos\xi}\,|_\omega\right|\,\mathrm{d}\omega\,\mathrm{d}^2\Omega = \frac{\omega^2}{c^2}\frac{\mathcal{N}_r^2}{v_g}\,\mathrm{d}\omega\,\mathrm{d}^2\Omega, \tag{1.117}$$

where

$$\mathcal{N}_r^2 \equiv \mathcal{N}^2\sqrt{1+\left(\frac{\partial\ln\mathcal{N}}{\partial\theta}\,|_\omega\right)^2}\left|\frac{\partial\cos\theta}{\partial\cos\xi}\,|_\omega\right| \tag{1.118}$$

defines the *ray refractive index* $\mathcal{N}_r$. A comparison between the inverse refractive index and the inverse ray refractive index is shown in figure 1.13. But this means that for the axisymmetric anisotropic medium the result (1.113) will have to be replaced by

$$\mathrm{d}^3\boldsymbol{r}\,\mathrm{d}^3\boldsymbol{k} = \frac{\omega^2}{c^2}\,\mathcal{N}_r^2\,\mathrm{d}^2\boldsymbol{\mathcal{A}}\cdot\hat{\boldsymbol{s}}\,\mathrm{d}t\,\mathrm{d}\omega\,\mathrm{d}^2\Omega = \text{constant along the ray,} \tag{1.119}$$

and therefore also (1.115) replaced by

$$\frac{\mathcal{I}_\omega}{\mathcal{N}_r^2} = \text{constant along the ray.} \tag{1.120}$$

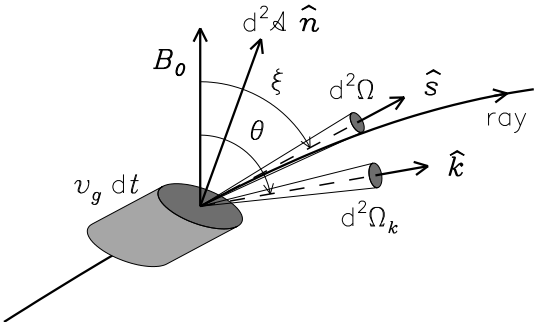

Figure 1.16: Energy transport in an anisotropic medium

We notice that for an isotropic medium the ray refractive index reduces to the ordinary refractive index, $\mathcal{N}_r = \mathcal{N}$. In dilute media like stellar atmospheres and at high enough frequencies the refractive indices $\mathcal{N}$ and $\mathcal{N}_r^2$ are both close to unity. The factors $\mathcal{N}^2$ and $\mathcal{N}_r^2$ in (1.115) and (1.120) are therefore often omitted in the literature. This omission is equivalent to the expectation that each ray travels along straight lines. Otherwise, in inhomogeneous media where $\mathcal{N}_r$ and $\mathcal{N}$ are functions of position $\boldsymbol{r}$, the rays will be curved lines, that is, the waves will be refracted or reflected by the medium.

A simple example will be illustrative.

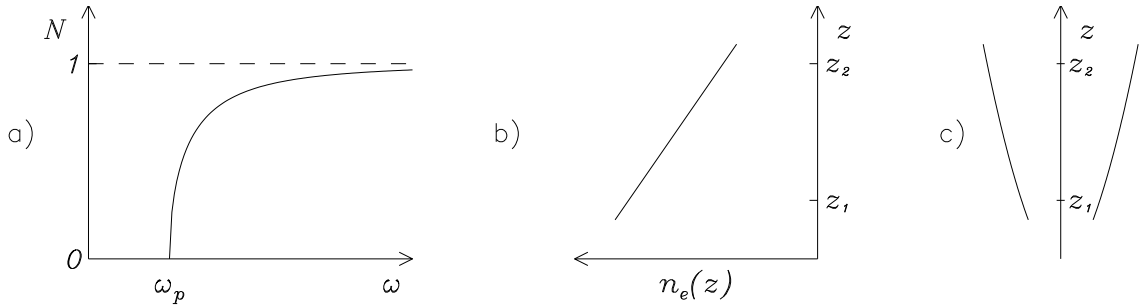

Figure 1.17: Ray refraction in plane-stratified ionized medium

Consider a plane-stratified ionized medium for which the electron density decreases with increasing z as illustrated in figure 1.17b. Let a ray of electromagnetic radiation fill the small solid angle element $\Delta\Omega_1 = \pi\theta_1^2$ centered on the positive z-axis at level z_1 where the refractive index is $\mathcal{N}_1$. As the ray propagates to level z_2 where the refractive index has changed to the value $\mathcal{N}_2$, the ray will be subject to Snell's law, requiring that the relation $\mathcal{N}_1 \sin\theta_1 = \mathcal{N}_2 \sin\theta_2$ is satisfied. For small θ this can be approximated by

$$\mathcal{N}_1\,\theta_1 = \mathcal{N}_2\,\theta_2. \tag{1.121}$$

This means that off-vertical rays will be slightly refracted as illustrated in figure 1.17c. The refractional effect is important when the wave frequency is comparable to the plasma frequency.

If the wave propagates without energy loss or gain, then $\mathcal{I}_\omega(z_1)\,\Delta\Omega_1 = \mathcal{I}_\omega(z_2)\,\Delta\Omega_2$. But from (1.121) it then also follows that

$$\frac{\mathcal{I}_\omega(z_1)}{\mathcal{N}^2(z_1)} = \frac{\mathcal{I}_\omega(z_2)}{\mathcal{N}^2(z_2)},$$

that is, the intensity of the beam and the refractive index of the medium are related to each other in a simple fashion. Formulated in an alternative way, the refractive index modulates the intensity in the beam. In the next section we will show that this simple result has a more general validity.

Refraction is but one of many effects that will determine the propagation of radiation fields in the presence of matter. The induced acceleration of free electrons by the electric field of the wave will also lead to generation of radiation traveling in different directions (Thompson scattering). Photons may suffer collisions with electrons and be deflected with a corresponding change in energy (Compton scattering). Often more importantly, photons may be absorbed by atoms or molecules, the excess energy later to be re-emitted in arbitrary direction and possibly at different frequencies in a spontaneous emission transition or as a copy of a triggering photon in a stimulated emission process. A study of radiative transport aims at explaining and predicting the cumulative effect of all these different processes.

Quiz 1.35 : Let $\boldsymbol{r}$ and $\boldsymbol{k}$ represent the central position and wave vector of a wave packet at time t. For $t' = t + \mathrm{d}t$, where $\mathrm{d}t$ is infinitesimal, the solution of the ray equations (1.105)-(1.106) can be written

$$\boldsymbol{r}' = \boldsymbol{r} + \frac{\partial\omega}{\partial\boldsymbol{k}}\mathrm{d}t$$
$$\boldsymbol{k}' = \boldsymbol{k} - \frac{\partial\omega}{\partial\boldsymbol{r}}\mathrm{d}t.$$

Show that the Jacobian for the transformation from $(\boldsymbol{r}, \boldsymbol{k})$ to $(\boldsymbol{r}', \boldsymbol{k}')$ is given by $J = 1 + \mathcal{O}((\mathrm{d}t)^2)$ and therefore that (1.109) is satisfied.

Quiz 1.36 : A ray of radiation with angular frequency ω enters at $z = 0$ a plane-stratified isotropic plasma,

$$\omega_p^2 = \begin{cases} \alpha z & 0 < z < L \\ \alpha L & L < z, \end{cases}$$

under an angle of incidence θ_0 (angle between the ray and the positive z-axis). Write down and solve the ray equations for the two cases 1) $\omega^2\cos^2\theta_0 < \alpha L$ and 2) $\omega^2\cos^2\theta_0 > \alpha L$. What is the geometrical shape of the rays? Show that Snell's law is satisfied, that is, $\mathcal{N}\sin\theta =$ constant along the ray. [Hint: Express $\sin^2\theta$ in terms of the components of the wave vector $\boldsymbol{k}$.] Write conditions 1) and 2) in terms of the refractive index $\mathcal{N}$ and interpret you result.

1.10 The Radiative Transport Equation

In the previous section we argued that in a time stationary and ideal refractive medium the quantity $\mathcal{I}_\omega/\mathcal{N}_r^2$ will remain constant as transported along the ray. In mathematical terms

this result is most conveniently formulated in the form of the differential equation

$$\hat{\boldsymbol{s}} \cdot \nabla \left(\frac{\mathcal{I}_\omega}{\mathcal{N}_r^2} \right) \equiv \frac{\mathrm{d}}{\mathrm{d}s} \left(\frac{\mathcal{I}_\omega}{\mathcal{N}_r^2} \right) = 0. \tag{1.122}$$

Here $\hat{\boldsymbol{s}}$ is a unit vector that varies along the ray according to the requirements of (1.105) and s is to be interpreted as measuring the arc length along the ray.

Several effects, however, enter to modify our idealized result. Due to the interaction with the medium through which the radiation travels, energy may be scattered out of the ray, energy may be absorbed by the medium, or the medium may have the ability to emit or redirect energy into the ray. These effects may be included by adding extra terms to the right hand side of (1.122). We let the *emissivity* j_ω represent the local contribution to the specific intensity $\mathcal{I}_\omega$ in the form of photons that are added to the beam per unit arc element along the ray. Likewise, it is reasonable to assume that the amount of energy that is scattered out of the beam or absorbed by the medium per unit arc element along the ray is proportional to the energy available in the beam, that is, in the form $-\alpha_\omega \mathcal{I}_\omega$. The coefficient α_ω is the *linear extinction coefficient*. This coefficient is sometimes written in the equivalent forms

$$\alpha_\omega \equiv \sigma_\omega n \equiv \kappa_\omega \rho. \tag{1.123}$$

Here σ_ω is the *extinction coefficient per particle* and κ_ω the *opacity* or *mass extinction coefficient* and n and ρ are the number density of the relevant particle and the mass density of the medium. With the contributions of emissivity and extinction added, the resulting *radiative transport equation* reads

$$\mathcal{N}_r^2 \frac{\mathrm{d}}{\mathrm{d}s} \left(\frac{\mathcal{I}_\omega}{\mathcal{N}_r^2} \right) = j_\omega - \alpha_\omega \mathcal{I}_\omega. \tag{1.124}$$

To make some preliminary acquaintance with the general properties of this equation, we shall assume the coefficients j_ω and α_ω to be known. For convenience we define two important quantities. The *source function*

$$S_\omega \equiv \frac{1}{\mathcal{N}_r^2} \frac{j_\omega}{\alpha_\omega} \tag{1.125}$$

will play an important role in the analysis of the emission from the medium. The other quantity is the differential *optical depth* $\mathrm{d}\tau_\omega$ related to the differential arc length $\mathrm{d}s$ by

$$\mathrm{d}\tau_\omega \equiv -\alpha_\omega(s)\, \mathrm{d}s. \tag{1.126}$$

Arc length s increases going along the ray, the corresponding optical depth $\tau_\omega(s)$ increases traversing the ray in the opposite direction and from the position of a given observer. The difference in optical depth between two points $s = 0$ and s along a given ray

$$\tau_\omega(0) - \tau_\omega(s) \equiv \int_0^s \alpha_\omega(s')\, \mathrm{d}s', \tag{1.127}$$

is a measure of the total extinction suffered by the beam along that ray segment. The optical depth $\tau_\omega(s)$ represents an alternative way of measuring lengths s in the medium. The geometry is illustrated in figure 1.18.

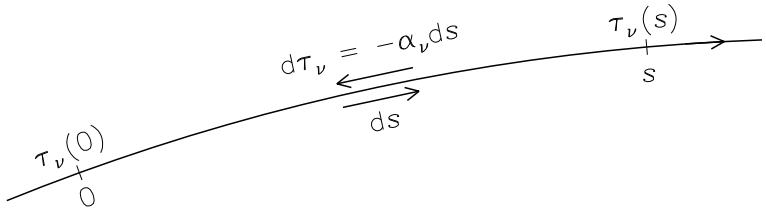

Figure 1.18: Geometric length and optical depth along a ray

In terms of source function S_ω and optical depth τ_ω the radiative transport equation (1.124) takes the simple form

$$\frac{\mathrm{d}}{\mathrm{d}\tau_\omega}\left(\frac{\mathcal{I}_\omega}{\mathcal{N}_r^2}\right) - \frac{\mathcal{I}_\omega}{\mathcal{N}_r^2} = -S_\omega \tag{1.128}$$

with the formal solution

$$\frac{\mathcal{I}_\omega}{\mathcal{N}_r^2}(s)\exp(-\tau_\omega(s)) = \frac{\mathcal{I}_\omega}{\mathcal{N}_r^2}(0)\exp(-\tau_\omega(0)) + \int_{\tau_\omega(s)}^{\tau_\omega(0)} S_\omega(\tau'_\omega)\exp(-\tau'_\omega)\,\mathrm{d}\tau'_\omega. \tag{1.129}$$

For a simplest possible application of the general solution (1.129) consider a homogeneous medium in which $\mathcal{N}_r$, j_ω and α_ω are all constants. In this case the beam will suffer no refraction. This means that the rays are straight lines. For frequencies well above characteristic frequencies of the medium, such as the plasma and gyro frequencies, the ray refractive index is also close to unity, $\mathcal{N}_r \approx 1$, and may therefore be discarded. The spesific intensity at an observer at position s then reduces to the simple form

$$\mathcal{I}_\omega(s) = \mathcal{I}_\omega(0)\exp(-\Delta\tau_\omega) + S_\omega\left(1 - \exp(-\Delta\tau_\omega)\right) \tag{1.130}$$

in terms of the corresponding quantity incident at $s = 0$ and the optical depth difference $\Delta\tau_\omega = \tau_\omega(0) - \tau_\omega(s)$ between positions $s = 0$ and s.

For the case of pure extinction ($j_\omega = 0$), the source term in (1.130) vanishes. The result allows for the physical interpretation: the probability that a given photon incident at optical depth τ_ω will make it to the observer before being removed from the beam is proportional to $\exp(-\tau_\omega)$. The *mean free optical depth* for photons arriving at the observer is therefore

$$\langle\,\tau_\omega\,\rangle \equiv \frac{\int_0^\infty \tau_\omega \exp(-\tau_\omega)\,\mathrm{d}\tau_\omega}{\int_0^\infty \exp(-\tau_\omega)\,\mathrm{d}\tau_\omega} = 1. \tag{1.131}$$

The corresponding *mean free path length* is

$$l_\omega = \frac{\langle\,\tau_\omega\,\rangle}{\alpha_\omega} = \frac{1}{\alpha_\omega}. \tag{1.132}$$

The inverse linear extinction coefficient α_ω thus represents the average distance a photon with the specified frequency will survive in the medium before it suffers a destruction or modification process.

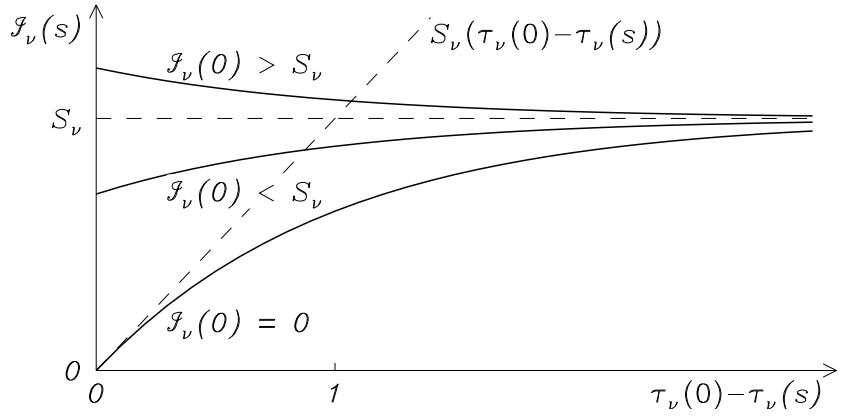

Figure 1.19: Solution of the radiative transport equation for homogeneous media

Solutions of (1.130) for different values of $\mathcal{I}_\omega(0)$ compared to S_ω are illustrated in figure 1.19. For an initial intensity $\mathcal{I}_\omega(0)$ less than the source function S_ω the intensity will increase along the ray, and vice versa. After passing an *optically thin medium*, $\tau_\omega \ll 1$, the incident intensity still shines through. In this limit the medium is not able to modify the incident radiation significantly. In the opposite limit, after traversing an *optically thick medium*, $\tau_\omega \gg 1$, the resulting intensity only depends on the source function of the medium. The incident population of photons in the beam is completely extinct, being scattered out of the beam or absorbed by the medium. The photons reaching the observer have all been generated in the medium, propagating along the beam.

At this point some general comments on the radiative transport equation and its properties will be appropriate. The equation assumes that macroscopic quantities like emissivity j_ω and extinction coefficient α_ω can be properly defined for the medium through which the radiation is propagating. Our next task will therefore be to discuss if or how these quantities may be determined in terms of microscopic processes relating to individual charged particles, atoms or molecules in the medium. Some of these processes may be given an adequate description in terms of classical physics, others will require a quantum mechanical treatment. For the time being we note that these processes may be conveniently divided into different main classes: absorption, scattering and conversion. The former class is characterized by photons being absorbed by the medium, leading to an increase in thermal or internal energy of the medium, later possibly being reemitted as new photons. In a scattering process the medium is left largely unmodified, but the directions of the involved photons are changed. For the latter process a significant frequency change of the involved photons is expected.

Quiz 1.37 : What are the physical units of the emissivity j_ω and the different forms of the extinction coefficient: α_ω, σ_ω or κ_ω?

Quiz 1.38 : It is sometimes convenient to express the radiative transport equations in terms of *optical path length* τ_ω defined through $\mathrm{d}\tau_\omega = +\alpha_\omega \mathrm{d}s$. What is the form of the radiative transport equation (1.128) and its formal solution (1.129) in term of this variable?

Chapter 2

Electromagnetic Radiation

In the presence of material media, for example a neutral or an ionized gas, an electromagnetic wave will be modified. Wave energy may be generated or lost in emission or absorption processes, and the wave may suffer diffraction or refraction effects. Emission and absorption processes are characterized as *bound-bound*, *bound-free* or *free-free* transitions according to if the electrons involved in the process before and after the transition are in bound or free states. A discussion of these basic processes involving bound states will require quantum mechanics and will be the subject of the following chapters. In the present chapter we discuss free-free transitions for which classical electromagnetic theory will suffice.

2.1 Electromagnetic Potentials

In chapter 1 some general properties of the Maxwell equations (1.1)-(1.4) were discussed. As we can see directly from these equations, the sources of electric and magnetic fields are electric current and charge densities (free current and charge densities, polarization current and charge densities or magnetization current density). To find the fields generated from an arbitrary distribution of currents and charges we therefore need to establish the general solution of the inhomogeneous Maxwell equations. This will therefore be our first task. To this end we first introduce the electromagnetic scalar and vector potentials $\Phi(\boldsymbol{r}, t)$ and $\boldsymbol{A}(\boldsymbol{r}, t)$.

Equation (1.3) is trivially satisfied if the magnetic field $\boldsymbol{B}$ is derived as the curl of a vector potential $\boldsymbol{A}$,

$$\boldsymbol{B} = \nabla \times \boldsymbol{A} \tag{2.1}$$

Substituting this expression into (1.2) leads to $\nabla \times (\boldsymbol{E} + \partial \boldsymbol{A}/\partial t) = 0$. Thus also equation (1.2) will be automatically satisfied if the electric field $\boldsymbol{E}$ is derived from a scalar potential Φ as

$$\boldsymbol{E} = -\nabla\Phi - \frac{\partial \boldsymbol{A}}{\partial t}. \tag{2.2}$$

It remains to find solutions for Φ and $\boldsymbol{A}$ for arbitrary charge and current densities ρ and $\boldsymbol{j}$ such that also (1.1) and (1.4) are satisfied. Substitution of (2.1) and (2.2) into these equations leads to

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\Phi + \frac{\partial}{\partial t}\left(\nabla \cdot \boldsymbol{A} + \frac{1}{c^2}\frac{\partial \Phi}{\partial t}\right) = -\frac{1}{\epsilon_0}\rho \tag{2.3}$$

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\boldsymbol{A} - \nabla\left(\nabla \cdot \boldsymbol{A} + \frac{1}{c^2}\frac{\partial \Phi}{\partial t}\right) = -\mu_0 \boldsymbol{j}. \tag{2.4}$$

In the first of these equation we have taken the liberty to add and subtract identical terms.

To produce a given set of electric and magnetic fields, the electromagnetic potentials Φ and $\boldsymbol{A}$ may always be chosen such that

$$\nabla \cdot \boldsymbol{A} + \frac{1}{c^2}\frac{\partial \Phi}{\partial t} = 0. \tag{2.5}$$

In fact, for given electromagnetic fields $\boldsymbol{E}$ and $\boldsymbol{B}$ an infinite set of different potentials Φ and $\boldsymbol{A}$ may be found. If Φ' and $\boldsymbol{A}'$ is one "proper" set of potentials, that is, potentials that leads to the correct fields $\boldsymbol{E}$ and $\boldsymbol{B}$, then

$$\Phi = \Phi' - \frac{\partial f}{\partial t}, \quad \boldsymbol{A} = \boldsymbol{A}' + \nabla f,$$

where f is an arbitrary function of $\boldsymbol{r}$ and t, is another proper set of potentials. An $f(\boldsymbol{r}, t)$ may always be found such that Φ and $\boldsymbol{A}$ satisfy (2.5). Electromagnetic potentials Φ and $\boldsymbol{A}$ chosen such as to satisfy (2.5) are said to satisfy the *Lorentz gauge*. Under Lorentz gauge the electromagnetic potentials satisfy the inhomogeneous wave equations

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\Phi = -\frac{1}{\epsilon_0}\rho \tag{2.6}$$

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\boldsymbol{A} = -\mu_0 \boldsymbol{j}. \tag{2.7}$$

A complete solution of (2.6)-(2.7) consists of a sum of a particular integral of the inhomogeneous equations and a general solution of the corresponding homogeneous equations. A physically acceptable particular integral is

$$\Phi(\boldsymbol{r}, t) = \frac{1}{4\pi\epsilon_0}\int \mathrm{d}^3\boldsymbol{r}' \, \frac{\rho(\boldsymbol{r}', t')}{|\,\boldsymbol{r} - \boldsymbol{r}'\,|} \tag{2.8}$$

$$\boldsymbol{A}(\boldsymbol{r}, t) = \frac{\mu_0}{4\pi}\int \mathrm{d}^3\boldsymbol{r}' \, \frac{\boldsymbol{j}(\boldsymbol{r}', t')}{|\,\boldsymbol{r} - \boldsymbol{r}'\,|}, \tag{2.9}$$

where the "retarded" time t' is defined by

$$t' = t - \frac{1}{c}\,|\,\boldsymbol{r} - \boldsymbol{r}'\,|\,. \tag{2.10}$$

The potentials (2.8) and (2.9), which are called the *retarded potentials*, satisfy the *causality principle*, the potentials at the field point $\boldsymbol{r}$ at time t reflects the charge and current densities at the source point $\boldsymbol{r}'$ at an earlier time t' determined such that information of these source densities propagating at the speed of light may just reach the field point $\boldsymbol{r}$ at the given time t. The geometry is illustrated in figure 2.1.

To prove that the retarded potentials, as given by (2.8) and (2.9), are proper solutions of the Maxwell equations it must be demonstrated that both the Lorentz condition (2.5) and the inhomogeneous wave equations (2.6) and (2.7) are satisfied. This is easily done through simple substitution. For the Lorentz condition we thus find

$$\begin{aligned}\nabla \cdot \boldsymbol{A} + \frac{1}{c^2}\frac{\partial \Phi}{\partial t} &= \frac{\mu_0}{4\pi}\int \mathrm{d}^3\boldsymbol{r}' \left\{-\nabla'\frac{1}{|\,\boldsymbol{r} - \boldsymbol{r}'\,|}\cdot \boldsymbol{j}(\boldsymbol{r}', t') + \left(\nabla t' \cdot \frac{\partial \boldsymbol{j}}{\partial t'} + \frac{\partial \rho}{\partial t'}\right)\frac{1}{|\,\boldsymbol{r} - \boldsymbol{r}'\,|}\right\} \\ &= \frac{\mu_0}{4\pi}\int \mathrm{d}^3\boldsymbol{r}' \left\{-\nabla' \cdot \frac{\boldsymbol{j}(\boldsymbol{r}', t')}{|\,\boldsymbol{r} - \boldsymbol{r}'\,|} + \left(\nabla'_{t'} \cdot \boldsymbol{j}(\boldsymbol{r}', t') + \nabla' t' \cdot \frac{\partial \boldsymbol{j}}{\partial t'} + \nabla t' \cdot \frac{\partial \boldsymbol{j}}{\partial t'} + \frac{\partial \rho}{\partial t'}\right)\frac{1}{|\,\boldsymbol{r} - \boldsymbol{r}'\,|}\right\}.\end{aligned}$$

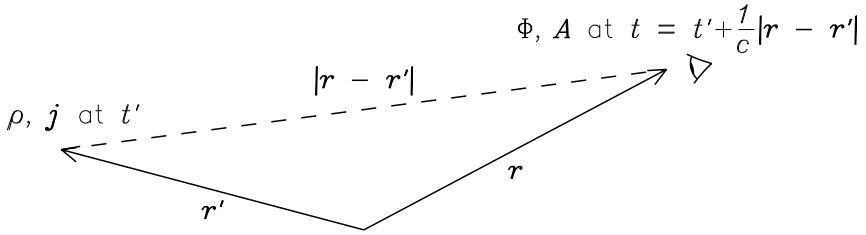

Figure 2.1: Causally connected source and field points

The symbol $\nabla'_{t'}$ here means derivation with respect to $\boldsymbol{r}'$ while keeping t' constant. The first term on the right hand side vanishes due to the Gauss integral theorem for spatially limited current densities. The third and fourth terms mutually cancel, while the sum of the second and fifth terms vanishes due to the conservation theorem for charge and current densities (1.12). Thus, the Lorentz condition (2.5) is indeed satisfied.

For the wave equation (2.6) we find

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\Phi = \frac{1}{4\pi\epsilon_0}\int \mathrm{d}^3\boldsymbol{r}'\left\{\rho\nabla^2\frac{1}{|\boldsymbol{r}-\boldsymbol{r}'|} + 2\nabla\frac{1}{|\boldsymbol{r}-\boldsymbol{r}'|}\cdot\nabla t'\frac{\partial\rho}{\partial t'}\right.$$
$$\left. + \left(\nabla\cdot(\nabla t'\frac{\partial\rho}{\partial t'}) - \frac{1}{c^2}\frac{\partial^2\rho}{\partial t'^2}\right)\frac{1}{|\boldsymbol{r}-\boldsymbol{r}'|}\right\}.$$

Noting that

$$\nabla^2\frac{1}{|\boldsymbol{r}-\boldsymbol{r}'|} = -4\pi\delta(\boldsymbol{r}-\boldsymbol{r}'),$$

where $\delta(\boldsymbol{r})$ is the Dirac δ–function (see section A.9), the first term on the right hand side provides for the desired right hand value. The remaining terms on the right hand side cancel identically. This is seen by making use of the identity

$$\nabla\cdot(\nabla t'\frac{\partial\rho}{\partial t'}) = \frac{1}{c}\frac{\partial\rho}{\partial t'}\nabla\cdot\frac{\boldsymbol{r}-\boldsymbol{r}'}{|\boldsymbol{r}-\boldsymbol{r}'|} + \frac{1}{c^2}\frac{\partial^2\rho}{\partial t'^2}.$$

and the results of section A.4. This result automatically implies that also each Cartesian component of the vector wave equation (2.7) are satisfied. Thus, we have shown that the retarded potentials (2.8) and (2.9), with the electromagnetic fields defined by (2.1) and (2.2), is the proper solution of the Maxwell equations for arbitrary sources ρ and $\boldsymbol{j}$.

Quiz 2.1 : Verify that a Lorenz gauge may always be found.

Quiz 2.2 : Verify by substitution that the vector potential $\boldsymbol{A}$ as given by (2.9) satisfies the wave equation (2.7).

Quiz 2.3 : Convince yourself that (2.9)–(2.8) with t' chosen as the "advanced" time $t' = t+ \mid \boldsymbol{r}-\boldsymbol{r}' \mid /c$ instead of (2.10) is also a formal solution of (2.6)–(2.7). Why must this solution be discarded on physical grounds?

2.2 Radiation from Point Charges

A point charge q moving along a trajectory $\boldsymbol{r}' = \boldsymbol{R}_q(t')$ with velocity $c\boldsymbol{\beta}(t') = \dot{\boldsymbol{R}}_q(t')$ represents charge and current densities

$$\rho(\boldsymbol{r}', t') = q\delta(\boldsymbol{r}' - \boldsymbol{R}_q(t')) \tag{2.11}$$

$$\boldsymbol{j}(\boldsymbol{r}', t') = qc\boldsymbol{\beta}(t')\delta(\boldsymbol{r}' - \boldsymbol{R}_q(t')). \tag{2.12}$$

The corresponding electromagnetic potentials under Lorentz gauge are found by substituting these source densities in (2.8) and (2.9) and performing the necessary volume integrals. The integrals get non-vanishing contributions only from points $\boldsymbol{r}'$ such that the argument of the δ-function vanishes,

$$\boldsymbol{r}''(\boldsymbol{r}') \equiv \boldsymbol{r}' - \boldsymbol{R}_q(t') = 0, \tag{2.13}$$

where t' is the retarded time as given by (2.10).

The volume integrals in (2.8) and (2.9) are now evaluated by making use of (A.74),

$$\int f(\boldsymbol{r}')\delta(\boldsymbol{r}''(\boldsymbol{r}'))\,\mathrm{d}^3\boldsymbol{r}' = \int \frac{f(\boldsymbol{r}')}{J(\boldsymbol{r}')}\delta(\boldsymbol{r}'')\,\mathrm{d}^3\boldsymbol{r}'' = \sum_i \frac{f(\boldsymbol{r}_i')}{J(\boldsymbol{r}_i')}\bigg|_{\boldsymbol{r}''(\boldsymbol{r}_i')=0} \tag{2.14}$$

where J is the Jacobian (A.75) corresponding to the change of integration variables from $\boldsymbol{r}'$ to $\boldsymbol{r}''$, and the sum on the right hand side of (2.14) indicates that contributions to the integral result from each point $\boldsymbol{r}'$ for which $\boldsymbol{r}'' = 0$. Because the speed of any charge is always less than the speed of light, one and only one such solution of equation (2.10) exists for any given $\boldsymbol{r}$ and t. Some algebra will show that in our case

$$J = \frac{1}{R}(R - \boldsymbol{R}\cdot\boldsymbol{\beta}) \tag{2.15}$$

where $\boldsymbol{R} \equiv \boldsymbol{r} - \boldsymbol{R}_q(t')$ and $R = |\boldsymbol{R}|$. The result of the integration is the *Lienard-Wiechert potentials*

$$\Phi(\boldsymbol{r}, t) = \frac{q}{4\pi\epsilon_0}\frac{1}{R - \boldsymbol{R}\cdot\boldsymbol{\beta}}\bigg|_{t'=t-\frac{1}{c}R(t')} \tag{2.16}$$

$$\boldsymbol{A}(\boldsymbol{r}, t) = \frac{q}{4\pi\epsilon_0 c}\frac{\boldsymbol{\beta}}{R - \boldsymbol{R}\cdot\boldsymbol{\beta}}\bigg|_{t'=t-\frac{1}{c}R(t')}. \tag{2.17}$$

The electric and magnetic fields generated by the point charge can now be found by differentiating the electromagnetic potentials with respect to $\boldsymbol{r}$ and t according to (2.1) and (2.2). To this end we need to remember that t' is a function of both $\boldsymbol{r}$ and t through the relation

$$t' = t - \frac{1}{c}R(t'). \tag{2.18}$$

Thus

$$\nabla = \nabla|_{t'} + \nabla t'\frac{\partial}{\partial t'} \quad \text{and} \quad \frac{\partial}{\partial t} = \frac{\partial t'}{\partial t}\frac{\partial}{\partial t'} \tag{2.19}$$

where $\nabla|_{t'}$ denotes that t' is to be considered constant under the differentiation. The expressions for $\nabla t'$ and $\partial t'/\partial t$ are found with the help of (2.18),

$$\left(1+\frac{1}{c}\frac{\partial R}{\partial t'}\right)\frac{\partial t'}{\partial t}=1 \tag{2.20}$$

$$\left(1+\frac{1}{c}\frac{\partial R}{\partial t'}\right)\nabla t'+\frac{1}{c}\nabla|_{t'}R=0. \tag{2.21}$$

Since

$$\frac{\partial R}{\partial t'}=\frac{\boldsymbol{R}}{R}\cdot\frac{\partial \boldsymbol{R}}{\partial t'}=-c\boldsymbol{\beta}\cdot\frac{\boldsymbol{R}}{R}$$

we find

$$\nabla=\nabla|_{t'}-\frac{1}{c}\frac{\boldsymbol{R}}{R-\boldsymbol{\beta}\cdot\boldsymbol{R}}\frac{\partial}{\partial t'}\quad\text{and}\quad\frac{\partial}{\partial t}=\frac{R}{R-\boldsymbol{\beta}\cdot\boldsymbol{R}}\frac{\partial}{\partial t'}$$

and therefore

$$\boldsymbol{E}(\boldsymbol{r},t)=-\nabla|_{t'}\Phi+\frac{1}{c}\frac{\boldsymbol{R}}{R-\boldsymbol{\beta}\cdot\boldsymbol{R}}\frac{\partial\Phi}{\partial t'}-\frac{R}{R-\boldsymbol{\beta}\cdot\boldsymbol{R}}\frac{\partial\boldsymbol{A}}{\partial t'} \tag{2.22}$$

$$\boldsymbol{B}(\boldsymbol{r},t)=\nabla|_{t'}\times\boldsymbol{A}-\frac{1}{c}\frac{\boldsymbol{R}}{R-\boldsymbol{\beta}\cdot\boldsymbol{R}}\times\frac{\partial\boldsymbol{A}}{\partial t'}. \tag{2.23}$$

The first term on the right hand side of (2.22) and (2.23) fall off as $1/R^2$ at large distances from the point charge. The other terms give contributions that fall off as $1/R$ and will thus dominate at large distances. These dominating parts of the electric and magnetic fields are called the *radiation fields*. We see that the radiation fields satisfy the relation

$$\boldsymbol{B}(\boldsymbol{r},t)=\frac{1}{c}\hat{\boldsymbol{n}}\times\boldsymbol{E}(\boldsymbol{r},t) \tag{2.24}$$

where $\hat{\boldsymbol{n}}=\boldsymbol{R}/R$ is taken at the retarded time. If we also note that

$$\frac{\partial}{\partial t'}\frac{1}{R-\boldsymbol{\beta}\cdot\boldsymbol{R}}=\frac{\dot{\boldsymbol{\beta}}\cdot\boldsymbol{R}}{(R-\boldsymbol{\beta}\cdot\boldsymbol{R})^2}+\mathcal{O}(\frac{1}{R^2}),$$

we find the resulting radiation electric field resulting from the accelerated point charge as

$$\begin{aligned}\boldsymbol{E}(\boldsymbol{r},t)&=\frac{q}{4\pi\epsilon_0 c}\left(\frac{\dot{\boldsymbol{\beta}}\cdot\boldsymbol{R}\,\boldsymbol{R}}{(R-\boldsymbol{\beta}\cdot\boldsymbol{R})^3}-\frac{R\,\dot{\boldsymbol{\beta}}}{(R-\boldsymbol{\beta}\cdot\boldsymbol{R})^2}-\frac{R\,\dot{\boldsymbol{\beta}}\cdot\boldsymbol{R}\,\boldsymbol{\beta}}{(R-\boldsymbol{\beta}\cdot\boldsymbol{R}\,R)^3}\right)\\&=\frac{q}{4\pi\epsilon_0 c}\frac{\hat{\boldsymbol{n}}\times((\hat{\boldsymbol{n}}-\boldsymbol{\beta})\times\dot{\boldsymbol{\beta}})}{(1-\hat{\boldsymbol{n}}\cdot\boldsymbol{\beta})^3R}\bigg|_{t'=t-R(t')/c}.\end{aligned} \tag{2.25}$$

Quiz 2.4 : Show that the Jacobian associated with the coordinate change from $\boldsymbol{r}'$ to $\boldsymbol{r}''$ as defined by (2.13) is given by (2.15). [*Hint:* Choosing one of the coordinate axis parallel with the particle velocity at the instantaneous retarded time simplifies your calculation.]

Quiz 2.5 : Verify (2.24) and (2.25) starting from (2.22) and (2.23).

2.3 Radiated Energy and Power Spectrum

The radiation fields are associated with an instantaneous energy flux given by the Poynting vector

$$\boldsymbol{P}(\boldsymbol{r},t) = \frac{1}{\mu_0}\boldsymbol{E}(\boldsymbol{r},t)\times\boldsymbol{B}(\boldsymbol{r},t) = \epsilon_0 c\hat{\boldsymbol{n}}E^2(\boldsymbol{r},t) \tag{2.26}$$

The energy radiated through a fixed area element $r^2\mathrm{d}\Omega$ in direction $\hat{\boldsymbol{r}}$ per unit time t is given by $\boldsymbol{P}(\boldsymbol{r},t)\cdot\hat{\boldsymbol{r}}r^2\mathrm{d}\Omega$. If the origin is chosen at the particles instantaneous position at time t', then $\hat{\boldsymbol{r}}=\hat{\boldsymbol{n}}$, $R=r$ and the radiated energy per solid angle and unit time is

$$\frac{\mathrm{d}P(t)}{\mathrm{d}\Omega} = \hat{\boldsymbol{n}}\cdot\boldsymbol{P}R^2 = \epsilon_0 c\left(\boldsymbol{E}(\boldsymbol{r},t)R(t')\right)^2 = \left.\frac{q^2}{16\pi^2\epsilon_0 c}\frac{(\hat{\boldsymbol{n}}\times((\hat{\boldsymbol{n}}-\boldsymbol{\beta})\times\dot{\boldsymbol{\beta}}))^2}{(1-\hat{\boldsymbol{n}}\cdot\boldsymbol{\beta})^6}\right|_{t'}. \tag{2.27}$$

To find the corresponding radiated energy per unit eigentime t' of the particle, we must multiply this expression with $\mathrm{d}t/\mathrm{d}t' = 1-\hat{\boldsymbol{n}}\cdot\boldsymbol{\beta}$, that is,

$$\frac{\mathrm{d}P(t')}{\mathrm{d}\Omega} = (1-\hat{\boldsymbol{n}}\cdot\boldsymbol{\beta})\frac{\mathrm{d}P(t)}{\mathrm{d}\Omega} = \left.\frac{q^2}{16\pi^2\epsilon_0 c}\frac{(\hat{\boldsymbol{n}}\times((\hat{\boldsymbol{n}}-\boldsymbol{\beta})\times\dot{\boldsymbol{\beta}}))^2}{(1-\hat{\boldsymbol{n}}\cdot\boldsymbol{\beta})^5}\right|_{t'}. \tag{2.28}$$

For the special case of a non-relativistic charged particle, $\beta\ll 1$, the last result reduces to

$$\frac{\mathrm{d}P(t')}{\mathrm{d}\Omega} = \left.\frac{q^2}{16\pi^2\epsilon_0 c}\left(\hat{\boldsymbol{n}}\times(\hat{\boldsymbol{n}}\times\dot{\boldsymbol{\beta}})\right)^2\right|_{t'}. \tag{2.29}$$

The total radiated energy per unit time from the accelerated point charge is found by integrating (2.28) over all directions. In the non-relativistic limit the result is the familiar Larmor formula

$$P(t') = \frac{q^2}{16\pi^2\epsilon_0 c}\dot{\boldsymbol{\beta}}^2\int\sin^2\theta\,\mathrm{d}\cos\theta\,\mathrm{d}\phi = \frac{q^2\dot{\boldsymbol{\beta}}^2}{6\pi\epsilon_0 c}. \tag{2.30}$$

In the general case the corresponding result can be shown to be

$$P(t') = \frac{q^2}{6\pi\epsilon_0 c}\frac{\dot{\boldsymbol{\beta}}^2-(\boldsymbol{\beta}\times\dot{\boldsymbol{\beta}})^2}{(1-\beta^2)^3}. \tag{2.31}$$

In many applications we are interested not only in the total energy radiated by the accelerated charged particle per unit time, but also the corresponding energy or power spectra, that is, the distribution of the radiated energy as a function of frequency. This can be found by making use of the Parseval theorem (1.47) or (1.51), as outlined in tables 1.1 and 1.2. Three cases may be identified: a radiation pulse extending over a finite time interval, a continuous radiation source observed over a time interval of length T or a periodic source with period T. When the Parseval theorem is applied to (2.27) for a finite radiation pulse the result is

$$\int\frac{\mathrm{d}P(t)}{\mathrm{d}\Omega}\,\mathrm{d}t = \int_0^\infty\frac{\mathrm{d}\mathcal{E}(\omega)}{\mathrm{d}\Omega}\,\mathrm{d}\omega = 4\pi\epsilon_0 c\int_0^\infty\left|\widetilde{[\boldsymbol{E}R]}(\omega)\right|^2\,\mathrm{d}\omega \tag{2.32}$$

where

$$\widetilde{[\boldsymbol{E}R]}(\omega) = \frac{1}{2\pi}\int\boldsymbol{E}(t)R(t')\,\exp(\iota\omega t)\,\mathrm{d}t. \tag{2.33}$$

$\mathrm{d}\mathcal{E}(\omega)/\mathrm{d}\Omega$ represents the energy radiated per unit solid angle and unit angular frequency interval. We shall refer to $\mathrm{d}\mathcal{E}(\omega)/\mathrm{d}\Omega$ as the *energy spectrum* of the received radiation pulse. Notice that we in our notation suppressed the possible directional dependence of the spectrum. For a continuous radiation source observed over a time interval of length T or a periodic source with period T we define instead the *power spectrum* $\mathrm{d}\mathcal{P}(\omega)/\mathrm{d}\Omega$ through

$$\frac{1}{T}\int_T \frac{\mathrm{d}P(t)}{\mathrm{d}\Omega}\,\mathrm{d}t = \int_0^\infty \frac{\mathrm{d}\mathcal{P}(\omega)}{\mathrm{d}\Omega}\,\mathrm{d}\omega \tag{2.34}$$

where for the continuous case

$$\frac{\mathrm{d}\mathcal{P}(\omega)}{\mathrm{d}\Omega} = \frac{1}{T}\frac{\mathrm{d}\mathcal{E}(\omega)}{\mathrm{d}\Omega} = \frac{4\pi\epsilon_0 c}{T}\left|\widetilde{[\boldsymbol{E}R]}_T(\omega)\right|^2 \tag{2.35}$$

with

$$\widetilde{[\boldsymbol{E}R]}_T(\omega) = \frac{1}{2\pi}\int_T \boldsymbol{E}(t)R(t')\exp(\iota\omega t)\,\mathrm{d}t, \tag{2.36}$$

and for the periodic case

$$\frac{\mathrm{d}\mathcal{P}(\omega)}{\mathrm{d}\Omega} = \sum_{n=1}^\infty \mathcal{P}_n\delta(\omega - \frac{2\pi n}{T}) = 2\epsilon_0 c\sum_{n=1}^\infty \left|\widetilde{[\boldsymbol{E}R]}_n\right|^2 \delta(\omega - \frac{2\pi n}{T}) \tag{2.37}$$

with

$$\widetilde{[\boldsymbol{E}R]}_n = \frac{1}{T}\int_T \boldsymbol{E}(t)R(t')\exp(\iota\frac{2\pi n}{T}t)\,\mathrm{d}t. \tag{2.38}$$

The expressions for the energy spectra, (2.32), (2.35) or (2.37), are complicated by the fact that the radiation field $\boldsymbol{R}(t)$ in (2.33), (2.36) or (2.38) depends on the retarded time t'. The expressions may, however, be simplified by making use of the far field approximation. The *far field approximation*, illustrated in figure 2.2, amounts to placing the coordinate origin in the neighborhood of the radiating particle and assuming the relevant orbit excursions to be small compared to the distance to the observer. This means that we may consider $\hat{\boldsymbol{n}}$ to be constant and approximate

$$R(t') = |\,\boldsymbol{r} - \boldsymbol{R}_q(t')\,| \approx r - \hat{\boldsymbol{n}}\cdot\boldsymbol{R}_q(t')$$

and therefore

$$t = t' + \frac{1}{c}R(t') \approx \frac{r}{c} + t' - \frac{1}{c}\hat{\boldsymbol{n}}\cdot\boldsymbol{R}_q(t').$$

When substituting this approximation into (2.33) or (2.36) and changing integration variable from t to t', that is, $\mathrm{d}t = (1 - \hat{\boldsymbol{n}}\cdot\boldsymbol{\beta})\,\mathrm{d}t'$, the resulting energy spectrum is

$$\frac{\mathrm{d}\mathcal{E}(\omega)}{\mathrm{d}\Omega} = \frac{q^2}{16\pi^3\epsilon_0 c}\left|\int \frac{\hat{\boldsymbol{n}}\times((\hat{\boldsymbol{n}}-\boldsymbol{\beta})\times\dot{\boldsymbol{\beta}})}{(1-\hat{\boldsymbol{n}}\cdot\boldsymbol{\beta})^2}\,\exp\iota\omega(t' - \frac{1}{c}\hat{\boldsymbol{n}}\cdot\boldsymbol{R}_q(t'))\,\mathrm{d}t'\right|^2 \tag{2.39}$$

where the time integral is to be taken over the retarded time interval corresponding to the observational interval.

The polarization of the emitted radiation is specified by the direction of the vector integral in (2.39). The intensity of radiation in a given polarization can thus be obtained by taking the scalar product of the desired unit polarization vector with the vector integrals before forming their absolute squares.

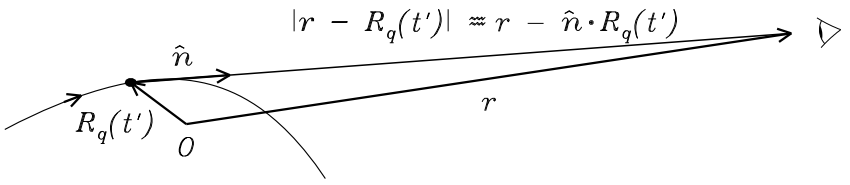

Figure 2.2: The far field approximation

Quiz 2.6 : The result (2.29) is sometimes referred to as the dipole radiation approximation. Can you explain why?

Quiz 2.7 : Check that (2.35) and (2.37) have identical physical units.

Quiz 2.8 : Compare the quantities $\mathcal{P}(\omega)$ of (1.69) and $\mathrm{d}\mathcal{P}(\omega)/\mathrm{d}\Omega$ of (2.32). What are their similarities and differences?

Quiz 2.9 : Verify (2.31). [*Hint:* Choose coordinate axis such that $\boldsymbol{\beta} \parallel (0,0,1)$ and $\dot{\boldsymbol{\beta}} \parallel (\sin\psi, 0, \cos\psi)$. The ϕ-integration is trivial, the θ-integration leads to integrals of the form

$$I_n = \int_{-1}^{1} \frac{\mathrm{d}\cos\theta}{(1-\beta\cos\theta)^n}$$

with

$$I_3 = \frac{2}{(1-\beta^2)^2}, \quad I_4 = \frac{2(3+\beta^2)}{3(1-\beta^2)^3}, \quad I_5 = \frac{2(1+\beta^2)}{(1-\beta^2)^4}.]$$

Quiz 2.10 : The equation of motion for a non-relativistic electron spiraling in a constant magnetic field $\boldsymbol{B}_0$, $m\dot{\boldsymbol{v}} = q\boldsymbol{v}\times\boldsymbol{B}_0$, implies that the kinetic energy of the electron remains constant, $\frac{1}{2}m\boldsymbol{v}^2 =$ constant. Can you reconcile this result with Larmor's formula (2.30)? What fraction of the (perpendicular) kinetic energy of the electron will be radiated away per gyro-period in a magnetic field of 1T?

Quiz 2.11 : Verify the identity

$$\frac{\hat{\boldsymbol{n}}\times((\hat{\boldsymbol{n}}-\boldsymbol{\beta}(t'))\times\dot{\boldsymbol{\beta}}(t'))}{(1-\hat{\boldsymbol{n}}\cdot\boldsymbol{\beta}(t'))^2} = \frac{\mathrm{d}}{\mathrm{d}t'}\left(\frac{\hat{\boldsymbol{n}}\times(\hat{\boldsymbol{n}}\times\boldsymbol{\beta}(t'))}{1-\hat{\boldsymbol{n}}\cdot\boldsymbol{\beta}(t')}\right). \tag{2.40}$$

Making use of this identity and an integration by parts, show that (2.39) for the special case of periodic orbits may be transformed into

$$\frac{\mathrm{d}\mathcal{E}(\omega)}{\mathrm{d}\Omega} = \frac{q^2\omega^2}{16\pi^3\epsilon_0 c}\left|\int \hat{\boldsymbol{n}}\times(\hat{\boldsymbol{n}}\times\boldsymbol{\beta})\,\exp\iota\omega(t'-\frac{1}{c}\hat{\boldsymbol{n}}\cdot\boldsymbol{R}_q(t'))\,\mathrm{d}t'\right|^2. \tag{2.41}$$

What requirements must the integration limits satisfy for this result to be valid?

Quiz 2.12 : By writing the radiation loss of a non-relativistic accelerated charged particle as $P(t') = -\boldsymbol{v}(t')\cdot\boldsymbol{F}_{react}(t')$, show that the accumulated radiation loss $\int P(t')\,\mathrm{d}t'$ can be accounted for with a *reaction force*

$$\boldsymbol{F}_{react}(t') = \frac{q^2}{6\pi\epsilon_0 c^3}\ddot{\boldsymbol{v}}(t'), \tag{2.42}$$

and a "corrected" equation of motion

$$m(\dot{\boldsymbol{v}} - \tau_0 \ddot{\boldsymbol{v}}) = \boldsymbol{F} \quad \text{with} \quad \tau_0 = \frac{q^2}{6\pi\epsilon_0 c^3 m}. \tag{2.43}$$

Compare τ_0 for an electron with the time needed for a photon to travel a distance equal to radius of a nucleus with a cross-section of 1 barn $\equiv 10^{-28}$ m^2.
With $\boldsymbol{F} = 0$ show that (2.43) allows for two different types of solution: $\dot{\boldsymbol{v}}(t') = 0$ and $\dot{\boldsymbol{v}}(t') \sim \exp(t/\tau_0)$. Why is the latter one refered to as a "run away" solution?
The "run away" type can be suppressed by writing the solution of (2.43) as

$$m\dot{\boldsymbol{v}}(t') = \int_0^\infty \mathrm{d}x \, \exp(-x) \, \boldsymbol{F}(t' + \tau_0 x).$$

Verify this solution. Do you see any problem with the causality principle?

2.4 Applications

With the results of section 2.3 we now have at our disposal the tools needed to discuss a suite of radiation problems that can be treated within classical physics. We therefore next turn to some of the most important applications. Common to these applications is that free, accelerated electrons are the source of the emitted radiation.[1] What will be different are the mechanisms responsible for the acceleration of the electron. We shall consider three important examples.

a) In a binary collision between an electron and an atom or an ion, the electron will suffer acceleration. The resulting electromagnetic radiation is referred to as Bremsstrahlung. We shall consider the case of a collision between a non-relativistic electron and an ion.

b) Magnetic fields are frequently encountered in astrophysics. An electron moving in such a field will follow a spiraling path. The resulting radiation is denoted cyclotron radiation. We consider the general case of a relativistic electron.[2]

c) Finally, the electric field of a propagating electromagnetic wave will set free electrons that it encounters into oscillations. The resulting radiation will act to scattered the energy in the original wave into other directions. The process is referred to as Thomson scattering.

Also other situations can be envisioned where free electrons are accelerated and thus give rise to electromagnetic radiation. Thus the collision between a photon and an electron give rise to an effect known as the Compton effect. Even the acceleration of the electron in a nuclear beta-decay will result in the emission of radiation. The latter two examples will require a quantum mechanical treatment.

2.4.1 Bremsstrahlung

Let us first conclude that in the collision between two non-relativistic charged particles with identical charge to mass ratio, no radiation is emitted. For this case we substitute the total electric current from the two particles in (2.29). This will only require $q\dot{\boldsymbol{\beta}}$ to be replaced with

$$\sum_i q_i \dot{\boldsymbol{\beta}}_i = \frac{q_i}{m_i} \frac{\mathrm{d}}{\mathrm{d}t} \sum_i m_i \boldsymbol{\beta}_i.$$

[1] Due to their much large mass and therefore much smaller accelerations, ions usually contribute insignificantly to the radiation

[2] For relativistic electrons this radiation is sometimes referred to as synchrotron radiation

From momentum conservation in the collision this expression vanishes and we have proved our statement. For relativistic electrons the result will be different. In fact, when the kinetic energy of the electrons are of the order of the rest mass energy m_0c^2, higher order quadrupole radiation from electron-electron encounters are comparable to that of electron-ion collisions.

We now consider the encounter of a non-relativistic electron with a much heavier ion. To lowest order we may consider the ion to be a fixed force center during the collision[3]. If the radiation energy loss suffered by the electron during the encounter is small compared with its kinetic energy, the electron will in the Coulomb force field of the ion experience an acceleration,

$$\dot{\boldsymbol{v}} = -\frac{Ze^2}{4\pi\epsilon_0 m}\frac{\boldsymbol{r}}{r^3}, \tag{2.44}$$

and therefore follow a hyperbolic orbit as illustrated in figure 2.3,

$$r(\psi) = \frac{D}{1+\cos\psi/\sin\frac{\chi}{2}}, \tag{2.45}$$

where ψ is the angle between $\boldsymbol{r}$ and the pericentrum direction $\hat{\boldsymbol{\mathcal{P}}}$. We note that $r(\psi)\to\infty$ for $\psi\to\pm(\chi+\pi)/2$.

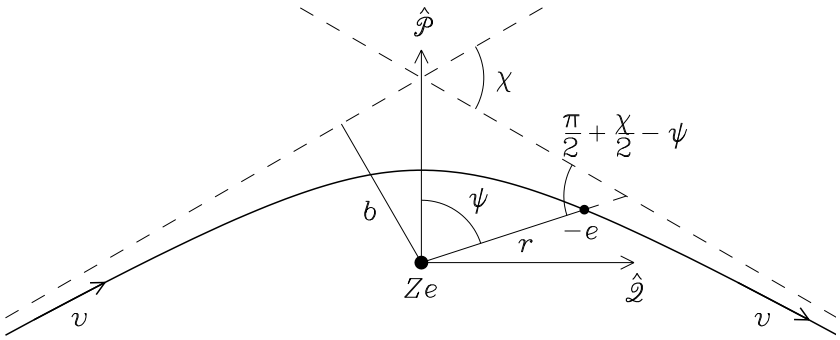

Figure 2.3: Hyperbolic encounter between electron and positive ion

The constant D and the total deflection χ suffered by the electron during the encounter are related to the electron speed v and the impact parameter b. The relationship is easily found by noting that

$$b = \lim_{\psi\to(\chi+\pi)/2} r(\psi)\sin(\frac{\pi}{2}+\frac{\chi}{2}-\psi) = D\tan\frac{\chi}{2},$$

and then making use of the angular momentum and energy conservation laws

$$\begin{gathered} bv = r^2\dot{\psi} = r_0v_0 \\ \frac{1}{2}mv^2 = \frac{1}{2}m(\dot{r}^2+r^2\dot{\psi}^2) - \frac{Ze^2}{4\pi\epsilon_0 r} = \frac{1}{2}mv_0^2 - \frac{Ze^2}{4\pi\epsilon_0 r_0}. \end{gathered} \tag{2.46}$$

[3]The effects of a finite ion mass M are easily taken into account by considering the equivalent one-body problem (see section 3.4) with reduced mass $\mu = mM/(m+M)$. Notice that $m\boldsymbol{r} = \mu\rho$ where $\rho = \boldsymbol{r}-\boldsymbol{R}$ is the position of the electron relative to the ion and we have chosen $\boldsymbol{R}_{CM} = (m\boldsymbol{r}+M\boldsymbol{R})/(m+M) = 0$.

Here $r_0 = r(\psi = 0)$ and v_0 are the pericentrum distance and the corresponding orbital speed. The result is

$$\tan\frac{\chi}{2} = \frac{b_0}{b} \quad \text{with} \quad b_0 = \frac{Ze^2}{4\pi\epsilon_0 m v^2}. \tag{2.47}$$

We note that b_0 is the impact parameter that leads to a 90° deflection.

Let us now return to our radiation problem. The instantaneous power loss by the non-relativistic electron is found by substituting (2.44) into Larmor's formula (2.30)

$$P(t') = \frac{e^2}{6\pi\epsilon_0 c^3}\left(\frac{Ze^2}{4\pi\epsilon_0 m}\right)^2 \frac{1}{r^4(t')} \tag{2.48}$$

The total energy loss during one encounter is given by

$$W = \int P(t')\,\mathrm{d}t' = \frac{e^2}{6\pi\epsilon_0 c^3}\left(\frac{Ze^2}{4\pi\epsilon_0 m}\right)^2 \int_{-\infty}^{\infty} \frac{\mathrm{d}t'}{r(t')^4} = \frac{e^2}{6\pi\epsilon_0 c^3}\left(\frac{Ze^2}{4\pi\epsilon_0 m}\right)^2 2\int_{r_0}^{\infty} \frac{\mathrm{d}r}{r^4\,\dot{r}}.$$

By substituting

$$\dot{r}^2 = v^2 + \frac{2Ze^2}{4\pi\epsilon_0 m}\frac{1}{r} - \frac{b^2v^2}{r^2}$$

from (2.46), the latter integral may be preformed analytically with the result

$$W = \frac{1}{2}mv^2\,\frac{4}{3Z}\left(\frac{v\,b_0}{c\,b}\right)^3\left(3\frac{b_0}{b} + \left(1 + 3\frac{b_0^2}{b^2}\right)\left(\frac{\pi}{2} + \arctan\frac{b_0}{b}\right)\right). \tag{2.49}$$

The result is seen to diverge in the small impact parameter limit. This is only to signal that in this limit quantum mechanical effects will be dominating. The electron is not allowed in a single encounter to radiate away more than its kinetic energy. We also remember that our result was derived in the non-relativistic limit, $v/c \ll 1$. In figure 2.4 the ratio of total radiated energy per encounter and the non-relativistic kinetic energy of the electron is plotted as a function of v/c and b_0/b. For the validity of our theory this ratio must remain small.

We next need to investigate the spectral distribution of this total energy loss. The term $\hat{\boldsymbol{n}} \cdot \boldsymbol{R}_q(t')/c$ in the exponent of (2.39) is of the order $\beta t'$ during the encounter and may for this application be neglected together with the other terms containing $\boldsymbol{\beta}$. The non-relativistic limiting form of (2.39),

$$\frac{\mathrm{d}\mathcal{E}(\omega)}{\mathrm{d}\Omega} = \frac{e^2}{16\pi^3\epsilon_0 c^3}\left|\hat{\boldsymbol{n}} \times \left(\hat{\boldsymbol{n}} \times \int_{-\infty}^{\infty} \dot{\boldsymbol{v}}\,\exp(\iota\omega t)\,\mathrm{d}t\right)\right|^2,$$

reduces after an integration over directions to

$$\mathcal{E}(\omega) = \frac{e^2}{6\pi^2\epsilon_0 c^3}\left|\int_{-\infty}^{\infty} \dot{\boldsymbol{v}}\,\exp(\iota\omega t)\,\mathrm{d}t\right|^2. \tag{2.50}$$

For an order of magnitude estimate of the effective duration τ of the encounter we may write $\tau = b/v$. In the high-frequency end of the spectrum, $(\omega\tau \gg 1)$, the exponential in (2.50) oscillates rapidly compared with the variation in $\dot{\boldsymbol{v}}$. The time integral and thus also the

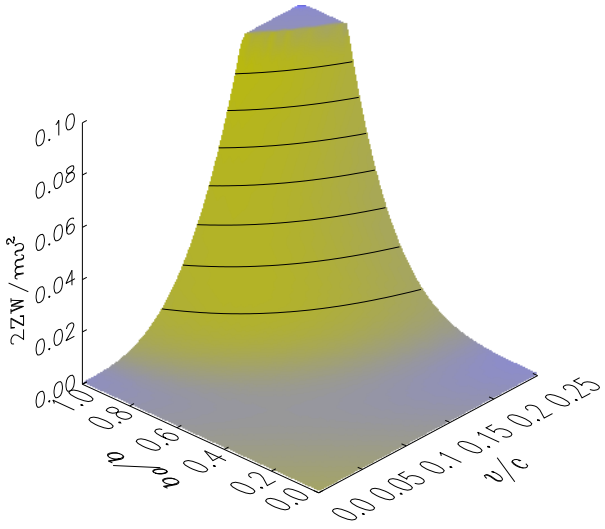

Figure 2.4: Total energy radiated during one encounter

radiation spectrum will therefore take small values. In the low frequency end of the spectrum ($\omega\tau \ll 1$), however, the exponential in (2.50) is approximately constant and therefore

$$\mathcal{E}(\omega) \approx \frac{e^2}{6\pi^2\epsilon_0 c^3} \, |\,\Delta \boldsymbol{v}\,|^2$$

where $\Delta\boldsymbol{v}$ is the vectorial change in $\boldsymbol{v}$ during the encounter. As seen from figure 2.3 and (2.47)

$$|\,\Delta \boldsymbol{v}\,|^2 = 4v^2 \sin^2\frac{\chi}{2} = 4v^2 \frac{b_0^2}{b_0^2 + b^2}.$$

The limiting forms of the Bremsstrahlung spectrum will therefore be

$$\mathcal{E}(\omega) \approx \begin{cases} \dfrac{2e^2v^2}{3\pi^2\epsilon_0 c^3}\dfrac{b_0^2}{b_0^2+b^2} & \omega\tau \ll 1 \\ 0 & \omega\tau \gg 1 \end{cases} \tag{2.51}$$

The change-over between the low and high frequency limits is expected near $\omega = v/b$.

For a quantitative discussion of the Bremsstrahlung radiation spectrum the effect of the exponential factor in (2.50) needs to be taken into account. To this end the expression for $\dot{\boldsymbol{v}}$ for the hyperbolic orbit may be expressed in terms of true anomaly F,[4]

$$\begin{aligned} \boldsymbol{r} &= b_0\left(\hat{\boldsymbol{\mathcal{P}}}(\eta - \cosh F) + \sqrt{\eta^2-1}\,\hat{\boldsymbol{\mathcal{Q}}}\sinh F\right) \\ t &= \frac{b_0}{v}(\eta \sinh F - F) \end{aligned} \tag{2.52}$$

where $\eta = \sqrt{1 + b^2/b_0^2}$ and $\hat{\boldsymbol{\mathcal{P}}}$ and $\hat{\boldsymbol{\mathcal{Q}}}$ are orthogonal unit vectors in the orbital plane of the electron, the former pointing toward pericentrum. The result after substitution in (2.50) is

$$\mathcal{E}(\omega) = \frac{2e^2v^2}{3\pi^2\epsilon_0 c^3}\frac{b_0^2}{b^2+b_0^2} H\left(\eta, \frac{\omega b_0}{v}\right) \tag{2.53}$$

[4]The correctness of expression (2.52) is easily checked by substitution into (2.44).

where the "correctional" factor

$$H\left(\eta, \frac{\omega b_0}{v}\right) = A_{\mathcal{P}}^2 + A_{\mathcal{Q}}^2 \tag{2.54}$$

with

$$A_{\mathcal{P}} = \eta \int_0^\infty \frac{\eta - \cosh F}{(\eta \cosh F - 1)^2} \cos \omega t \, \mathrm{d}F \tag{2.55}$$

$$A_{\mathcal{Q}} = \eta\sqrt{\eta^2 - 1} \int_0^\infty \frac{\sinh F}{(\eta \cosh F - 1)^2} \sin \omega t \, \mathrm{d}F. \tag{2.56}$$

$A_{\mathcal{P}}$ and $A_{\mathcal{Q}}$ are functions of η and $\omega b_0/v$ that need to be evaluated numerically. In figure 2.5a the "correctional" factor H has been plotted as a function of $\omega b_0/v$ for different impact parameter ratios b_0/b. For $\omega b_0/v \ll 1$ the "correctional" factor reduces to unity as expected. For small impact parameters b, however, the radiation spectrum is significantly enhanced and shifted toward higher frequencies. The shift toward higher radiation frequencies with decreasing impact parameter b is a natural consequence of the fact that close encounters have the shorter effective collision time. We should, however, keep in mind that in this limit the validity of our derivation is being challenged by quantum mechanical effects.

The result (2.53) refers to the radiation spectrum produced by a single encounter of an electron moving with speed v with one ion for an impact parameter b. To find the radiation emitted for an average electron per unit time in a plasma with N_i ions per unit volume, we have to multiply with $N_i v \, 2\pi b \, \mathrm{d}b$, integrate over b, and finally average over the distribution $f_{\boldsymbol{r}v}$ of electron speeds. For local thermal equilibrium at temperature $\mathcal{T}$ we have from (6.37)

$$f_{\boldsymbol{r}v}(v) = \left(\frac{m}{2\pi\mathcal{T}}\right)^{3/2} 4\pi v^2 \exp\left(-\frac{mv^2}{2\mathcal{T}}\right).$$

Noting that $b \, \mathrm{d}b = b_0^2 \eta \, \mathrm{d}\eta$ the averaged Bremstrahlung power spectrum per electron is

$$\overline{\mathcal{P}}(\omega) = \frac{8}{\sqrt{\pi}} \frac{e^2\overline{v}^2}{3\pi\epsilon_0 c^3} N_i \pi \bar{b}_0^2 \overline{v} \int_0^\infty \mathrm{d}x^2 \exp(-x^2) G\left(\frac{\omega \bar{b}_0}{\overline{v} x^3}\right) \tag{2.57}$$

with

$$G(y) = \int_1^\infty \frac{\mathrm{d}\eta}{\eta} H(\eta, y). \tag{2.58}$$

In this expression we recognize the second term as the low frequency limit of the energy spectrum of a single encounter for a thermal electron $\overline{v} = \sqrt{2\mathcal{T}/m}$ with impact parameter $\bar{b}_0 = Ze^2/(4\pi\epsilon_0 m\overline{v}^2)$. The third term is the number of ions a thermal electron will encounter per time unit with impact parameter less than $\bar{b}_0$ The G function appearing in the velocity integral is an approximate constant for large value of $\omega b_0/v$, effectively a factor 2, as shown in figure 2.5b. The average one electron power spectrum $\overline{\mathcal{P}}(\omega)$ is therefore approximately flat, contrary to the energy spectrum $\mathcal{E}(\omega)$ for a single encounter. The reason for this fact is that energetic small impact parameter encounters are few and therefore do not dominate the average result.

There are, however, limitations to the validity of this result both at high and low frequencies. At the high frequency end quantum mechanical corrections come into play when the

impact parameter b approaches the de Broglie wavelength of the electron $\lambda_{\rm dB} = h/mv$, that is where the photon energy $\hbar\omega$ approaches the kinetic energy of the electron. In this limit a correctional Gaunt factor will be required, equivalent to increasing the lower limit of the η-integral in (2.58). At the low frequency end the effective Debye shielding of the Coulomb force field of the ion by surrounding electrons in a plasma must be taken into account. This sets an upper limit to the η-integral corresponding to impact parameter equal to the Debye length $\lambda_D = \sqrt{\epsilon_0 \mathcal{T}/(ne^2)}$ where n is the electron density of the plasma.

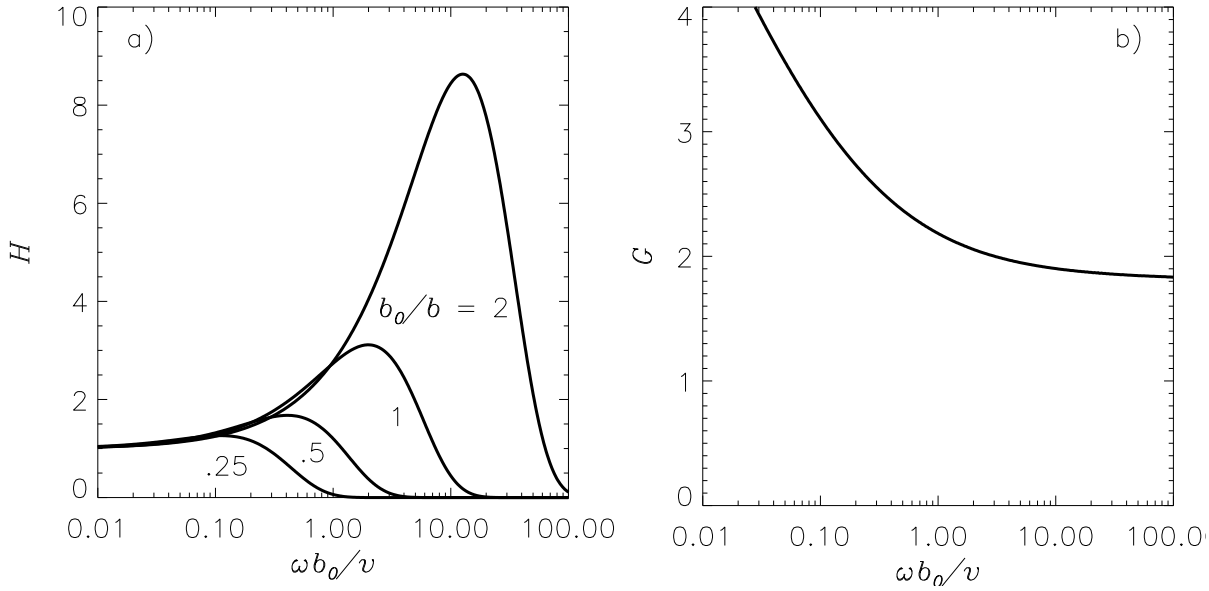

Figure 2.5: Bremsstrahlung from electron-ion encounters: a) relative shape of the "correctional" spectral factor $H(\eta, \omega b_0/v)$, (2.54), for different impact parameter ratios b_0/b, and b) shape factor $G(\omega b_0/v)$, (2.58), for the one electron power spectrum in a thermal plasma.

An important general observation is that the orbital planes for individual encounters in a plasma will be oriented "at random". When adding up the contributions from all encounters we thus expect the total emitted radiation from the plasma to be unpolarized. With respect to polarization, Bremsstrahlung differs significantly from cyclotron radiation to be discussed in the next subsection.

2.4.2 Cyclotron Radiation

We next turn to the radiation resulting from an electron suffering acceleration in a magnetic field. The Lorentz force $-e\boldsymbol{v} \times \boldsymbol{B}_0$ acting on an electron moving in a magnetic field $\boldsymbol{B}_0$ will give rise to a spiraling orbit for the electron. We will refer to the ensuing emitted radiation as cyclotron radiation, irrespective of the electron energy. Due to the common occurrence of magnetic fields in cosmos, cyclotron radiation constitutes a common contributor to the total radiation observed.

In figure 2.6 the *radiation diagram* is given for an electron spiraling in a constant magnetic field $\boldsymbol{B}_0$. The parallel velocity is set to zero. The spiraling electron is indicated in the lower

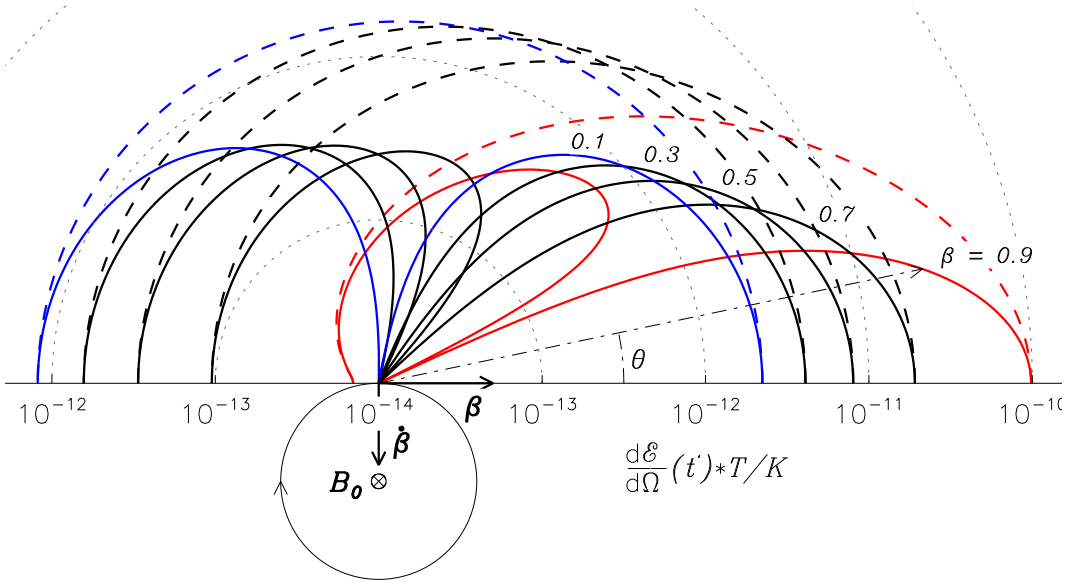

Figure 2.6: Radiation diagram for a spiraling point particle

part of the figure. The plot gives the distribution of the radiated energy per unit solid angle according to (2.28), emitted during one gyro period $T = 2\pi/\omega_B$. The radiated energy is measured in units of the relativistic kinetic energy $K = m_0c^2(\gamma - 1)$. Here $\gamma = 1/\sqrt{1-\beta^2}$, m_0 is the rest mass of the electron and $\omega_B = eB_0/\gamma m_0$. Results are given as a function of the polar angle θ relative to the direction of the instantaneous velocity $\boldsymbol{\beta}$ and for different values of β for a magnetic field strength $B_0 = 1$ T. Full curves correspond to the radiation diagram in the $(\boldsymbol{\beta}, \dot{\boldsymbol{\beta}})$ plane, dashed curves correspond to the radiation diagram in the $(\boldsymbol{\beta}, \boldsymbol{\beta} \times \dot{\boldsymbol{\beta}})$ plane. For clarity the curves for the lowest and highest energies ($\beta = 0.1$ and $\beta = 0.9$) have been given separate colors. Note the logarithmic scale. The polar diagram with the chosen normalization scales linearly with the strength of the magnetic field B_0. If the magnetic field is reduced from 1T to .1T, the scale of the radiation diagram should thus also be reduced with a factor 10.

In the non-relativistic limit the radiation pattern is dipolar with the maximum radiated power in the directions perpendicular to $\dot{\boldsymbol{\beta}}$. With increasing β the radiated power increases and the radiation diagram tilts in the velocity forward direction. For a highly relativistic electron the radiation takes the shape of a pencil beam of angular width $\theta_{\rm rms}$ varying inversely with γ, $\theta_{rms} \approx .27/\gamma$.

To find the corresponding energy spectrum the actual charged particle trajectory need to be substituted in (2.39) and the remaining integral performed. Before taking on this task, however, let us study the problem in a semi-quantitative manner by making use of basic properties of the Fourier transform. For the case of a highly relativistic electron spiraling in a magnetic field with an angular frequency ω_B, a distant observer will be illuminated only as the pencil shaped radiation beam with typical angular width $2\theta_{\rm rms}$ swings in her direction. That is, the observer will, if properly placed, receive a series of short radiation pulses of duration Δt. This duration may be estimated by noting that seen from the particle the observer will

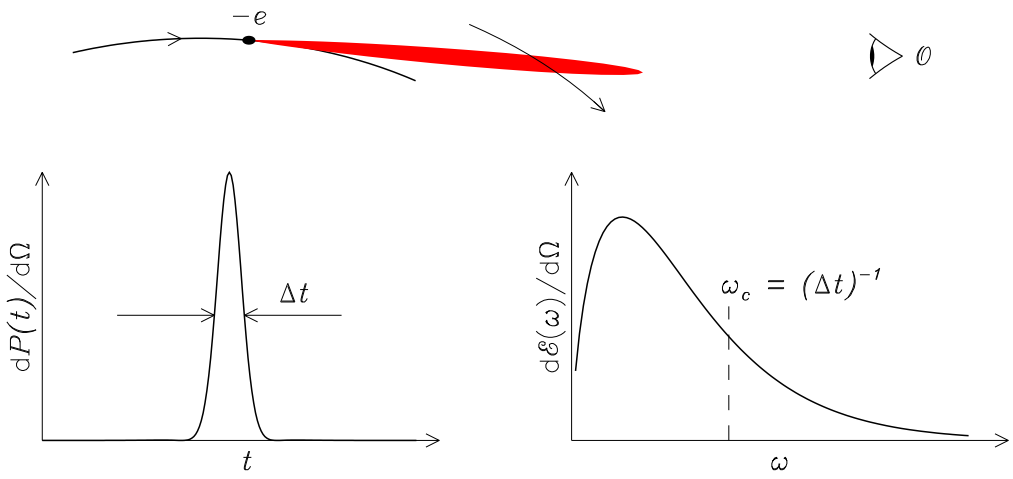

Figure 2.7: Observed frequency spectrum from a spiraling charged particle

be located within the radiation beam for a time interval

$$\Delta t' \approx \frac{2\pi}{\omega_B}\frac{2\theta_{\rm rms}}{2\pi} = \frac{2\theta_{\rm rms}}{\omega_B}.$$

The corresponding time interval for the observer is

$$\Delta t \approx (1-\beta)\Delta t' \approx \frac{\Delta t'}{2\gamma^2} \approx \frac{\theta_{\rm rms}}{\gamma^2\omega_B}.$$

A radiation pulse of duration Δt will give rise to a frequency spectrum with significant components up to the order of a critical frequency $\omega_c \sim (\Delta t)^{-1}$, (see quiz 1.20). The result is illustrated in figure 2.7 for a $\gamma = 5$ electron for which $\omega_c \approx 500\omega_B$. The radiation spectrum from the spiraling, relativistic electron thus extends to frequencies far exceeding the fundamental gyro frequency ω_B of the electron.

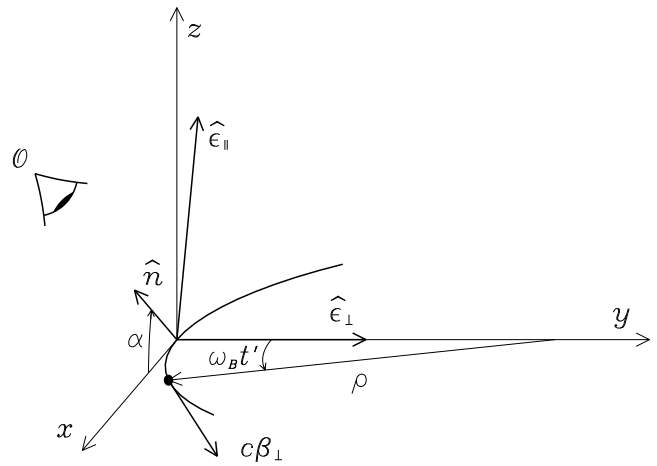

Figure 2.8: Observer and radiating particle geometry

For the more quantitative discussion of the radiation spectrum let us consider a relativistic electron spiraling in the xy-plane with orbital radius $\rho = c\beta_\perp/\omega_B$, passing through the origin at the retarded time $t' = 0$,

$$\boldsymbol{R}_q(t') = \rho(\hat{\boldsymbol{x}}\,\sin\tau + \hat{\boldsymbol{y}}\,(1-\cos\tau)) \quad \text{with} \quad \tau = \omega_B t'.$$

The electron is viewed by an observer located in direction $\hat{\boldsymbol{n}}$, without loss of generality chosen to lie in the xz-plane and making an angle α with the x-axis. The geometry is illustrated in figure 2.8. Substitution gives

$$t' - \hat{\boldsymbol{n}}\cdot\boldsymbol{R}_q(t')/c = \frac{1}{\omega_B}(\tau - \beta_\perp\cos\alpha\,\sin\tau) \equiv \frac{1}{\omega_B}\Phi(\tau) \tag{2.59}$$

and

$$\frac{\hat{\boldsymbol{n}}\times((\hat{\boldsymbol{n}}-\boldsymbol{\beta})\times\dot{\boldsymbol{\beta}})}{(1-\hat{\boldsymbol{n}}\cdot\boldsymbol{\beta})^2} = \omega_B\beta_\perp\,(g_\perp\,\hat{\boldsymbol{\epsilon}}_\perp + \sin\alpha\,g_\parallel\,\hat{\boldsymbol{\epsilon}}_\parallel) \tag{2.60}$$

with

$$g_\perp(\tau) = \frac{\beta_\perp\cos\alpha - \cos\tau}{(1-\beta_\perp\cos\alpha\,\cos\tau)^2}$$
$$g_\parallel(\tau) = -\frac{\sin\tau}{(1-\beta_\perp\cos\alpha\,\cos\tau)^2}.$$

We notice that $g_\perp(-\tau) = g_\perp(\tau)$ and $g_\parallel(-\tau) = -g_\parallel(\tau)$. As expected, for small α and large β ($\beta \approx 1$) the functions $g_\perp(\tau)$ and $g_\parallel(\tau)$ are both strongly peaked around $\tau = 0$.

The unit vectors $\hat{\boldsymbol{\epsilon}}_\perp$ and $\hat{\boldsymbol{\epsilon}}_\parallel$ constitute together with $\hat{\boldsymbol{n}}$ a right-handed orthogonal vector triplet with $\hat{\boldsymbol{\epsilon}}_\perp$ perpendicular to the plane containing $\boldsymbol{B}_0$ and $\hat{\boldsymbol{n}}$ (see figure 2.8).

With (2.59) and (2.60) substituted in (2.39) we find

$$\frac{\mathrm{d}\mathcal{E}(\omega)}{\mathrm{d}\Omega} = \frac{e^2\beta_\perp^2\omega_B^2}{4\pi^2\epsilon_0 c}\,|\,A_\perp(\omega)\,\hat{\boldsymbol{\epsilon}}_\perp + A_\parallel(\omega)\,\hat{\boldsymbol{\epsilon}}_\parallel\,|^2 \tag{2.61}$$

where

$$A_\perp(\omega) = \int_0^{\pi/2} g_\perp\,\cos\left(\frac{\omega}{\omega_B}\,\Phi(\tau)\right)\,\mathrm{d}\tau \tag{2.62}$$

$$A_\parallel(\omega) = \iota\sin\alpha\int_0^{\pi/2} g_\parallel\,\sin\left(\frac{\omega}{\omega_B}\,\Phi(\tau)\right)\,\mathrm{d}\tau. \tag{2.63}$$

The shape of the spectrum confirms the result of our previous semi-quantitative discussion. In fact, the frequency spectrum plotted in figure 2.7 was calculated from (2.61) for an observer $\mathcal{O}$ located in the plane of the spiraling electron, $\alpha = 0$.

In our calculation of the radiation spectrum we considered only one radiation pulse. In reality, the received radiation will be in the form of a series of similar pulses with a repetition rate $\omega_B/2\pi$. According to the theory of Fourier transform of periodic signals outlined in table 1.2, we should therefore expect to see a spectrum composed of discrete lines at integral multiples of the fundamental angular frequency ω_B. We can simply convert to this format by considering the angular distribution of radiated power in the nth harmonic of the fundamental frequency

$$\frac{\mathrm{d}\mathcal{P}_n}{\mathrm{d}\Omega} = \frac{\omega_B^2}{2\pi}\,\frac{\mathrm{d}\mathcal{E}}{\mathrm{d}\Omega}(\omega = n\omega_B). \tag{2.64}$$

The factor $\omega_B/2\pi$ is the repetition rate and converts energy to power while the last factor ω_B converts from per frequency unit to per harmonic.

In the total cyclotron radiation spectrum from relativistic electrons in a magnetized plasma we do not expect to see discrete spectral lines. The electrons in the plasma will have a broad distribution of kinetic energies. This will give rise to a corresponding spreading out of the spectral lines from each individual electron, resulting in a continuous total cyclotron power spectrum. For a quantitative evaluation the individual electron result will have to be averaged over the relativistic electron velocity distribution in the plasma.

The polarization of the cyclotron radiation from relativistic electrons differ significantly from that of Bremsstrahlung. The polarization of the radiation from an individual electron is given by the direction of the vector sum in (2.61). From (2.62)-(2.63) the dominating polarization of the emitted radiation from each individual electron is seen to be in the $\hat{\boldsymbol{\epsilon}}_\perp$-direction (see figure 2.8). This will therefore also be the case for the total cyclotron radiation from the magnetized plasma. We thus expect observed cyclotron radiation spectra to have a significant linear polarization in the direction perpendicular to the projection of the ambient magnetic field.

Quiz 2.13 : Instead of a single charge q spiraling in a magnetic field $\boldsymbol{B}$, a set of N such charges move with fixed relative positions around the same circle. Show that the total resulting power from N charges radiated into the n.th harmonic of ω_B is given by

$$\frac{\mathrm{d}P_n^{(N)}}{\mathrm{d}\Omega} = \frac{\mathrm{d}P_n^{(1)}}{\mathrm{d}\Omega}\left|\sum_{j=1}^{N}\exp(\imath n\theta_j)\right|^2$$

where $\mathrm{d}P_n^{(1)}/\mathrm{d}\Omega$ represents the radiation from a single charge and θ_j is the angular position of charge j at time $t = 0$. What are the limiting forms of this result for the case that i) the charges are uniformly spaced around the circle, ii) the charges clump together in one location, and iii) the charges are randomly distributed around the circle? Can you give a qualitative explanation of this result? [*Hint:* Use (1.50).]

Quiz 2.14 : In the Crab nebula we find electrons with energies up to 10^{12} eV moving in magnetic fields of 10^{-8} T. What is the fundamental frequency ω_B and the critical frequency ω_c for this case? What type of polarization do you expect for the resulting radiation? Discuss. Estimate the time needed for one electron to decrease its energy from 10^{12} eV to 10^{11} eV when no energy replenishing mechanism is provided.

2.4.3 Thomson scattering

Free electrons in a plasma will be set into oscillatory motions by the electric field of an incident electromagnetic wave. This will according to (2.29) cause the electron to radiate. The energy radiated will be taken from the incoming wave. The net result is that energy is taken from the wave and reradiated into different directions, that is, the energy in the wave is being scattered by the free electrons. The process is being referred to as Thomson scattering. We want to estimate the effectiveness of this mechanism. The situation is illustrated in figure 2.9.

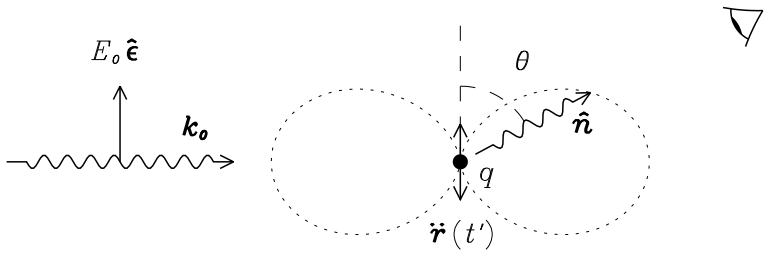

Figure 2.9: Thomson scattering of an electromagnetic wave

Let us assume the incident wave to be plane polarized,

$$\boldsymbol{E}(\boldsymbol{r},t) = E_i\hat{\boldsymbol{\epsilon}}_i\cos(\boldsymbol{k}_i\cdot\boldsymbol{r}-\omega_i t). \tag{2.65}$$

In a dilute plasma and for high enough frequencies the dispersion relation can be approximated by the the vacuum relation $\omega_i = k_i c$. We consider an electron moving in a straight line in the absence of the wave, $\boldsymbol{R}_q(t) = \boldsymbol{R}_0 + \boldsymbol{v}t$. In the presence of the wave the electron will experience an oscillatory perturbation in its orbit governed by

$$\dot{\boldsymbol{v}}(t) = -\frac{eE_i\hat{\boldsymbol{\epsilon}}_i}{m}\cos((\boldsymbol{k}_i\cdot\boldsymbol{v}-\omega_i)t). \tag{2.66}$$

For this result we neglected a non-interesting constant phase term.

Substitution of (2.66) into (2.29) gives the differential scattered power

$$\frac{\mathrm{d}P_s(t)}{\mathrm{d}\Omega} = \frac{e^4E_i^2}{16\pi^2\epsilon_0 m^2c^3}\,|\hat{\boldsymbol{n}}\times(\hat{\boldsymbol{n}}\times\hat{\boldsymbol{\epsilon}}_i)|^2\cos^2(\omega_i-\boldsymbol{k}_i\cdot\boldsymbol{v})t'.$$

The time averaged result can be written in the form

$$\frac{\mathrm{d}\overline{P_s}}{\mathrm{d}\Omega} = P_i\,r_0^2\,|\hat{\boldsymbol{n}}\times(\hat{\boldsymbol{n}}\times\hat{\boldsymbol{\epsilon}}_i)|^2 \tag{2.67}$$

where $P_i = \frac{1}{2}c\epsilon_0E_i^2$ is the incident energy flux and

$$r_0 \equiv \frac{e^2}{4\pi\epsilon_0 mc^2} = 2.8179409\cdot10^{-15}\mathrm{m} \tag{2.68}$$

is the classical electron radius. The cross section for scattering from a single electron is thus essentially given by the classical cross section of the electron, $\sigma_T = \frac{8\pi}{3}r_0^2 = 6.65246\cdot10^{-29}$ m^2. The radiation diagram for the Thomson scattering process, that is, the magnitude of $\mathrm{d}\overline{P_s}/\mathrm{d}\Omega$ as a function of the angle θ between $\hat{\boldsymbol{n}}$ and $\boldsymbol{k}_i$ is indicated in figure 2.9 by the dotted ovals. The polarization $\boldsymbol{\epsilon}_s$ of the scattered radiation is determined by the polarization of the incident wave, that is, $\hat{\boldsymbol{\epsilon}}_s \parallel \hat{\boldsymbol{n}}\times(\hat{\boldsymbol{n}}\times\hat{\boldsymbol{\epsilon}}_i)$.

To find the corresponding power spectrum of the scattered radiation resort could be made to the general formalism (2.39). In the present case an alternative approach is to remark that the electron in its unperturbed rest frame will experience a harmonic oscillation at the Doppler shifted frequency $\omega' = \omega_i - \boldsymbol{k}_i\cdot\boldsymbol{v}$. This will also be the frequency of the scattered radiation

as seen from the moving electron. The observer will experience an additional Doppler shift in the recieved frequency, $\omega_s = \omega' + \boldsymbol{k}_s \cdot \boldsymbol{v} = \omega_i + \boldsymbol{K} \cdot \boldsymbol{v}$ where $\boldsymbol{K} = \boldsymbol{k}_s - \boldsymbol{k}_i$ is the scattering wave vector and $\boldsymbol{k}_s = \frac{\omega_s}{c}\hat{\boldsymbol{n}}$ the scattered wave vector. For a Maxwellian velocity distribution (6.35) for the electrons

$$f_{\boldsymbol{rv}}(\boldsymbol{v}) = \left(\frac{m}{2\pi\mathcal{T}}\right)^{3/2} \exp\left(-\frac{m\boldsymbol{v}^2}{2\mathcal{T}}\right),$$

the expected scattered power spectrum is therefore

$$\begin{aligned}\frac{\mathrm{d}\mathcal{P}_s(\omega)}{\mathrm{d}\Omega} &= P_i\, r_0^2\, |\hat{\boldsymbol{n}} \times (\hat{\boldsymbol{n}} \times \hat{\boldsymbol{\epsilon}}_i)|^2 \int \mathrm{d}^3\boldsymbol{v}\, f_{\boldsymbol{rv}}(\boldsymbol{v})\, \delta(\omega - \omega_s) \\ &= P_i\, r_0^2\, |\hat{\boldsymbol{n}} \times (\hat{\boldsymbol{n}} \times \hat{\boldsymbol{\epsilon}}_i)|^2 \left(\frac{m}{2\pi\mathcal{T}K^2}\right)^{1/2} \exp\left(-\frac{m(\omega-\omega_i)^2}{2\mathcal{T}K^2}\right) \qquad (2.69)\end{aligned}$$

with $K = 2k_i \sin\frac{\theta}{2}$.

The width of the spectrum is seen to be determined by the thermal velocity of the electrons

$$\Delta\omega_{FWHH} = \sqrt{\frac{8\ln 2\,\mathcal{T}}{m}}K. \qquad (2.70)$$

The result is valid as long as correlations between electrons can be neglected. Due to Debye shielding effects in the plasma this is correct only if $K\lambda_D \gg 1$ where $\lambda_D = \sqrt{\epsilon_0\mathcal{T}/Ne^2}$ is the Debye length.

The Solar corona is made visible during occultations through Solar radiation scattered by free electrons.

Quiz 2.15 : Discuss conditions for the validity of the approximation in (2.66). What role will the magnetic field in the incoming wave play?

Quiz 2.16 : Convince yourself that the result (2.69) can be derived from a consideration of the t'-dependence of the integrand $\dot{\boldsymbol{v}}(t')\exp\iota\omega(t' - \hat{\boldsymbol{n}}\cdot\boldsymbol{R}_q(t')/c)$ for the moving electron in (2.39).

Quiz 2.17 : An electron is performing oscillations around the origin under the influence of a restoring harmonic oscillator force $\boldsymbol{F} = -\omega_0^2\,\boldsymbol{r}$. Including also the radiation reaction force $\boldsymbol{F}_{react}$ as given in (2.42), verify that the equation of motion is

$$m(\ddot{\boldsymbol{r}} - \tau_0\,\dddot{\boldsymbol{r}}) = -\omega_0^2\,\boldsymbol{r}$$

with physically relevant solution $\boldsymbol{r}(t) \sim \exp(-\iota\omega t)$ where

$$\omega \approx \omega_0 - \iota\frac{\Gamma}{2} \quad \text{and} \quad \Gamma = \tau_0\,\omega_0^2.$$

Show that that the initial energy in the oscillation decays proportional to $\exp(-\Gamma t)$ and that the initial energy loss rate ΓW_0 equals the power radiated by the oscillator. Show that the power $\mathcal{P}(\omega)$ radiated per unit angular frequency when the radiation reaction is included is related to the power $\overline{P}$ radiated in the absence of the reaction force by

$$\mathcal{P}(\omega) = \frac{1}{\pi}\,\frac{\Gamma/2}{(\omega-\omega_0)^2 + (\Gamma/2)^2}\,\overline{P}. \qquad (2.71)$$

Evaluate the effective width of the spectrum.

Quiz 2.18 : Estimate the fraction of the Solar radiation flux that will be scattered per unit volume element by free electrons in the corona. What do you predict about the polarization of the scattered radiation?

Quiz 2.19 : Compare Bremsstrahlung and Thomson scattering from the Solar corona.

Chapter 3

Spectra of One–Electron Atoms

By the end of the nineteenth century, spectroscopy had provided detailed and very accurate evidence that any given atom or molecule was only able to emit or absorb radiation at a particular set of wavelengths. Thus, in 1885 Balmer pointed out that the wavelengths of four of the hydrogen lines in the visible part of the spectrum could be expressed through the simple formula

$$\lambda = \text{constant}\frac{n_1^2}{n_1^2 - 4} \quad \text{for} \quad n_1 = 3,\, 4,\, 5,\, 6.$$

This result was rearranged and extended by Rydberg to read

$$\frac{1}{\lambda} = R_H \left(\frac{1}{2^2} - \frac{1}{n_1^2} \right) \quad \text{for} \quad n_1 = 3,\, 4,\, 5,\, 6,\, \cdots \tag{3.1}$$

where $R_H = 1.096776 \cdot 10^7\ \text{m}^{-1}$ is the *Rydberg constant* for hydrogen. The quantity $\sigma = 1/\lambda$ is called the *repetence*. In spectroscopic literature the unit Kayser $= 1\ \text{cm}^{-1}$ is often used for this quantity. The empirical formula (3.1) gave the wavelengths of all lines in the Balmer series with a relative accuracy better that 10^{-5}.

Additional series of lines of the hydrogen spectrum were soon found to be described accurately in terms of the more general formula

$$\frac{1}{\lambda} = R_H \left(\frac{1}{n^2} - \frac{1}{n_1^2} \right) \quad \text{for} \quad n_1 = n+1,\, n+2,\, \cdots \tag{3.2}$$

with n being a positive integer number: Lyman, Balmer, Paschen, Brackett and Pfund series for $n = 1,\, 2,\, 3,\, 4$ and 5, respectively.

These experimental results indicated that the hydrogen atom was only allowed to exist in one of a discrete set of energy states,

$$W_n = -hc\, R_H \frac{1}{n^2} \quad \text{for} \quad n = 1,\, 2,\, 3,\, \cdots, \tag{3.3}$$

and that emission and absorption of radiation was thus restricted to radiation with wavelengths λ satisfying

$$\frac{1}{\lambda} = \frac{W_{n_1} - W_n}{hc}.$$

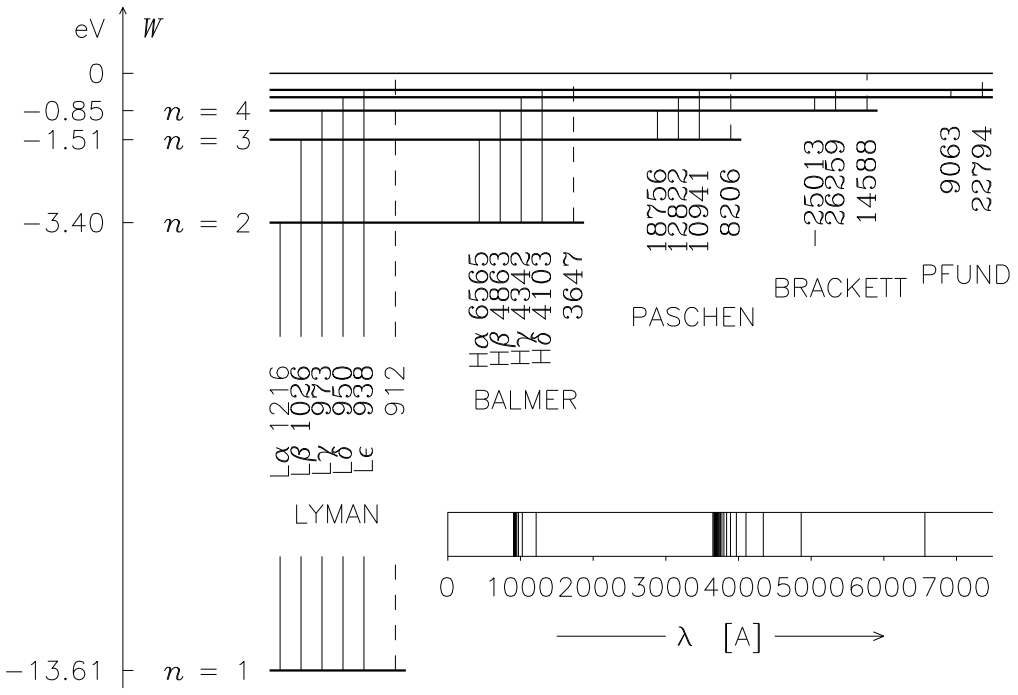

Figure 3.1: The hydrogen spectrum

Available energy levels of the hydrogen atom according to (3.3) with corresponding transitions are illustrated in figure 3.1. In the same figure, the spectrum insert shows the actual position of the spectral lines corresponding to the Lyman and the Balmer series.

These results were in striking contrast to contemporary physics of that time. Classical physics was not able to explain why the hydrogen atom was only allowed to exist in discrete energy states. The spectroscopic results thus pointed to the necessity of improvements in the laws of physics. Bohr (1913) provided a partial answer, but it was only through the introduction of quantum mechanics by Schrödinger, Heisenberg, Pauli and their colleagues in Göttingen during the 1920's that a satisfactory new theory was finally established. Any understanding of atomic or molecular spectra is only possible through the formalism of *quantum mechanics*.

The interpretation of the radiation received from different celestial bodies and the extraction of the underlying physics lies at the very heart of astrophysics. This involves the study of literally millions of different spectral lines. Most of these are of atomic origin, but many are also of molecular origin. A complete discussion of all aspects of the physics of spectral line formation clearly lies outside the scope of our discussion. We shall have to be satisfied by studying some of the more fundamental facts. In this chapter we review the formalism of quantum mechanics as applied to the spectra of one-electron atoms. In the next chapter we proceed to a discussion of the corresponding spectra from many-electron atoms, while chapter 5 contains a review some typical aspects of molecular spectra.

3.1 Quantum Mechanics

In classical mechanics the motion $\boldsymbol{r}(t)$ of a particle with mass m is determined by the Newton equation of motion

$$m\frac{\mathrm{d}^2\boldsymbol{r}}{\mathrm{d}t^2} = \boldsymbol{F}, \tag{3.4}$$

where $\boldsymbol{F}$ is the total force acting on the particle. If the total force is conservative, that is, if it is derivable from a potential, $\boldsymbol{F} = -\nabla U(\boldsymbol{r}) \equiv -\partial U(\boldsymbol{r})/\partial \boldsymbol{r}$, a mathematically equivalent form of Newton's equation of motion is the *Hamilton equations*

$$\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t} = \frac{\partial \mathcal{H}(\boldsymbol{r},\boldsymbol{p})}{\partial \boldsymbol{p}} \tag{3.5}$$

$$\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t} = -\frac{\partial \mathcal{H}(\boldsymbol{r},\boldsymbol{p})}{\partial \boldsymbol{r}} \tag{3.6}$$

where $\boldsymbol{p} = m\boldsymbol{v}$ is the particle momentum and

$$\mathcal{H}(\boldsymbol{r},\boldsymbol{p}) = \frac{\boldsymbol{p}^2}{2m} + U(\boldsymbol{r}) \tag{3.7}$$

is the *Hamiltonian*. In (3.5)-(3.6) the Hamiltonian $\mathcal{H}(\boldsymbol{r},\boldsymbol{p})$ is to be considered a function of independent variables $\boldsymbol{r}$ and $\boldsymbol{p}$. Physically we recognize $\mathcal{H}(\boldsymbol{r},\boldsymbol{p})$ as the total energy of the particle. For time-independent force fields, the total energy is a constant of motion, that is,

$$\mathcal{H}(\boldsymbol{r},\boldsymbol{p}) = W \tag{3.8}$$

where W is a constant depending on the initial conditions imposed on the system. With forces varying with time the Hamiltonian will also depend on time explicitly, $\mathcal{H}(\boldsymbol{r},\boldsymbol{p},t)$. In this case the Hamiltonian is not a constant of motion.

At first glance the formulation (3.5)-(3.6) may seem more complicated than the original Newton's equation of motion (3.4). Mathematically, however, Hamilton's formulation is a rather powerful one. The formalism is easily generalized from the simple point particle problem to any complicated mechanical system ruled by conservative forces. Any advanced discussion of mechanics will have the Hamiltonian or the closely related Lagrangian formulation as its starting point. These aspects fall outside the scope of our discussion. The important point to note is that the prescription of a single function $\mathcal{H}(\boldsymbol{r},\boldsymbol{p})$, together with the corresponding set of rules (3.5)-(3.6), completely describes the dynamics of the system.

Classical mechanics as represented by (3.4) or (3.5)-(3.6) provides a precise description of the motion of macroscopic systems. Spectroscopy, however, clearly demonstrated that these laws failed when it came to atomic or molecular levels. Classical mechanics provided no mechanism that would lead to the formation of a discrete set of spectral lines. In fact, classical mechanics could not even predict the existence of stable atoms. An electron in orbit around a nucleus experiences a continuous acceleration and should according to classical theory radiate electromagnetic waves, the electron orbit finally collapsing into the nucleus.

One of the greatest breakthroughs in physics therefore came when Schrödinger was able to give a recipe for how to generalize the laws of mechanics to be valid also for atomic systems. Indeed, Schrödinger realized that the natural mathematical formulation for systems existing in only a discrete set of states, like a vibrating violin string of length L, mass per unit length

μ and tension S only provides a discrete set of possible tones (frequencies), is an eigenvalue formulation. For the particular vibrating string example,

$$S\frac{\partial^2 y}{\partial x^2} = \mu\ddot{y}$$

with boundary condition $y(x,t) = 0$ for $x = 0, L$ and $\ddot{y} = -\omega^2 y$ for a consistent vibrating state, the eigenvalue formulation is

$$-S\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} = \mu\omega^2 y,$$

with solutions $\omega = 2\pi n\sqrt{S/\mu}/L$ and $y(x) = a\sin 2\pi nx/L$ where n is a positive integer.

Any physical system with a Hamiltonian $\mathcal{H}$ of the form (3.7) can be recast into an eigenvalue formulation by replacing the momentum vector $\boldsymbol{p}$ with a first order spatial differential operator. The mechanics formulation introduced by Schrödinger is quantum mechanics. The quantum mechanical "recipe" consists of a few simple steps summarized in table 3.1.

A few comments on this set of rules may be appropriate. For time independent problems the classical statement (3.8) of the total energy as a constant of motion is in quantum mechanics replaced with the stationary Schrödinger equation (3.10). This is an eigenvalue problem, where the total energy W and the wave function $\Psi(\boldsymbol{r})$ are eigenvalues and eigenfunctions of the Hamilton operator $H(\boldsymbol{r}, \boldsymbol{p})$. Physically acceptable solutions $\Psi(\boldsymbol{r}, t)$ will be found for a discrete set of values for the total energy W. This will be repeatedly demonstrated in the following and is indeed the property needed in order to accommodate the experimental results of atomic spectroscopy. Some useful operator definitions and theorems are listed in table 3.2.

We recognize the left-hand side of the normalization condition (3.9) as the scalar product of the wave function $\Psi(\boldsymbol{r})$ with itself. The normalization condition implies that for physically acceptable wave functions $|\,\Psi(\boldsymbol{r})\,|^2$ may at most have integrable singularities and must approach zero fast enough as $r \to \infty$. Since Schrödinger's equation (3.10) is linear in Ψ, the normalization condition (3.9) may always be satisfied as long as a square integrable solution is found. This normalization is necessary for interpreting $|\,\Psi(\boldsymbol{r}, t)\,|^2\ \mathrm{d}^3\boldsymbol{r}$ as the probability for finding the particle at time t in the volume element $\mathrm{d}^3\boldsymbol{r}$ centered on $\boldsymbol{r}$. The probability concept in quantum mechanics means that we have to abandon the classical idea of the electron as a point particle with an electron smeared out as charge cloud.

Quantum mechanics represents a generalization of classical mechanics. From the mathematical and interpretational point of view, quantum mechanics is the more "complete" theory. In contrast to classical theory, quantum mechanics provides correct predictions at the atomic level. And, in the "infinite mass" limit, quantum mechanical can be shown to reduce to the classical mechanics result.

In the next sections we review the application of quantum mechanics to the simplest atomic system, the hydrogen or hydrogen-like atom, consisting of a nucleus with charge Ze surrounded by one electron with charge $-e$. Examples of such atoms are H I, He II, Li III, C VI and Fe XXVI. We have here made use of the standard notation where the Roman number added to the chemical element symbol indicates the ionization state of the atom. I refers to the atom in its neutral state, II represents the first ionization state, and so on. We shall later argue that even the alkali type atoms Na I, Mg II, K I and Ca II may in many respects be considered as examples of one-electron atoms.

For a quantum mechanical analysis of a given system, (if a classical analogue exists) first establish the classical Hamiltonian $\mathcal{H}(\boldsymbol{r}, \boldsymbol{p}, t)$.

Replace the momentum vector $\boldsymbol{p}$ with the corresponding momentum operator $\boldsymbol{p} = -\iota\hbar\nabla$, and thus transform the classical Hamiltonian function $\mathcal{H}$ into the corresponding Hamiltonian operator H.

Replace the classical concept of a precise particle position $\boldsymbol{r}(t)$ with the (for bound states) normalized *wave function* $\Psi(\boldsymbol{r}, t)$,

$$\int |\Psi(\boldsymbol{r}, t)|^2 \, \mathrm{d}^3\boldsymbol{r} = 1, \tag{3.9}$$

and interpret $|\Psi(\boldsymbol{r}, t)|^2$ as the *probability density* for finding the system at position $\boldsymbol{r}$ at time t.

For time-stationary problems determine the wave function Ψ as the solution of the *stationary Schrödinger equation*

$$H(\boldsymbol{r}, \boldsymbol{p})\Psi(\boldsymbol{r}) = W\Psi(\boldsymbol{r}) \tag{3.10}$$

with suitable boundary conditions imposed.

For non-stationary problems solve the *time-dependent Schrödinger equation*

$$\iota\hbar\frac{\partial}{\partial t}\Psi(\boldsymbol{r}, t) = H(\boldsymbol{r}, \boldsymbol{p}, t)\Psi(\boldsymbol{r}, t). \tag{3.11}$$

Any physical quantity (observable) $\mathcal{Q}$ is represented by a corresponding operator Q. Possible results of a measurement of the value of the physical quantity $\mathcal{Q}$ are the set of eigenvalues q of the eigenvalue equation, $Q\Phi = q\Phi$ where the eigenfunctions Φ need not be simultaneous eigenfunctions of the Schrödinger equation.

Two physical quantities $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are *simultaneously measurable* if the corresponding operators Q_1 and Q_2 have identical eigenfunctions Φ. This will be possible only if the two operators *commute*,

$$[Q_1, Q_2] \equiv Q_1Q_2 - Q_2Q_1 = 0.$$

The evolution of the *mean value* of Q, defined as $\langle\, Q \,\rangle \equiv (\Psi, Q\Psi)$, is determined by

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\, Q \,\rangle = \frac{\iota}{\hbar}\langle\, [H, Q] \,\rangle + \langle\, \frac{\partial Q}{\partial t} \,\rangle. \tag{3.12}$$

Table 3.1: The quantum mechanical recipe

We consider operators Q acting on functions $\phi(\boldsymbol{r})$ and $\psi(\boldsymbol{r})$, sufficiently differentiable and approaching zero fast enough as $|\,\boldsymbol{r}\,|\to\infty$ for the following expressions to exist.

The *scalar product* of ϕ and ψ is defined as $(\phi,\psi)\equiv\int\phi^*(\boldsymbol{r})\psi(\boldsymbol{r})\,\mathrm{d}^3\boldsymbol{r}$.

The functions ϕ and ψ are *orthogonal* if $(\phi,\psi)=0$.

The *adjoint operator* $Q^\dagger$ is defined by $(Q^\dagger\phi,\psi)\equiv(\phi,Q\psi)$ for any choice of ϕ and ψ. The operator Q is *self-adjoint* or *Hermitian* if $Q^\dagger=Q$.

The eigenvalues λ of the Hermitian operator Q are real-valued.
Proof: With $Q\psi=\lambda\psi$ we find $\lambda^*(\psi,\psi)=(Q\psi,\psi)=(\psi,Q\psi)=\lambda(\psi,\psi)$.

The eigenfunctions ψ_1, ψ_2 of an Hermitian operator Q corresponding to different eigenvalues λ_1 and λ_2 are orthogonal.
Proof: With $\lambda_2(\psi_2,\psi_1)=(Q\psi_2,\psi_1)=(\psi_2,Q\psi_1)=\lambda_1(\psi_2,\psi_1)$ and $\lambda_1\neq\lambda_2$ it follows that $(\psi_2,\psi_1)=0$.

Eigenfunctions belonging to identical eigenvalues may be chosen to be orthogonal.
Proof: If $Q\psi_1=\lambda\psi_1$, $Q\psi_2=\lambda\psi_2$ and $(\psi_1,\psi_2)=K(\psi_1,\psi_1)\neq0$ then ψ_1 and $\psi_2'=\psi_2-K\psi_1$ is an alternative orthogonal set of eigenfunctions for the given eigenvalue.

Table 3.2: Useful operator definitions and theorems

Quiz 3.1 : Verify that (3.5)-(3.6) with (3.7) are equivalent to (3.4).

Quiz 3.2 : Show that the Hamiltonian $\mathcal{H}(\boldsymbol{r},\boldsymbol{p})$ is a constant of motion for a time-independent force field, that is, $\mathrm{d}\mathcal{H}(\boldsymbol{r},\boldsymbol{p})/\mathrm{d}t=0$. What is the corresponding result for $\mathcal{H}(\boldsymbol{r},\boldsymbol{p},t)$?

Quiz 3.3 : Prove that

$$\boldsymbol{p}=-\imath\hbar\nabla\quad\text{and}\quad H=\frac{\boldsymbol{p}^2}{2m}+U(r)$$

are both Hermitian operators. [*Hint:* Perform partial integrations, using (A.33), (A.34) or (A.50), and assume that the functions ϕ and ψ appearing in $(\phi,\boldsymbol{p}\psi)$ and $(\phi,H\psi)$ both approach zero sufficiently fast as $|\,\boldsymbol{r}\,|\to\infty$.]

Quiz 3.4 : Make use of (3.11) and the Hermitian property of H to verify (3.12).

Quiz 3.5 : To interpret $|\,\Psi\,|^2$ as a probability density it is necessary that $\Psi(\boldsymbol{r},t)$ remains normalized, that is, $\int|\,\Psi(\boldsymbol{r},t)\,|^2\;\mathrm{d}^3\boldsymbol{r}=1$ for all times. Verify that this is the case.

Quiz 3.6 : Show that

$$[H,\boldsymbol{r}]=-\frac{\imath\hbar}{m}\,\boldsymbol{p}. \tag{3.13}$$

Then make use of (3.12) to prove the *Ehrenfest theorem*

$$m\frac{\mathrm{d}}{\mathrm{d}t}\langle\,\boldsymbol{r}\,\rangle=\langle\,\boldsymbol{p}\,\rangle\quad\text{and}\quad\frac{\mathrm{d}}{\mathrm{d}t}\langle\,\boldsymbol{p}\,\rangle=-\langle\,\nabla U(r)\,\rangle. \tag{3.14}$$

Do you see similarities with the corresponding classical equations of motion?

Quiz 3.7 : Let A and B be Hermitian operators. Define the *standard deviation* of A as $\langle\,\Delta A\,\rangle \equiv \sqrt{(\Psi,(A-\langle\,A\,\rangle)^2\Psi)} \equiv \sqrt{\big(\Psi,A'^2\Psi\big)}$ and similarly for B. Verify the following steps in the derivation of the *Heisenberg uncertainty relation* :

$$(\Psi,[A,B]\Psi)=(\Psi,[A',B']\Psi)=(A'\Psi,B'\Psi)-(B'\Psi,A'\Psi)=2\iota\,\mathrm{Im}\,\big((A'\Psi,B'\Psi)\big)$$

and therefore

$$\frac{1}{2}\,|\,(\Psi,[A,B]\Psi)\,|\leq|\,(A'\Psi,B'\Psi)\,|\leq\langle\,\Delta A\,\rangle\langle\,\Delta B\,\rangle. \tag{3.15}$$

Interprete your result. Choose $A=x$ and $B=p_x$ as a specific example.

3.2 The One-Electron Atom

We consider now a one electron atom with nuclear charge Ze. The electron mass m is much smaller than the corresponding nuclear mass M. In the lowest order approximation the nucleus may therefore be considered fixed at the origin of the coordinate system, not participating in any motion. In the classical picture the electron at position $\boldsymbol{r}(t)$ then moves in a central electrostatic force field with corresponding potential

$$U(r)=-\frac{Ze^2}{4\pi\epsilon_0 r}. \tag{3.16}$$

The Hamiltonian operator is then

$$H_0\equiv-\frac{\hbar^2}{2m}\nabla^2+U(r). \tag{3.17}$$

With a spherically symmetric force field it is convenient to write the Schrödinger equation (3.10) in terms of spherical coordinates

$$-\frac{\hbar^2}{2mr^2}\left[\frac{\partial}{\partial r}(r^2\frac{\partial\Psi}{\partial r})+\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta\frac{\partial\Psi}{\partial\theta})+\frac{1}{\sin^2\theta}\frac{\partial^2\Psi}{\partial\varphi^2}\right]+U(r)\Psi=W\Psi. \tag{3.18}$$

We are looking for "bounded" solutions of the wave function Ψ. This means, in addition to requiring $|\,\Psi\,|^2$ to be finite everywhere, that the wave function vanishes at infinity, $\Psi\to 0$ as $r\to\infty$. The solution is found by making use of the method of separation of variables, looking for a solution of the form

$$\Psi(\boldsymbol{r})=\Xi(r)\Theta(\theta)\Phi(\varphi). \tag{3.19}$$

From the mathematical point of view, equation (3.18) is known to have unique solutions. If a solution of the form (3.19) can be found, satisfying the physically relevant boundary conditions, then this solution is *the only* possible solution.

Substituting (3.19) into (3.18) and dividing by $\Phi(\varphi)$ leads to an equation of the form

$$A(r,\theta)+B(r,\theta)\,\left[\frac{1}{\Phi}\frac{\mathrm{d}^2\Phi}{\mathrm{d}\varphi^2}\right]=0.$$

The explicit expression of the functions A and B may be found from (3.18). For our present needs it is sufficient to note their formal variable dependence, that is, that both are independent of the azimuthal angle φ. For this equation to be satisfied for any given values of r and θ and for all values of φ, it is necessary that

$$\frac{1}{\Phi}\frac{\mathrm{d}^2\Phi}{\mathrm{d}\varphi^2} = \text{constant} = -m_\ell^2.$$

The uniqueness condition, $\Phi(\varphi + 2\pi) = \Phi(\varphi)$, requires the separation constant to be equal to the negative square of an integer number m_ℓ. We shall refer to m_ℓ as the *azimuthal quantum number*. The solution is therefore

$$\Phi_{m_\ell}(\varphi) = \exp(\imath m_\ell \varphi). \tag{3.20}$$

The functions (3.20) constitute a complete, orthogonal set of functions with normalization

$$\int_0^{2\pi} \Phi^*_{m'_\ell}(\varphi)\Phi_{m_\ell}(\varphi)\,\mathrm{d}\varphi = 2\pi\delta_{m'_\ell,m_\ell}$$

where $\delta_{m'_\ell,m_\ell}$ is the Kronecker δ. Complete here means that any continuous and square integrable function, periodic over $(0, 2\pi)$, can be expanded in terms of base functions of the type Φ_{m_ℓ}. From Fourier theory we know that this is the case (see summary in table 1.2).

With $\Phi(\varphi)$ known, the Schrödinger equation divided by Ψ/r^2 may now be written as

$$A(r) + \frac{1}{\Theta}\left[\frac{1}{\sin\theta}\frac{\mathrm{d}}{\mathrm{d}\theta}(\sin\theta\frac{\mathrm{d}\Theta}{\mathrm{d}\theta}) - \frac{m_\ell^2}{\sin^2\theta}\,\Theta\right] = 0.$$

The explicit expression for $A(r)$ is again irrelevant for our argument. We only note that $A(r)$ is independent of θ. For this equation to be satisfied for any given r and for all values of θ, we must require

$$\frac{1}{\Theta}\left[\frac{1}{\sin\theta}\frac{\mathrm{d}}{\mathrm{d}\theta}(\sin\theta\frac{\mathrm{d}\Theta}{\mathrm{d}\theta}) - \frac{m_\ell^2}{\sin^2\theta}\,\Theta\right] = \text{constant} = -\ell(\ell+1). \tag{3.27}$$

Physically, $\Theta(\theta)$ must remain finite in the interval $[0, \pi]$. It can be shown that to satisfy this requirement, the separation constant must take the particular form $-\ell(\ell+1)$ with ℓ being a non-negative integer (see quiz 3.8). We will refer to ℓ as the *angular quantum number*. Equation (3.27) is known as the *associated Legendre differential equation*. The relevant solutions are the *associated Legendre polynomials* of first kind

$$\Theta_{\ell m_\ell}(\theta) = P_\ell^{|m_\ell|}(\cos\theta). \tag{3.28}$$

Some useful properties of the associated Legendre polynomials P_ℓ^m are listed in table 3.3. We note in particular that $P_\ell^{|m_\ell|}(\cos\theta)$ is non-vanishing only for

$$|\,m_\ell\,| \leq \ell. \tag{3.29}$$

The associated Legendre polynomials $P_\ell^m(\cos\theta)$ and the complex exponentials $\exp(\imath m\varphi)$ find a practical combination in the *spherical harmonics* functions $Y_\ell^m(\theta, \varphi)$. Their definition

Associate Legendre polynomials of first kind – generating function

$$P_\ell^m(x) = \frac{(-1)^m}{2^\ell \ell!}(1-x^2)^{m/2}\,\frac{\mathrm{d}^{\ell+m}}{\mathrm{d}x^{\ell+m}}(x^2-1)^\ell, \quad \ell, m \geq 0 \tag{3.21}$$

Some lower order polynomials

$$P_0^0 = 1, \qquad P_1^0 = x, \quad P_1^1 = -\sqrt{1-x^2}$$
$$P_2^0 = \frac{1}{2}(3x^2-1), \quad P_2^1 = -3x\sqrt{1-x^2}, \quad P_2^2 = 3(1-x^2)$$

Recurrence relations $(\ell \geq 1)$

$$(2\ell+1)\sqrt{1-x^2}P_\ell^{m-1} = P_{\ell-1}^m - P_{\ell+1}^m, \quad m \geq 1 \tag{3.22}$$
$$(2\ell+1)xP_\ell^m = (\ell+m)P_{\ell-1}^m - (\ell-m+1)P_{\ell+1}^m \tag{3.23}$$
$$\begin{aligned}(2\ell+1)\sqrt{1-x^2}P_\ell^{m+1} = {} & (\ell-m)(\ell-m+1)P_{\ell+1}^m \\ & -(\ell+m)(\ell+m+1)P_{\ell-1}^m\end{aligned} \tag{3.24}$$
$$P_\ell^{m+1} = -\frac{2mx}{\sqrt{1-x^2}}P_\ell^m - (\ell+m)(\ell-m+1)P_\ell^{m-1}, \quad m \geq 1 \tag{3.25}$$

Normalization

$$\int_{-1}^{1} P_\ell^m(x)P_{\ell'}^m(x)\,\mathrm{d}x = \frac{2}{2\ell+1}\,\frac{(\ell+m)!}{(\ell-m)!}\,\delta_{\ell,\ell'} \tag{3.26}$$

Table 3.3: Associate Legendre functions

and some of their properties are summarized in table 3.4. The Y_ℓ^m-functions are normalized, they are orthogonal with respect to both indices, and they form a complete set, that is, any function of the polar angles θ and ϕ may be expanded in terms of $\{Y_\ell^m\}$[1].

The radial part $\Xi(r)$ of the wave function is finally determined by the equation

$$\frac{1}{r^2}\frac{\mathrm{d}}{\mathrm{d}r}(r^2\frac{\mathrm{d}\Xi}{\mathrm{d}r}) + \left[\frac{2m}{\hbar^2}(W-U(r)) - \frac{\ell(\ell+1)}{r^2}\right]\Xi = 0. \tag{3.33}$$

It is here convenient to introduce the dimensionless variable

$$\rho = 2\beta r \quad \text{with} \quad \beta^2 = -\frac{2mW}{\hbar^2}. \tag{3.34}$$

Then, making the substitution

$$\Xi(r) = \rho^\ell \exp(-\frac{\rho}{2})u(\rho),$$

it is seen that $u(\rho)$ will have to satisfy the *associated Laguerre differential equation*

$$\rho\frac{\mathrm{d}^2u}{\mathrm{d}\rho^2} + (\mu+1-\rho)\frac{\mathrm{d}u}{\mathrm{d}\rho} + (\nu-\mu)u = 0 \tag{3.35}$$

[1]In the literature also the notation $Y_{\ell m}$ for the spherical harmonics will be found.

Spherical harmonics functions – definition

$$Y_\ell^m(\theta,\varphi) = \sqrt{\frac{2\ell+1}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}}\, P_\ell^m(\cos\theta)\,\mathrm{e}^{\iota m\varphi}, \quad m \geq 0 \tag{3.30}$$

$$Y_\ell^{-m}(\theta,\varphi) = (-)^m Y_\ell^{m*}(\theta,\varphi) \tag{3.31}$$

Some lower order functions

$$Y_0^0 = \sqrt{\frac{1}{4\pi}}, \qquad Y_1^0 = \sqrt{\frac{3}{4\pi}}\cos\theta, \quad Y_1^{\pm1} = \mp\sqrt{\frac{3}{8\pi}}\sin\theta\,\mathrm{e}^{\pm\iota m\varphi}$$

$$Y_2^0 = \sqrt{\frac{5}{16\pi}}(3\cos^2\theta - 1), \quad Y_2^{\pm1} = \mp\sqrt{\frac{15}{8\pi}}\sin\theta\cos\theta\,\mathrm{e}^{\pm\iota m\varphi}, \quad Y_2^{\pm2} = \sqrt{\frac{15}{32\pi}}\sin^2\theta\,\mathrm{e}^{\pm2\iota m\varphi}$$

Recurrence relations

$$\cos\theta\, Y_\ell^m = \sqrt{\frac{(\ell+m+1)(\ell-m+1)}{(2\ell+1)(2\ell+3)}}\, Y_{\ell+1}^m + \sqrt{\frac{(\ell+m)(\ell-m)}{(2\ell+1)(2\ell-1)}}\, Y_{\ell-1}^m$$

$$\sin\theta\,\mathrm{e}^{\pm\iota\varphi}\, Y_\ell^m = \mp\sqrt{\frac{(\ell\pm m+1)(\ell\pm m+2)}{(2\ell+1)(2\ell+3)}}\, Y_{\ell+1}^{m\pm1} \pm \sqrt{\frac{(\ell\mp m)(\ell\mp m-1)}{(2\ell+1)(2\ell-1)}}\, Y_{\ell-1}^{m\pm1}$$

Orthonormalization

$$\int Y_{\ell'}^{m'*}(\theta,\varphi)\, Y_\ell^m(\theta,\varphi)\,\mathrm{d}^2\mathbf{\Omega} = \delta_{\ell',\ell}\,\delta_{m',m} \quad (\mathrm{d}^2\mathbf{\Omega} = \mathrm{d}\cos\theta\,\mathrm{d}\varphi) \tag{3.32}$$

Completeness relation

$$\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} Y_\ell^{m*}(\theta',\varphi')\, Y_\ell^m(\theta,\varphi) = \delta(\mathbf{\Omega}' - \mathbf{\Omega}).$$

Table 3.4: Spherical harmonics

with

$$\mu = 2\ell+1, \quad \nu = n+\ell \quad \text{and} \quad n = -\frac{Ze^2\beta}{8\pi\epsilon_0 W}. \tag{3.36}$$

A solution such that $\Xi(r)$ remains finite at the origin and vanishes at infinity is only possible if n is an integer satisfying the condition $n > \ell$. The solution for $u(\rho)$ is then the *associated Laguerre polynomials* $L_{\nu-\mu}^{(\mu)}(\rho)$ defined in table 3.5. The Laguerre polynomial $L_n(x) = L_n^{(0)}(x)$ is a polynomial of degree n with leading term $(-1)^n x^n/n!$. The associate Laguerre polynomial[2] $L_{n-\ell-1}^{(2\ell+1)}(x)$ is therefore a polynomial of degree $N = n-\ell-1$ with leading term $(-1)^N x^N/N!$ and having N zeroes in the interval $(0,\infty)$.

[2]Several definitions of associate Laguerre polynomials regarding subscripts, signs and normalizations are found in the literature. This therefore also leads to different values for the normalization constants $C_{n\ell}$ in (3.48). The present choice is in accordance with Abramowitz and Stegun: Handbook of Mathematical Functions.

Laguerre polynomials – generating function

$$L_\nu(x) = \frac{1}{\nu!}\exp(x)\,\frac{\mathrm{d}^\nu}{\mathrm{d}x^\nu}\left(x^\nu \exp(-x)\right) \tag{3.37}$$

Associate Laguerre polynomials

$$L_\nu^{(\mu)}(x) = (-1)^\mu \frac{\mathrm{d}^\mu}{\mathrm{d}x^\mu} L_{\nu+\mu}(x) \tag{3.38}$$

Some lower order polynomials

$$L_0 = 1, \quad L_1 = -x+1, \quad L_2 = \frac{1}{2}x^2 - 2x + 1, \quad L_3 = -\frac{1}{6}x^3 + \frac{3}{2}x^2 - 3x + 1$$
$$L_4 = \frac{1}{24}x^4 - \frac{2}{3}x^3 + 3x^2 - 4x + 4, \quad L_5 = -\frac{1}{120}x^5 + \frac{5}{24}x^4 - \frac{5}{3}x^3 + 5x^2 - 5x + 1$$
$$L_0^{(1)} = 1, \quad L_1^{(1)} = -x+2, \quad L_2^{(1)} = \frac{1}{2}x^2 - 3x + 3, \quad L_3^{(1)} = -\frac{1}{6}x^3 + 2x^2 - 6x + 4$$
$$L_0^{(3)} = 1, \quad L_1^{(3)} = -x+4, \quad L_2^{(3)} = \frac{1}{2}x^2 - 5x + 10, \quad L_3^{(3)} = -\frac{1}{6}x^3 + 3x^2 - 15x + 20$$

Recurrence relation

$$(\nu+1)L_{\nu+1}^{(\mu)} = (2\nu + \mu + 1 - x)L_\nu^{(\mu)} - (\nu+\mu)L_{\nu-1}^{(\mu)} \tag{3.39}$$

Normalization

$$\int_0^\infty x^{\mu+1}\exp(-x)L_\nu^{(\mu)}(x)L_{\nu'}^{(\mu)}(x)\,\mathrm{d}x = (2\nu+\mu+1)\frac{(\nu+\mu)!}{\nu!}\,\delta_{\nu,\nu'} \tag{3.40}$$

Table 3.5: Laguerre polynomials

Combining (3.34) and (3.36), the allowed energy states of the one-electron atom is seen to be

$$W_n = -\frac{1}{2}mc^2\frac{(Z\alpha)^2}{n^2} = -hc\,R_\infty\frac{Z^2}{n^2}, \tag{3.41}$$

where

$$\alpha = \frac{e^2}{4\pi\epsilon_0\hbar c} = \frac{1}{137.03599} \tag{3.42}$$

is the dimensionless *fine structure constant*, and

$$R_\infty = \frac{mc^2}{2hc}\,\alpha^2 = 1.0973732\cdot 10^7\ \mathrm{m}^{-1} \tag{3.43}$$

is the Rydberg constant for the infinitely massive nucleus. This result may be compared with the experimental value of the Rydberg constant $R_H = 1.096776\cdot 10^7\ \mathrm{m}^{-1}$. The relative error of $6\cdot 10^{-4}$ will be reduced to less than $2\cdot 10^{-7}$ when the effect of a finite mass of the nucleus is taken into account (see section 3.4). The stretching factor β in (3.34) is found to reduce to

$$\beta = \frac{Z}{na_B} \tag{3.44}$$

in terms of the *Bohr radius*.

$$a_B = \frac{4\pi\epsilon_0\hbar^2}{me^2} = 5.29177249 \cdot 10^{-11} \text{ m} \tag{3.45}$$

The radial part of the wave function depends on two quantum numbers, the *principal quantum number* $n > 0$ and the angular quantum number $\ell < n$,

$$\Xi_{n\ell}(r) = R_{n\ell}(\frac{2Zr}{na_B}) \tag{3.46}$$

where

$$R_{n\ell}(\rho) = \rho^\ell \exp(-\frac{\rho}{2})\, L^{(2\ell+1)}_{n-\ell-1}(\rho). \tag{3.47}$$

The result of applying the Schrödinger equation (3.10) to the one-electron atom is therefore that the wave functions available to the electron are

$$\Psi_{n\ell m_\ell}(\boldsymbol{r}) = C_{n\ell}\, R_{n\ell}(\frac{2Zr}{na_B})\, Y^{m_\ell}_\ell(\theta,\varphi) \tag{3.48}$$

with the normalization constant

$$C_{n\ell} = \left\{ \left(\frac{2Z}{na_B}\right)^3 \frac{(n-\ell-1)!}{2n(n+\ell)!} \right\}^{1/2}. \tag{3.49}$$

The wave function (3.48) depends on three integer quantum numbers n, ℓ and m_ℓ, satisfying the condition

$$n > \ell \geq |\, m_\ell \,| \,. \tag{3.50}$$

The energy levels (3.41) available to the electron only depend on the principal quantum number n. This means that different electron states may correspond to the identical energy. To the energy level W_n there are n^2 different states for the electron, that is, we have n^2-*fold energy degeneracy*. With the subsequent introduction of the electron spin we shall see that the energy degeneracy is actually $2n^2$-fold.

In figure 3.2 axial cuts of some of the lower order $|\,\Psi_{n\ell m_\ell}(\boldsymbol{r})\,|^2$ are given. Since such cuts are symmetric about the xyplane, only the $z > 0$ part of the cuts are displayed. The full elecron probability density distributions are found by rotating each of the cuts aroud the z-axis. The color scale is normalized to the maximum density for each plot.

In figure 3.3 the radial distribution of the electron density,

$$\begin{aligned}\rho_r(r) &= \int |\,\Psi_{n\ell m_\ell}(\boldsymbol{r})\,|^2\, r^2\, \mathrm{d}\cos\theta\, \mathrm{d}\varphi \\ &= \left(\frac{2Z}{na_B}\right)^3 \frac{(n-\ell-1)!}{2n(n+\ell)!} r^2 R^2_{n\ell}\left(\frac{2Z}{na_B}r\right),\end{aligned}$$

is plotted as a function of Zr/a_B for some of the lowest order electron states. We note that $\rho_r(r)\,\mathrm{d}r$ represents the probability for finding the electron in a spherical shell of radius r and thickness $\mathrm{d}r$. In the lowest energy state, $n = 1$, $\ell = 0$, the electron will normally be found in the neighborhood of the spherical shell of radius $r = a_B/Z$. With increasing n the electron

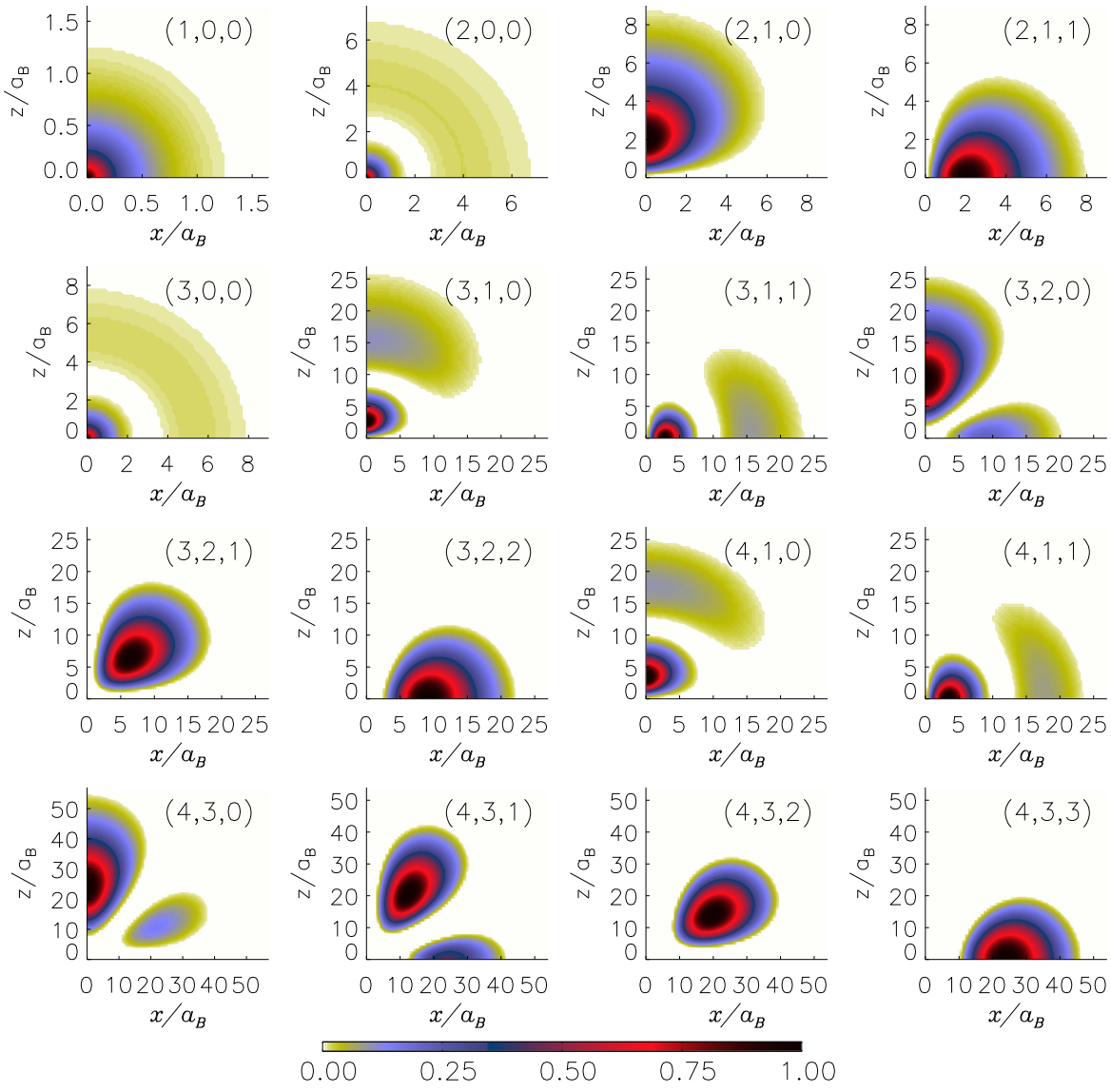

Figure 3.2: Axial cuts through some lower order electron probability densities

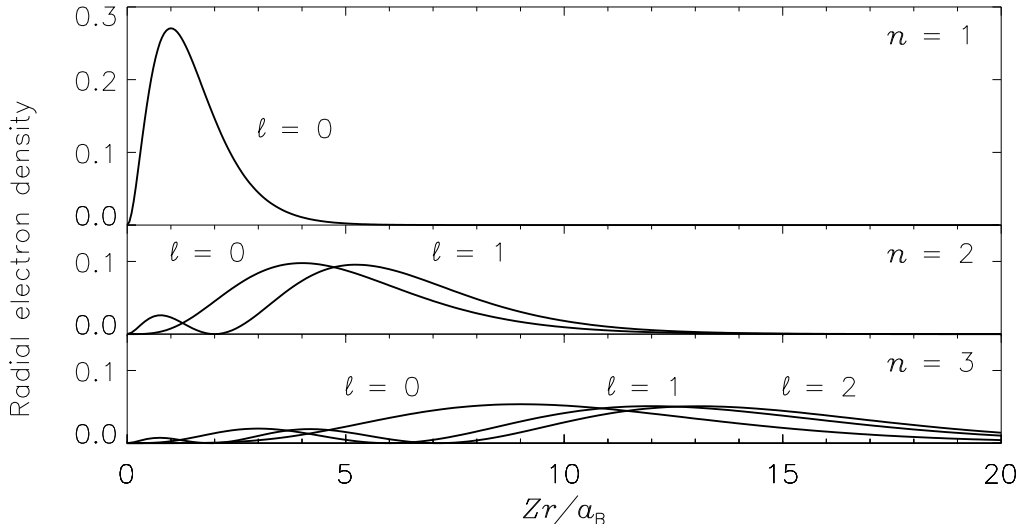

Figure 3.3: Radial distribution of the electron cloud

moves away from the nucleus with the average radius increasing as $r = n^2 a_B/Z$. At the same time the electron cloud spreads over a larger radial range. The average radial distance rule agrees with the classical point of views. Thus, if we assume the electron to move in circular orbits with centripetal acceleration in accordance with the acting Coulomb force,

$$m\frac{v^2}{r} = \frac{\partial}{\partial r}U(r) = -\frac{U}{r},$$

and require the total energy to be given by the quantum mechanical value (3.41),

$$H_0 = \frac{\boldsymbol{p}^2}{2m} + U(r) = \frac{1}{2}U(r) = W_n,$$

then the radius r of the electron orbit is indeed found to satisfy $r = n^2 a_B/Z$.

With the radial electron density given, the average values of different powers of r,

$$\langle\, r^k \,\rangle_{n\ell} = \int r^k \,|\, \Psi_{n\ell m_\ell}(\boldsymbol{r}) \,|^2 \; r^2 \,\mathrm{d}r \,\mathrm{d}\cos\theta \,\mathrm{d}\varphi, \tag{3.51}$$

may be calculated. For reference purposes we conclude this section by including table 3.6 containing the results for some selected powers k. Note that the $k = -3$ result is only valid for $\ell \neq 0$.

k	1	-1	-2	-3
$\left(\frac{Z}{a_B}\right)^k \langle\, r^k \,\rangle_{n\ell}$	$\frac{3}{2}n^2 - \frac{1}{2}\ell(\ell+1)$	n^{-2}	$n^{-3}(\ell+\frac{1}{2})^{-1}$	$n^{-3}\left[\ell(\ell+\frac{1}{2})(\ell+1)\right]^{-1}$

Table 3.6: Average values of some selected powers of r

Quiz 3.8 : Show that (3.27) for $m_\ell = 0$ reduces to

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[(1-x^2)\frac{\mathrm{d}}{\mathrm{d}x}\Theta\right] - \lambda\Theta = 0,$$

where $x = \cos\theta$ and λ is the separation constant. Argue that solutions may be written in the form

$$\Theta(x) = \cdots + a_k x^k + a_{k+2}x^{k+2} + \cdots$$

with

$$\frac{a_{k+2}}{a_k} = \frac{k(k+1)+\lambda}{(k+1)(k+2)}.$$

Thus, verify that non-singular solutions for Θ in the interval $x = [-1, 1]$ is only possible if $\lambda = -\ell(\ell+1)$, where ℓ is a non-negative integer number. [The infinite power series with $a_{k+2}/a_k \to 1$ as $k \to \infty$ diverges as $x^2 \to 1$.]

Quiz 3.9 : With $\vartheta = (x^2-1)^\ell$ and $x = \cos\theta$ show that

$$(1-x^2)\frac{\mathrm{d}\vartheta}{\mathrm{d}x} + 2\ell x\,\vartheta = 0.$$

By differentiating this equation $\ell + 1$ times, show that

$$[(1-x^2)\frac{\mathrm{d}^2}{\mathrm{d}x^2} - 2x\frac{\mathrm{d}}{\mathrm{d}x} + \ell(\ell+1)]\frac{\mathrm{d}^\ell\vartheta}{\mathrm{d}x^\ell} = 0,$$

and that $\mathrm{d}^\ell(x^2-1)^\ell/\mathrm{d}x^\ell$ is a solution of (3.27) for $m_\ell = 0$ (see table 3.3).

Quiz 3.10 : Make use of the properties listed in table 3.3 to show that the functions $P_\ell^{|m_\ell|}(\cos\theta)\exp(\imath m_\ell\varphi)$ will have even or odd symmetry under the transformation $\boldsymbol{r} \to -\boldsymbol{r}$ (that is, $\cos\theta \to -\cos\theta$ and $\varphi \to \varphi + \pi$) if ℓ is an even or odd integer, respectively. [This result will be needed for the discussion of selection rules.]

Quiz 3.11 : Verify (3.41) and (3.44).

Quiz 3.12 : Verify that for a given n there are n^2 different combinations of ℓ and m_ℓ allowed.

Quiz 3.13 : Make use of the recurrence relations for the Laguerre polynomials (see table 3.5) to prove by induction that $L_n^{(1)}(0) = n + 1$.

3.3 Physical Interpretation of Quantum Numbers

The quantum numbers n, ℓ and m_ℓ allow for an immediate physical interpretation. We have already seen the association between allowed energy levels and the quantum number n. The quantum number n is therefore also referred to as the principal quantum number. The quantum numbers ℓ and m_ℓ are both related to the angular momentum of the electron in its orbit.

The classical definition of angular momentum is $\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p}$. In quantum mechanics the same definition applies, but $\boldsymbol{L}$ is now an operator. The components of $\boldsymbol{L}$ as well as $\boldsymbol{L}^2$ all commute with the Hamiltonian (3.17), $[H_0, \boldsymbol{L}] = 0$ and $[H_0, \boldsymbol{L}^2] = 0$. $\boldsymbol{L}^2$ and the component

of $\boldsymbol{L}$ along an arbitrary chosen z-axis similarly commute, $[\boldsymbol{L}^2, L_z] = 0$. The components of $\boldsymbol{L}$ do not, however, commute among themselves, instead it is easily shown that

$$[L_x, L_y] = \imath\hbar L_z \tag{3.52}$$

and similar relations derived by cyclic permutations of the indices. According to the quantum mechanical recipe of table 3.1, $\boldsymbol{L}^2$ and L_z are both simultaneously measurable with H, but not the full $\boldsymbol{L}$-vector. In fact, simple calculations show that

$$\boldsymbol{L}^2\Psi_{n\ell m_\ell} = \hbar^2\ell(\ell+1)\Psi_{n\ell m_\ell} \tag{3.53}$$

$$L_z\Psi_{n\ell m_\ell} = \hbar m_\ell \Psi_{n\ell m_\ell}. \tag{3.54}$$

The angular quantum number ℓ thus determines the magnitude of the orbital angular momentum. The azimuthal quantum number m_ℓ represents the component of the orbital angular momentum along the (arbitrarily chosen) z-axis. For a given integer value of ℓ the requirement $\ell \geq \mid m_\ell \mid$ allows for $2\ell + 1$ different integer values for m_ℓ. Possible directions for the angular momentum vector for a given choice of z-axis are illustrated for the case $\ell = 2$ in figure 3.4. In particular, it is seen that the angular momentum vector will never fall along the z-axis. If it was to fall along the z-axis, L_x and L_y would both vanish while at the same time also L_z would take a definite value. This is in conflict with the fact that the components of the orbital angular momentum vector $\boldsymbol{L}$ are not simultaneously measurable quantities.

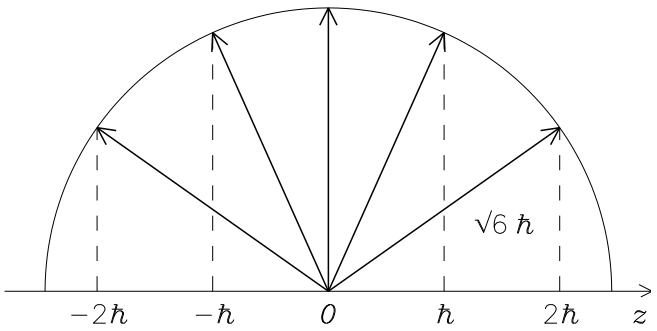

Figure 3.4: Quantization of the angular momentum vector, $\ell = 2$

Quiz 3.14 : In spherical coordinates show that

$$\boldsymbol{L} = -\imath\hbar\boldsymbol{r}\times\nabla = -\imath\hbar\left(\hat{\boldsymbol{\varphi}}\frac{\partial}{\partial\theta} - \hat{\boldsymbol{\theta}}\frac{1}{\sin\theta}\frac{\partial}{\partial\varphi}\right), \tag{3.55}$$

$$L_z = -\imath\hbar\frac{\partial}{\partial\varphi}, \tag{3.56}$$

$$\boldsymbol{L}^2 = -\imath\hbar\boldsymbol{r}\cdot(\nabla\times\boldsymbol{L}) = -\hbar^2\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta\frac{\partial}{\partial\theta}) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}\right], \tag{3.57}$$

and

$$-\frac{\hbar^2}{2m}\nabla^2 = -\frac{\hbar^2}{2mr^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{\boldsymbol{L}^2}{2mr^2}. \tag{3.58}$$

Remember that these operators all operate on scalar wave functions $\Psi(\boldsymbol{r})$ and make use of formulas listed in section A.2.3.

Quiz 3.15 : Make use of the results of quiz 3.14 to show that the operators H_0, $\boldsymbol{L}^2$ and L_z all commute,

$$[H_0, L_z] = 0, \quad [H_0, \boldsymbol{L}^2] = 0, \quad \text{and} \quad [\boldsymbol{L}^2, L_z] = 0.$$

Quiz 3.16 : Verify (3.52), (3.53) and (3.54).

Quiz 3.17 : With Ψ given by (3.48), show that the orbital angular momentum vector $\boldsymbol{L}$ as given by (3.55) is *not* measurable, that is, it is not possible to find a constant vector $\boldsymbol{l}$ such that $\boldsymbol{L}\Psi = \boldsymbol{l}\Psi$.

Quiz 3.18 : The operators L_+ and L_- defined by

$$L_\pm \equiv L_x \pm \iota L_y \tag{3.59}$$

have the property of raising or lowering the m-index of the spherical harmonics,

$$L_\pm Y_\ell^m = \hbar\sqrt{\ell(\ell+1) - m(m\pm 1)}\, Y_\ell^{m\pm 1}. \tag{3.60}$$

Show that

$$\begin{aligned} L_\pm &= \hbar \exp(\pm\iota\varphi)\left(\pm\frac{\partial}{\partial\theta} + \iota\cot\theta\frac{\partial}{\partial\varphi}\right) \\ &= \hbar \exp(\pm\iota\varphi)\left(\mp\sqrt{1-x^2}\frac{\partial}{\partial x} + \iota\frac{x}{\sqrt{1-x^2}}\frac{\partial}{\partial\varphi}\right) \end{aligned}$$

where $x = \cos\theta$. Next show that

$$L_+ P_\ell^m(x)\exp(\iota m\phi) = \hbar P_\ell^{m+1}(x)\exp(\iota(m+1)\phi), \tag{3.61}$$

$$L_- P_\ell^m(x)\exp(\iota m\phi) = \hbar(\ell+m)(\ell-m+1)P_\ell^{m-1}(x)\exp(\iota(m-1)\phi) \tag{3.62}$$

and thus verify that (3.60) is correct.

Quiz 3.19 : Show that $[\boldsymbol{L}^2, L_\pm] = 0$ and $[L_z, L_\pm] = \pm\hbar L_\pm$ where $L_\pm$ is defined by (3.59). If Ψ is an eigenfunction of $\boldsymbol{L}^2$ and L_z with eigenvalues $\hbar^2\lambda$ and $\hbar\mu$, respectively, then $L_\pm\Psi$ is either an eigenfunction of $\boldsymbol{L}^2$ and L_z with eigenvalues $\hbar^2\lambda$ and $\hbar(\mu\pm 1)$, or the null function. From the indentity

$$\boldsymbol{L}^2 = L_\mp L_\pm + L_z^2 \pm \hbar L_z \tag{3.63}$$

prove that the eigenvalues of $\boldsymbol{L}^2$ must be of the form $\hbar^2\mu_{\max}(\mu_{\max}+1)$ where $\mu_{\max}$ is the maximum value allowed for μ?

3.4 The Isotope Effect

In the discussion of the one-electron atom in section 3.2 several different effects were neglected. The nucleus was assumed to have infinite mass, the electron spin was not taken into account, effects due to external fields were not included, and so on. We must therefore consider this discussion only as a lowest order theory. In particular, we shall refer to (3.17) as the zeroth order Hamiltonian H_0 in the following. We shall now consider a number of correctional or additional effects, one at the time, starting with the *isotope effect*. The isotope effect arises because the nuclear mass M_n is finite and therefore that the nucleus will respond slightly to the motion of the electron. This response is reflected in the two-body Hamiltonian

$$H = \frac{\boldsymbol{p}_n^2}{2M_n} + \frac{\boldsymbol{p}_e^2}{2m} + U(|\,\boldsymbol{r}_e - \boldsymbol{r}_n\,|), \tag{3.64}$$

where $\boldsymbol{r}_n$ and $\boldsymbol{r}_e$ refer to the position vectors of the nucleus and the surrounding electron. The corresponding momentum vectors are $\boldsymbol{p}_n = M_n\dot{\boldsymbol{r}}_n$ and $\boldsymbol{p}_e = m\dot{\boldsymbol{r}}_e$.

The two-body central force problem can, however, always be reduced to an equivalent one-body problem for an equivalent particle of the *reduced mass*

$$\mu = M_n m/(M_n + m) \tag{3.65}$$

and position vector $\boldsymbol{r} = \boldsymbol{r}_e - \boldsymbol{r}_n$, in addition to a uniform motion of the center-of-mass of the system, $\boldsymbol{R} = (M_n\boldsymbol{r}_n + m\boldsymbol{r}_e)/(M_n + m)$. This conclusion follows from the fact that the two-body Hamiltonian (3.64) can be written as

$$H = \frac{\boldsymbol{P}^2}{2M} + \frac{\boldsymbol{p}^2}{2\mu} + U(r) \tag{3.66}$$

with $M = M_n + m$, $\boldsymbol{P} = M\dot{\boldsymbol{R}}$ and $\boldsymbol{p} = \mu\dot{\boldsymbol{r}}$ (see quiz 3.21).

The quantum mechanical recipe (see table 3.1) is to replace the momentum vectors of the equivalent particle and the center-of-mass system by operators $\boldsymbol{p} = -\imath\hbar\nabla_{\boldsymbol{r}}$ and $\boldsymbol{P} = -\imath\hbar\nabla_{\boldsymbol{R}}$, respectively. The Schrödinger equation, now a partial differential equation for the wave function in six independent variables $\Psi = \Psi(\boldsymbol{R}, \boldsymbol{r}, t)$, naturally separates into equations in $\boldsymbol{r}$ and $\boldsymbol{R}$ by the assumption $\Psi = \Psi_{\boldsymbol{r}}(\boldsymbol{r})\Psi_{\boldsymbol{R}}(\boldsymbol{R}, t)$. The first part of the problem, the equivalent one-body part,

$$(-\frac{\hbar^2}{2\mu}\nabla_{\boldsymbol{r}}^2 + U(r))\Psi_{\boldsymbol{r}}(\boldsymbol{r}) = W\Psi_{\boldsymbol{r}}(\boldsymbol{r}),$$

is identical to our one-electron atom with fixed nucleus (section 3.2). The only change necessary to incorporate the effect of a finite nuclear mass is therefore to replace the electron mass m in (3.17) by the reduced mass μ as given by (3.65). The corrected discrete energy levels for the atom are given by

$$W_n = -hc\,\frac{M_n}{M}R_\infty\frac{Z^2}{n^2}. \tag{3.67}$$

The theoretical value of the modified Rydberg constant $M_nR_\infty/M = 1.0967758 \cdot 10^7$ m^{-1} for the H I atom now agrees with the experimental value to within a relative error $2 \cdot 10^{-7}$. This is a very satisfactory result for any physical theory! The result is important not only for the precision achieved, but also because the result provides a method for determining astrophysical isotope abundances. The energy levels and therefore also the spectra of different isotopes of a given element are slightly shifted relative to each other.

The second and implicitly time-dependent part of the problem, the center-of-mass motion of the atom,

$$-\frac{\hbar^2}{2M}\nabla^2_{\boldsymbol{R}}\Psi_{\boldsymbol{R}}(\boldsymbol{R}) = W_{CM}\Psi_{\boldsymbol{R}}(\boldsymbol{R}),$$

is significantly different in that it allows for arbitrary (continuous) values of the energy W_{CM}. This part of the problem is considered in quiz 3.23.

Quiz 3.20 : Masses m_1 and m_2 at $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ act on each other with equal and opposite forces $\boldsymbol{F}$ depending on their distance $r = |\,\boldsymbol{r}_1 - \boldsymbol{r}_2\,|$. The equations of motion are

$$m_1\ddot{\boldsymbol{r}}_1 = \boldsymbol{F}(|\,\boldsymbol{r}_1 - \boldsymbol{r}_2\,|)$$
$$m_2\ddot{\boldsymbol{r}}_2 = -\boldsymbol{F}(|\,\boldsymbol{r}_1 - \boldsymbol{r}_2\,|).$$

Introduce the center-of-mass and relative vectors $\boldsymbol{R} = (m_1\boldsymbol{r}_1 + m_2\boldsymbol{r}_2)/(m_1 + m_2)$ and $\boldsymbol{r} = \boldsymbol{r}_1 - \boldsymbol{r}_2$ as new variables and show that the two-body problem reduces to a uniform motion of the center-of-mass in addition to a *one-body* problem for a particle with reduced mass $\mu = m_1m_2/(m_1 + m_2)$ in a spherically symmetric force field $\boldsymbol{F}(r)$.

Quiz 3.21 : Verify that (3.66) is equivalent to (3.64).

Quiz 3.22 : When going from H I to D I what is the relative difference in the energy levels, $(W_n^{\mathrm{DI}} - W_n^{\mathrm{HI}})/W_n^{\mathrm{HI}}$? [D represents deuterium, that is ^{2}H.] What is the corresponding value when going from ^{3}He II to ^{4}He II?

Quiz 3.23 : Argue that the elementary dispersive plane wave,

$$\psi_{\boldsymbol{R}} \sim \exp(\iota\boldsymbol{k}\cdot\boldsymbol{R} - \iota\frac{\hbar k^2}{2M}t), \quad \text{for any } \boldsymbol{k},$$

is a formal, but strictly speaking, physically unacceptable solution of the Schrödinger equation for the center-of-mass motion of the atom, whereas the superposition

$$\Psi_{\boldsymbol{R}}(\boldsymbol{R},t) = \int A(\boldsymbol{k})\exp(\iota\boldsymbol{k}\cdot\boldsymbol{R} - \iota\frac{\hbar k^2}{2M}t)\,\mathrm{d}^3\boldsymbol{k}$$

is acceptable. If $\Psi_{\boldsymbol{R}}(\boldsymbol{R}, t = 0) \sim \exp(-\boldsymbol{R}^2/(2\alpha) + \iota\boldsymbol{k}_0\cdot\boldsymbol{R})$ what is $\langle\,\boldsymbol{R}\,\rangle_{t=0}$ and $\langle\,\boldsymbol{P}\,\rangle_{t=0}$? What do you predict for the values of $\langle\,\boldsymbol{R}\,\rangle$ and $\langle\,\boldsymbol{P}\,\rangle$ at later times? How do the phase and group velocities of the elementary dispersive plane wave compare with the classical center-of-motion velocity? [Hint: See quiz 3.6.]

Quiz 3.24 : For the case studied in quiz 3.23 show that $A(\boldsymbol{k}) \sim \exp(-\alpha(\boldsymbol{k} - \boldsymbol{k}_0)^2/2)$ and therefore

$$\Psi_{\boldsymbol{R}}(\boldsymbol{R},t) \sim \exp\left(-\frac{(\boldsymbol{R} - \hbar\boldsymbol{k}_0t/M)^2}{2(\alpha + \iota\hbar t/M)} + \iota\boldsymbol{k}_0\cdot\boldsymbol{R} - \iota\frac{\hbar\boldsymbol{k}_0^2}{2M}t\right).$$

If $\langle\,(\Delta\boldsymbol{R})^2\,\rangle_{t=0} = a_B^2$, how long will it take for the uncertainty in $\boldsymbol{R}$ to double? Is the quantum mechanical spreading effect important? [Hint: See table 1.1.]

3.5 An External Magnetic Field

From the classical point of view, the motion of the electron in a closed orbit in the central force field of the nucleus represents a magnetic dipole moment $\boldsymbol{m}_L = I\boldsymbol{\mathcal{A}}$, where I is the equivalent current carried by the electron in its orbit, $I = e/T$ and T is the orbital period of the electron. $\boldsymbol{\mathcal{A}}$ is the area of the orbit, with the positive direction in accordance with the right-hand rule, that is, $\boldsymbol{\mathcal{A}}$ points in the direction traveled by a right hand screw when rotated in the direction of the orbital current. As illustrated in figure 3.5, this area can be expressed as

$$\boldsymbol{\mathcal{A}} = \int \mathrm{d}\boldsymbol{\mathcal{A}} = -\frac{1}{2}\int_0^T \boldsymbol{r} \times \boldsymbol{v}\,\mathrm{d}t = -\frac{T}{2m}\boldsymbol{L}.$$

The latter step follows since $\boldsymbol{L} = \boldsymbol{r} \times m\boldsymbol{v}$ is a constant of motion.

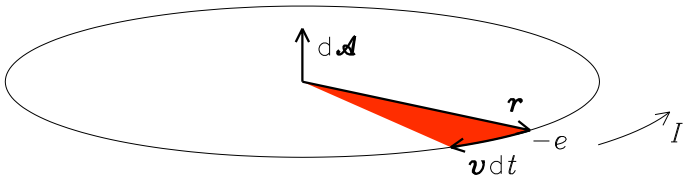

Figure 3.5: Orbital magnetic dipole moment

The *orbital magnetic dipole moment* of the electron is thus

$$\boldsymbol{m}_L = -\frac{e}{2m}\boldsymbol{L}, \tag{3.68}$$

that is, a constant of proportionality $e/2m$ relates the orbital magnetic dipole moment $\boldsymbol{m}_L$ to the orbital angular momentum $\boldsymbol{L}$ of the electron. The relation is sometimes written in the form $\boldsymbol{m}_L = \mu_B\boldsymbol{L}/\hbar$ where $\mu_B = e\hbar/(2m) = 9.2740155 \cdot 10^{-24}$ Js is the *Bohr magneton*. We note that when the isotope effect is included, the reduced electron mass should be used in this expression.

In an external magnetic field the orbital magnetic dipole moment represents an additional potential energy term in the Hamiltonian (see quiz 3.26)

$$H_{BL} = -\boldsymbol{m}_L \cdot \boldsymbol{B} = \frac{e}{2m}\boldsymbol{B} \cdot \boldsymbol{L}. \tag{3.69}$$

Quantum mechanically, H_{BL} should be treated as an extra operator term to be included in the Hamiltonian, $H = H_0 + H_{BL}$. The zero-order Hamiltonian H_0 is given by (3.17). With the z-axis along the magnetic field and L_z and H_0 simultaneously measurable, the wave function does not change with the introduction of the extra term. The energy levels do, however, take new values

$$W = W^{(0)} + \Delta W_{BL}$$

with $W^{(0)}$ given by (3.41), and, using (3.54),

$$\Delta W_{BL} = \hbar\,\frac{eB}{2m}\,m_\ell = \hbar\omega_L m_\ell. \tag{3.70}$$

Here $\omega_L = eB/2m$ is the angular *Larmor frequency* of the electron in the external magnetic field. This result would mean that the energy degeneracy discussed in section 3.2 would be

partially lifted by the external magnetic field, with the energy level n splitting into $2n-1$ equidistant levels. Unfortunately, this result *does not* generally fit with experiments (although it will be recovered from the more accurate theory under certain conditions for many-electron atoms or for very strong magnetic fields). To rectify the result it is necessary to take into account that electrons also possess internal spin.

Quiz 3.25 : For one-electron atoms with "spin-less" electrons in an external magnetic field $\boldsymbol{B}$, determine the energy degeneracy of the energy level W_{nm_ℓ}.

Quiz 3.26 : Fill in the missing steps in the following derivation of the force $\boldsymbol{F}$ acting on a small current loop, carrying a current I and with magnetic moment $\boldsymbol{m}_L = I\boldsymbol{\mathcal{A}}$, in an external magnetic field $\boldsymbol{B}$ (satisfying $\nabla\cdot\boldsymbol{B}=0$ and $\nabla\times\boldsymbol{B}=0$):

$$\begin{aligned}\boldsymbol{F} = I\oint \mathrm{d}\boldsymbol{\ell}\times\boldsymbol{B} = I\int(\mathrm{d}^2\boldsymbol{\mathcal{A}}\times\nabla)\times\boldsymbol{B} = I\int[\,\mathrm{d}^2\boldsymbol{\mathcal{A}}\times(\nabla\times\boldsymbol{B})+\cdots]\\ = \boldsymbol{m}_L\cdot\nabla\boldsymbol{B} = -\boldsymbol{m}_L\times(\nabla\times\boldsymbol{B})+\cdots = \nabla(\boldsymbol{m}_L\cdot\boldsymbol{B}).\end{aligned}$$

Argue that the potential energy of a magnetic moment $\boldsymbol{m}_L$ in a magnetic field $\boldsymbol{B}$ is given by $H_B = -\boldsymbol{m}_L\cdot\boldsymbol{B}$. [*Hint*: Use the generalized Stoke integral theorem (A.51).]

3.6 The Electron Spin

The Stern-Gerlach experiment (1921) in which a collimated beam of silver atoms were shot between the pole pieces of a strong permanent magnet toward a photographic plate, as illustrated in figure 3.6, gave an unexpected result. Between the pole pieces an electrically neutral silver atom possessing a magnetic moment $\boldsymbol{m}_L$, for instance due to the orbital motion of the outermost valence electron, should be subject to a force $\boldsymbol{F} = \boldsymbol{m}_L\cdot\nabla\boldsymbol{B}$ (see quiz 3.26). However, instead of a diffuse blackening of the photographic plate due to randomly oriented forces on the silver atoms due to individually oriented orbital magnetic moments in the strongly inhomogeneous magnetic field, a distinct two-banded pattern was observed. The experiment was explained by Goudsmit and Uhlenbeck (1925) by assigning to the electron a magnetic dipole moment $\boldsymbol{m}_S$ associated with a *spin angular momentum* $\boldsymbol{S}$ of the valence electron of the silver atom. The Hermitian spin operator $\boldsymbol{S}$ is independent of the position $\boldsymbol{r}$ of the electron, but operates in an internal spin parameter space of the electron. Assuming, in analogy with the orbital magnetic dipole moment, a linear relationship between the spin magnetic dipole moment and the spin operator, and that the spin operator behaves similarly to the orbital angular momentum $\boldsymbol{L}$ as regards quantization rules, that is,

$$\boldsymbol{S}^2\,|\,\Psi\,\rangle = \hbar^2 s(s+1)\,|\,\Psi\,\rangle \tag{3.71}$$

$$S_z\,|\,\Psi\,\rangle = \hbar m_s\,|\,\Psi\,\rangle, \tag{3.72}$$

the splitting of the beam could be understood if the corresponding quantum numbers s and m_s were only allowed to take the values $s=\frac{1}{2}$ and $m_s=\pm\frac{1}{2}$.

In (3.71) and (3.72) we introduced the notation $|\,\Psi\,\rangle$ for the wave function. This is done as a reminder that the wave function, in addition to the previous space-dependent part $\Psi_{n\ell m_\ell}(\boldsymbol{r})$, now also contains a factor describing the spin state of the electron. This extra factor

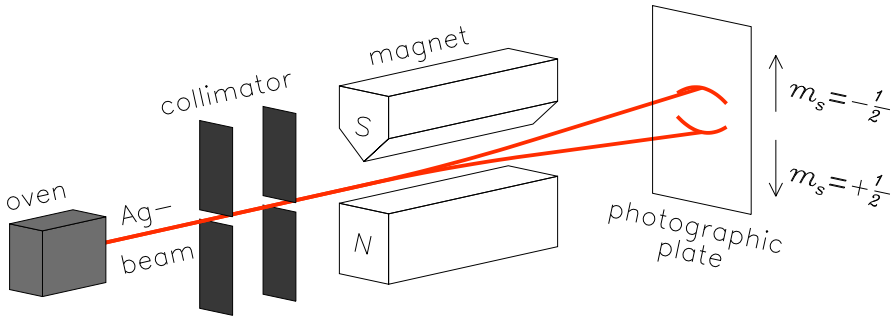

Figure 3.6: Schematics of the Stern-Gerlach experiment

is sometimes referred to as χ_{sm_s} or alternatively as α or β corresponding to "spin up" or "spin down". To specify a particular wave function the values of the quantum numbers involved must be prescribed. With the electron spin included, this means the specification of five quantum numbers: n, ℓ, m_ℓ, s and m_s. When an explicit indication of the quantum numbers is needed we will make use of the notation $|\,\Psi\,\rangle = |\,n\,\ell\,m_\ell\,s\,m_s\,\rangle$. In terms of the the spin factors α and β, we may identify $|\,n\,\ell\,m_\ell\,\frac{1}{2}\,\frac{1}{2}\,\rangle = \alpha\Psi_{n\ell m_\ell}(\boldsymbol{r})$ and $|\,n\,\ell\,m_\ell\,\frac{1}{2}\,-\frac{1}{2}\,\rangle = \beta\Psi_{n\ell m_\ell}(\boldsymbol{r})$. In the subsequent discussion we shall also find it convenient to represent a complete set of quantum numbers by a single symbol, that is, let $|\,i\,\rangle$, $|\,f\,\rangle$ or $|\,q\,\rangle$ indicate different quantum states.

With $s = \frac{1}{2}$, the spin state space consists of only two different states, in accordance with the Stern-Gerlach experiment. The spin factors α and β may therefore be represented by two-dimensional column matrices, for instance,

$$\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}.$$

The scalar product of α with itself now reduces to the matrix product $\alpha^\dagger\alpha$ between α and the adjoint matrix with $\alpha^\dagger = (\alpha_1^*, \alpha_2^*)$.

We shall in the following find it convenient to write the normalization condition for the wave function $|\,\Psi\,\rangle$, in the form

$$\langle\,\Psi\,|\,\Psi\,\rangle \equiv (|\,\Psi\,\rangle, |\,\Psi\,\rangle) = 1. \tag{3.73}$$

For the case that $|\,\Psi\,\rangle = \alpha\Psi_{n\ell m_\ell}$, this condition takes the explicit form

$$\langle\,\Psi\,|\,\Psi\,\rangle = \alpha^\dagger\,\alpha \int |\,\Psi_{n\ell m_\ell}(\boldsymbol{r})\,|^2\ \mathrm{d}^3\boldsymbol{r} = 1.$$

With $\Psi_{n\ell m_\ell}(\boldsymbol{r})$ normalized, this leads to the corresponding normalization requirements on the spin factors,

$$\alpha^\dagger\,\alpha = \beta^\dagger\,\beta = 1.$$

The wave functions corresponding to different eigenvalues are orthogonal. This is a consequence of the Hermitian property of the spin operators (see table 3.2). This leads to the additional requirement

$$\alpha^\dagger\,\beta = \beta^\dagger\,\alpha = 0,$$

and therefore that we may choose the representations

$$\alpha = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \beta = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{3.74}$$

In the following discussion we shall find it convenient, for arbitrary Hermitian operators Q, also to introduce the notation

$$\langle\, \Phi \mid Q \mid \Psi \,\rangle \equiv (Q \mid \Phi\,\rangle, |\, \Psi\,\rangle) = (|\, \Phi\,\rangle, Q \mid \Psi\,\rangle). \tag{3.75}$$

The new notation does not indicate the position of the "comma" in the scalar product. For Hermitian operators Q this is allowed since the results of letting the operator act to the "right" or to the "left" are identical.

In section 3.5 a linear relationship between the magnetic dipole moment $\boldsymbol{m}_L$ and the angular momentum $\boldsymbol{L}$ associated with the orbital motion of the electron was demonstrated. A similar relationship has been established between the magnetic dipole moment $\boldsymbol{m}_S$ due to the spinning electron and the electron spin vector $\boldsymbol{S}$,

$$\boldsymbol{m}_S = -g_s \frac{e}{2m} \boldsymbol{S}. \tag{3.76}$$

The dimensionless factor g_s is called the *gyro-magnetic ratio* of the electron. It turns out to that $g_s \approx 2$. We shall see that this difference in the proportionally factors between the orbital and spin magnetic momenta and the orbital and spin angular momenta leads to some complications.

For the proper discussion of the effects of electron spin, including the magnitude of the gyro-magnetic ratio, g_s, the Schrödinger equation must be replaced by the relativistic Dirac equation[3]. Such a discussion will, however, fall outside our scope. For the present purposes it will suffice to note that to lowest order the Hamiltonian H_0 and therefore the corresponding energy levels are independent of the electron spin. This means that the degeneracy of the energy level n as discussed in section 3.2 should be increased by a factor 2 to $2n^2$ when including electron spin.

In an external magnetic field $\boldsymbol{B}$, the spin magnetic dipole moment $\boldsymbol{m}_S$ of the electron will give rise to an additional potential energy term in the Hamiltonian of the form

$$H_{BS} = -\boldsymbol{m}_S \cdot \boldsymbol{B} \approx \frac{e}{m} \boldsymbol{B} \cdot \boldsymbol{S}. \tag{3.77}$$

We shall return to the effects of electron spin in an external magnetic field in section 3.13.

Quiz 3.27 : Prove that the spin eigenvectors α and β must be orthogonal.

Quiz 3.28 : Show that if the spin states α and β are represented by the vectors (3.74) and the components of the spin operator $\boldsymbol{S} = \hbar\boldsymbol{\sigma}/2$ by the two-by-two Pauli spin matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -\iota \\ \iota & 0 \end{pmatrix} \quad \text{and} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{3.78}$$

[3]Dirac's relativistic electron equation predicts $g_s = 2$. Including quantum electrodynamic (QED) effects the corrected value is $g_s = 2.002319314$, in close agreement with experimental values.

then

$$\boldsymbol{S}^2\alpha = \frac{3}{4}\hbar^2\alpha, \quad \boldsymbol{S}^2\beta = \frac{3}{4}\hbar^2\beta, \quad S_z\alpha = \frac{1}{2}\hbar\,\alpha \quad \text{and} \quad S_z\beta = -\frac{1}{2}\hbar\,\beta,$$

in accordance with (3.71) and (3.72).

Quiz 3.29 : A simple calculation will show that the orbital angular momentum operator $\boldsymbol{L} = -\iota\hbar\boldsymbol{r} \times \nabla$ satisfies the commutator rule

$$[L_x, L_y] \equiv L_xL_y - L_yL_x = \iota\hbar L_z.$$

Show that the spin angular operator $\boldsymbol{S} = \hbar\boldsymbol{\sigma}/2$ with $\boldsymbol{\sigma}$ represented by (3.78) satisfies the corresponding matrix commutator rule

$$[S_x, S_y] \equiv S_xS_y - S_yS_x = \iota\hbar S_z.$$

Quiz 3.30 : Show that the spin operators $S_\pm \equiv S_x \pm \iota S_y$ satisfy the following relations,

$$S_+ \left\{ \begin{array}{c} \alpha \\ \beta \end{array} \right\} = \hbar \left\{ \begin{array}{c} 0 \\ \alpha \end{array} \right\} \quad \text{and} \quad S_- \left\{ \begin{array}{c} \alpha \\ \beta \end{array} \right\} = \hbar \left\{ \begin{array}{c} \beta \\ 0 \end{array} \right\}, \tag{3.79}$$

that is, $S_\pm$ acts to raise or lower the m_s value. Note that this is in accordance with the results for $L_\pm$ (see quiz 3.18).

Quiz 3.31 : Verify that

$$\boldsymbol{L} \cdot \boldsymbol{S} = L_zS_z + \frac{1}{2}(L_+S_- + L_-S_+) \tag{3.80}$$

Quiz 3.32 : An electron is placed in a constant magnetic field $\boldsymbol{B} = B\hat{\boldsymbol{z}}$. Taking account only of the spin part (3.77) of the Hamiltonian and letting $\boldsymbol{S} = \hbar\boldsymbol{\sigma}/2$ with $\boldsymbol{\sigma}$ given by (3.78), show that the time-dependent wave function $\mid \psi(t) \rangle$, the eigenfunction of the time-dependent Schrödinger equation (3.11), may be written as

$$\mid \psi(t) \rangle = c_\alpha \exp(-\iota\frac{\omega_0}{2}t)\,\alpha + c_\beta \exp(\iota\frac{\omega_0}{2}t)\,\beta,$$

where c_α and c_β are complex constants and $\omega_0 = eB/m$. Thus show that

$$\begin{aligned} \langle S_x \rangle \equiv \langle \psi(t) \mid S_x \mid \psi(t) \rangle &= \mathrm{Re}\,(c_\alpha^* c_\beta\,\hbar\,\exp(\iota\omega_0 t)) \\ \langle S_y \rangle \equiv \langle \psi(t) \mid S_y \mid \psi(t) \rangle &= \mathrm{Im}\,(c_\alpha^* c_\beta\,\hbar\,\exp(\iota\omega_0 t)) \\ \langle S_z \rangle \equiv \langle \psi(t) \mid S_z \mid \psi(t) \rangle &= \frac{\hbar}{2}\left(\mid c_\alpha \mid^2 - \mid c_\beta \mid^2\right) \\ \langle S_x \rangle^2 + \langle S_y \rangle^2 + \langle S_z \rangle^2 &= \frac{\hbar^2}{4}. \end{aligned}$$

Argue that the average spin vector $\langle \boldsymbol{S} \rangle$ of the electron behaves "classically", processing as a spinning top around the direction of the magnetic field.

3.7 Total Angular Momentum

The orbital momentum $\boldsymbol{L}$ and the spin momentum $\boldsymbol{S}$ both represent angular momenta. In classical physics we may add such vectors in order to obtain the total angular momentum of the system. This holds true also in quantum mechanics. The rules for how to quantize the different angular momenta are, however, more complicated. The complications arise because the different components of the $\boldsymbol{L}$ and $\boldsymbol{S}$ operators do not commute among themselves.

The two groups of operators, H_0, $\boldsymbol{L}^2$, L_z and $\boldsymbol{S}^2$, S_z operate in different spaces, the former ones in the continuous configuration space $\boldsymbol{r}$, the latter ones in the discrete spin parameter space. The two groups of operators therefore trivially commute. Thus, $\boldsymbol{L}^2$, L_z, $\boldsymbol{S}^2$ and S_z may be quantized simultaneously with the Hamiltonian H_0 with corresponding eigenfunctions $|\, n\,\ell\, m_\ell\, s\, m_s \,\rangle$. The quantization rules imply that for a given value of $\boldsymbol{L}^2$ only certain values of L_z are allowed, that is, only certain directions of $\boldsymbol{L}$ are allowed. The same argument applies to $\boldsymbol{S}$ and even holds true for the *total angular momentum* operator

$$\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S}.$$

It can be shown (see quiz 3.33) that also the group of operators $\boldsymbol{L}^2$, $\boldsymbol{S}^2$, $\boldsymbol{J}^2$ and J_z may be quantized simultaneously with H_0, namely

$$\begin{aligned} \boldsymbol{L}^2\,|\,\Psi\,\rangle &= \hbar^2\ell(\ell+1)\,|\,\Psi\,\rangle \\ \boldsymbol{S}^2\,|\,\Psi\,\rangle &= \hbar^2 s(s+1)\,|\,\Psi\,\rangle \\ \boldsymbol{J}^2\,|\,\Psi\,\rangle &= \hbar^2 j(j+1)\,|\,\Psi\,\rangle \\ J_z\,|\,\Psi\,\rangle &= \hbar m_j\,|\,\Psi\,\rangle, \end{aligned} \tag{3.81}$$

where j and m_j are new quantum numbers replacing the former orbital and spin azimuthal quantum numbers m_ℓ and m_s. The new quantum numbers must satisfy the constraints

$$j = |\,\ell - s\,|, |\,\ell - s\,| + 1, \cdots, \ell + s \tag{3.82}$$

and

$$m_j = -j, -j+1, \cdots, j. \tag{3.83}$$

The constraint (3.82) is a statement of the fact that the sum of two vectors takes its minimum and maximum lengths when the two vectors are anti-parallel and parallel, respectively. The wave function $|\,\Psi\,\rangle$ as given in (3.81) is specified in terms of the five quantum numbers n, ℓ, s, j and m_j and shall also be denoted $|\, n\,\ell\, s\, j\, m_j \,\rangle$.

Wave functions of the type $|\, n\,\ell\, s\, j\, m_j \,\rangle$ are expressible as linear combinations of the wave functions of the type $|\, n\,\ell\, m_\ell\, s\, m_s \,\rangle$, and vice versa. In fact, for $s = \frac{1}{2}$ the following result can be shown to be valid

$$|\, n\,\ell{=}0\, s\, j{=}\frac{1}{2}\, m_j \,\rangle = |\, n\,\ell{=}0\, m_\ell{=}0\, s\, m_s{=}m_j \,\rangle \tag{3.84}$$

$$\begin{aligned} |\, n\,\ell\, s\, j{=}\ell{\pm}\frac{1}{2}\, m_j \,\rangle &= \sqrt{\frac{\ell \pm m_j + \frac{1}{2}}{2\ell+1}}\; |\, n\,\ell\, m_\ell{=}m_j{-}\frac{1}{2}\, s\, m_s{=}\frac{1}{2} \,\rangle \\ &\pm \sqrt{\frac{\ell \mp m_j + \frac{1}{2}}{2\ell+1}}\; |\, n\,\ell\, m_\ell{=}m_j{+}\frac{1}{2}\, s\, m_s{=}{-}\frac{1}{2} \,\rangle \quad \text{for} \quad \ell > 0. \end{aligned} \tag{3.85}$$

With $\ell = 0$ we have $j = s$ and $m_j = m_s$. The former result is therefore trivially true. The latter result can be justified by noting that

- for given $\ell > 0$ and $s = \frac{1}{2}$ the only allowed values for j are $j = \ell \pm \frac{1}{2}$,
- to produce a given m_j we must require $m_\ell + m_s = m_j$,
- the expansion coefficients in (3.85) ensure that $|\, n\ell sj{=}\ell{\pm}\frac{1}{2}\, m_j \,\rangle$ are orthonormal,
- the m_ℓ values actually referred to on the right hand side of (3.85) satisfy $|\, m_\ell |{\leq}\, \ell$.

The coefficients of the expansion (3.85) are known as *Clebsch-Gordan* coefficients.

Quiz 3.33 : Show that $[H_0, \boldsymbol{J}^2]$, $[\boldsymbol{L}^2, \boldsymbol{J}^2]$ and $[\boldsymbol{S}^2, \boldsymbol{J}^2]$ all vanish and thus argue that H_0, $\boldsymbol{L}^2$, $\boldsymbol{S}^2$, $\boldsymbol{J}^2$ and J_z are simultaneously measurable.

Quiz 3.34 : Show that the $|\, n\ell sjm_j \,\rangle$ wave functions, as expressed by (3.84)-(3.85), are normalized if the wave functions $|\, n\ell m_\ell s m_s \,\rangle$ are orthonormal. Demonstrate that $|\, n\ell s\, j = \ell + \frac{1}{2}\, m_j \,\rangle$ and $|\, n\ell s\, j = \ell - \frac{1}{2}\, m_j \,\rangle$ are orthogonal.

Quiz 3.35 : Make use of (3.80), (3.60) and (3.79) to show that the wavefunction (eigenfunction) $|\, n\ell s\, j = \ell{+}\frac{1}{2}\, m_j \,\rangle$ as given by (3.85) corresponds to an eigenvalue $\hbar^2 j(j{+}1)$ for $\boldsymbol{J}^2$ with $j = \ell + \frac{1}{2}$.

3.8 Spectroscopic Notation

It has become common practice to denote different electron states in the atom in terms of the n, ℓ, s, j, m_j set of quantum numbers in the form

$$n\ell\ {}^{2S+1}L_J^{o,e}. \tag{3.86}$$

The first part of this form, $n\ell$, will be referred to as the *electron configuration*, ${}^{2S+1}L$ as the *spectral term* and ${}^{2S+1}L_J$ as the *spectral level*. In this notation the principal quantum number n and the spin and total angular momentum quantum numbers $S = s$ and $J = j$ are given by their numerical values. The orbital angular momentum quantum numbers ℓ and $L = \ell$ are indicated by letters. In table 3.7 the letters corresponding to different values of ℓ are given. For L the corresponding capital letters are used. The superscripts o or e indicate the parity of the wave function regarding the transformation $\boldsymbol{r} \to -\boldsymbol{r}$. If the wave function changes sign during this transformation, the parity is said to be odd. This is indicated by the superscript o. If the wave function remains unchanged during the mirroring transformation, the parity is even. This is indicated by the superscript e – or more commonly by suppressing the subscript. The latter tradition is unfortunate because sometimes also the superscript o is omitted.

ℓ	0	1	2	3	4	5
symbol	s	p	d	f	g	h

Table 3.7: Spectroscopic terminology

For the one-electron atom, the terminology (3.86) may seem unnecessary complicated. First of all, the orbital angular momentum values are indicated twice, once with ℓ and once with L. Additionally, from quiz 3.10 it follows that the parity of the wave function will be

even or odd according to whether the ℓ-value is an even or odd integer. Finally, for the electron $S = \frac{1}{2}$ and therefore $2S + 1 = 2$ always. We shall refer to this value 2 as indicating a doublet state in the following. The reason for this seemingly unnecessary duplication of notation will become clear when we discuss many-electron atoms in the following chapter. On the contrary we note that the value of the azimuthal quantum number m_j is not indicated by the terminology (3.86).

The *ground state* of the one-electron atom, corresponding to the lowest energy level, is characterized by quantum numbers $n = 1$, $\ell = 0$ and $s = j = \frac{1}{2}$ and will therefore be denoted

$$1\mathrm{s}\,^2\mathrm{S}_{1/2}.$$

The first *excited states* have $n = 2$, $\ell = 0$ or 1, $s = \frac{1}{2}$ and $j = \frac{1}{2}$ or $\frac{3}{2}$. The corresponding electron states are

$$2\mathrm{s}\,^2\mathrm{S}_{1/2}, \quad 2\mathrm{p}\,^2\mathrm{P}^o_{1/2}, \quad 2\mathrm{p}\,^2\mathrm{P}^o_{3/2}.$$

Quiz 3.36 : Express the $n = 3$ states in spectroscopic notation.

3.9 Transition Rates

Except for the brief introductory discussion of the effects of an external magnetic field in section 3.5, we have so far considered the atom as an isolated system without any interaction with the outside world. We have demonstrated that the atom may only exist in a discrete set of energy states. The next problem to be discussed is how the atom may jump between different energy states with the accompanied emission or absorption of electromagnetic radiation. To study these phenomena it is necessary to include an additional time-dependent term in the Hamiltonian of the system. We shall limit our discussion to purely radiative transitions. That is, we shall only consider emission and absorption processes where a single atom is involved in the interaction with the electromagnetic field. In dense gases additional effects may appear as a result of interactions (collisions) with neighboring atoms.

In the search for the extra term H' of the Hamilton operator describing the interaction between the one-electron atom and an external electromagnetic wave, we shall be guided by the corresponding classical system. Together with the positive nucleus at the origin, the electron at position $\boldsymbol{r}$ represents a time-varying electric dipole moment

$$\boldsymbol{m}_E = -e\boldsymbol{r}.$$

In the electric field $\boldsymbol{E}$ of an external plane wave, this electric dipole has a potential energy

$$H' = -\boldsymbol{m}_E \cdot \boldsymbol{E} = e\boldsymbol{r} \cdot \boldsymbol{E}(\boldsymbol{r}, t) \tag{3.87}$$

with

$$\boldsymbol{E}(\boldsymbol{r}, t) = \frac{E_0}{2} \left(\hat{\boldsymbol{\epsilon}} \exp(\iota \boldsymbol{k} \cdot \boldsymbol{r} - \iota\omega t) + \hat{\boldsymbol{\epsilon}}^* \exp(-\iota \boldsymbol{k} \cdot \boldsymbol{r} + \iota\omega t) \right). \tag{3.88}$$

Here E_0 is the (real) amplitude of the electric field while $\hat{\boldsymbol{\epsilon}}$ is the (possibly complex) polarization vector.

The extra term (3.87) makes the total Hamiltonian

$$H = H_0 + H'$$

time dependent[4]. It is therefore necessary to replace the stationary Schrödinger equation (3.10) with the corresponding time-dependent Schrödinger equation (3.11), here rewritten as

$$\imath\hbar\frac{\partial}{\partial t}\,|\,\Psi(t)\,\rangle = H\,|\,\Psi(t)\,\rangle \tag{3.89}$$

with the time dependence of the wave function $|\,\Psi\,\rangle$ explicitly indicated. We shall seek solutions of (3.89) in the form of the expansion

$$|\,\Psi(t)\,\rangle = \sum_q c_q(t)\,\exp(-\frac{\imath}{\hbar}\,W_q\,t)\;|\,q\,\rangle, \tag{3.90}$$

where $c_q(t)$ are time-dependent coefficients and where $|\,q\,\rangle$ and W_q are the eigenfunctions and eigenvalues of the zeroth order problem,

$$H_0\,|\,q\,\rangle = W_q\,|\,q\,\rangle.$$

We are allowed to seek solutions of (3.89) of the form (3.90) because the set of eigenfunctions $|\,q\,\rangle$ of the zeroth order problem constitute a complete set. The form of the solution we are seeking also allows for an immediate physical interpretation. If the atom in the absence of any external wave field is prepared in a particular state $|\,i\,\rangle$ at time $t = 0$,

$$c_q(t=0) = \begin{cases} 1 & \text{for} \quad q = i \\ 0 & \text{otherwise,} \end{cases}$$

then the atom will remain in this state,

$$c_q(t) = c_q(t=0) \quad \text{for all } q\text{'s.}$$

In the presence of a perturbing external electromagnetic wave field, the atom may decide to jump to another state. In this case $H' \neq 0$ and the coefficients $c_q(t)$ will evolve with time. In particular, the probability that the atom will be making a transition from state $|\,i\,\rangle$ to another state $|\,j\,\rangle$ by time t is given by $|\,c_j(t)\,|^2$. A schematic illustration is given in figure 3.7. The ticks on the q-axis represent different quantum states of the atom, while the length $|\,c_q(t)\,|^2$ of the vertical arrows indicate the probability of finding the atom in each of these states at time t. The transition probability per unit time or the *transition rate* from state i to state j is therefore

$$w_{ji} = \frac{\mathrm{d}}{\mathrm{d}t}\,|\,c_j\,|^2\,. \tag{3.91}$$

With this interpretation we are only left with the problem of solving for $c_j(t)$. We substitute the expansion (3.90) into (3.89) and form the scalar product of the resulting equation with $|\,j\,\rangle$. Taking into account the orthogonality of different eigenstates of the unperturbed atom, $\langle\, j \mid q\,\rangle = \delta_{jq}$, and defining

$$\omega_{jq} \equiv \frac{W_j - W_q}{\hbar},$$

[4]The more general discussion of the interaction between the atom and an external electromagnetic field will start by expressing the Hamiltonian in terms of the electromagnetic potentials Φ and $\boldsymbol{A}$,

$$H(\boldsymbol{r},\boldsymbol{p}) = \frac{1}{2m}(\boldsymbol{p} - q\boldsymbol{A})^2 + q\Phi.$$

The results of the two different approaches coincide except at very strong fields.

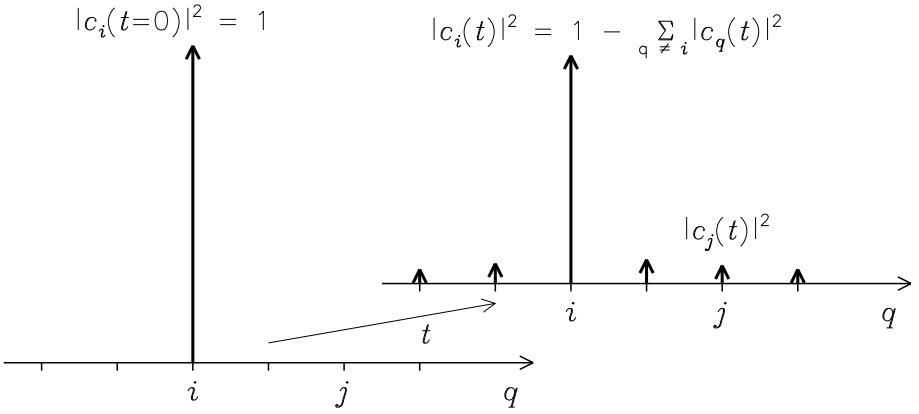

Figure 3.7: State probabilities at different times

the result of this operation is

$$\iota\hbar\,\dot{c}_j(t) = \sum_q c_q(t)\,\exp(\iota\omega_{jq}\,t)\,\langle\, j \mid e\boldsymbol{E}(\boldsymbol{r},t)\cdot\boldsymbol{r} \mid q\,\rangle. \tag{3.92}$$

We note that the right hand side expansion is an infinite sum over all states of the zeroth order system.

Two approximations lead to considerable simplifications. First, the wavelength λ represents a typical scale length for variations of the electric field $\boldsymbol{E}$ of the electromagnetic wave, the corresponding length for the wave function $|\,q\,\rangle$ is the Bohr radius a_B. With $\lambda \gg a_B$ the electric field $\boldsymbol{E}$ may be considered approximately constant over the atom, that is, we may replace $\boldsymbol{E}(\boldsymbol{r},t)$ in (3.92) with $\boldsymbol{E}(\boldsymbol{r}=0,t)$. This is a good approximation for visible as well as parts of the ultraviolet wavelengths. We shall refer to this approximation as the *electric dipole approximation*. Secondly, while the transition probability from the initial state i to any other state q remains small, $c_i \approx 1$ and $|\,c_q\,| \ll 1$ for $q \neq i$, only the term i in the sum on the right hand side of (3.92) needs to be retained. The time dependence of $\dot{c}_j(t)$ for $j \neq i$ is now of the simple form $\exp(\iota(\omega_{ji} \pm \omega)t)$ and may easily be integrated to give

$$\begin{aligned} c_j(t) = {} & \frac{eE_0}{2\hbar}\,\langle\, j \mid \hat{\boldsymbol{\epsilon}}\cdot\boldsymbol{r} \mid i\,\rangle\,\frac{1-\exp(\iota(\omega_{ji}-\omega)t)}{\omega_{ji}-\omega} \\ & + \frac{eE_0}{2\hbar}\,\langle\, j \mid \hat{\boldsymbol{\epsilon}}^*\cdot\boldsymbol{r} \mid i\,\rangle\,\frac{1-\exp(\iota(\omega_{ji}+\omega)t)}{\omega_{ji}+\omega}. \end{aligned} \tag{3.93}$$

The result depends on the frequency ω of the electromagnetic wave. Significant interactions between the electron and the external wave will take place only if one of the two numerators vanishes, that is, if

$$\omega \approx \pm\omega_{ji}. \tag{3.94}$$

In an absorption process, $W_j > W_i$, the first term of (3.93) contributes, in an emission process the second term is the important one. We recognize (3.94) as the frequency condition for photons involved in the transition of the atom from one state to another.

One final step is needed in order to calculated the transition rate w_{ji} according to (3.91). In (3.93) the electromagnetic wave was assumed to be strictly monochromatic. This is according

to our discussion in section 1.6 an idealization. To get the physically relevant transition rate we should integrate the idealized transition probability $|\, c_j(t)\,|^2$ over a frequency interval containing the frequency satisfying the frequency condition (3.94). For this purpose we replace the factor $|\,E_0\,|^2$ appearing in $|\,c_j(t)\,|^2$ according to the prescription

$$\frac{\epsilon_0}{2} E_0^2 c \to \mathcal{I}_\omega(\omega)\,\mathrm{d}\omega\,\mathrm{d}^2\Omega, \tag{3.95}$$

where $\mathcal{I}_\omega$ is the specific intensity of the radiation field. The left hand side expression represents the energy flux of a monochromatic plane wave (3.88) propagating in the given direction $\hat{\boldsymbol{k}}$ (see (1.31)). The right hand side expression represents the energy transported in the more realistic external radiation field per unit time and unit area within the angular frequency interval $\mathrm{d}\omega$ and within the solid angle interval $\mathrm{d}^2\Omega$ in the wave vector direction $\hat{\boldsymbol{k}}$. Notice that we in our notation for the spectral intensity $\mathcal{I}_\omega(\omega)$ suppressed the directional dependence of the intensity.

Assuming $\mathcal{I}_\omega(\omega)$ to be approximately constant over the actual frequency interval, noting that $|\,1-\exp(iyt)\,|^2 = 4\sin^2(yt/2)$ and making use of the integral identity

$$\int_{-\infty}^{\infty} \frac{\sin^2 x}{x^2}\,\mathrm{d}x = \pi,$$

the transition probability for a transition from state $|\, i\,\rangle$ to state $|\, j\,\rangle$ (satisfying $\omega_{ij} > 0$) during time t reduces to

$$|\,c_j(t)\,|^2 = \frac{4\pi^2\alpha}{\hbar}\,\mathcal{I}_\omega(|\,\omega_{ji}\,|)\,\mathrm{d}^2\Omega\;\; |\langle\, j\;|\;\hat{\boldsymbol{\epsilon}}\cdot\boldsymbol{r}\;|\;i\,\rangle|^2\;\; t,$$

where α is the fine-structure constant (3.42). The corresponding *transition rate per unit solid angle* induced by the external radiation field is therefore, according to (3.91), given by

$$w_{ji} = \frac{4\pi^2\alpha}{\hbar}\,\mathcal{I}_\omega(|\,\omega_{ji}\,|)\;\; |\langle\, j\;|\;\hat{\boldsymbol{\epsilon}}\cdot\boldsymbol{r}\;|\;i\,\rangle|^2\;. \tag{3.96}$$

The transition rate is seen to be proportional with the specific intensity of the external radiation field at the frequency satisfying the frequency condition (3.94). We notice that because of the substitution (3.95) the physical unit of the transition rate w changed to per unit solid angle. In addition, the result contains the factor $\langle\, j\;|\;\hat{\boldsymbol{\epsilon}}\cdot\boldsymbol{r}\;|\;i\,\rangle$, given as the product of the matrix element of the electric dipole moment $\boldsymbol{m}_E = -e\boldsymbol{r}$ of the atom between the initial and final states and the polarization $\hat{\boldsymbol{\epsilon}}$ of the radiation field. The result represents the interaction between the electromagnetic wave and the electric dipole moment of the atom and is therefore said to represent the electric dipole transition rates. For the corresponding result for states $|\, j\,\rangle$ satisfying $\omega_{ji} < 0$ we only have to replace $\hat{\boldsymbol{\epsilon}}$ with $\hat{\boldsymbol{\epsilon}}^*$.

To interpret the result (3.96) in greater detail, let us now consider two atomic states a and b with $W_a < W_b$. In a transition from the lower energy state a to the higher state b a photon with the proper energy $\hbar\omega_{ba}$ is absorbed by the atom. The transition rate per unit solid angle for this *absorption* process is given by

$$w_{ba}^{abs} = \frac{4\pi^2\alpha}{\hbar}\,\mathcal{I}_\omega(\omega_{ba})\;\; |\langle\, b\;|\;\hat{\boldsymbol{\epsilon}}\cdot\boldsymbol{r}\;|\;a\,\rangle|^2\;. \tag{3.97}$$

As expected, the absorption rate depends on the number of photons in the appropriate energy range passing by the atom per unit time as represented by the specific intensity $\mathcal{I}_\omega(\omega_{ba})$.

It may be more surprising that the same statement also applies for the corresponding expression for the reverse process in which the atom jumps from the higher energy state b to the lower state a while getting rid of its excess energy through the emission of an extra photon with identical energy and direction of propagation. The transition rate per unit solid angle is

$$w_{ab}^{st.e} = \frac{4\pi^2\alpha}{\hbar}\,\mathcal{I}_\omega(\omega_{ba})\;|\langle\, b \mid \hat{\boldsymbol{\epsilon}}\cdot\boldsymbol{r} \mid a\,\rangle|^2\,. \tag{3.98}$$

We did here make use of the fact that $\langle\, a \mid \hat{\boldsymbol{\epsilon}}^*\cdot\boldsymbol{r} \mid b\rangle^* = \langle\, b \mid \hat{\boldsymbol{\epsilon}}\cdot\boldsymbol{r} \mid a\,\rangle$. Because this transition rate depends on the number of incident photons similar to the emitted photon, the latter process will be referred to as *stimulated emission*. The absorption and the stimulated emission processes are illustrated schematically in figure 3.8.

The transition rates for absorption and stimulated emission, (3.97) and (3.98), are seen to be equal. Still, stimulated emission is usually less intense than absorption. This is a result of the fact that the number of atoms in the higher energy state $\mathcal{N}_b$ is usually much less than the number of atoms in the lower energy state $\mathcal{N}_a$.

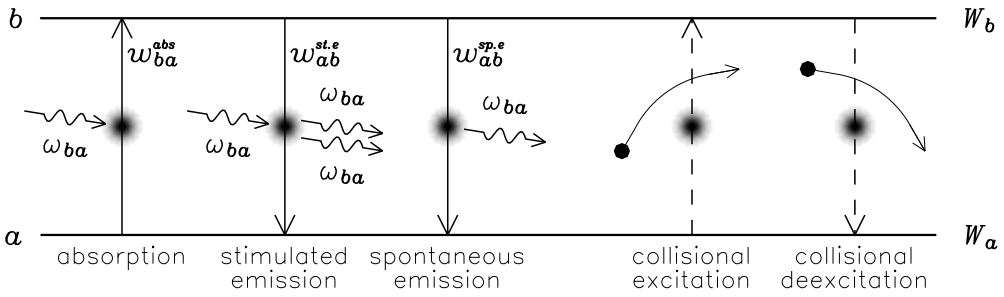

Figure 3.8: Absorption, stimulated emission and spontaneous emission processes between two atomic states a and b with energies W_a and W_b

In addition to the absorption and stimulated emission processes, the atom may also take part in a *spontaneous emission* process between the two atomic states and for which the transition rate $w_{ab}^{sp.e}$ is independent of the presence of the external radiation field. The expression for the spontaneous emission rate can be derived by invoking the *principle of detailed balance* for a gas in thermodynamic equilibrium with a radiation field. The principle states that to be able to maintain the thermodynamic equilibrium distributions of *particles and fields* in the presence of transition processes, it is necessary that the total number of transitions in either direction between any two atomic states are equal.

Consider the gas and radiation field to be in thermodynamic equilibrium at temperature $\mathcal{T} = \kappa T$. Here κ is the Boltzmann constant and $\mathcal{T}$ and T are the temperature in energy units or degrees Kelvin, respectively. Let $\mathcal{N}_a$ and $\mathcal{N}_b$ denote the number of atoms occupying the two states a and b. The number of transitions from state a to b per unit time and per unit solid angle is given by $w_{ba}^{abs}\mathcal{N}_a$. The number of reverse transitions is $(w_{ab}^{st.e} + w_{ab}^{sp.e})\mathcal{N}_b$. According to the principle of detailed balance these transition rates must be equal, that is,

$$w_{ab}^{sp.e} = w_{ba}^{abs}\frac{\mathcal{N}_a}{\mathcal{N}_b} - w_{ab}^{st.e}. \tag{3.99}$$

In thermal equilibrium the ratio of the number of atoms in the two atomic states a and b is determined by the energy difference between these states in accordance with the Boltzmann relation (see section 6.7)

$$\frac{\mathcal{N}_b}{\mathcal{N}_a} = \exp(-\frac{\hbar\omega_{ba}}{\mathcal{T}}). \tag{3.100}$$

The specific intensity in each polarization of the radiation field in thermal equilibrium is given by the Planck radiation function (see section 6.15)

$$\mathcal{I}_\omega(\omega) = \mathcal{B}_\omega(\mathcal{T}) = \frac{\hbar}{(2\pi)^3 c^2} \frac{\omega^3}{\exp(\frac{\hbar\omega}{\mathcal{T}}) - 1}. \tag{3.101}$$

Substitution of (3.100)-(3.101) into (3.99) now leads to the transition rate for spontaneous emission rate per unit solid angle and polarization $\hat{\boldsymbol{\epsilon}}$ as

$$w_{ab}^{sp.e} = \frac{\alpha\omega_{ba}^3}{2\pi c^2} \mid \langle\, b \mid \hat{\boldsymbol{\epsilon}} \cdot \boldsymbol{r} \mid a \,\rangle \mid^2 . \tag{3.102}$$

Also the spontaneous emission process is indicated schematically in figure 3.8.

The expressions (3.97), (3.98) and (3.102) all refer to transition rates per unit solid angle and for a specified polarization. The total spontaneous emission rate for arbitrary polarization and directions of propagation for the emitted photon, $\mathcal{W}_{ab}^{sp.e}$, is found by summing over orthogonal polarizations for a given direction of propagation and then integrating over all directions. The result is

$$\mathcal{W}_{ab}^{\mathrm{sp.e}} = \frac{4\alpha\,\omega_{ba}^3}{3c^2} \mid \boldsymbol{r}_{ab} \mid^2, \tag{3.103}$$

where $\boldsymbol{r}_{ab} \equiv \langle\, a \mid \boldsymbol{r} \mid b \,\rangle$. Similarly, for unpolarised and isotropic radiation the corresponding total absorption and stimulated emission rates $\mathcal{W}_{ab}^{abs}$ and $\mathcal{W}_{ab}^{st.e}$ are found by again summing over orthogonal polarizations and integrating over all directions of propagation for the photons involved. The results are

$$\mathcal{W}_{ba}^{\mathrm{abs}} = \mathcal{W}_{ab}^{\mathrm{st.e}} = 4\pi \frac{4\pi^2\alpha}{3\hbar} \mathcal{I}_\omega(\omega_{ba}) \mid \boldsymbol{r}_{ab} \mid^2, \tag{3.104}$$

where $\mathcal{I}_\omega(\omega_{ba})$ now refers to the total specific intensity regardless of polarization.

Oscillator strengths

Transition rates are often expressed in terms of oscillator strengths. The *oscillator strength* f_{ki} for a transition from state $|\, i \,\rangle$ to state $|\, k \,\rangle$ is defined by

$$f_{ki} \equiv \frac{2m}{3\hbar}\, \omega_{ki}\, |\, \boldsymbol{r}_{ki} \,|^2, \tag{3.105}$$

where m is the electron mass, $\omega_{ki} \equiv (W_k - W_i)/\hbar$ and $\boldsymbol{r}_{ki} \equiv \langle\, k \mid \boldsymbol{r} \mid i \,\rangle$. The oscillator strength is a dimensionless quantity. We note that $f_{ki} > 0$ for absorption and $f_{ki} < 0$ for emission processes. The reason for introducing this quantity is that it satisfies a convenient normalisation condition, the *Thomas-Reiche-Kuhn sum rule*

$$\sum_k f_{ki} = 1. \tag{3.106}$$

The summation is here taken over all quantum states $|\,k\,\rangle$ available for the atom, including the continuum. The sum rule is useful in that it sets natural limits to the values to be expected for the different transition rates.

To prove this result we note that

$$|\,\boldsymbol{r}_{ki}\,|^2 = \langle\, i \mid \boldsymbol{r} \mid k \,\rangle \cdot \langle\, k \mid \boldsymbol{r} \mid i \,\rangle,$$

$$(W_k - W_i)\langle\, i \mid \boldsymbol{r} \mid k \,\rangle = -\langle\, i \mid (H\boldsymbol{r} - \boldsymbol{r}H) \mid k \,\rangle = \frac{\iota\hbar}{m}\langle\, i \mid \boldsymbol{p} \mid k \,\rangle,$$

and similarly

$$(W_k - W_i)\langle\, k \mid \boldsymbol{r} \mid i \,\rangle = -\frac{\iota\hbar}{m}\langle\, k \mid \boldsymbol{p} \mid i \,\rangle.$$

We did here make use of (3.13). Adding the last two expressions we may now write

$$f_{ki} = \frac{\iota}{3\hbar}\{\langle\, i \mid \boldsymbol{p} \mid k \,\rangle \cdot \langle\, k \mid \boldsymbol{r} \mid i \,\rangle - \langle\, i \mid \boldsymbol{r} \mid k \,\rangle \cdot \langle\, k \mid \boldsymbol{p} \mid i \,\rangle.$$

The collection of all orthonormal wave functions for the hydrogen atom, including both bound states and the continuum, form a *complete set*. This means that

$$\sum_k \mid k \,\rangle\langle\, k \mid = I, \tag{3.107}$$

where I is the identity operator. We thus find

$$\sum_k f_{ki} = \frac{\iota}{3\hbar}\{\langle\, i \mid \boldsymbol{p} \cdot \boldsymbol{r} - \boldsymbol{r} \cdot \boldsymbol{p} \mid i \,\rangle = I$$

in accordance with (3.106) since $\boldsymbol{p} \cdot \boldsymbol{r} - \boldsymbol{r} \cdot \boldsymbol{p} = -\iota\, 3\hbar$.

Initial config	Final config	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$\sum_{n=5}^{\infty}$	Continuum
1s	np	—	0.416	0.079	0.029	0.041	0.435
2s	np	—	—	0.435	0.103	0.111	0.351
2p	ns	-0.139	—	0.014	0.003	0.003	0.008
2p	nd	—	—	0.696	0.122	0.109	0.183

Table 3.8: Averaged oscillator strengths for selected configurations of one-electron atoms

In table 3.8 the oscillator strengths for the two lowest energy levels for one-electron atoms, averaged over azimuthal quantum numbers are given,

$$\overline{f}_{n'\ell',n\ell} \equiv \frac{1}{2\ell+1}\sum_{m'_\ell,m_\ell} f_{n'\ell'm'_\ell,n\ell m_\ell}. \tag{3.108}$$

The averaged oscillator strengths trivially also satisfy the sum rule.

The transitions rates (3.97), (3.98) and (3.102) all refer to transitions between discrete, negative energy eigenstates of the atom. We refer to such transitions as *bound-bound transitions*. These lead to the formation of line spectra. An atom may also be involved in a

photo-ionization transition in which the energy of the absorbed photon is sufficient to tear an electron from the atom. A *radiative recombination transition* is the opposite process. Here a free electron is captured by an ion while the excess energy is radiated away. Characteristic of these *bound-free transitions* is that the corresponding spectra are continuous. We also speak of *free-free transitions* in which the electron involved in the transition is in a free state both before and after the event. Examples are the Bremsstrahlung emission, the cyclotron radiation or the Thompson scattering by free electrons accelerated in the electric field of heavy ions, in an external magnetic field or in the electric field of an electromagnetic wave, respectively. We discussed these processes in chapter 2.

Quiz 3.37 : Why is it for the quantum mechanical treatment of the extra term H' not sufficient to substitute the complex expression (1.26) for the electric field of the plane wave?

Quiz 3.38 : With the wave function $|\,\Psi(t)\,\rangle$ defined by (3.90), show that the normalization condition $\langle\,\Psi(t)\mid\Psi(t)\,\rangle = 1$ leads to

$$\sum_q |\,c_q(t)\,|^2 = 1.$$

Argue that $|\,c_q(t)\,|^2$ can be interpreted as a probability, the probability that the atom at time t occupies state q.

Quiz 3.39 : Express the total transition rates for absorption, and stimulated and spontaneous emission (3.103)- (3.104) in terms of oscillator strengths.

Quiz 3.40 : Make use of the fact that any wave with polarization $\hat{\boldsymbol{\epsilon}}$ can be written as a sum of two plane-polarized waves with mutually orthogonal polarizations $\hat{\boldsymbol{\epsilon}}_1$ and $\hat{\boldsymbol{\epsilon}}_2$ to derive the expressions (3.103)-(3.104) for the total transition rates (regardless of the directions of propagation of the photons involved) between state $|\,b\,\rangle$ and state $|\,a\,\rangle$ for arbitrary (both) polarization(s).

[*Hint:* Note that

$$\mathcal{W}_{ab}^{\text{sp.e}} = \int (w_{ab}^{\text{sp.e}}(\hat{\boldsymbol{\epsilon}}_1) + w_{ab}^{\text{sp.e}}(\hat{\boldsymbol{\epsilon}}_2))\,\mathrm{d}^2\Omega,$$

further note that $\langle\,a\mid\hat{\boldsymbol{\epsilon}}\cdot\boldsymbol{r}\mid b\,\rangle = \hat{\boldsymbol{\epsilon}}\cdot\boldsymbol{r}_{ab}$, let θ denote the angle between $\boldsymbol{k}$ and $\boldsymbol{r}_{ab}$, choose $\hat{\boldsymbol{\epsilon}}_1$ in the $(\boldsymbol{k},\boldsymbol{r}_{ab})$-plane and $\hat{\boldsymbol{\epsilon}}_2$ perpendicular to this plane, and then perform the $\mathrm{d}^2\Omega$ integration.]

Quiz 3.41 : Einstein (1916) studied the relation between absorption, stimulated and spontaneous emission in thermal equilibrium through the rate equations

$$\dot{N}_{ba} = B_{ba}N_a\rho(\omega_{ba})$$
$$\dot{N}_{ab} = A_{ab}N_b + B_{ab}N_b\rho(\omega_{ba}),$$

requiring detailed balance, $\dot{N}_{ba} = \dot{N}_{ab}$. Here N_a, N_b are the number of atoms in states a and b, $\dot{N}_{ab}$, $\dot{N}_{ba}$ are the number of atoms making transitions from b to a and from a to b per unit time, and $\rho(\omega_{ba}) \equiv u_\omega(\omega_{ba})$ is the energy density in the electromagnetic radiation field at the relevant frequency, regardless of polarization. The

coefficients A_{ab}, B_{ab} and B_{ba} are known as the Einstein coefficients for spontaneous and stimulated emission and absorption, respectively.

Show with reference to (3.99)-(3.101) that

$$B_{ba} = B_{ab} \quad \left[= \frac{\mathcal{W}_{ba}^{\text{abs}}}{\rho(\omega_{ba})} \right] \tag{3.109}$$

$$A_{ab} = \frac{\hbar \omega_{ba}^3}{\pi^2 c^3} B_{ab} \quad \left[= \mathcal{W}_{ab}^{\text{sp.e}} \right]. \tag{3.110}$$

3.10 Selection Rules and Atomic Lifetimes

To estimate the strength of the spectral line feature corresponding to atomic transitions between states i and f, it is necessary to know the value of the transition rate w_{fi}. A transition from state i to another state f will be called an *allowed* transition if $w_{fi} > 0$. A *forbidden* transition is one for which $w_{fi} = 0$. Fortunately, simple rules can be established to decide for which transitions the transition rate takes non-vanishing values. The rules we are looking for are known as the *selection rules* for electric dipole transitions.

We shall again consider the one-electron atom with n, ℓ, m_ℓ, s and m_s as quantum numbers. Let the change in these quantum numbers associated with the transition between the two states be denoted by Δn, $\Delta \ell$, Δm_ℓ, Δs and Δm_s. The set n, ℓ, s, j and m_j would have been a more proper choice of quantum numbers at this point. With the help of the relations (3.84) - (3.85) it will, however, be easy to transform results from one set of quantum numbers to the other.

For the spin part of the wave function it is trivially seen that $\Delta s = 0$. The electron will be a spin $s = \frac{1}{2}$ particle before and after the transition. Furthermore, since $\hat{\boldsymbol{\epsilon}} \cdot \boldsymbol{r}$ is independent of electron spin, the orthogonality of the spin part of the wave functions immediately leads to the requirement $\Delta m_s = 0$.

The unperturbed wave functions have even or odd parity under the coordinate transformation $\boldsymbol{r} \to -\boldsymbol{r}$. If $|\, i \,\rangle$ and $|\, f \,\rangle$ have identical parities then the integrand of $\langle\, f \mid \boldsymbol{r} \mid i \,\rangle$ is an odd function of $\boldsymbol{r}$ and the integral will vanish. Necessary for a non-vanishing transition rate is therefore that the initial and final electron states have different parities.

For plane polarized waves we may choose, $\hat{\boldsymbol{\epsilon}} = \hat{\boldsymbol{z}}$ and therefore $\hat{\boldsymbol{\epsilon}} \cdot \boldsymbol{r} = r \cos\theta$. The angular part of the integral (3.97) reduces to

$$\int_{-1}^{1} \mathrm{d}\cos\theta \int_0^{2\pi} \mathrm{d}\varphi \, \exp(-\iota \Delta m_\ell \varphi) \, \cos\theta \, P_{\ell+\Delta\ell}^{|m_\ell + \Delta m_\ell|}(\cos\theta) \, P_\ell^{|m_\ell|}(\cos\theta).$$

The properties of the complex exponential and those of the associated Legendre polynomials, listed in table 3.3, will show that the angular integrals will take non-vanishing values only if

$$\Delta m_\ell = 0 \quad \text{and} \quad \Delta \ell = \pm 1. \tag{3.111}$$

For circularly polarized waves we choose instead, $\hat{\boldsymbol{\epsilon}} = (\hat{\boldsymbol{x}} \pm \iota \hat{\boldsymbol{y}})/\sqrt{2}$ and therefore $\hat{\boldsymbol{\epsilon}} \cdot \boldsymbol{r} = r \sin\theta \exp(\pm \iota\varphi)$. The corresponding integrals are

$$\int_{-1}^{1} \mathrm{d}\cos\theta \int_0^{2\pi} \mathrm{d}\varphi \, \exp(-\iota(\Delta m_\ell \mp 1)\varphi) \, \sin\theta \, P_{\ell+\Delta\ell}^{|m_\ell + \Delta m_\ell|}(\cos\theta) \, P_\ell^{|m_\ell|}(\cos\theta).$$

These integrals will take non-vanishing values only if

$$\Delta m_\ell = \pm 1 \quad \text{and} \quad \Delta \ell = \pm 1. \tag{3.112}$$

We notice (see quiz 3.10) that the requirement $\Delta\ell = \pm 1$ represents a strengthening of the requirement that the initial and final electron states must have different parities.

Combining the above results with the relations connecting $\mid nj\ell sm_j \rangle$ and $\mid n\ell m_\ell sm_s \rangle$ eigenstates of the atom, (3.84) and (3.85), the general set of *selection rules* for electric dipole radiation transitions of the one-electron atom may now be formulated.

The one-electron atom will absorb or emit radiation in electron transitions satisfying the following conditions:

- $\Delta s = 0$
- *initial and final states have different parities*
- $\Delta\ell = \pm 1$
- $\Delta j = 0, \pm 1$
- $\Delta m_j = 0, \pm 1$.

In figure 3.9 a common way of illustrating possible transitions have been given. Here the energy levels for the H I atom have been ordered according to *spectral term*, ^{2S+1}L. The energy levels have been given in units of Kayser [cm^{-1}] and electron volts [eV] with the energy level of the ground state as origin of the ordinate axis. Examples of allowed transitions between different states of the atom, satisfying $\Delta\ell = \pm 1$, have been indicated, some with the corresponding wavelength in units of ångstrom [Å]. A diagram of this type is called a *Grotrian diagram* and is often met with in the literature. The usefulness of such diagrams will be seen in chapter 4 where we will be discussing properties and spectra of many-electron atoms.

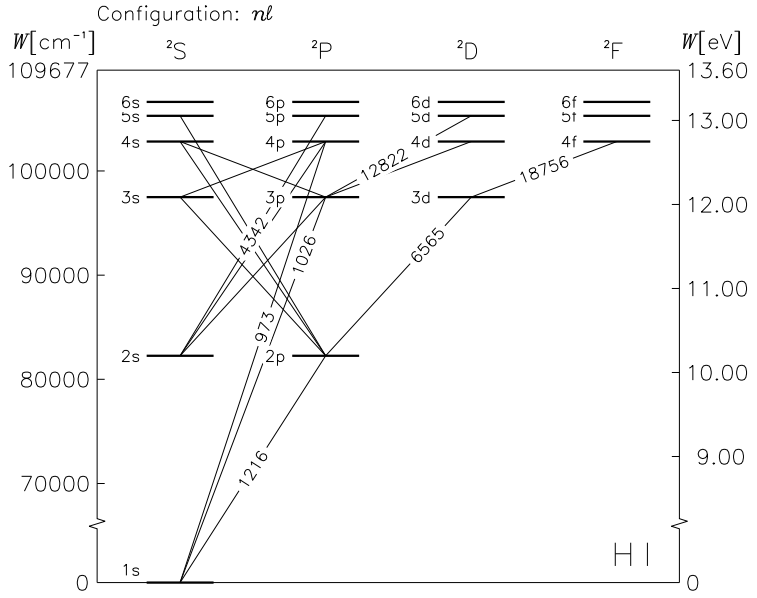

Figure 3.9: Grotrian diagram showing energy levels for different spectral terms and examples of allowed transitions

A transition for which any one of the above requirements is not fulfilled represents a forbidden transition. This does not mean that the corresponding spectral line may not be observed in nature. A forbidden transition only means that according to the electric dipole approximation the transition may not occur. A forbidden transition may still give rise to a *forbidden line* in the spectrum when higher order effects are taken into account. For instance, in the *electric quadrupole approximation*, the $\boldsymbol{r}$-dependence of the wave electric field $\boldsymbol{E}(\boldsymbol{r},t)$ in (3.92) is taken into account to first order in $\boldsymbol{r}$ through the expansion

$$\exp(\imath \boldsymbol{k}\cdot\boldsymbol{r}) \approx 1 + \imath \boldsymbol{k}\cdot\boldsymbol{r}.$$

This gives rise to the appearance of matrix elements of the type $\langle\, f \mid \boldsymbol{k}\cdot\boldsymbol{r}\boldsymbol{r} \mid i\,\rangle$. For these elements to take non-vanishing values, the initial and final electron states must have identical parities. Similar conclusions apply for magnetic contributions to the magnetic dipole transition rate. In addition, effects due to inter-atomic collisions may give non-vanishing transition rates. It is, however, normally found that transition rates for forbidden transitions will be smaller than those for allowed transitions.

Left to itself, shielded from external radiation fields, an excited atom will tend to decay to some lower energy state of the atom. The rate at which this decay takes place is determined by the sum of the spontaneous emission rates from the given excited state i to all lower energy states f,

$$\mathcal{W}_i = \sum_f \mathcal{W}_{fi}^{sp.e}. \tag{3.113}$$

If $N_i(t)$ atoms are prepared in state i at time t, then during the time interval $\mathrm{d}t$ a number $N_i(t)\mathcal{W}_i\mathrm{d}t$ of these will decay to lower energy states. This leads to the differential equation for the remaining number of atoms in state i,

$$\frac{\mathrm{d}}{\mathrm{d}t}N_i(t) = -N_i(t)\mathcal{W}_i,$$

with the solution $N_i(t) = N_i(t=0)\exp(-\mathcal{W}_i t)$. During every time interval of length

$$\Delta t = \tau_i \equiv 1/\mathcal{W}_i, \tag{3.114}$$

the number of the original atoms remaining in state i is reduced by a factor e^{-1}. The quantity τ_i represents the *natural lifetime* of the state i.

For atomic states for which allowed transitions to lower energy states exist, natural lifetimes are typically found in the range 10^{-8} to 10^{-6} s. If allowed transitions to lower states do not exit, the natural lifetimes may be orders of magnitude longer. Such states are referred to as *meta-stable states*. In the presence of external radiation fields or inter-atomic collisions typical lifetimes of the meta-stable states will be shortened.

Quiz 3.42 : List all allowed transitions between the $n = 2$ and $n = 3$ states.

Quiz 3.43 : Determine the meta-stable states of one-electron atoms.

Quiz 3.44 : Make use of the orthogonality property of the angular part of the wave-functions together with the recurrence relations for the associated Legendre functions, as given in table 3.3, to verify (3.111) and (3.112).

Quiz 3.45 : Argue for the selection rules $\Delta j = 0, \pm 1$ and $\Delta m_j = 0, \pm 1$ given the selection rules for $\Delta\ell$, Δm_ℓ and Δm_s. [Hint: Make use of (3.85) and notice that $\langle\, n+\Delta n\, \ell+\Delta\ell\, s\, j+\Delta j\, m_j+\Delta m_j \mid \hat{\boldsymbol{\epsilon}}\cdot\boldsymbol{r} \mid n\, \ell\, s\, j\, m_j \,\rangle$ appears as a difference of two terms. Convince yourself that these two terms cancel, for instance, for $\Delta\ell = 1$, $\Delta m_j = 0$ and $\Delta j = 2$ by making explicit use of the normalization factors for Y_ℓ^m and P_ℓ^m given in tables 3.3 and 3.4.]

Quiz 3.46 : How does transition rates $\mathcal{W}_{ab}^{\text{sp.e.}}$ and natural lifetimes τ_i scale with nuclear charge number Z for given atomic states a, b, i? [*Hint:* How does $|\,\boldsymbol{r}_{ab}\,|^2$ scale with Z?]

3.11 Spectral Line Formation

With the selection rules at hand, let us now study the formation of spectral lines for one-electron atoms. To lowest order the energy levels only depend on the principal quantum number n. The ground state is $1\mathrm{s}\,^2\mathrm{S}_{1/2}$. The lowest excited states are $2\mathrm{s}\,^2\mathrm{S}_{1/2}$ and $2\mathrm{p}\,^2\mathrm{P}^o_{1/2,3/2}$, all corresponding to the identical energy to lowest order. Transitions between the ground state and the two p-states are both allowed. The corresponding spectral line formed is called the *resonance line*. For hydrogen the resonance line is found at $\lambda 1216$ and is referred to as the L_α-line of the Lyman series. A transition between the ground state and the $2\mathrm{s}\,^2\mathrm{S}_{1/2}$ state is forbidden since these states both have even parity. The latter state is therefore a meta-stable state, for hydrogen with a natural lifetime of 0.12 s.

For transitions between the $n = 1$ and $n = 3$ energy levels, only the $3\mathrm{p}\,^2\mathrm{P}^o_{1/2,3/2}$ states may participate. These transitions give rise to the L_β-line in the Lyman series at $\lambda 1026$. The $3\mathrm{s}\,^2\mathrm{S}_{1/2}$ and $3\mathrm{d}\,^2\mathrm{D}_{3/2,5/2}$ states are not meta-stable states as these may participate in transitions with the $2\mathrm{p}\,^2\mathrm{P}^o_{1/2,3/2}$ states, leading to the H_α-line of the Balmer series at $\lambda 6563$. Transitions between the $2\mathrm{p}\,^2\mathrm{P}^o_{1/2}$ and the $3\mathrm{d}\,^2\mathrm{D}_{5/2}$ states are not allowed as these would correspond to $\Delta j = \pm 2$. In a similar manner the selection rules may be used to determine allowed transitions between any set of energy levels. In fact, it will be seen that there will exist allowed transitions between any given set of energy levels, as indicated in figure 3.1 for the hydrogen atom. This statement is true as long as the energy levels only depend on the principal quantum number n.

The polarization of the emitted or absorbed radiation is determined by the value of Δm_j for the transition. If $\Delta m_j = 0$, the polarization is linear. For $\Delta m_j = \pm 1$ the radiation is circularly polarized. This conclusion follows directly from arguments leading to the establishment of the selection rules in section 3.10.

The strength of a given spectral line depends on the value of the corresponding transition rate, together with the number of atoms available for participating in the transition. For an absorption process this means the number of atoms in the lower energy state, for emission processes it is the number of atoms in the higher energy state that is the relevant quantity. The distribution of atoms over different electron states will be discussed in chapter 6. Suffice it here to say that the fraction of atoms in the higher energy states increases with temperature.

If an electromagnetic radiation field with a continuous frequency or wavelength spectrum on its way toward an observer traverses a gaseous medium, a certain fraction of the photons with the appropriate energy will be lost due to absorption processes. If the absorbed photons are not replaced by photons created in the corresponding emission processes, the observer will see a deficit in the spectrum at frequencies corresponding to radiation transitions in the gas.

An *absorption line* in the original continuous spectrum will result. The first observations of spectral features were of this type. With the help of a glass prism, Fraunhofer was able to demonstrate that the Solar radiation consisted of a continuous spectrum with a number of dark lines (Fraunhofer lines). The lines form as the continuous spectrum radiation emerging from the photosphere on its way passes the cooler chromosphere. If, on the other hand, the number of photons emitted by the gas exceeds the number of absorbed photons an *emission line* superposed on the continuous spectrum will be formed. This situation arises for instance in scattered light observation from the hot, but tenuous coronal gas outside the Solar limb.

Observed absorption or emission line features have finite widths. Different processes contribute to this line broadening. At this point we shall discuss three of the basic mechanisms, leading to distictively different spectral line profiles. Subsequent sections will discuss other mechanisms.

3.11.1 Natural line profile

Due to the decay rate of the excited state $|\,i\,\rangle$ as depicted in figure 3.10a, the $c_i(t)$ coefficient in (3.92) will not remain constant as assumed for the derivation of (3.93). Instead, for the next order approximation it must be assumed that the state $|\,i\,\rangle$ component of the total wavefunction (3.90) will vary with time as

$$\chi(t) = \exp\left(-\iota\frac{W_i}{\hbar}t - \frac{t}{2\tau_i}\right) \quad \text{for} \quad t > 0,$$

where τ_i is the natural lifetime (3.114) of the state $|\,i\,\rangle$. The corresponding Fourier transform (1.40) is

$$\tilde{\chi}(\frac{W}{\hbar}) = \frac{1}{2\pi}\int_0^\infty \exp\left(\iota\frac{W-W_i}{\hbar}t - \frac{t}{2\tau_i}\right)\,\mathrm{d}t = \frac{\iota}{2\pi}\left(\frac{W-W_i}{\hbar} + \frac{\iota}{2\tau_i}\right)^{-1}. \tag{3.115}$$

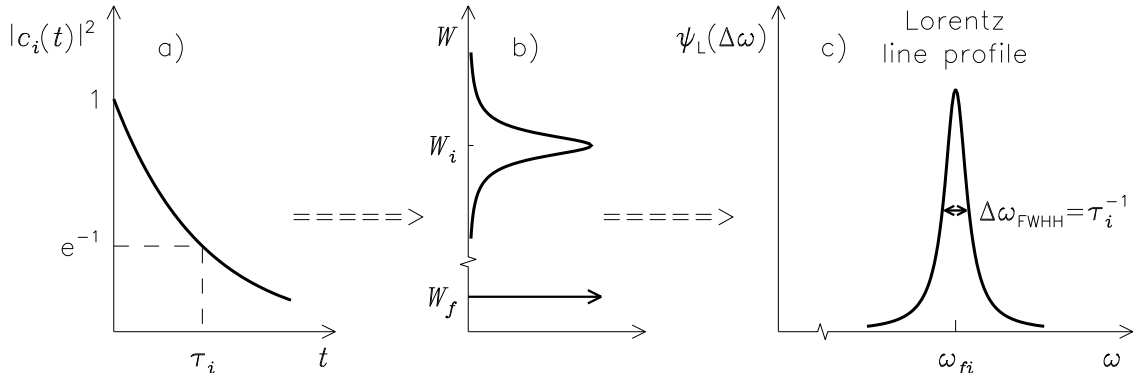

Figure 3.10: Formation of the Lorentz spectral line profile

The energy probability distribution of the excited state, given by the absolute square of the Fourier amplitude (3.115), and illustrated in figure 3.10b, is in accordance with the Heisenberg uncertainty principle: the energy uncertainty increasing inversely proportional

with the natural lifetime of the excited state. Each of these energies will contribute to the resulting spectral line and therefore leading to a spectral line profile

$$\psi_L(\Delta\omega) = \frac{1}{2\pi\tau_i}\left((\Delta\omega)^2 + \frac{1}{4\tau_i^2}\right)^{-1} \tag{3.116}$$

illustrated in figure 3.10c. We here defined $\Delta\omega = \omega - \omega_{fi}$. This profile is refered to as the *Lorentz line profile*. Notice that spectral line profiles are normalized, $\int \psi(\Delta\omega)\,\mathrm{d}(\Delta\omega) = 1$. The full width at half height of the Lorentz line profile is seen to be given by

$$\Delta\omega_{\text{FWHH}} = \frac{1}{\tau_i}. \tag{3.117}$$

In the above discussion we assumed the natural lifetime of the final state $|\, f \,\rangle$ to be infinite and the corresponding energy probability distribution therefore δ-function like. With allowance for a finite natural lifetime τ_f also for the final state, the effective lifetime τ_{fi} to be used in (3.116) is defined by

$$\frac{1}{\tau_{fi}} = \frac{1}{\tau_i} + \frac{1}{\tau_f}. \tag{3.118}$$

3.11.2 Collisional or pressure broadening

The natural line profile results from the natural finite lifetime for atoms in a given state. Collisions between atoms in a gas resulting in radiationless transitions of the colliding atoms between different atomic states lead to a similar effect. The combined effect due to natural lifetime and collisions is described by replacing the inverse natural lifetime τ_i^{-1} for state $|\, i \,\rangle$ by an effective inverse lifetime $\nu_i + \tau_i^{-1}$ in (3.116). Here ν_i represents the collision frequency for collisions leading to radiationless transitions out of state $|\, i \,\rangle$. The collision frequency, that is the average number of collisions per unit time suffered by our atom, can be expressed as $\nu_i = \sigma_i n \langle\, v \,\rangle$ where the effective collision cross section σ_i only depends on atomic properties, where n is the number density of collision partners, and where $\langle\, v \,\rangle = \sqrt{8\mathcal{T}/\pi m}$ represents a typical relative velocity between our atom and its collision partners (see section 6.5).

Collisions between atoms in a gas will thus lead to a broadening of spectral lines, but the line profile will remain Lorentzian. The collision frequency ν_i increases with increasing pressure in the gas. The broadening of spectral lines due to collisions is therefore referred to as *collisional* or *pressure broadening*. Collisional broadening will normally dominate the natural line width. In fact the natural line width will only be observed for gases in the limit of vanishing pressure. In stellar interiors high gas pressure and therefore strong collisional broadening effects contributes to the production of the continuous radiation spectra from stars.

3.11.3 Thermal Doppler broadening

Relative motion between the atoms in the emitting or absorbing medium is another mechanism contributing to the line profile. The Doppler effect arises when there is a relative motion between the emitting or absorbing atom and the observer. If the atom in its rest frame is emitting photons with angular frequency ω_{fi}, then an observer, toward which the atom is moving with a line-of-sight velocity v_ℓ, will observe a shifted frequency

$$\omega = \omega_0\left(1 + \frac{v_\ell}{c}\right). \tag{3.119}$$

If all emitting atoms within the field of view are moving away from the observer, the spectral line will be shifted toward smaller frequencies (red-shifted). If the emitting atoms are moving toward the observer, the line will be shifted toward higher frequencies (blue-shifted). But if there is a spread in line-of-sight velocities of the emitting atoms within the field of view, a broadening of the line will result. Identical conclusions apply to absorption lines. The effect is referred to as *Doppler broadening* and the corresponding line profile as the *Doppler line profile*.

As a quantitative example consider a gas with a Maxwellian distribution of line-of-sight velocities (see figure 3.11)

$$f_{\boldsymbol{r}v_\ell}(v_\ell) \sim \exp\left(-\frac{Mv_\ell^2}{2\mathcal{T}}\right).$$

Here M is the atomic mass and $\mathcal{T}$ kinetic temperature in energy units ($\mathcal{T} = \kappa T$). The fraction $f_{\boldsymbol{r}v_\ell}(v_\ell)\mathrm{d}v_\ell$ of atoms with line-of-sight velocities in the range $(v_\ell, v_\ell + \mathrm{d}v_\ell)$ will contribute to the broadened spectral line at the relative frequency shift $\Delta\omega/\omega_0 = (\omega - \omega_0)/\omega_0 = v_\ell/c$. The spectral line profile will therefore in this case take a Gaussian shape

$$\psi_D(\Delta\omega) = \frac{1}{\omega_0}\sqrt{\frac{Mc^2}{2\pi\mathcal{T}}}\exp\left(-\frac{Mc^2}{2\mathcal{T}}\left(\frac{\Delta\omega}{\omega_0}\right)^2\right). \tag{3.120}$$

The full width at half height $\Delta\omega_{\mathrm{FWHH}}$ ($\Delta\lambda_{\mathrm{FWHH}}$) is given by

$$\frac{\Delta\lambda_{\mathrm{FWHH}}}{\lambda_0} \approx \frac{\Delta\omega_{\mathrm{FWHH}}}{\omega_0} = \sqrt{\frac{8\ln 2\ \mathcal{T}}{Mc^2}}. \tag{3.121}$$

In figure 3.11 we assumed that the radiating gas had no net line-of-sight velocity relative to the observer. With a net line-of-sight velocity superposed on the thermal motions of the individual radiating atoms the line profile will suffer a corresponding shift in frequency (wavelength).

Doppler broadening will result from thermal motions in a gas, from inhomogeneous macroscopic motions such as produced from turbulence or waves in the gas, or from the rotational motion of astronomical objects.

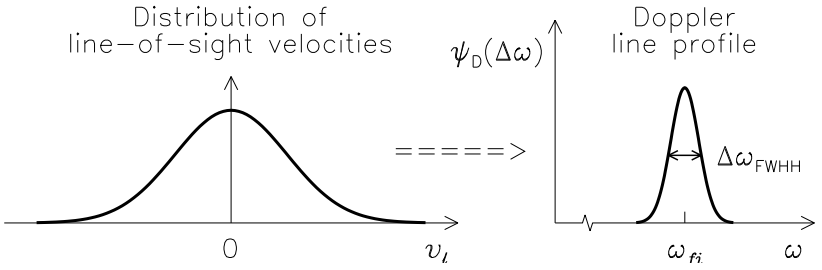

Figure 3.11: Distribution of line-of-sight velocities leading to spectral line broadening

3.11.4 The Voigt line profile

In the previous subsections the line profiles due to natural lifetime and collisions and due to the Doppler effect were considered independently of each other. These two effects have

distictively different origins and give rise to distinctively different line profiles, the Doppler effect giving rise to a broad central core and narrow wings, the Lorentz profile due to natural lifetime and collisions having the opposite characteristics. In reality these two effects will work together in forming the spectral line profile. The total effect is found by forming the convolution product of the two profiles,

$$\psi_V(\Delta\omega) = \psi_L(\Delta\omega) \otimes \psi_D(\Delta\omega) \equiv \int_{-\infty}^{\infty} \psi_L(\Delta\omega')\, \psi_D(\Delta\omega - \Delta\omega')\, \mathrm{d}\Delta\omega' \tag{3.122}$$

The line profile at a certain frequency shift $\Delta\omega$ is the sum of products of contributions from the two profiles with a sum of individual frequency shifts equal to the given $\Delta\omega$ value. The combined profile is known as the *Voigt line profile*. In figure 3.12 the result is plotted for the case where the Lorentz and the Doppler profiles have identical full widths at half height. For a narrower Lorentz profile the Voigt profile approaches the Doppler profile, for a wider Lorentz profile, the Voigt profile approaches the Lorentz profile. Generally, the core of the total line profile is dominated by the Doppler profile, while the wings are determined by the Lorentz profile.

Line profiles are regularly used to infer density and temperature of the emitting or absorbing medium, temperature mainly from the form of the central core of the line, density from the wings. Characteristic Doppler broadening effects also result from stellar rotations and from waves or turbulent motions. But there are also other mechanisms that will contribute to the final line shape. Some of the more important of these include fine structure splitting, Zeeman and Stark effects. We shall discuss these effects in the following sections.

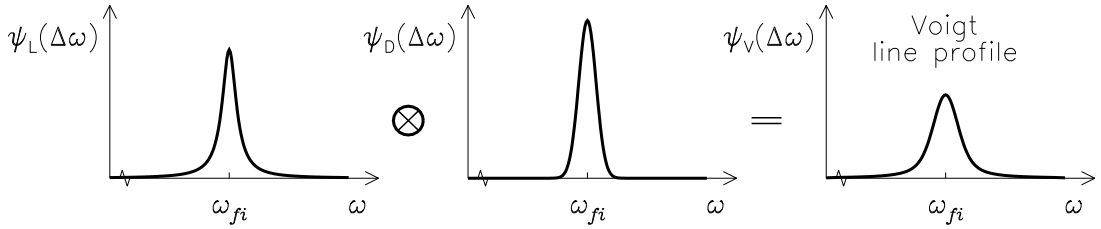

Figure 3.12: Voigt spectral line profile

Quiz 3.47 : A distant star with radius R equal to the Solar radius $R_\odot$ has a rotational axis perpendicular to the line of sight. The rotational period is $P = 24$ hours. The surface temperature of the star is $T = 6000$ K. Calculate the total width of the H_α-line resulting from the stellar rotation and compare with thermal Doppler broadening. How is your result modified if the angle θ between the rotational axis and the line of sight is changed?

Quiz 3.48 : A distant star with radius R performs solid body rotation with angular velocity Ω perpendicular to the line of sight. Show that the relative rotational Doppler displacement over the cross section of the star is given as $\Delta\omega/\omega_0 = \Omega x/c$ where x is a coordinate perpendicular to the line of sight and the rotational axis. Next show

that the spectral line profile due to stellar rotation can be expressed as

$$\psi_R(\Delta\omega) = \frac{1}{\omega_0}\frac{c}{\Omega}\frac{2}{\pi R^2}\sqrt{R^2 - \left(\frac{c}{\Omega}\frac{\Delta\omega}{\omega_0}\right)^2}. \tag{3.123}$$

What is the corresponding $\Delta\omega_{\text{FWHH}}/\omega_0$? [*Hint:* Assume that $\psi_R(\Delta\omega)\,\mathrm{d}(\Delta\omega) \sim \mathrm{d}A_{\Delta\omega}$ where $\mathrm{d}A_{\Delta\omega}$ is the area of the star cross section that produces a rotational Doppler displacement within $(\Delta\omega, \Delta\omega + \mathrm{d}(\Delta\omega))$.]

Quiz 3.49 : How would you evaluate the combined spectral line profile $\psi(\Delta\omega)$ due to thermal, collisional and rotational broadening?

Quiz 3.50 : The spectral line profile ψ can be given either as a function of $\Delta\omega$ or alternatively as a function of $\Delta\lambda$, that is, $\psi(\Delta\omega)$ or $\psi(\Delta\lambda)$. What is the relation between these functions?

3.12 Fine Structure Splitting

At this point we have concluded the lowest order discussion of one-electron atomic spectra. It will now be necessary to consider some smaller but still important additional effects leading to modification and splitting of the zeroth order energy levels. These include fine structure splitting due to relativistic and spin-orbit interaction effects in the atom, Zeeman and Stark effects due to external magnetic and electric fields, and finally also hyperfine effects due to the existence of nuclear spin. These effects may each give rise to splitting, shifting or broadening of spectral lines. We start with fine structure splitting.

For a stringent discussion of this effect, the relativistic Dirac electron equation is needed. As this falls outside our scope, we shall have to be satisfied with a semi-empirical discussion. For this purpose let us again consider the one-electron atom as a classical system. The electron is moving in its orbit with velocity $\boldsymbol{v}$ in the electrostatic field from the nucleus,

$$\boldsymbol{E} = \frac{Ze}{4\pi\epsilon_0}\frac{\boldsymbol{r}}{r^3}.$$

An observer moving with velocity $\boldsymbol{v}$ in an electric field $\boldsymbol{E}$ will "see" a magnetic field

$$\boldsymbol{B}' = -\frac{1}{c^2}\boldsymbol{v}\times\boldsymbol{E} = -\frac{1}{mc^2}\boldsymbol{p}\times\frac{Ze}{4\pi\epsilon_0}\frac{\boldsymbol{r}}{r^3} = \frac{Ze}{4\pi\epsilon_0 mc^2}\frac{\boldsymbol{L}}{r^3}. \tag{3.124}$$

In this magnetic field $\boldsymbol{B}'$, the spin magnetic moment $\boldsymbol{m}_S = -e\boldsymbol{S}/m$ of the electron due to the electron spin $\boldsymbol{S}$ will give rise to an extra potential energy term in the Hamiltonian

$$H_{B'S} = -\boldsymbol{m}_S\cdot\boldsymbol{B}' = \frac{Ze^2}{4\pi\epsilon_0 m^2c^2}\frac{\boldsymbol{L}\cdot\boldsymbol{S}}{r^3}.$$

A rigorous relativistic treatment of the motion of the spinning electron in its bounded orbit will introduce one additional contribution of the identical form, only half the size and of the opposite sign. The effect is known as *Thomas precession*. Combining these two effects we have the *spin-orbit term* of the Hamiltonian

$$H_{LS} = \frac{Ze^2}{8\pi\epsilon_0 m^2c^2}\frac{\boldsymbol{L}\cdot\boldsymbol{S}}{r^3}. \tag{3.125}$$

We must also include a first order relativistic kinetic energy correction to the Hamiltonian. Expanding the relativistic kinetic energy expression we find

$$\sqrt{m^2c^4+\boldsymbol{p}^2c^2}-mc^2 = mc^2\left(1+\frac{\boldsymbol{p}^2}{2m^2c^2}-\frac{\boldsymbol{p}^4}{8m^4c^4}+\cdots-1\right) = \frac{\boldsymbol{p}^2}{2m}-\frac{\boldsymbol{p}^4}{8m^3c^2}+\cdots .$$

The first term is the non-relativistic kinetic energy. The second term is the *relativistic mass term* of the Hamiltonian. We write it in the form

$$H_{RM} = -\frac{\boldsymbol{p}^4}{8m^3c^2} = -\frac{1}{2mc^2}\left(H_0+\frac{Ze^2}{4\pi\epsilon_0 r}\right)^2. \tag{3.126}$$

The relativistic treatment will show that there is still an extra term of the Hamiltonian that needs to be included. The term is referred to as the *Darwin term* and reads

$$H_D = \frac{\pi\hbar^2}{2m^2c^2}\frac{Ze^2}{4\pi\epsilon_0}\delta(\boldsymbol{r}). \tag{3.127}$$

The sum of the relativistic mass, the spin-orbit and the Darwin terms is known as the *fine structure term* of the Hamiltonian

$$H_{FS} = H_{RM}+H_{LS}+H_D. \tag{3.128}$$

The eigenfunctions of the Hamiltonian $H = H_0+H_{FS}$ differ from those of the unperturbed Hamiltonian H_0. The effects of the fine structure term (3.128) may, however, still be evaluated from the knowledge of the eigenfunctions of the unperturbed Hamiltonian by using first order stationary perturbation theory. Stationary perturbation theory is reviewed in table 3.9.

As discussed in section 3.6, the energy levels W_n of the zero order Hamiltonian H_0 are $2n^2$-fold degenerate. We also note that H_{LS} and therefore also H_{FS} do not commute with $\boldsymbol{L}$ or $\boldsymbol{S}$. This means that the $\{|\,n\ell m_\ell s m_s\,\rangle\}$-set of eigenfunctions is not the most optimum basis set for the perturbation analysis. However, making use of

$$\boldsymbol{L}\cdot\boldsymbol{S} = \frac{1}{2}(\boldsymbol{J}^2-\boldsymbol{L}^2-\boldsymbol{S}^2),$$

we note that H_{FS} induces no cross-coupling between eigenfunctions $\{|\,n\ell sjm_j\,\rangle\}$ belonging to the degenerate eigenvalue $W_n^{(0)}$, that is,

$$\langle\, n\ell' sj'm_j'\,|\, H_{FS}\,|\, n\ell sjm_j\,\rangle \neq 0 \quad \text{only if } \ell'=\ell,\ j'=j \text{ and } m_j'=m_j.$$

This means that the simplified form (3.136) of the degenerate perturbation theory in table 3.9 can be applied. The energy shift introduced by H_{FS} is therefore given by

$$\begin{aligned}\Delta W_{FS} &= \langle\, n\ell sjm_j\,|\, H_{FS}\,|\, n\ell sjm_j\,\rangle \\ &= -\frac{1}{2mc^2}\left(W_n^2+2W_n\left(\frac{Ze^2}{4\pi\epsilon_0}\right)\langle\,\frac{1}{r}\,\rangle_{n\ell}+\left(\frac{Ze^2}{4\pi\epsilon_0}\right)^2\langle\,\frac{1}{r^2}\,\rangle_{n\ell}\right) \\ &\quad+\frac{Ze^2\hbar^2}{16\pi\epsilon_0 m^2c^2}\left(j(j+1)-\ell(\ell+1)-s(s+1)\right)\langle\,\frac{1}{r^3}\,\rangle_{n\ell}\,(1-\delta_{\ell 0}) \\ &\quad+\frac{\pi\hbar^2}{2m^2c^2}\frac{Ze^2}{4\pi\epsilon_0}\,|\,\Psi_{n00}(0)\,|^2\,\delta_{\ell 0}.\end{aligned}$$

Let a complete set of orthonormal eigenvectors $\{|\, q^{(0)}\,\rangle\}$ with corresponding eigenvalues $\{W_q^{(0)}\}$ of the Hamiltonian H_0 be known. We are seeking approximations for the eigenvectors $|\, q\,\rangle$ and eigenvalues W_q of the Hamiltonian $H = H_0 + H'$,

$$H\,|\, q\,\rangle = W_q\,|\, q\,\rangle, \tag{3.129}$$

assuming H' to be "small". The approximate solution is:

For *non-degenerate eigenvalues* $W_q^{(0)}$, $W_q = W_q^{(0)} + \Delta W_q$ and $|\, q\,\rangle = |\, q^{(0)}\,\rangle + |\,\Delta q\,\rangle$ with

$$\Delta W_q = \langle\, q^{(0)}\,|\, H'\,|\, q^{(0)}\,\rangle + \sum_{p\neq q} \frac{|\,\langle\, p^{(0)}\,|\, H'\,|\, q^{(0)}\,\rangle\,|^2}{W_q^{(0)} - W_p^{(0)}} + \mathcal{O}((H')^3) \tag{3.130}$$

$$|\,\Delta q\,\rangle = \sum_{p\neq q} \frac{\langle\, p^{(0)}\,|\, H'\,|\, q^{(0)}\,\rangle}{W_q^{(0)} - W_p^{(0)}}\;|\, p^{(0)}\,\rangle + \mathcal{O}((H')^2). \tag{3.131}$$

If $W_q^{(0)}$ is a *g-fold degenerate eigenvalue* with corresponding orthonormal eigenvectors $|\, q_j^{(0)}\,\rangle$, $j = 1, \cdots, g$, choose normalized sets of coefficients $\{c_j\}$, $\sum_j |\, c_j\,|^2 = 1$, with

$$|\, q\,\rangle = \sum_{j=1}^{g} c_j\,|\, q_j^{(0)}\,\rangle + |\,\Delta q\,\rangle \quad \text{and} \quad W_q = W_q^{(0)} + \Delta W_q, \tag{3.132}$$

such that $|\,\Delta q\,\rangle$ remains "small" for H' "small". Solutions for $\{c_j\}$ must satisfy

$$\sum_{j=1}^{g} \Big(\langle\, q_i^{(0)}\,|\; H'\;|\, q_j^{(0)}\,\rangle - \Delta W_q\;\delta_{ij}\Big)\; c_j = 0, \quad i = 1, \cdots, g. \tag{3.133}$$

The different solutions for ΔW_q are found from the consistency relation

$$\det |\langle\, q_i^{(0)}\,|\; H'\;|\, q_j^{(0)}\,\rangle - \Delta W_q\;\delta_{ij}| = 0 \tag{3.134}$$

with corresponding

$$|\,\Delta q\,\rangle = \sum_{j=1}^{g} c_j \sum_{p\neq q} \frac{\langle\, p^{(0)}\,|\; H'\;|\, q_j^{(0)}\,\rangle}{W_q^{(0)} - W_p^{(0)}}\;|\, p^{(0)}\,\rangle + \mathcal{O}((H')^2), \tag{3.135}$$

where $|\, p^{(0)}\,\rangle$ are eigenvectors of H_0 *not* belonging to the degenerate eigenvalue $W_q^{(0)}$.
In particular, if H' introduces no cross-coupling between eigenvectors $\{|\, q_j^{(0)}\,\rangle\}$, that is $\langle\, q_i^{(0)}\,|\; H'\;|\, q_j^{(0)}\,\rangle = 0$ for $i \neq j$, then simplified solutions are valid,

$$\Delta W_{qi} = \langle\, q_i^{(0)}\,|\; H'\;|\, q_i^{(0)}\,\rangle + \mathcal{O}((H')^2), \quad i = 1, \cdots, g. \tag{3.136}$$

Table 3.9: Stationary perturbation theory

Average values of $1/r$, $1/r^2$ and $1/r^3$ over the zeroth order wave functions are listed in table 3.6. The central value of the zero-order wave function $\Psi_{n00}(\boldsymbol{r})$ can be found from (3.48) (see also quiz 3.13),

$$|\,\Psi_{n00}(0)\,|^2 = \frac{Z^3}{\pi a_B^3 n^3}. \tag{3.137}$$

Substituting these values we find

$$\Delta W_{FS} = \frac{1}{2}mc^2\frac{(Z\alpha)^4}{n^3}\left(\frac{3}{4n} - \frac{1}{\ell+\frac{1}{2}} + \frac{j(j+1)-\ell(\ell+1)-s(s+1)}{2\ell(\ell+\frac{1}{2})(\ell+1)}(1-\delta_{\ell 0}) + \delta_{\ell 0}\right). \tag{3.138}$$

The first two terms in the parenthesis of (3.138) represent the effects of the relativistic mass correction (3.126). The $\delta_{\ell 0}$ is the Darwin term and the remaining $1-\delta_{\ell 0}$ term is spin-orbit coupling. Still another simplification can be made in the result. Making use of the fact that $s = \frac{1}{2}$ and $j = \ell \pm \frac{1}{2}$, the total fine structure energy correction term (3.138) can be written

$$\Delta W_{FS} = \frac{1}{2}mc^2\,\frac{(Z\alpha)^4}{n^3}\left(\frac{3}{4n} - \frac{1}{j+\frac{1}{2}}\right). \tag{3.139}$$

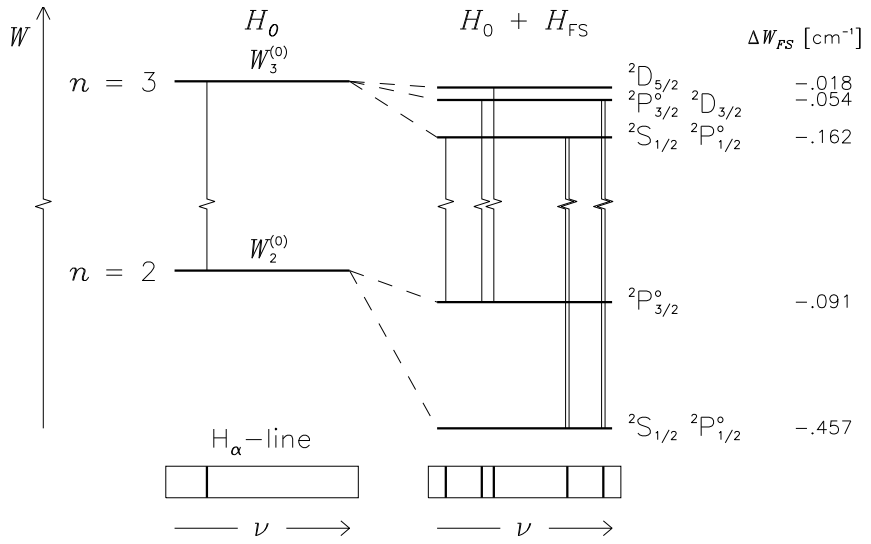

Figure 3.13: Fine structure splitting of the $n = 2$ and $n = 3$ H I energy levels.

Fine structure splitting of the energy levels of one-electron atoms will lead to the corresponding splitting of the spectral lines. As an example consider $n = 2$ and $n = 3$ energy levels of H I. Transitions between these levels give rise to the H_α-line of the Balmer series at $\lambda 6563$. The fine structure splitting of these energy levels compared with the corresponding zeroth order result is illustrated in figure 3.13. The energy levels are now functions of n and j. Identical values of j can be achieved through two different combinations of ℓ and $s = \frac{1}{2}$. Some

energy degeneracy is therefore still left. The splitting of the energy levels is seen to diminish with increasing n. In the figure the spectroscopic notation for the different electron states corresponding to the different energy levels are also given. The energy differences among the $n = 2$ and the $n = 3$ levels are drawn to the same scale, while the distances between $n = 2$ and $n = 3$ levels are off-scale.

The selection rules derived in section 3.10 are still valid. These rules limit the allowed radiation transitions between the two given n-level states to seven. Allowed transitions are illustrated by vertical lines in figure 3.13. Since two pairs of these transitions are between identical energy levels, only five spectral lines will result. We thus see that the H_α-line in fact consists of five individual lines. These are indicated in the lower part of figure 3.13. The maximum separation of these five individual lines is .2 Å for the H I atom. According to (3.121) this corresponds to the FWHH Doppler broadening at a temperature $T \approx 1800$ K.

The fine structure result (3.139) tells that the energy levels for given n and j are identical irrespective of the value of ℓ. Very accurate experiments as well as calculations have shown that this is not quite correct. The corresponding effect is referred to as *Lamb shift* and leads to the splitting of the $^2S_{1/2}$ and $^2P_{1/2}$ levels for $n = 2$ for hydrogen by about 1/10 of the fine structure splitting of the same level. The discussion of this effect falls outside our scope.

Quiz 3.51 : Estimate typical magnitudes of the magnetic field $\boldsymbol{B}'$.

Quiz 3.52 : Show that $\boldsymbol{L} \cdot \boldsymbol{S}$ does not commute with neither $\boldsymbol{L}$ nor $\boldsymbol{S}$. Why does this fact make the $\{|\, n\ell m_\ell s m_s \,\rangle\}$-basis set unfavourable for the perturbation analysis?

Quiz 3.53 : Show that expressions (3.138) and (3.139) are indentical with $j = \ell \pm \frac{1}{2}$.

Quiz 3.54 : Calculate the energies of the fine structure splitted $n = 2$ and $n = 3$ levels of the hydrogen atom in units of eV and in repetence units (Kayser). Calculate the (maximum) fine structure splitting of the H_α-line in ångstrom.

Quiz 3.55 : Maximum fine structure splitting of the hydrogen H_α-line is $\Delta\lambda_{FS} \approx .2$ Å. The equivalent temperature $T^{\text{equiv}}(\text{H I})$, determined by requiring the thermal width to satisfy $\Delta\lambda_{\text{FWHH}} = \Delta\lambda_{FS}$, is $T^{\text{equiv}}(\text{H I}) \approx 1800$ K. We want to find the corresponding equivalent temperature for the H_α-line from the $(Z-1)$-th ionized state of the element ${}^N_Z\text{X}$. Show that $\Delta\lambda_{FS}/\lambda \sim Z^2$ and therefore $T^{\text{equiv}}({}^N_Z\text{X}) = NZ^4\, T^{\text{equiv}}(\text{H I})$. What is $T^{\text{equiv}}({}^4\text{He})$?

Quiz 3.56 : To verify (3.131) and (3.130) substitute $H = H_0 + H'$, $W_q = W_q^{(0)} + \Delta W$ and $|\, q \,\rangle = |\, q^{(0)} \,\rangle + |\, \Delta q \,\rangle$ in (3.129) and assume H', ΔW and $|\, \Delta q \,\rangle$ all to be "small". Show that for $|\, q \,\rangle$ to be normalized it will be required that $\langle\, q^{(0)} \,|\, \Delta q \,\rangle = 0 + \mathcal{O}((H')^2)$. Left-multiply (3.129) with $\langle\, q^{(0)} \,|$ and verify that $\Delta W = \langle\, q^{(0)} \,|\, H' \,|\, q^{(0)} \,\rangle + \mathcal{O}((H')^2)$. Next verify (3.131) by expanding $|\, \Delta q \,\rangle$ in terms of the unperturbed eigenfunctions. Finally verify (3.130).

Quiz 3.57 : With $H' = H_{FS}$, $|\, q^{(0)} \,\rangle = |\, n\ell s j m_j \,\rangle$ and $\Delta W = \Delta W_{FS}$ show that the corresponding $|\, \Delta q \,\rangle = |\, \Delta q_{FS} \,\rangle$ can be written in the form

$$|\, \Delta q_{FS} \,\rangle = \sum_{n' \neq n} \beta_{n'}\, |\, n' \ell s j m_j \,\rangle.$$

What is the expression for $\beta_{n'}$?

3.13 The Zeeman Effect

We now return to the effect of an external magnetic field $\boldsymbol{B}$ on the energy levels of the atom. In astrophysical context this is an important effect which will allow us to infer the strength and structure of stellar and interstellar magnetic fields from the observation of the shape and polarization of spectral lines.

An introductory study of this effect was made in section 3.5, that time disregarding the effects of electron spin. It is now necessary to investigate the total effect of the external magnetic field, taking into account both the orbital and the spin magnetic momenta of the electron as described by the perturbing term

$$H_B = H_{BL} + H_{BS} \approx \frac{eB}{2m}(L_z + 2S_z) = \frac{eB}{2m}(J_z + S_z). \tag{3.140}$$

The study of the effects of an external magnetic field is complicated by the fact that we may not "turn off" the omnipresent fine structure splitting effect. We shall therefore be satisfied by restricting our discussion mainly to two limiting cases, the weak field limit in which the external magnetic field effect is dominated by the fine structure effect, and the strong field limit where the opposite relation is valid. Both limits find applications in astrophysics. We start with the weak field limit.

3.13.1 The weak field limit

In section 3.12 we learned that when taking fine structure effect into account, the $|\, n\,\ell\, s\, j\, m_j \,\rangle$ set of eigenfunctions is the proper choice. When we therefore now want to include the additional and assumed smaller effect of a weak external magnetic field, we have to evaluate the contribution from H_B to an energy level for given quantum numbers n and j with this set of wave functions as zeroth order states. The evaluation of the contribution from the J_z term is trivial due to the quantization rules (3.81). To evaluate the effect of the S_z term we need to check whether S_z introduces any cross coupling among wave functions $|n\ell s j m_j\,\rangle$ with the specified values of n and j. Making use of the expansion of the wave functions $|\, n\,\ell\, s\, j\, m_j \,\rangle$ in terms of $|\, n\,\ell\, m_\ell\, s\, m_s \,\rangle$ as given by (3.84) and (3.85) it is easily seen that this is not the case and therefore that the simplified degenerate perturbation theory (3.136) is sufficient. The result is that the additional contribution to the energy levels from the external magnetic field is given as

$$\begin{aligned}
\Delta W_B &= \frac{eB}{2m} \langle\, n\,\ell\, s\, j\, m_j \,|\, J_z + S_z \,|\, n\,\ell\, s\, j\, m_j \,\rangle \\
&= \left(1 \pm \frac{1}{2\ell+1}\right) \hbar \frac{eB}{2m}\, m_j \quad \text{for} \quad j = \ell \pm \frac{1}{2} \\
&= g\, \hbar\omega_L\, m_j
\end{aligned} \tag{3.141}$$

where

$$g = 1 + \frac{j(j+1) - \ell(\ell+1) + s(s+1)}{2j(j+1)} \tag{3.142}$$

is the *Landé g-factor* and $\omega_L = eB/2m$ the angular Larmor frequency of the electron. The effect of the external magnetic field is seen to split any energy level into $2j+1$ equidistant levels with an energy separation increasing linearly with the strength of the external magnetic

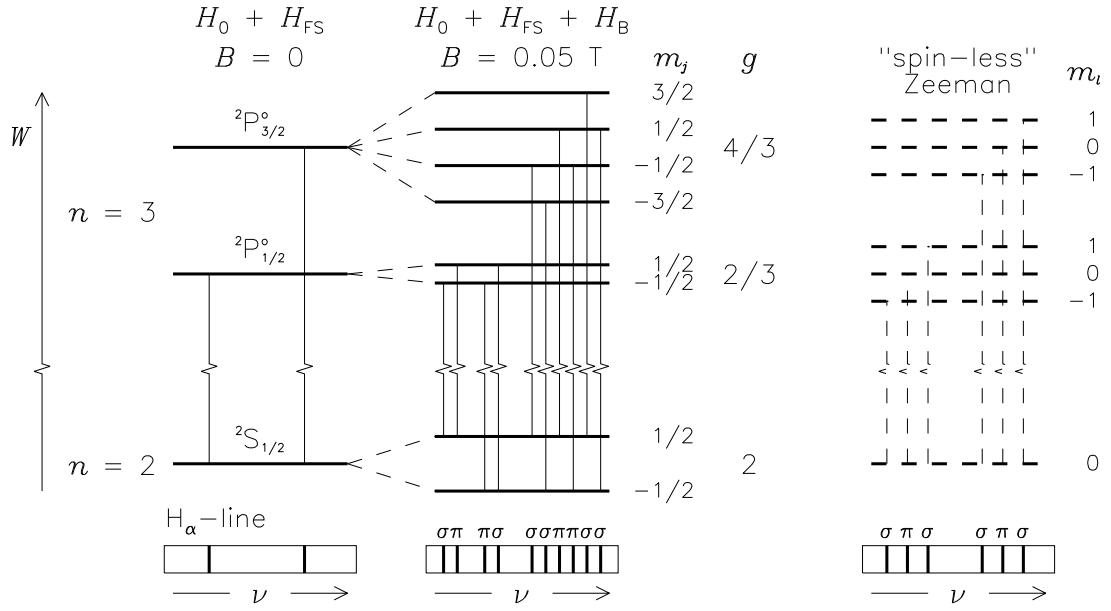

Figure 3.14: Anomalous Zeeman splitting of selected $n = 2$ and $n = 3$ H I energy levels. The corresponding normal Zeeman splitting for "spin-less" electrons are indicated to the right (dashed lines).

field. The energy separation, however, also depends on the values of both j and ℓ through the presence of the Landé g-factor.

In figure 3.14 the splitting of some selected $n = 2$ and $n = 3$ energy levels due to an external magnetic field $B = .05$ T have been indicated. Transitions between the chosen energy levels (for $\boldsymbol{B} = 0$) are producing the two highest frequency components of the fine structure splitted H_α-line (see figure 3.13). In the presence of an external magnetic field the $^2S_{1/2}$ energy level splits into two levels with a Landé factor $g = 2$. Also the $^2P^o_{1/2}$ energy level splits into two levels, but this time with $g = 2/3$. The $^2P^o_{3/2}$ energy level splits into four individual levels with the Landé factor $g = 4/3$. In the figure, the splittings among the $n = 2$ and among the $n = 3$ levels are drawn to scale. The distances between the $n = 2$ and $n = 3$ levels are off scale.

The splitting behavior described differs significantly both regarding the number of and the energy difference between splitted levels from that predicted on the basis of "spin-less" electrons as given in (3.70). The corresponding "spin-less" result is indicated to the right in figure 3.14.

In figure 3.14 we have also indicated allowed radiative transitions between the given energy levels and the relative position of the corresponding spectral lines. The spectral line pattern for "spin-less" electrons is for historical reasons referred to as *normal Zeeman effect*. In the present context this case would seem to have only "academic" interest. In the next chapter, however, where we will be discussing the physics of many-electron atoms, this case will find physical applications. The realistic case, for which the electron spin is taken into account, is referred to as *anomalous Zeeman effect*. In the figure, also the polarization (plane (π) or circular (σ)) of the radiation in the different spectral lines are indicated.

The result (3.141) is a valid approximation as long as the effect of the external magnetic

field is small compared to the fine structure effect. That is, the result (3.141) is valid as long as the external magnetic field is small compared to the apparent magnetic field (3.124) "seen" by the electron in its orbit in the electric field from the nucleus. For hydrogen atoms this means magnetic fields of order of magnitude less than 1 T.

3.13.2 The strong field limit

In the opposite limit, in which the external magnetic field is much larger than the apparent magnetic field seen by the electron in its orbit, it is better to make use of the $|\, n\,\ell\, m_\ell\, s\, m_s \,\rangle$ set of eigenfunctions as zero order wave functions and consider the fine structure splitting effect as a small perturbation. The contribution to the energy levels from the external magnetic field is this time

$$\Delta W_B = \hbar\omega_L(m_\ell + 2m_s). \tag{3.143}$$

The relativistic kinetic energy correctional term ΔW_{RM} and the Darwin term ΔW_D remain identical to the corresponding terms in (3.138), depending only on the quantum numbers n and ℓ. It is easily checked that $\boldsymbol{L}\cdot\boldsymbol{S}$ does not introduce any cross coupling between $|\, n\,\ell\, m_\ell\, s\, m_s \,\rangle$ eigenfunctions with given n and $m_\ell + 2m_s$. The spin-orbit contribution may thus be evaluated as

$$\Delta W_{LS} = \langle\, n\ell m_\ell s m_s \,|\, H_{LS} \,|\, n\ell m_\ell s m_s \,\rangle = \begin{cases} 0 & \ell = 0 \\ \dfrac{Ze^2\hbar^2}{8\pi\epsilon_0 m^2c^2}\, m_\ell\, m_s \,\langle\, \dfrac{1}{r^3} \,\rangle_{n\ell} & \ell \neq 0. \end{cases}$$

We did here make use of the fact (see quiz 3.31) that

$$\boldsymbol{L}\cdot\boldsymbol{S} = L_zS_z + \frac{1}{2}(L_+S_- + L_-S_+)$$

where $L_\pm = L_x \pm \imath L_y$ and $S_\pm = S_x \pm \imath S_y$. According to the results of quiz 3.18 and 3.30, the contributions from the two last terms vanish. After substituting the average value of $1/r^3$ from table 3.6, the spin-orbit contribution for $\ell \neq 0$, together with the relativistic mass correction form (3.138), thus reduces to

$$\Delta W_{FS} = \frac{1}{2}mc^2\frac{(Z\alpha)^4}{n^3}\left[\frac{3}{4n} - \frac{\ell(\ell+1) - m_\ell\, m_s}{\ell(\ell+\frac{1}{2})(\ell+1)}\right]. \tag{3.144}$$

The fine structure modification of the strong field Zeeman effect (3.143) is traditionally referred to as the *Paschen-Back effect*. The result is illustrated in figure 3.15 for the $2s$ and $3p$ states for a magnetic field strength $B = 1$ T. For comparison the zero order energy levels and the fine structure splitting of these levels without an external magnetic field are shown to the left. The rightmost panel shows the strong field Zeeman splitting neglecting fine structure effects while the remaining panel (second from the right) also includes the fine structure effect (Paschen-Back effect).

The energy difference between the Zeeman splitted $\ell = 0$ levels are identical for all values of n. For $\ell > 0$ the spin-orbit contribution leads to deviations from this equal distance rule. This will be reflected in the spectral lines formed. Allowed transitions between the energy levels displayed in figure 3.15 are illustrated by vertical lines. The selection rules derived in section 3.10 are still valid. In particular, transitions may only occur between energy levels

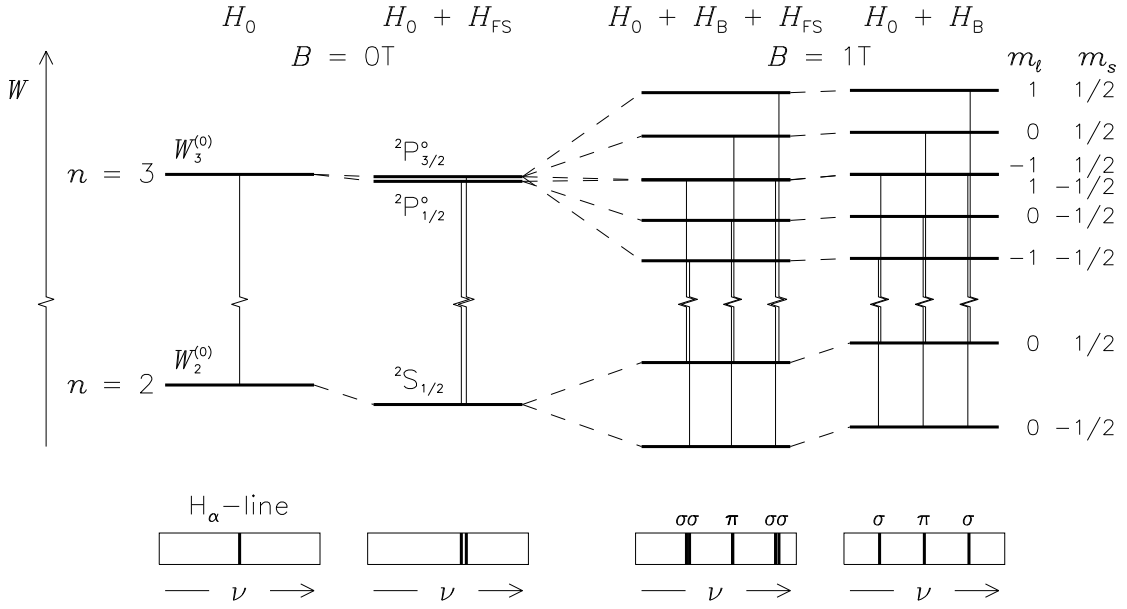

Figure 3.15: Pachen-Back effect for selected $n = 2$ and $n = 3$ energy levels

corresponding to identical m_s values. This reduces the number of allowed transitions to six, the two $m_\ell = 0 \leftrightarrow m_\ell = 0$ transitions having identical frequencies. The resulting five spectral lines are illustrated in the bottom panel of figure 3.15. The π- and σ-lines correspond to $\Delta m_\ell = 0$ and $\Delta m_\ell = \pm 1$, respectively.

For intermediate external magnetic field strengths, a more careful perturbation expansion must be performed. In particular, the degeneracy of energy levels and the fact that some of the cross-coupling terms in (3.134) are non-vanishing must be taken into account. It will then be seen how the weak external field result illustrated in figure 3.14 is continuously transformed into the strong field limit of figure 3.15.

3.13.3 Intermediate magnetic field strength

For intermediate magnetic field strengths it is necessary to treat the fine structure and the Zeeman terms together, that is consider $H' = H_{FS} + H_B$. This case is slightly more complicated as it is here no longer possible to avoid making full scale use of degenerate perturbation theory (see table 3.9). We choose to make use of the $|\, n\ell sjm_j \,\rangle$ set of eigenvectors of H_0 and restrict our discussion to the $n = 2$ energy levels. Higher n-values can be treated similarly, but at the expense of having to evaluate $2n^2$ dimensional determinants.

Writing $|\, n\ell sjm_j \,\rangle = |\, \ell jm_j \,\rangle$ for $n = 2$, $s = \frac{1}{2}$ and ordering the $n = 2$ eigenvectors as $\{|\, q_k^{(0)} \,\rangle, k = 1, \ldots, 8\} = \{|\, 0\tfrac{1}{2} -\tfrac{1}{2} \,\rangle, |\, 0\tfrac{1}{2}\tfrac{1}{2} \,\rangle, |\, 1\tfrac{1}{2} -\tfrac{1}{2} \,\rangle, |\, 1\tfrac{1}{2}\tfrac{1}{2} \,\rangle, |\, 1\tfrac{3}{2} -\tfrac{3}{2} \,\rangle, |\, 1\tfrac{3}{2} -\tfrac{1}{2} \,\rangle, |\, 1\tfrac{3}{2}\tfrac{1}{2} \,\rangle, |\, 1\tfrac{3}{2}\tfrac{3}{2} \,\rangle\}$ we have already evaluated the necessary $\langle\, q_i^{(0)} \,|\, H_{FS} \,|\, q_k^{(0)} \,\rangle$ terms. Non-zero results are only found for $i = k$. For the corresponding $\langle\, q_i^{(0)} \,|\, H_B \,|\, q_j^{(0)} \,\rangle$ terms we make use of the Clebsch-Gordan expansion of the $|\, n\ell sjm_j \,\rangle$ eigenvectors in terms of $|\, n\ell m_\ell sm_s \,\rangle$ eigenvectors as given by (3.84)-(3.85) in order to evaluate the effect of S_z. This time non-zero off-diagonal terms are found for (i, k) combinations (3,6) and (4,7). The first order degenerate perturbation

result after substitution in (3.134) is

$$\det\begin{vmatrix} W_1-\beta & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & W_1+\beta & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & W_1-\frac{\beta}{3} & 0 & 0 & \frac{\sqrt{2}\beta}{3} & 0 & 0 \\ 0 & 0 & 0 & W_1+\frac{\beta}{3} & 0 & 0 & \frac{\sqrt{2}\beta}{3} & 0 \\ 0 & 0 & 0 & 0 & W_2-2\beta & 0 & 0 & 0 \\ 0 & 0 & \frac{\sqrt{2}\beta}{3} & 0 & 0 & W_2-\frac{2\beta}{3} & 0 & 0 \\ 0 & 0 & 0 & \frac{\sqrt{2}\beta}{3} & 0 & 0 & W_2+\frac{2\beta}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & W_2+2\beta \end{vmatrix}$$

$$= \left((W_1-\tfrac{\beta}{3})(W_2-\tfrac{2\beta}{3})-\tfrac{2\beta^2}{9}\right)\left((W_1+\tfrac{\beta}{3})(W_2+\tfrac{2\beta}{3})-\tfrac{2\beta^2}{9}\right)$$

$$(W_1^2-\beta^2)(W_2^2-4\beta^2)=0, \tag{3.145}$$

where

$$W_1 = \frac{1}{2}mc^2\,\frac{(Z\alpha)^4}{n^3}\left(\frac{3}{4n}-1\right)-\Delta W,$$

$$W_2 = \frac{1}{2}mc^2\,\frac{(Z\alpha)^4}{n^3}\left(\frac{3}{4n}-\frac{1}{2}\right)-\Delta W,$$

and $\beta = \hbar\omega_L$. We recognize the first term of W_1 and W_2 as the fine structure contribution to ΔW for the two j-values involved. The 8 different solutions for the energy perturbation ΔW as functions of the magnetic field strength B are plotted in figure 3.16, the six $2p\,^2P^o_{3/2}$ and $2p\,^2P^o_{1/2}$ levels with black and blue lines, and the two $2s\,^2S_{1/2}$ with red lines. In the left hand part of the figure the result coincides with the weak field limit results. In the right hand part of the figure the trend toward the five equidistant energy levels of the strong field limit result can be seen.

It is useful to compare the relative importance of the fine structure and Zeeman terms. The former scales like Z^4/n^3 the latter like B. What can be considered a "weak" or "strong" magnetic field case thus depends on the mass of the atom (through the value of Z) and the energy level considered. From figure 3.16 we see that for $Z = 1$ and $n = 2$ the "intermediate" field strength is typically $B = 1$ T. For $Z = 1$ and $n = 3$ the typical "intermediate" field strength is reduced by more than a factor 3. In astrophysical context magnetic fields with strengths $B \gg 1$ T is often found in certain compact stars (some white dwarfs, pulsars, $\cdots$) .

3.13.4 Physical visualization of angular momentum coupling

The two limiting cases of the Zeeman effect may be visualized through the following "physical" model. In an external magnetic field, a magnetic moment $\boldsymbol{m}$ will experience a torque $\boldsymbol{m}\times\boldsymbol{B}$. If the magnetic dipole is associated with a parallel angular momentum, this momentum vector will perform a precessional motion around the magnetic field, analogous to the precessional motion of a spinning top in a gravitational field. During this motion, the component of the angular momentum along the magnetic field remains constant. The quantization rule for the

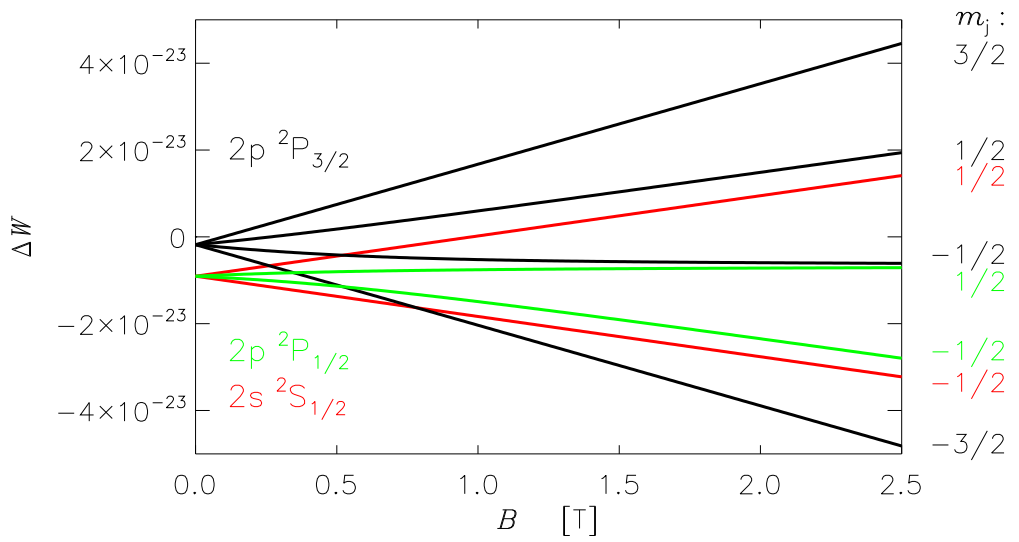

Figure 3.16: Combined fine structure and Zeeman splitting as functions of magnetic field B for $Z = 1$ and $n = 2$.

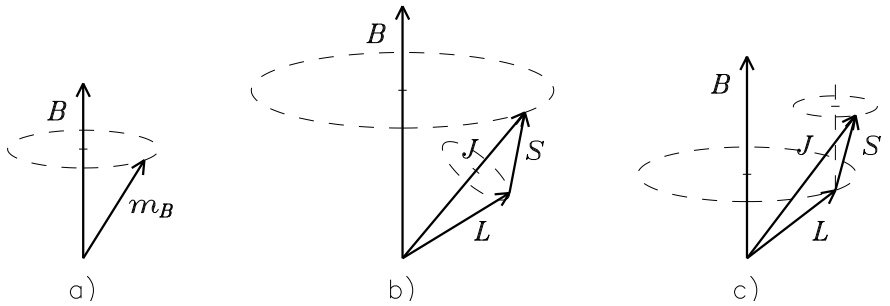

Figure 3.17: Precessing motions in a magnetic field

normal Zeeman effect, where only one of the vectors $\boldsymbol{L}$ or $\boldsymbol{S}$ is different from zero, conforms with this classical result. The situation is illustrated in figure 3.17a.

The situation is different when both $\boldsymbol{L}$ and $\boldsymbol{S}$ are non-vanishing. In the weak external magnetic field limit the spin-orbit interaction represents the stronger coupling between these vectors. In this case the total angular momentum $\boldsymbol{J}$ performs the precessional motion around the magnetic field as discussed above. The individual vectors $\boldsymbol{L}$ and $\boldsymbol{S}$ are at the same time performing a precessional motion around the total angular momentum vector $\boldsymbol{J}$. We thus would expect that the effect of the components of $\boldsymbol{S}$ perpendicular to $\boldsymbol{J}$ in H_B to average out. The situation is illustrated in figure 3.17b. In "classical" terms the net value of H_B may then be written

$$\frac{e}{2m}\boldsymbol{B}\cdot(\boldsymbol{J}+\boldsymbol{J}\frac{\boldsymbol{J}\cdot\boldsymbol{S}}{\boldsymbol{J}^2}) = \frac{eB}{2m}J_z(1+\frac{\boldsymbol{J}^2-\boldsymbol{L}^2+\boldsymbol{S}^2}{2\boldsymbol{J}^2}) \tag{3.146}$$

Here we made use of the identity $\boldsymbol{J}^2-\boldsymbol{L}^2+\boldsymbol{S}^2 = 2\boldsymbol{J}\cdot\boldsymbol{S}$. The second term of the left hand side of (3.146) is just the component of $\boldsymbol{S}$ along $\boldsymbol{J}$ in accordance with the interpretation offered above. If the squared angular momenta on the right hand side of (3.146) are now replaced with their quantum mechanical values, we see that our interpretation is in accordance with

the result (3.141)-(3.142).

In the strong external magnetic field limit, the coupling between the $\boldsymbol{L}$ and $\boldsymbol{S}$ vectors is not strong enough to prevent the individual vectors from precessing on their own around the magnetic field. The average value of each vector is equal to the component of these vectors along $\boldsymbol{B}$ in accordance with (3.143). The situation is illustrated in figure 3.17c. In this case the spin-orbit interaction term with its product $\boldsymbol{L} \cdot \boldsymbol{S}$ must be evaluated as the product of the average values of each of the two vectors $\boldsymbol{L}$ and $\boldsymbol{S}$

$$\boldsymbol{L} \cdot \boldsymbol{S} \rightarrow \hat{\boldsymbol{B}}(\hat{\boldsymbol{B}} \cdot \boldsymbol{L}) \cdot \hat{\boldsymbol{B}}(\hat{\boldsymbol{B}} \cdot \boldsymbol{S}).$$

The justification for the above "physical" model is primarily as a "memorization" rule. Still, the model may be a useful tool in certain situations.

3.13.5 Polarization and directional effects

Without an external magnetic field, the atom has no preferred direction. The strength of any spectral line feature formed must therefore be independent of the direction of the radiation absorbed or emitted by the atom. For an atom in an external field this is no longer true. The field imposes a preferred direction for the wave function and we must expect a directional effect for the strengths of spectral lines.

To study this effect let the external field be parallel with the z-axis and let the wave vector $\boldsymbol{k}$ of the electromagnetic wave be lying in the xz-plane, making an angle χ with the z-axis. The polarization vector $\hat{\boldsymbol{\epsilon}}$ must be perpendicular to $\boldsymbol{k}$. In figure 3.18a two plane polarized directions are given, $\hat{\boldsymbol{\epsilon}}_{\parallel}$ in the plane containing $\boldsymbol{B}$ and $\boldsymbol{k}$, and $\hat{\boldsymbol{\epsilon}}_{\perp}$ perpendicular to this plane. An arbitrary wave polarization can be constructed from a suitable linear combination of these two plane polarized directions. We note from the figure that

$$\hat{\boldsymbol{\epsilon}}_{\parallel} \cdot \boldsymbol{r} = r(\sin\chi\cos\theta - \cos\chi\sin\theta\cos\varphi), \tag{3.147}$$

$$\hat{\boldsymbol{\epsilon}}_{\perp} \cdot \boldsymbol{r} = r\sin\theta\sin\varphi. \tag{3.148}$$

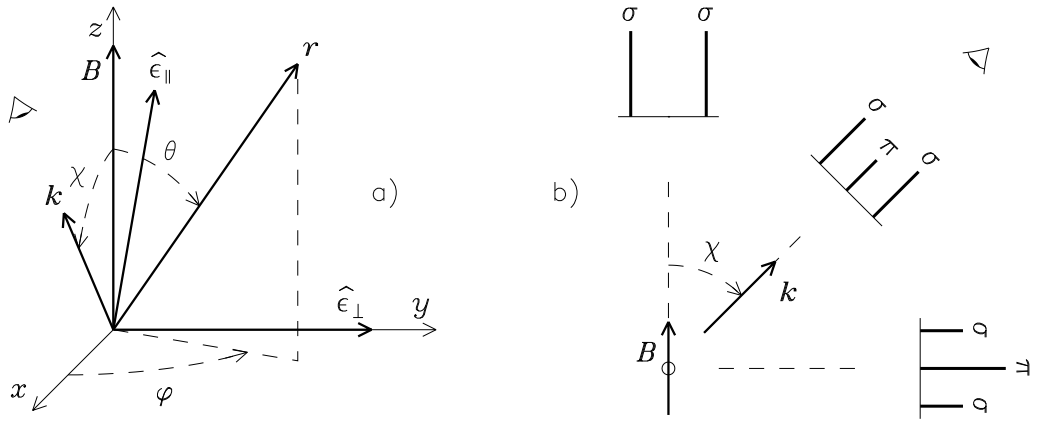

Figure 3.18: Zeeman effect – directional dependence

The strength of the spectral line resulting from a transition from state $|\, i \,\rangle$ to state $|\, f \,\rangle$ depends on the value of the integral $\langle\, f \,|\, \hat{\boldsymbol{\epsilon}} \cdot \boldsymbol{r} \,|\, i \,\rangle$ through the expressions for the transition

rates w_{fi} for absorption and emission as given in (3.97) and (3.98). On the basis of the discussion of selection rules in section 3.10, we see that the first term of $\hat{\boldsymbol{\epsilon}}_{\parallel} \cdot \boldsymbol{r}$ will give non-vanishing contributions to these integrals only if $\Delta m_\ell = 0$ and $\Delta \ell = \pm 1$. $\hat{\boldsymbol{\epsilon}}_{\perp} \cdot \boldsymbol{r}$ and the second term of $\hat{\boldsymbol{\epsilon}}_{\parallel} \cdot \boldsymbol{r}$ both give non-vanishing contributions only if $\Delta m_\ell = \pm 1$ and $\Delta \ell = \pm 1$. This means, first of all, that the external $\boldsymbol{B}$-field does not lead to any modification of the selection rules as stated in section 3.10.

Secondly, we may easily inspect how the strength of a given spectral line will vary with the direction χ of the electromagnetic radiation, that is, with the angle χ with respect to the local magnetic field under which the absorbing or emitting atoms are viewed. The π-line with $\Delta m_\ell = 0$ corresponds to a radiation field polarized in the plane containing $\boldsymbol{k}$ and $\boldsymbol{B}$. The intensity of this line will vary with the angle χ as

$$I_\pi(\chi) = I_{\pi\parallel} \sin^2 \chi \tag{3.149}$$

where $I_{\pi\parallel}$ is a constant. When observing along the external magnetic field $\boldsymbol{B}$ the π-line vanishes.

The contributions from the parallel and perpendicular polarizations to the σ-lines with $\Delta m_\ell = \pm 1$ may be added to produce an elliptical polarized spectral line with intensity varying as

$$I_\sigma(\chi) = I_{\sigma\perp} + I_{\sigma\parallel} \cos^2 \chi \tag{3.150}$$

where $I_{\sigma\perp}$ and $I_{\sigma\parallel}$ are both constants. It can be shown (see quiz 3.62) that $I_{\sigma\perp} = I_{\sigma\parallel}$. In particular, for $\chi = \pi/2$ the σ-lines correspond to a radiation field polarized perpendicular to $\boldsymbol{k}$ and $\boldsymbol{B}$. At other angles the σ-line will be elliptically polarized. The directional dependence of the Zeeman lines are schematically illustrated in figure 3.18b.

Making use of the exact location and the polarization properties of the Zeeman lines, it is possible to infer not only the strength of stellar magnetic fields, but also the direction of these fields. This conclusion requires that the individual Zeeman lines can be resolved, which again depends on the temperature of the absorbing or emitting medium and the mass of the atom involved.

Quiz 3.58 : Verify (3.141).

Quiz 3.59 : What are the Landé g-factors for the $^2\mathrm{D}_{3/2}$ and $^2\mathrm{D}_{5/2}$ levels? Compare the splitting of these levels with that of the $^2\mathrm{P}^o_{3/2}$ level given in figure 3.14.

Quiz 3.60 : The fine structure splitting of the 2p $^2\mathrm{P}^o_{1/2}$ and 3s $^2\mathrm{S}_{1/2}$ and 3d $^2\mathrm{D}_{3/2}$ energy levels of the H I atom are identical to those of 2s $^2\mathrm{S}_{1/2}$ and 3p $^2\mathrm{P}^o_{1/2}$ and 3p $^2\mathrm{P}^o_{3/2}$ levels, respectively. The anomalous Zeeman splitting of the latter set of energy levels were studied in figure 3.14. What are the Zeeman splitting of the former set of levels? What are the corresponding Zeeman splitting of the fine structure lines?

Quiz 3.61 : In the strong field limit, what is the degeneracy g of the different energy levels $W = W_n^{(0)} + \Delta W_B$ with ΔW_B given by (3.143)? Argue why the extra contribution to the energy levels from the spin-orbit term H_{LS} can be calculated using the simplified degenerate perturbation result (3.136).

Quiz 3.62 : Verify that

$$\int_0^{2\pi} \sin\varphi \, \exp(\pm\imath\varphi) \, \mathrm{d}\varphi = \pm\imath \int_0^{2\pi} \cos\varphi \, \exp(\pm\imath\varphi) \, \mathrm{d}\varphi.$$

From (3.147) and (3.148) then argue that $I_{\sigma\perp} = I_{\sigma\|}$ in (3.150).

Quiz 3.63 : The weak field result (3.141) should properly contain $\langle\, q_{FS}^{(0)} \mid J_z + S_z \mid q_{FS}^{(0)} \,\rangle$ with $|\, q_{FS}^{(0)} \,\rangle = |\, n\ell sjm_j \,\rangle + |\, \Delta q_{FS} \,\rangle$ (see quiz 3.57). Convince yourself that first order terms in $|\, \Delta q_{FS} \,\rangle$ cancel and therefore that (3.141) is correct.

3.14 The Stark Effect

The *Stark effect* concerns the question of how an external electric field $\boldsymbol{E}$ modifies the energy levels of the atom and therefore also the spectral lines produced. For a constant external electric field $\boldsymbol{E} = E\hat{\boldsymbol{z}}$ the additional potential energy term of the Hamiltonian is equal to the scalar product of the external electric field with the electric dipole moment $\boldsymbol{m}_E = -e\boldsymbol{r}$ of the atom,

$$H_E = -\boldsymbol{m}_E \cdot \boldsymbol{E} = eEz. \tag{3.151}$$

For external fields much less than the typical internal electric field seen by the electron, the effect of the additional term may be analyzed using standard perturbation theory as summarized in table 3.9. For simplicity we assume the external fields to be large enough that fine structure effects may be neglected compared with the electric field effect. For the $n = 2$ energy level to be discussed below, this means that the external electric field must exceed 10^6 V/m. With these assumptions the result from perturbation theory will be of the general form

$$W = W^{(0)} + \alpha E + \frac{1}{2}\beta E^2 + \cdots . \tag{3.152}$$

The first and second order terms are referred to as the linear and quadratic Stark effects. From a classical point of view, the linear term represents the potential energy of a *permanent* electric dipole $\boldsymbol{m}_{Ep}$ in the external electric field. The quadratic term represents the potential energy $\frac{1}{2}\boldsymbol{m}_{Ei} \cdot \boldsymbol{E}$ of a dipole moment *induced* by the external field, $\boldsymbol{m}_{Ei} = \beta\boldsymbol{E}$.

From a classical point of view we would not expect a free atom to possess a permanent electric dipole moment. We would thus expect the linear Stark effect to vanish. For non-degenerate energy states this holds true also from the quantum mechanical point of view, the matrix elements $\langle\, n\ell m_\ell \mid z \mid n\ell m_\ell \,\rangle$ vanish for any wave function $|\, n\ell m_\ell \,\rangle$ with even or odd parity[5]. There is thus no linear Stark splitting of the ground state energy level. For degenerate energy states where it is necessary to make use of degenerate perturbation theory, however, the conclusion will be different.

As an example consider the four-fold degenerate $n = 2$ energy level with wave functions $|\, 2\ell m_\ell \,\rangle$, $\ell = 0, 1$. (We choose the ordering $\{|\, 200\,\rangle, |\, 21-1\,\rangle, |\, 210\,\rangle, |\, 211\,\rangle\}$.) From the discussion of selection rules, we know that the matrix elements $\langle\, n\ell' m'_\ell \mid z \mid n\ell m_\ell \,\rangle$ are non-vanishing only if $m'_\ell = m_\ell$ and $\ell' = \ell \pm 1$. In fact, substituting the explicit form for the wave functions and performing the required integrals, we find the only non-vanishing element to be

$$\langle\, 210 \,|z|\, 200 \,\rangle = -3\,\frac{a_B}{Z}. \tag{3.153}$$

[5]Our choice of wave function notation reflects the fact that since H_E is independent of electron spin, spin effects may be disregarded in the present context.

We are thus left with the problem of solving the equation

$$\det \begin{vmatrix} -\Delta W & 0 & a & 0 \\ 0 & -\Delta W & 0 & 0 \\ a & 0 & -\Delta W & 0 \\ 0 & 0 & 0 & -\Delta W \end{vmatrix} = \Delta W^2(\Delta W^2 - a^2) = 0,$$

with $a = -3a_B eE/Z$. The energy level is seen to split into three equidistant levels with energy separation a. The situation is illustrated in figure 3.19, where also the corresponding expressions for the perturbed wave functions are given.

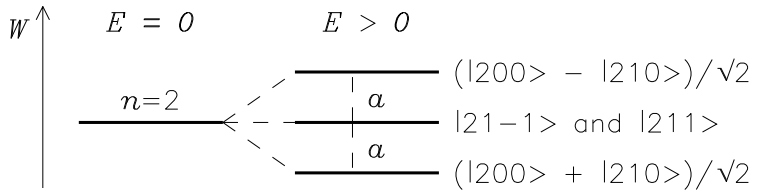

Figure 3.19: Linear Stark splitting of the $n = 2$ energy level

Similarly, the non-vanishing matrix elements for nine-fold degenerate $n = 3$ energy level are

$$\langle\, 310 \,|\, z \,|\, 300 \,\rangle = -3\sqrt{6}, \quad \langle\, 320 \,|\, z \,|\, 310 \,\rangle = -3\sqrt{3}, \quad \langle\, 32\,\pm 1 \,|\, z \,|\, 31\,\pm 1 \,\rangle = -\frac{9}{2}, \qquad (3.154)$$

all in units of a_B/Z, with corresponding energy splitting $\Delta W = 0, \pm\frac{3}{2}\,a, \pm 3\,a$. Generally, the linear Stark effect splits the energy level n of the one-electron atom into $2n-1$ equidistant energy levels with the distance between subsequent levels increasing with n.

As noted for the $n = 2$ energy level, the perturbed wave functions in the presence of an external electric field will no longer necessarily have a definite parity (even or odd). This means that ℓ will no longer be a "good" quantum number, nor will the selection rules of section 3.10 remain valid. In particular, because of the mixing of the Ψ_{200} and Ψ_{210} states, the meta-stable $2s$ state will be "quenched" by an external electric field with the lifetime being significantly shortened compared to its "natural" value of 0.12 s.

The quadratic Stark effect will for one-electron atoms, and for electric field strengths normally encountered, be much smaller than the corresponding linear effect. The ground state, for which the linear effect vanishes, represents one exception. Quadratic Stark effect will be discussed in section 4.9.2.

Electric fields naturally occur in ionized gases. Strong electric micro-fields regularly arise during collisions (close approaches) between neutral atoms and charged particles. Radiation transitions taking place during such collisions will be subject to the Stark effect, giving rise to a general broadening of spectral lines. The effect is referred to as *Stark broadening*. The detailed line profile depends on the probability distribution function for the strength of such electric micro-fields in the plasma. The line width increases with pressure, partly because the average inter-particle distance decreases with increasing gas density and partly because collisions tend to become "harder" with increasing temperature. In figure 3.20 a schematic illustration of the Stark effect during a close approach between an atom and an ion is given. The form of the ion orbit relative to the atom depends on the total charge of the atom. In the

figure the atom is assumed have lost all but one of its electrons. Radiative transitions in the atom taking place during the close encounter will have different wavelengths dependent on the time-varying distance $r(t)$ between the atom and the ion. The Stark effect in ionized gases thus contributes to a general broadening mechanism for spectral lines. The combined effect of thermal, collisional and Stark broadening mechanisms taking place in hot and dense stellar interiors contribute to the formation of the continuous radiation spectrum leaving stellar photospheres, that is, the combined effect of many neighbouring individual broad spectral lines gives the appearence of a continuous spectrum.

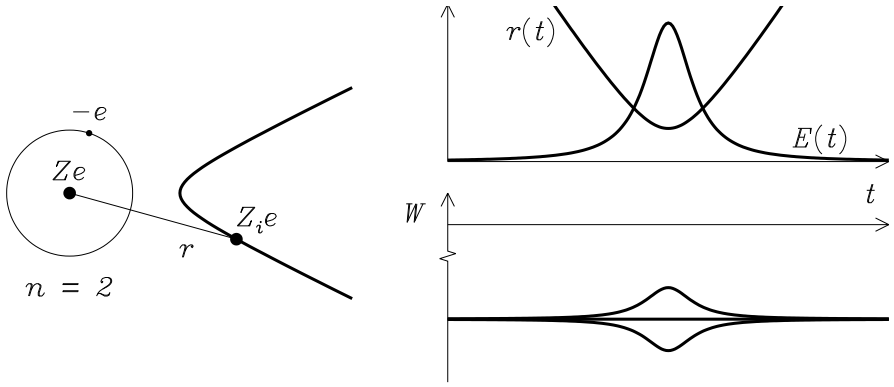

Figure 3.20: Time-varying Stark effect during close encounter between an atom and an ion

In the discussion above we have assumed the strength of the external field to be relatively modest for a linear perturbation expansion to be valid. Additional effects are to be expected with increasing field strengths. External field strengths comparable to the electric field normally seen by the electron within the atom may lead to field ionization of the atom. This is illustrated schematically in figure 3.21 The dashed line represents the electrostatic potential due to the nuclear charge. The solid line is the modified potential in the presence of a constant external field,

$$U = \frac{Ze}{4\pi\epsilon_0}\frac{1}{r} - eEz,$$

plotted along an axis through the nucleus and parallel with the external field. The external field is seen to reduce the effective bounding potential with an amount δU. This will lead to a corresponding reduction in the effective ionization potential of the atom and a destabilization of high-n energy states of the atom. This effect has importance for the evaluation of the partition function for internal states of the atom (see section 6.7). For strong enough external fields even the stability of the ground state of the atom may be lost.

Quiz 3.64 : Argue why you, for a given constant external electric field E, would expect the distance between split energy levels to increase with increasing principal quantum number n.

Quiz 3.65 : A singly charged ion approaches an H I atom in the $n = 2$ state to within a distance $d = 10$ Å. Calculate the maximum modification of the $n = 2$ energy level of the atom induced by the ion. What is the corresponding width of the resonance line?

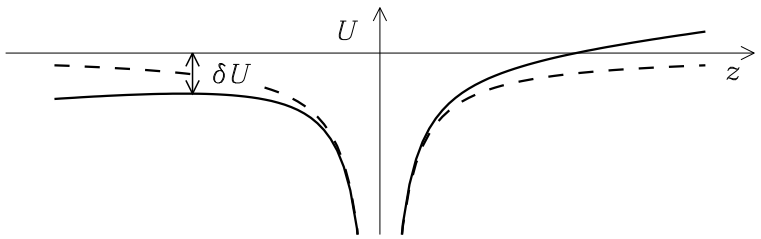

Figure 3.21: Modification of bounding potential by an external field

Quiz 3.66 : With non-vanishing matrix elements as given by (3.154) derive the corresponding splitting ΔW of the $n = 3$ energy level.

3.15 Nuclear Spin and Hyperfine Effects

We shall end our review of first order effects for the one-electron atom by a brief discussion of nuclear spin effects. This is an effect that will give us a handle on how to investigate the density distribution of neutral atomic hydrogen in the galaxy.

Electrons, protons and neutrons are all spin 1/2 particles. The spin of protons and neutrons in the nucleus combine to form the total nuclear spin angular momentum $\boldsymbol{I}$. The nuclear spin angular momentum will satisfy quantization rules similar to those familiar for $\boldsymbol{L}$, $\boldsymbol{S}$ or $\boldsymbol{J}$. The nuclear spin number I will be half-integer or integer valued for odd or even nucleon numbers, respectively. The electrically charged and spinning nucleus represents a nuclear magnetic moment

$$\boldsymbol{m}_I = g_I \frac{e}{2M} \boldsymbol{I}, \tag{3.155}$$

where M is the nuclear mass and g_I is the *nuclear gyro-magnetic ratio.* The latter is a dimensionless number of order unity. Spin numbers I and nuclear gyro-magnetic ratios g_I for some selected nuclei are listed in table 3.10.

Nucleus	Spin I	gyro-magnetic ratio g_I
proton	1/2	5.5883
neutron	1/2	-3.8263
$^{2}_{1}$H	1	0.85742
$^{3}_{2}$He	1/2	-4.255
$^{4}_{2}$He	0	-
$^{12}_{6}$C	0	-

Table 3.10: Nuclear spin and gyro-magnetic ratio

The magnetic field generated by the nuclear magnetic moment $\boldsymbol{m}_I$ will be interacting with the orbital and spin angular momenta $\boldsymbol{L}$ and $\boldsymbol{S}$ of the electron to produce an extra potential energy term H_I of the Hamiltonian. For simplicity we shall only consider the effects of nuclear spin on $\boldsymbol{L} = 0$ electron states. For this case the interaction term reduces to

$$H_I = -\boldsymbol{m}_S \cdot \boldsymbol{B}_I \tag{3.156}$$

where the electron spin magnetic moment $\boldsymbol{m}_S$ is given by (3.76) and the magnetic field $\boldsymbol{B}_I$ produced by the nuclear magnetic dipole moment $\boldsymbol{m}_I$ can be derived from the vector potential

$$\boldsymbol{A}_I(\boldsymbol{r}) = \frac{\mu_0}{4\pi}\boldsymbol{m}_I \times \nabla\frac{1}{r}, \tag{3.157}$$

that is,

$$\boldsymbol{B}_I(\boldsymbol{r}) = \nabla \times \boldsymbol{A}_I(\boldsymbol{r}) = \frac{\mu_0}{4\pi}\left(\boldsymbol{m}_I\nabla^2\frac{1}{r} - \boldsymbol{m}_I\cdot\nabla\nabla\frac{1}{r}\right). \tag{3.158}$$

We note that the result is singular at the dipole location.

The sum of the nuclear and electron spin angular momenta constitutes the *total atomic angular momentum*

$$\boldsymbol{F} = \boldsymbol{I} + \boldsymbol{S}. \tag{3.159}$$

$\boldsymbol{F}$ and F_z obey the familiar set of angular momentum quantization rules with corresponding quantum numbers F and M_F satisfying the constraints

$$F = |\, I - s \,|, I + s$$

and

$$M_F = -F, -F+1, \cdots, F.$$

Even the selection rules take the familiar form

$$\Delta F = 0, \pm 1,$$

the transition $F = 0 \longleftrightarrow F = 0$ being excluded.

The shift in the energy levels of $\boldsymbol{L} = 0$ states due to the interaction between the nuclear and electron spin can now be calculated using first order perturbation theory,

$$\Delta W_I = -\langle\, \boldsymbol{m}_S \cdot \boldsymbol{B}_I \,\rangle = \frac{\mu_0}{4\pi}\, g_I \frac{e^2}{2mM}\, \langle\, (\boldsymbol{S}\cdot\nabla)(\boldsymbol{I}\cdot\nabla)\frac{1}{r} - \boldsymbol{S}\cdot\boldsymbol{I}\,\nabla^2\frac{1}{r}\,\rangle. \tag{3.160}$$

In Cartesian coordinates the volume integral part of the first term reduces to

$$\sum_{i,j} S_i I_j \int |\,\Psi_{n00}(r)\,|^2\, \frac{\partial^2}{\partial x_i \partial x_j}\frac{1}{r}\,\mathrm{d}^3\boldsymbol{r}.$$

For $i \neq j$ the integrands are odd functions and the volume integrals thus vanish identically. The volume integrals of terms with $i = j$ are all equal,

$$\int |\,\Psi_{n00}(r)\,|^2\, \frac{\partial^2}{\partial x_i^2}\frac{1}{r}\,\mathrm{d}^3\boldsymbol{r} = \frac{1}{3}\int |\,\Psi_{n00}(r)\,|^2\, \nabla^2\frac{1}{r}\,\mathrm{d}^3\boldsymbol{r}.$$

Combining both terms of (3.160) and making use of the relations

$$\nabla^2\frac{1}{r} = -4\pi\delta(\boldsymbol{r}) \tag{3.161}$$

$$\boldsymbol{S}\cdot\boldsymbol{I} = \frac{1}{2}(\boldsymbol{F}^2 - \boldsymbol{I}^2 - \boldsymbol{S}^2), \tag{3.162}$$

the expression for the energy shift reduces to

$$\begin{aligned}\Delta W_I &= \frac{2\mu_0}{3}\, g_I \frac{e^2}{2mM} \left\langle \frac{1}{2}(\boldsymbol{F}^2 - \boldsymbol{I}^2 - \boldsymbol{S}^2)\, \delta(\boldsymbol{r}) \right\rangle \\ &= \frac{\mu_0}{6} g_I \frac{e^2\hbar^2}{mM} \left(F(F+1) - I(I+1) - s(s+1)\right)\, |\,\Psi_{n00}(0)\,|^2 \,.\end{aligned}$$

Here

$$\Psi_{n00}(0) = \frac{Z^3}{\pi a_B^3 n^3} \left(\frac{\mu}{m}\right)^3 \tag{3.163}$$

is the value of the wave function (3.48) at the origin, but corrected for the isotope effect, that is, the electron mass m in the expression (3.45) for the Bohr radius a_B has effectively been replaced with the reduced electron mass μ as given by (3.65).

For the H I atom we have $I = 1/2$ and therefore $F = 0$ or $F = 1$. The two values of F give rise to a splitting of the $l = 0$ energy levels. The energy separation between the $F = 0$ and $F = 1$ levels of ground state ($n = 1$) of the H I atom is thus given by

$$\delta W_I = hc\, R_\infty \alpha^2 \cdot \frac{8}{3} g_I \frac{m}{M} \left(\frac{\mu}{m}\right)^3 . \tag{3.164}$$

A radiation transition between these two levels is allowed and results in the measured radio wave frequency

$$\nu = (1420405751.800 \pm 0.028)\ \text{Hz}.$$

This frequency is one of the most accurately measured quantities in physics. First order perturbation theory as represented by (3.164), predicts the experimental result with a relative error better than 0.1 %. The radiation is normally referred to as the 21 cm radiation or the "spin-flip" line and forms the basis of a method for the mapping of the distribution of H I in spiral arms of the Milky Way and other galaxies.

Chapter 4

Spectra of Many-Electron Atoms

We proceed now to a discussion of the physics of many-electron atoms and their spectra. We shall recognize many of the basic features seen for the one-electron atom. We shall also see important differences. We shall even have to postulate a new principle, the Pauli exclusion principle, in order to be able to account for the properties of the many-electron atom.

4.1 The Pauli Exclusion Principle

For many-electron atoms the zero order Hamiltonian will contain a sum of potential energy contributions resulting from the mutual interaction between the different electrons

$$H = \sum_{i=1}^{N} H_0(i) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} U'(i,j). \tag{4.1}$$

Here N is the number of electrons in the atom,

$$H_0(i) = \frac{\boldsymbol{p}_i^2}{2m} + U(r_i) \quad \text{with} \quad U(r_i) = -\frac{Ze^2}{4\pi\epsilon_0 r_i} \tag{4.2}$$

is the one-electron Hamiltonian for electron number i, and

$$U'(i,j) = \frac{e^2}{4\pi\epsilon_0} \frac{1}{|\,\boldsymbol{r}_i - \boldsymbol{r}_j\,|} \tag{4.3}$$

is the interaction energy between electrons i and j at positions $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$. In (4.1) and (4.2) the nucleus was assumed to be fixed at the origin of the coordinate system. For neutral atoms and positive ions the nuclear charge number Z satisfies the requirement $Z \geq N$. This is the most common case. However, also some negative ions with $Z < N$ play important roles in astrophysics (H^- with $Z = 1$ and $N = 2$ is an important contributor to the extinction coefficient in stellar atmospheres). Relativistic effects (the relativistic kinetic energy correction, spin-orbit and spin-spin interactions and effects caused by nuclear spin) together with the effects due to external electric and magnetic fields were neglected in the Hamiltonian (4.1). We shall return to these first order effects in subsequent sections.

The allowed energy levels W of the atom is again found as the eigenvalues of the stationary Schrödinger equation

$$H\,|\,\Psi(1,2,\cdots,N)\,\rangle = W\,|\,\Psi(1,2,\cdots,N)\,\rangle. \tag{4.4}$$

The wave functions $|\,\Psi(1,2,\cdots,N)\,\rangle$ are now functions of the coordinates and spins of all the electrons in the atom.

The Hamiltonian H for a system of N indistinguishable particles must be invariant under an interchange of coordinates and spins of any two particles. The Schrödinger equation (4.4) as such does not impose any particular requirement on the behavior of the wave function under this operation. It was, however, found that for the predictions of this equation to be in agreement with observations for many-electron systems, it was necessary to impose the *Pauli exclusion principle* as an independent requirement. This principle, (the *Pauli principle* for short) requires the wave function $|\,\Psi(1,2,\cdots,N)\,\rangle$ to be antisymmetric with respect to the interchange of any pair of electron coordinates and spins,

$$|\,\Psi(1,\cdots,i,\cdots,j,\cdots,N)\,\rangle = -\,|\,\Psi(1,\cdots,j,\cdots,i,\cdots,N)\,\rangle. \tag{4.5}$$

An equivalent formulation of this principle is to say that two electrons cannot occupy identical states in the atom. If two electrons were to occupy identical states, the wave function would not change sign during the permutation of these two electrons, in contrast to the requirements of the principle.

Other symmetry properties of the wave function follow from the explicit form of the Hamiltonian. In particular, the Hamiltonian (4.1) is invariant with respect to the mirroring of all the electrons about the origin, $\boldsymbol{r}_i \to -\boldsymbol{r}_i$ for $i = 1,\cdots,N$. The wave function will reflect this fact by either remaining unchanged or by changing sign under this transformation. An electronic state corresponding to a wave function that does not change under the mirroring transformation will be said to have *even parity*. A state corresponding to a wave function that changes sign under this transformation will have *odd parity*.

4.2 The Central Field Approximation

The presence of the interaction energy terms (4.3) in the Hamiltonian makes the many-electron Schrödinger equation (4.4) inaccessible to exact analytical treatment, even in the present zero order theory. Approximation methods have been developed instead. In the *central field approximation*, solutions are sought by considering each electron to move in an effective central force field that includes the attraction from the nucleus together with an average repulsive effect between any selected electron and the other $N-1$ "average" electrons. With $V(r)$ representing the corresponding *effective potential*, the Hamiltonian may now be written in the form

$$H = H_c + H_{EC} \tag{4.6}$$

where

$$H_c = \sum_{i=1}^{N} \left(\frac{\boldsymbol{p}_i^2}{2m} + V(r_i) \right) \tag{4.7}$$

is the *central field Hamiltonian*. The remaining *electron correlation* term

$$H_{EC} = \sum_{i=1}^{N} (U(r_i) - V(r_i)) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} U'(|\,\boldsymbol{r}_i - \boldsymbol{r}_j\,|), \tag{4.8}$$

hopefully, represents a small correctional term with the proper choice of $V(r)$.

Seen from our selected electron, the other $N-1$ electrons will act to shield the nuclear charge. Close to the nucleus this shielding of the nuclear charge will be negligible and we expect the effective potential to take the form

$$V(r_i) \to -\frac{Ze^2}{4\pi\epsilon_0 r_i} \quad \text{as} \quad r_i \to 0.$$

In the opposite limit, the other electrons are expected to occupy the space between the selected electon and the nucleus. For this case we expect the shielding to have maximum effect and therefore

$$V(r_i) \to -\frac{(Z-N+1)e^2}{4\pi\epsilon_0 r_i} \quad \text{as} \quad r_i \to \infty.$$

The choice of the proper form of $V(r_i)$ for intermediate radii for a given state of the atom represents a demanding task to which we shall not enter. Still, on the basis of simple reasoning, we are able to gain valuable physical insight into the properties of the many-electron atom, even without the knowledge of the detailed form of $V(r_i)$.

In the central field approximation the eigenfunction $|\,\Psi(1,2,\cdots,N)\,\rangle$ of the N electron Schrödinger equation

$$H_c\,|\,\Psi(1,2,\cdots,N)\,\rangle = W_c\,|\,\Psi(1,2,\cdots,N)\,\rangle \tag{4.9}$$

reduces to a product of N one-electron eigenfunctions $|\,\psi(i)\,\rangle$ satisfying

$$\left(-\frac{\hbar^2}{2m}\nabla_i^2 + V(r_i)\right)|\,\psi(i)\,\rangle = w_i\,|\,\psi(i)\,\rangle. \tag{4.10}$$

We shall refer to the one-electron eigenfunctions $|\,\psi(i)\,\rangle$ as *spin-orbitals*. The corresponding energy W_c of the atom in this approximation equals the sum of the individual contributions from each electron, $W_c = \sum_i w_i$.

The eigenvalue problem (4.10) is identical to the Schrödinger equation (3.18) for the one-electron atom, except for the fact that the Coulomb potential $U(r_i) = -Ze^2/4\pi\epsilon_0 r_i$ is replaced by the effective potential $V(r_i)$. We may thus conclude that the spin part of the spin-orbitals $|\,\psi(i)\,\rangle$ is identical to that of the eigenfunction of the one-electron atom, that is ξ_{sm_s} or alternatively α or β. The spatial part of the spin-orbital will depend on three quantum numbers n_i, ℓ_i and $m_{\ell i}$, that is, the spatial part may be written in the form $\psi_{n_i\ell_i m_{\ell i}}(\boldsymbol{r}_i)$. The angular part of this function is identical to that of the wave function (3.48) of the one-electron atom. The radial part, however, will be different. In particular, the simple Rydberg formula (3.41), in which the energy only depends on the principal quantum number n_i, is only valid for the Coulomb potential $U(r_i)$. For the effective potential $V(r_i)$, corresponding to an extended charge distribution from the nucleus and the $N-1$ "average" electrons, the energy w_i will not only be a function of the principal quantum number n_i, but also depend on the shape of the electron orbital through the orbital quantum number ℓ_i, $w_i = w_i(n_i,\ell_i)$. With reference to figure 3.3, an electron with a given value of n_i will on the average be closer to the nucleus, the smaller the value of ℓ_i. The closer to the nucleus, the less effective will the negative charge distribution from the other electrons be in shielding the charge of the nucleus. Thus, the smaller the value of ℓ_i, the larger is the effective nuclear charge seen by the electron, and the stronger is the electron bounded to the nucleus. We conclude that $w_i(n_i,\ell_i)$ is an increasing function with respect to increasing values of both n_i and ℓ_i. This conclusion is important for any understanding of the properties of many-electron atoms.

The situation is illustrated schematically in figure 4.1. To the left the charge distribution resulting from the nuclear charge Ze and the spherically symmetric charge distribution from the $N-1$ "average" electrons is superposed a comparison between the corresponding effective potential energy curve $V(r)$ and the Coulomb potential $U(r)$ from a nuclear charge Ze. To the right, the quantum mechanical energy levels W_n and $w_{n\ell}$ corresponding to $U(r)$ and $V(r)$ are compared.

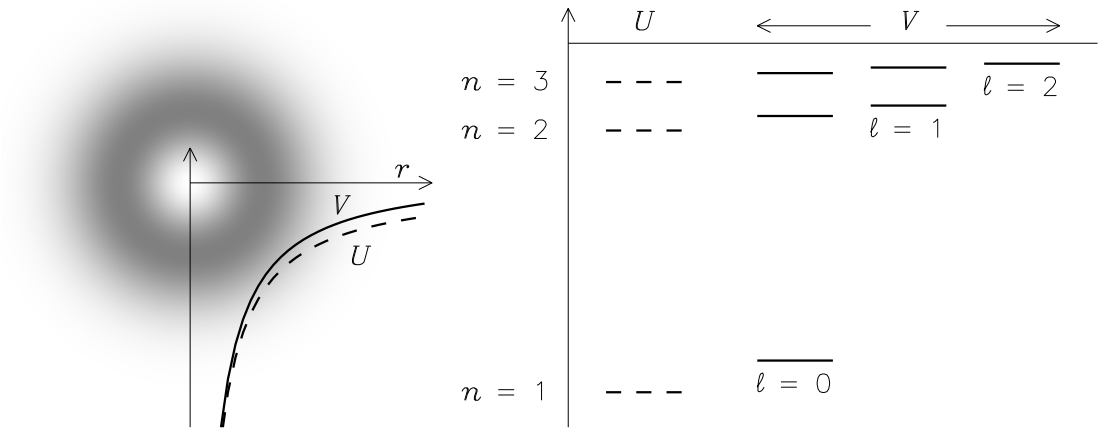

Figure 4.1: A comparison of energy levels for potentials $U(r)$ and $V(r)$.

With the energy contributions w_i from each electron depending on the principal and orbital quantum numbers (n_i, ℓ_i), the total energy of the atom in the central field approximation is determined by prescribing the list of these quantum numbers for the N electrons involved,

$$n_1\ell_1\, n_2\ell_2 \cdots n_N\ell_N.$$

We shall refer to this list as the *electron configuration* of the atom. Electrons having identical principal quantum numbers n_i are said to belong to the same electron *shell*. Different shells are denoted by capital letters K, L, M, N, $\cdots$. Electrons having identical principal and orbital angular quantum numbers n_i and ℓ_i are said to belong to the same *subshell*. The ℓ_i-values for different subshells are indicated by letters s, p, d, f, $\cdots$. Shell and subshell nomenclature are illustrated in table 4.1.

The solution of the N electron central field problem as outlined above suffers from one defect. It does not comply with the Pauli principle. In particular, this principle will limit the maximum number of electrons allowed in each shell and subshell. The maximum number of electrons in a given subshell is $2(2\ell+1)$. A given shell may hold $2n^2$ electrons. In table 4.1 these maximum numbers are indicated. The apparent Pauli principle defect of the central field approximation is rectified by expressing the solution as a fully anti-symmetric sum of products of one-electron wave functions in the form of a *Slater determinant*

$$|\,\Psi(1,2,\cdots,N)\,\rangle = \frac{1}{\sqrt{N!}} \begin{vmatrix} |\,\psi_{\sigma_1}(1)\,\rangle & \cdots & |\,\psi_{\sigma_1}(N)\,\rangle \\ \vdots & & \vdots \\ |\,\psi_{\sigma_N}(1)\,\rangle & \cdots & |\,\psi_{\sigma_N}(N)\,\rangle \end{vmatrix}. \tag{4.11}$$

Shell	n	ℓ	Designation of electrons	Electrons in subshell	Electrons in shell
K	1	0	1s	2	2
L	2	0	2s	2	8
		1	2p	6	
M	3	0	3s	2	18
		1	3p	6	
		2	3d	10	
N	4	0	4s	2	32
		1	4p	6	
		2	4d	10	
		3	4f	14	

Table 4.1: Electron shells and subshells

Here σ in the notation $|\,\psi_\sigma(i)\,\rangle$ denotes the complete set of quantum numbers for the one-electron wave function for electron number i. We notice that the set of quantum numbers of the one-electron wave functions are identical for each line of the Slater determinant, but different from line to line. The set of quantum numbers for the different lines of the Slater determinant must also be in accordance with the chosen electron configuration of the atom.

A wave function in the form of a Slater determinant (4.11) automatically satisfies the Pauli principle. If two electrons were to occupy the same state, two lines in the determinant would be identical and the determinant therefore vanish. Similarly, an interchange of electrons i and j corresponds to the permutation of the corresponding columns in the determinant. This would mean that the total wave function changes sign, again in accordance with the Pauli principle.

The factor $1/\sqrt{N!}$ in the definition of the Slater determinant (4.11) ensures the normalization of the many-electron wave function, given that each of the one-electron wave functions are orthonormal. In fact, the number of terms of the determinant is $N!$, these terms being mutually orthogonal since in any product of different such terms, the set of quantum numbers σ for the different electrons will be different.

We shall return with additional comments on the Slater determinant and its properties in section 4.3.

Quiz 4.1 : Make use of the fact (see quiz 3.10) that $P_\ell^{|m_\ell|}(\cos\theta)\exp(\imath m_\ell\varphi)$ is an even or odd function under the transformation $\boldsymbol{r}\to-\boldsymbol{r}$ if ℓ is an even or odd integer to show that $|\,\Psi(1,2,\cdots,N)\,\rangle$ as given by (4.11), has even or odd parity if $\ell_1+\ell_2+\cdots+\ell_N$ is an even or odd integer. Here $\ell_1,\ell_2,\cdots,\ell_N$ are the orbital angular quantum numbers of the different one-electron wave functions $|\,\psi(i)\,\rangle$.

Quiz 4.2 : Verify that the Slater determinant (4.11) is an eigenfunction of H_c as given by (4.7) with eigenvalue $W_c=\sum_i w_i$.

4.3 Angular Momenta and their Summation

The individual orbital and spin angular momenta $\boldsymbol{L}_i$ and $\boldsymbol{S}_i$ of the electrons in the central field approximation are all quantized in the familiar manner, (3.53), (3.54), (3.71) and (3.72).

The total orbital and total spin angular momenta $\boldsymbol{L} = \sum_i \boldsymbol{L}_i$ and $\boldsymbol{S} = \sum_i \boldsymbol{S}_i$ commute with the central field Hamiltonian H_c,

$$[H_c, \boldsymbol{L}] = 0 \quad \text{and} \quad [H_c, \boldsymbol{S}] = 0.$$

The quantities $\boldsymbol{L}^2$, L_z, $\boldsymbol{S}^2$ and S_z also commute among themselves and may therefore be quantized simultaneously with H_c,

$$\boldsymbol{L}^2 \,|\, \Psi(1,2,\cdots,N) \,\rangle = \hbar^2 L(L+1) \,|\, \Psi(1,2,\cdots,N) \,\rangle \tag{4.12}$$

$$L_z \,|\, \Psi(1,2,\cdots,N) \,\rangle = \hbar M_L \,|\, \Psi(1,2,\cdots,N) \,\rangle \tag{4.13}$$

$$\boldsymbol{S}^2 \,|\, \Psi(1,2,\cdots,N) \,\rangle = \hbar^2 S(S+1) \,|\, \Psi(1,2,\cdots,N) \,\rangle \tag{4.14}$$

$$S_z \,|\, \Psi(1,2,\cdots,N) \,\rangle = \hbar M_S \,|\, \Psi(1,2,\cdots,N) \,\rangle. \tag{4.15}$$

The quantum numbers L and $M_L = -L, -L+1, \cdots, L$ are integer valued. The maximum value of L is $\ell_1 + \ell_2 + \cdots + \ell_N$. Similar rules apply for the spin quantum numbers S and M_S, but these are integer or half-integer valued for even or odd numbers N of electrons in the atom.

The Slater determinant (4.11) is automatically an eigenfunction of the linear operators $L_z = \sum_i L_{zi}$ and $S_z = \sum_i S_{zi}$ with eigenvalues $\hbar M_L = \hbar \sum_i m_{\ell i}$ and $\hbar M_S = \hbar \sum_i m_{si}$. A single Slater determinant is, however, not automatically an eigenfunction of $\boldsymbol{L}^2$ and $\boldsymbol{S}^2$. In order to comply with the requirements set by the quantization of the total orbital and total spin angular momenta, it may thus be necessary to consider eigenfunctions $|\, \Psi(1,\cdots,N) \,\rangle$ in the form of linear combinations of Slater determinants.

In analogy with the results for the one-electron atom, we may also choose to work in terms of eigenfunctions allowing for the alternative simultaneous quantization of $\boldsymbol{L}^2$, $\boldsymbol{S}^2$, $\boldsymbol{J}^2$ and J_z where

$$\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S} \tag{4.16}$$

is the total angular momentum of the electrons. With this choice the quantization rules represented by the relations (4.13) and (4.15) are to be replaced by

$$\boldsymbol{J}^2 \,|\, \Psi(1,2,\cdots,N) \,\rangle = \hbar^2 J(J+1) \,|\, \Psi(1,2,\cdots,N) \,\rangle \tag{4.17}$$

$$J_z \,|\, \Psi(1,2,\cdots,N) \,\rangle = \hbar M_J \,|\, \Psi(1,2,\cdots,N) \,\rangle. \tag{4.18}$$

The corresponding integer or half-integer valued quantum numbers J and M_J satisfy

$$J = |\, L-S \,|,\, |\, L-S \,| + 1, \cdots, L+S \tag{4.19}$$

$$M_J = -J, -J+1, \cdots, J.$$

In the following we shall refer to the eigenfunctions $|\, 1,2,\cdots,N \,\rangle$ of the N electron Schrödinger equation when quantized simultaneously with $\boldsymbol{L}^2$, L_z, $\boldsymbol{S}^2$, S_z or alternatively $\boldsymbol{L}^2$, $\boldsymbol{S}^2$, $\boldsymbol{J}^2$, J_z as $|\, LM_LSM_S \,\rangle$ or $|\, LSJM_J \,\rangle$, respectively. Any eigenfunction of the $LSJM_J$ type may be expressed as a linear combination of eigenfunctions of LM_LSM_S type. The corresponding one-electron case was discussed in section 3.7.

Quiz 4.3 : Verify that the Slater determinant (4.11) is an eigenfunction of L_z and S_z with eigenvalues $\hbar M_L = \hbar \sum_i m_{\ell i}$ and $\hbar M_S = \hbar \sum_i m_{si}$.

Quiz 4.4 : For a two-electron atom show that $\boldsymbol{L}^2 = \boldsymbol{L}_1^2 + \boldsymbol{L}_2^2 + 2L_{z1}L_{z2} + L_{+1}L_{-2} + L_{-1}L_{+2}$ and similarly for $\boldsymbol{S}^2$. [*Hint:* Compare with (3.80). Properties of the $L_\pm$ and $S_\pm$ operators are given in (3.59) and (3.79)]

Quiz 4.5 : For a two-electron atom show that $(\alpha_1\beta_2-\beta_1\alpha_2)/\sqrt{2}$, $\beta_1\beta_2$, $(\alpha_1\beta_2+\beta_1\alpha_2)/\sqrt{2}$, and $\alpha_1\alpha_2$ are all eigenfunctions of $\boldsymbol{S}^2$ with eigenvalues $\hbar^2 S(S+1)$, the first one with $S = 0$, the three latter ones with $S = 1$. What are the corresponding eigenvalues of S_z? [*Hint:* Make use of the results of quiz 4.4]

Quiz 4.6 : The ground state of a two-electron atom has the electron configuration $1s^2$ and two available spin-orbitals $\alpha\,\psi_{100}(\boldsymbol{r})$ and $\beta\,\psi_{100}(\boldsymbol{r})$. Show that the Slater determinant is given by $(\alpha_1\beta_2 - \beta_1\alpha_2)/\sqrt{2}\,\psi_{100}(1)\psi_{100}(2)$, and that it is an eigenfunction of the angular momentum operators $\boldsymbol{L}^2$, L_z, $\boldsymbol{S}^2$ and S_z. What are the eigenvalues?

Quiz 4.7 : The lowest excited state of the two-electron atom has the electron configuration 1s2s and available spin-orbitals $\alpha\,\psi_{100}(\boldsymbol{r})$, $\beta\,\psi_{100}(\boldsymbol{r})$, $\alpha\,\psi_{200}(\boldsymbol{r})$, and $\beta\,\psi_{200}(\boldsymbol{r})$. Show that two of the four Slater determinants possible, namely

$$\alpha_1\alpha_2\,(\psi_{100}(1)\psi_{200}(2) - \psi_{200}(1)\psi_{100}(2))/\sqrt{2} \quad \text{and}$$
$$\beta_1\beta_2\,(\psi_{100}(1)\psi_{200}(2) - \psi_{200}(1)\psi_{100}(2))/\sqrt{2},$$

both are eigenfunctions of the angular momentum operators $\boldsymbol{L}^2$, L_z, $\boldsymbol{S}^2$ and S_z. What are the corresponding eigenvalues? Show that linear combinations of the two remaining Slater determinants are necessary in order to produce two additional eigenfunctions. What are the correct form of these eigenfunctions and the corresponding eigenvalues? [*Hint:* Make use of the results of quiz 4.5]

4.4 Electron Correlation and Fine Structure Effects

The central field approximation of section 4.2 with the simplified Hamiltonian $H = H_c$ represents a zero order theory. For a more detailed description of the properties of many-electron atoms, additional effects represented by extra terms in the Hamiltonian must be taken into account. The electron correlation term H_{EC} is one such effect. Another internal correctional term is the spin-orbit interaction term

$$H_{LS} = \frac{1}{2m^2c^2}\sum_{i=1}^{N}\frac{1}{r_i}\frac{\mathrm{d}V(r_i)}{\mathrm{d}r_i}\boldsymbol{L}_i\cdot\boldsymbol{S}_i. \tag{4.20}$$

The effect of these additional terms on the energy levels of the atom can be investigated using standard perturbation theory as summarized in table 3.9. How this perturbation calculation is carried out depends on the relative importance of the correctional terms. The situation resembles the discussion of the competing fine structure and Zeeman effects in section 3.13. For the present problem two limiting cases may be identified. The case most often encountered is where H_{EC} dominates H_{LS}. This case occurs for small and intermediate values of the charge number Z. We shall refer to this case as *L-S coupling* case. The opposite case, where H_{LS} dominates H_{EC}, arises for large Z and is known as *j-j coupling* case. This particular dependence on the charge number Z is expected on the basis of our discussion of the fine structure effect for one-electron atoms in chapter 3. Here we found a Z^4 dependence

for the fine structure effect. From the form of the H_{EC} term, a weaker Z dependence is expected for the electron correlation effect.

As the notation indicates, in the former case we may consider the addition of angular momenta to proceed by first forming the total orbital and total spin angular momenta with quantum numbers L and S, and then add these to find the total angular momentum with quantum number J. In the latter case, the individual orbital and spin angular momenta first add to form individual total angular momenta with quantum numbers j, these then add to form the total angular momentum with quantum number J. We shall only consider the L-S coupling case in the following.

The first step of a perturbation calculation for the L-S coupling case consists in obtaining approximate eigenfunctions and eigenvalues for the Hamiltonian $H = H_c + H_{EC}$. The energy levels W_c of the zero order central field approximation are degenerate since they are independent of the values of $m_{\ell i}$ and m_{si} for the individual electrons. Finding the energy levels for the Hamiltonian $H = H_c + H_{EC}$ from a given unperturbed energy level W_c thus involves the diagonalization of H_{EC} with respect to the subspace of degenerate states belonging to the eigenvalue W_c.

The perturbed eigenvalues W are characterized by the total orbital and the total spin quantum numbers L and S. This follows from the fact that the electron correlation term H_{EC} commutes with both $\boldsymbol{J}$ and $\boldsymbol{S}$ and therefore also with $\boldsymbol{L}$. The eigenvalue W must, however, be independent of the azimuthal quantum numbers $M_L = \sum_i m_{\ell i}$ and $M_S = \sum_i m_{si}$. This is a consequence of the fact that the energy levels of the atom in the absence of external fields must be independent of orientations.

It has been found empirically that this energy dependence on L and S is normally correctly described by the *Hund rules*:

- *For a given electron configuration, the state with the larger total spin quantum number S has the lower energy.*
- *For a given electron configuration and a given S, the state with the larger total orbital quantum number L has the lower energy.*

The last step of the perturbation calculation involves the evaluation of the energy contribution from the spin-orbit term H_{LS}, making use of the approximate eigenfunctions for the Hamiltonian (4.6). H_{LS} does not commute with $\boldsymbol{L}$ or $\boldsymbol{S}$. It is therefore advantageous to shift from wave functions of type $|\, LM_LSM_S \,\rangle$ of the Hamiltonian $H_c + H_{EC}$ to wave functions of type $|\, LSJM_J \,\rangle$. Indeed, it can be shown that

$$\langle\, LSJM_J |\, H_{LS} \,|\, LSJM_J \,\rangle = \frac{1}{2} A\,[J(J+1) - L(L+1) - S(S+1)]$$

where A is a constant depending on the electron configuration of the atom. We thus see that the spin-orbit interaction introduces a J dependent correctional term for the energy levels. Keeping the electron configuration and the values of L and S constant, the energy difference between energy levels corresponding to adjacent J values may therefore be expressed as

$$W(J) - W(J-1) = \frac{1}{2} A\,[J(J+1) - (J-1)J] = AJ. \tag{4.21}$$

The result 4.21 is known as the *Landé interval rule*. It has been established empirically for atoms with one incompletely filled subshell that $A > 0$ if the incompletely filled subshell is less

than half filled, and that $A < 0$ if this subshell is more than half filled. In the former case the state with the lower total angular quantum number J has the lower energy (*normal multiplet*), in the latter case the state with the larger J has the lower energy (*inverted multiplet*):

- *For an electron configuration with one incompletely filled subshell and given total S and L quantum numbers, the state with the larger total angular quantum number J has the higher (lower) energy if the incompletely filled subshell is less (more) than half filled.*

We illustrate the above results with two examples. Consider the two-electron configuration $n\mathrm{p}\,n'\mathrm{p}$. With different principal quantum numbers n and n', the Pauli principle is satisfied whatever the choice of the remaining individual quantum numbers. With 6 different placements of one electron in each of the two p-subshells, the given electron configuration will support 36 different atomic states. Addition of the individual orbital and spin angular momenta shows that the possible values for L and S are

$$L = 0, 1, 2$$
$$S = 0, 1$$

and therefore

$$J = 0, 1, 2, 3.$$

The energy levels for this configuration in the lowest order central field approximation, in the central field approximation with electron correlation effects, and in the central field approximation with both electron correlation and fine structure effects included, are illustrated schematically in figure 4.2. The figure does not pretend to represent distances between energy levels correctly, only the relative ordering of the levels.

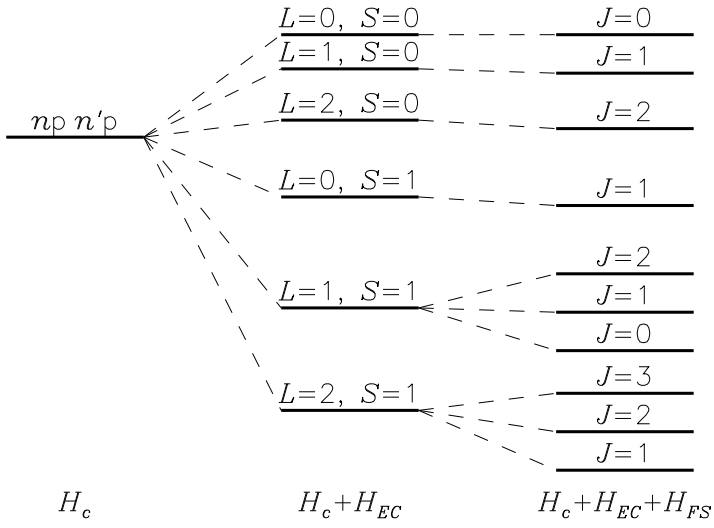

Figure 4.2: Splitting of the $n\mathrm{p}\,n'\mathrm{p}$ configuration by first order effects

As a second example consider the electron configuration $1s^2 2s^2 2p^2$ where the exponents indicate the number of electrons in the different subshells. This configuration will include the

ground state of the carbon atom. The electrons in the two filled s-subshells do not contribute to L or S. This conclusion is valid for the electrons of any filled subshell. The conclusion can be understood by noting that the sum of the individual azimuthal quantum numbers for the electrons in the filled subshell, $m_\ell = 0, \pm 1, \cdots, \pm\ell$, does not give any net contribution to M_L, and therefore no net contribution to L. The similar argument applies for spin. We thus conclude that:

- *The electrons of any filled subshell give no net contribution to the values of L, S or J.*

Possible values for L and S resulting from the 2p-electrons are $L = 0, 1, 2$ and $S = 0, 1$. However, only selected combinations of these values are allowed according to the Pauli principle. In fact, of the 36 different combinations of the two pairs of (m_ℓ, m_s) quantum numbers, only $6 \cdot 5/2 = 15$ are allowed. The first electron may choose freely among the 6 different one-electron wave functions available in the np-subshell, the next electron have only 5 different choices left, and we must divide by 2 because the electrons are indistinguishable.

To select combinations of L and S values allowed by the Pauli principle, the following observations are useful. First, the total number of different quantum states resulting from the allowed set of (L, S) values, should be in accordance with the number of different states based on the (m_ℓ, m_s) type of reasoning. Second, the L_z and S_z operators are linear combinations of the corresponding individual operators, $M_L = \sum_{i=1}^{N} m_{\ell i}$ and $M_S = \sum_{i=1}^{N} m_{si}$. The maximum value for M_L or M_S allowed by the Pauli principle will also be the maximum allowed value for L or S. At the same time, for a given allowed (L, S) pair, every combination of values of $M_L \in [-L, \cdots, L]$ and $M_S \in [-S, \cdots, S]$ shall be accounted for. A given allowed combination of M_L, M_S values thus only indicates that the corresponding allowed (L, S) pair will satisfy $L \geq \mid M_L \mid$, $S \geq \mid M_S \mid$. Valid strategies for identifying two allowed (L, S) pairs can now be formulated: 1) Find the maximum allowed M_S value and the corresponding maximum value of M_L. 2) Find the maximum value of M_L and the corresponding maximum allowed value of M_S. These combinations are both allowed combinations of L and S. Remaining allowed combinations of L and S values can often be identified on the basis of the number of quantum states now not accounted for.

In the present case the largest possible value of $\sum m_\ell$ is 2. This value will require anti-parallel spins, $\sum m_s = 0$. Thus, $(L = 2, S = 0)$ is the only allowed combination with $L = 2$. Parallel spins, $\sum m_s = 1$ will require different m_ℓ values, that is, the maximum allowed value of $\sum m_\ell$ is 1. Therefore, $(L = 1, S = 1)$ is an allowed combination. Together with $(L = 0, S = 0)$ these combinations represent all 15 allowed states. With the 2p-subshell less that half filled, the $L = 1$, $S = 1$, $J = 0$ state will have the lowest energy (normal multiplet). The resulting energy levels for the $1s^2 2s^2 2p^2$ configuration are illustrated schematically in figure 4.3. The figure only indicates the correct ordering of energy levels, not the correct distance between levels.

The electron configuration $1s^2 2s^2 2p^4$ will allow for the identical combinations of L, S and J. This is seen by noting that the problem of placing 4 electrons in the 6 possible states of the 2p-subshell according to the Pauli principle is equivalent to the problem of locating the two empty states. This time the 2p-subshell is more than half filled. The $L = 1$, $S = 1$, $J = 2$ state therefore now represents the lowest energy (inverted multiplet).

For the more general case the following observation is useful. The Pauli principle applies to electrons belonging to the any incompletely filled subshell (*equivalent electrons*). To find

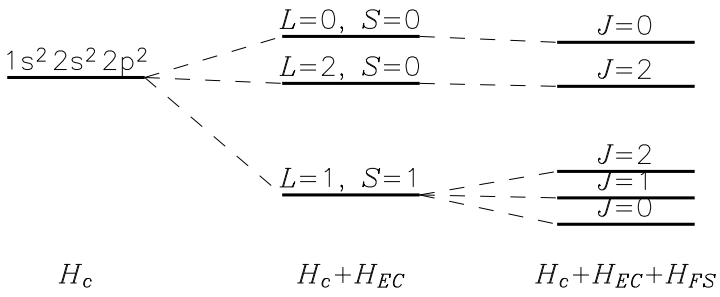

Figure 4.3: Splitting of the central field approximation energy level of the $1s^2 2s^2 2p^2$ configuration by electron correlation and fine structure effects

allowed combinations of the quantum numbers L and S for any given electron configuration (under L-S coupling) proceed by:

- *arrange electrons of incompletely filled subshells into groups of equivalent electrons and individual non-equivalent electrons,*
- *for each group of equivalent electrons apply the Pauli principle to find corresponding allowed combinations of L and S,*
- *combine groups of equivalent electrons or individual non-equivalent electrons one by one by applying the ordinary rules for the addition of angular momenta (unconstrained by the Pauli principle).*

We illustrate this strategy by considering the electron configuration $np^2 n'p$. The two np electrons form an equivalent group. Allowed combinations of quantum numbers L and S for this group were identified in the second example dicussed above. Adding the remaining $n'p$ electron to each of these combinations leads to the following end result:

$L = 2$, $S = 0$ gives $L = 1, 2, 3$ and $S = 1/2$.
$L = 1$, $S = 1$ gives $L = 0, 1, 2$ and $S = 1/2, 3/2$.
$L = 0$, $S = 0$ gives $L = 1$ and $S = 1/2$.

In table 4.2 a summary of allowed L, S combinations for electron configurations np^k and nd^k are given in terms of standard spectroscopic notation (see section 4.5). Note that the underlined spectral terms need to be counted multiple times in order to account for all possible atomic states.

Quiz 4.8 : Show that the number of allowed states for the two electron configurations np^2 and $np\,n'p$ with $n \neq n'$ can be accounted for in terms of the individual quantum numbers $n_i \ell_i m_{\ell i} s_i m_{si}$, in terms of the total quantum numbers $LM_L SM_S$ or in term of $LSJM_J$.

Electron configuration	Spectral terms
$n\mathrm{p}^2$, $n\mathrm{p}^4$	$^1\mathrm{S}$ $^1\mathrm{D}$ $^2\mathrm{P}$
$n\mathrm{p}^3$	$^2\mathrm{P}$ $^2\mathrm{D}$ $^4\mathrm{S}$
$n\mathrm{d}^2$, $n\mathrm{d}^8$	$^1\mathrm{S}$ $^1\mathrm{D}$ $^1\mathrm{G}$ $^3\mathrm{P}$ $^3\mathrm{F}$
$n\mathrm{d}^3$, $n\mathrm{d}^7$	$^2\mathrm{P}$ $^2\underline{\mathrm{D}}$ $^2\mathrm{F}$ $^2\mathrm{G}$ $^2\mathrm{H}$ $^4\mathrm{P}$ $^4\mathrm{F}$
$n\mathrm{d}^4$, $n\mathrm{d}^6$	$^1\underline{\mathrm{S}}$ $^1\underline{\mathrm{D}}$ $^1\mathrm{F}$ $^1\underline{\mathrm{G}}$ $^1\mathrm{I}$ $^3\underline{\mathrm{P}}$ $^3\mathrm{D}$ $^3\underline{\mathrm{F}}$ $^3\mathrm{G}$ $^3\mathrm{H}$ $^5\mathrm{D}$
$n\mathrm{d}^5$	$^2\mathrm{S}$ $^2\mathrm{P}$ $^2\underline{\mathrm{D}}$ $^2\underline{\mathrm{F}}$ $^2\underline{\mathrm{G}}$ $^2\mathrm{H}$ $^2\mathrm{I}$ $^4\mathrm{P}$ $^4\mathrm{D}$ $^4\mathrm{F}$ $^4\mathrm{G}$ $^6\mathrm{S}$

Table 4.2: Allowed spetral terms for selected electron configurations

4.5 Spectroscopic Notation and the Periodic System

The spectroscopic notation for quantum states introduced in section 3.8 for one-electron atoms is easily extended to many-electron atoms in the case of L-S coupling. The complete notation

$$n_1\ell_1\, n_2\ell_2 \cdots n_N\ell_N\, {}^{2S+1}L_J^{o,e}$$

is seen to consist of two parts. The electron configuration part is the list of N individual principal and orbital quantum numbers n_i and ℓ_i for the individual electrons of the atom. The individual quantum numbers n_i in this list are represented by their numerical values, the ℓ_i quantum numbers are written with symbols s, p, d, f, $\cdots$ in accordance with table 3.7. The presence of several electrons in the same subshell is indicated by exponents. Thus, the notation $n\ell^k$ indicates that k electrons occupy the $n\ell$-subshell. The electron configuration list is sometimes compressed by suppressing the notation for electrons of filled subshells, that is, $1\mathrm{s}^2 2\mathrm{s}^2 2\mathrm{p}^2$ may be written compactly as $2\mathrm{p}^2$. The electron configuration reflects properties of the central field approximation part, H_c, of the total Hamilton operator.

The second part of the spectroscopic notation indicates the manner in which the individual angular momenta add to form the corresponding total angular momenta. The expression ${}^{2S+1}L$, where the total spin quantum number S is given in terms of its numerical value and the total orbital quantum number L is represented by capital symbols S, P, D, F, $\cdots$ again in accordance with table 3.7, represents a *spectral term* and reflects properties of the electron correlation part , H_{EC}, of the total Hamilton operator.

A *spectral level* is indicated by adding the value of J as an index to a spectral term, ${}^{2S+1}L_J$. The J-index reflects the splitting of the energy levels due to fine structure effects. A *multiplet* consists of all possible spectral levels for a given spectral term. The superscript $2S+1$ of the spectral term defines the *multiplicity* of the state. With $L > S$ the rules for the addition of angular momenta will allow for $2S+1$ different values for J, $J = |L-S|, |L-S|+1, \cdots, L+S$. The multiplicity in this case equals the number of members in a given multiplet. If $L < S$, however, only $2L+1$ different values for J are allowed. Also in this case the value $2S+1$ is referred to as the multiplicity of the state. Quantum states with $2S+1 = 1, 2, 3, \cdots$ are referred to as *singlet*, *doublet*, *triplet*, $\cdots$ states.

The second part of the spectroscopic notation is completed by adding the parity indicator as a superscript to the spectral level, o for odd parity states and e or, more commonly, no superscript for even parity states. The parity of the wave function is determined by the sum of the individual angular quantum numbers of the electrons in the atom, $\ell_1 + \ell_2 + \cdots + \ell_N$. The parity is odd or even as this sum is an odd or even integer number (see quiz 4.1).

Let us now see how the spectroscopic notation apply for the ground states of the elements

of the periodic system. The He I atom is the simplest many-electron atom. In the ground state, both electrons occupy the K-shell, that is, both electrons have $n = 1$ and $\ell = 0$. The electron spins must be anti-parallel in order to satisfy the Pauli principle. This means that the He I ground state is characterized by $L = 0$, $S = 0$ and $J = 0$ and is thus denoted $1s^2\ {}^1S_0$, or ${}^1S_0$ for short. The sum of the individual orbital quantum numbers is an even number, corresponding to even parity of the state.

		K	L		M			N	Ground State
Z	Element	1s	2s	2p	3s	3p	3d	4s	Spectral Level
1	H	1							${}^2S_{1/2}$
2	He	2							${}^1S_0$
3	Li	2	1						${}^2S_{1/2}$
4	Be	2	2						${}^1S_0$
5	B	2	2	1					${}^2P^o_{1/2}$
6	C	2	2	2					${}^3P_0$
7	N	2	2	3					${}^4S^o_{3/2}$
8	O	2	2	4					${}^3P_2$
9	F	2	2	5					${}^2P^o_{3/2}$
10	Ne	2	2	6					${}^1S_0$
11	Na	2	2	6	1				${}^2S_{1/2}$
12	Mg	2	2	6	2				${}^1S_0$
13	Al	2	2	6	2	1			${}^2P^o_{1/2}$
14	Si	2	2	6	2	2			${}^3P_0$
15	P	2	2	6	2	3			${}^4S^o_{3/2}$
16	S	2	2	6	2	4			${}^3P_2$
17	Cl	2	2	6	2	5			${}^2P^o_{3/2}$
18	Ar	2	2	6	2	6			${}^1S_0$
19	K	2	2	6	2	6		1	${}^2S_{1/2}$
20	Ca	2	2	6	2	6		2	${}^1S_0$
21	Sc	2	2	6	2	6	1	2	${}^2D_{3/2}$
22	Ti	2	2	6	2	6	2	2	${}^3F_2$
23	V	2	2	6	2	6	3	2	${}^4F_{3/2}$
24	Cr	2	2	6	2	6	5	1	${}^7S_3$
25	Mn	2	2	6	2	6	5	2	${}^6S_{5/2}$
26	Fe	2	2	6	2	6	6	2	${}^5D_4$
27	Co	2	2	6	2	6	7	2	${}^4F_{9/2}$
28	Ni	2	2	6	2	6	8	2	${}^3F_4$
29	Cu	2	2	6	2	6	10	1	${}^2S_{1/2}$
30	Zn	2	2	6	2	6	10	2	${}^1S_0$

Table 4.3: Electron configuration and spectral level of the ground state of selected elements

The electron configuration of the ground state of the other elements in the periodic system may be understood on the basis of simple energy arguments. The energy of an electron spin-orbital generally increases with increasing principal quantum number n. For a given principal quantum number the energy increases with increasing orbital quantum number ℓ. With

increasing atomic number Z, we thus expect the electrons to fill in the available shells in the order of increasing n and subshells in the order of increasing ℓ. The results for the lower atomic number elements are displayed in table 4.3. A few exceptions to the general rule stated above are readily visible. Potassium (K) reveals the first irregularity. A new shell (the N-shell) is started even though the 3d-subshell is not yet completed. For the transitions to chromium (Cr) and copper (Cu) the opposite effect is observed, two electrons move into the 3d-subshell, leaving only one electron in the 4s-subshell. Additional irregularities are found in the higher atomic number elements of less interest in the astrophysical context.

The ground state spectral levels, also given in table 4.3, may be understood with the help of the Pauli principle, the Hund rules and the Landé interval rule. That is, for a given electron configuration: 1) identify the states with the largest S allowed by the Pauli principle, 2) among these states look for the states with largest L again in accordance with the Pauli principle, and finally 3) among the remaining states choose the state with the smallest (largest) J for normal (inverted) multiplets.

Elements for which all occupied subshells are filled are singlet ${}^1\mathrm{S}_0$ states. Elements with one electron outside filled subshells form normal multiplets and are doublet ${}^2\mathrm{S}_{1/2}$, ${}^2\mathrm{P}^o_{1/2}$ or ${}^2\mathrm{D}_{3/2}$ states depending on the ℓ-value of this electron. Elements with $n\mathrm{p}^5$ configuration allow for identical L, S and J values as the $n\mathrm{p}^1$ configuration. These elements, however, form inverted multiplets. The ground state spectral level for the $n\mathrm{p}^5$ configuration is therefore ${}^2\mathrm{P}^o_{3/2}$. The $n\mathrm{p}^2$ configuration was discussed in section 4.4. Minimum energy occurs for the triplet ${}^3\mathrm{P}_0$ state. The corresponding $n\mathrm{p}^4$ configurations form inverted multiplets with the triplet ${}^3\mathrm{P}_2$ state having minimum energy.

For atomic states with more than one partially filled subshell containing together one or more equivalent groups of electrons, it is necessary to indicate the spectral terms for each such group in the spectral notation. This is done by including these spectral terms in parenthesis in the electron configuration list. As an example, the electron configuration $1\mathrm{s}^2 2\mathrm{s}^2 2\mathrm{p}^3 n\ell$ of O I has one equivalent group, $2\mathrm{p}^3$. This group may have spectral terms ${}^2\mathrm{P}$, ${}^2\mathrm{D}$ or ${}^4\mathrm{S}$, all with odd parity. One possible excited atomic state of O I is thus $1\mathrm{s}^2 2\mathrm{s}^2 2\mathrm{p}^3({}^4\mathrm{S}^o) n\mathrm{p}\,{}^2\mathrm{P}_J$ with $J = 1/2, 3/2$.

Before proceeding to a further discussion of energy levels of excited states and radiation transitions between different levels, it will be useful to review the selection rules for many-electron atoms.

Quiz 4.9 : Explain the ground state spectral levels of K I, Ca I, Sc I, N I, Ni I and Fe I. Also check the parity of these ground states.

Quiz 4.10 : An excited He I atom has one electron in the K-shell and one in the M-shell. What are the selections of quantum numbers possible for this atom? What are the corresponding spectroscopic notations?

Quiz 4.11 : Assuming L-S coupling, derive possible spectral terms ${}^{2S+1}L$ and spectral levels ${}^{2S+1}L_J$ for the electron configurations $n\mathrm{d}^2$, $n\mathrm{p}^3$ and $n\mathrm{p}^4$.

Quiz 4.12 : List spectral terms for the electron configurations $n\mathrm{p}\,n'\mathrm{p}\,n''\mathrm{p}$ and $n\mathrm{p}^2\,n'\mathrm{s}\,n'\mathrm{p}$ where n, n' and n'' are all different.

4.6 Summary of Selection Rules

Transition rates and selection rules for radiative transitions in one-electron atoms were discussed in sections 3.9 and 3.10. The ability of the atom to jump between different quantum states by absorbing or emitting photons was explained as a time-dependent interaction between the electric field $\boldsymbol{E}$ of the electromagnetic wave and the electric dipole moment $\boldsymbol{m}_E = -e\boldsymbol{r}$ of the atom. For many-electron atoms the only difference is that the electric field this time interacts with the total electric dipole moment resulting from N electrons, $\boldsymbol{m}_E = -e(\boldsymbol{r}_1 + \boldsymbol{r}_2 + \cdots + \boldsymbol{r}_N)$. The analysis is analogous to that of the one-electron atom and will not be repeated. Except for a few minor differences, the one-electron result is generalized by replacing the angular momenta of the single electron with the corresponding total angular momenta of the N electrons.

The transition rate for a transition between two given states can be written in the form of a series, with the first term (the electric dipole approximation) corresponding to neglecting the spatial variation of the wave electric field over the atom. The second term (the electric quadrupole approximation) takes into account first order effects due to this variation. There are also other minor contributions (magnetic dipole approximation) to the transition rate that sometimes need to be taken into account[1]. Normally, however, the electric dipole approximation represents the dominating contribution to transition rates. A given transition is thus called an allowed transition if the contribution to the transition rate from the electric dipole approximation term is non-vanishing.

In table 4.4 the selection rules for radiation transitions in many-electron atoms (assuming L-S coupling) for the electric dipole approximation as well as for the electric quadrupole and magnetic dipole approximations are listed. ΔL, ΔS and ΔJ represent the changes in the total orbital, total spin and total total angular momenta of the atom associated with the transition. The symbol $(0 \nleftrightarrow 0)$ appearing in the table, for instance in connection with the total orbital angular momentum requirement $\Delta L = 0, \pm 1$, means that transitions between states both with $L = 0$ are *not* to be included.

From the table it is seen that transitions between different multiplets, $\Delta S \neq 0$, cannot occur according to any of the approximations listed. This is a result of the fact that the interaction term of the Hamiltonian involving the electric field of the electromagnetic wave is independent of electron spin. If corresponding spectral lines are still found in the spectrum, so-called *inter-combination lines*, this indicates that collisions between atoms or between atoms and free electrons have been active in the formation of the line.

For allowed transitions the two electron states involved must exhibit different parities. In particular, if only one electron is involved in the transition, the change in the angular quantum number $\Delta\ell$ of this electron must be ± 1. If more than one electron participate in the given transition, then $\Delta L = 0, \pm 1$ except that $L = 0 \rightarrow L = 0$ is not allowed, and *in addition* the sum $\sum_i \ell_i$ must change by an odd number. In contrast, electric quadrupole and magnetic dipole effects contribute to the transition rate only if the parities of the states involved are identical.

With the selection rules at hand, we are now ready to return to a discussion of excited states of some selected many-electron atoms and their transitions.

[1] The magnetic dipole contribution arises from the interaction of the magnetic moment of the electrons with the magnetic field of the electromagnetic wave.

Electric dipole (allowed)	Electric quadrupole (forbidden)	Magnetic dipole (forbidden)
$\Delta S = 0$	$\Delta S = 0$	$\Delta S = 0$
Parity change	No parity change	No parity change
$\Delta L = 0, \pm 1$ ($0 \nleftrightarrow 0$)	$\Delta L = 0, \pm 1, \pm 2$ ($0 \nleftrightarrow 0$, $0 \nleftrightarrow 1$)	$\Delta L = 0$
$\Delta J = 0, \pm 1$ ($0 \nleftrightarrow 0$)	$\Delta J = 0, \pm 1, \pm 2$ ($0 \nleftrightarrow 0$, $\frac{1}{2} \nleftrightarrow \frac{1}{2}$, $0 \nleftrightarrow 1$)	$\Delta J = 0, \pm 1$ ($0 \nleftrightarrow 0$)
$\Delta M_J = 0, \pm 1$	$\Delta M_J = 0, \pm 1, \pm 2$	$\Delta M_J = 0, \pm 1$
One electron jump $\Delta \ell = \pm 1$	One or no electron jump $\Delta \ell = 0, \pm 2$	No electron jump $\Delta n = 0$, $\Delta \ell = 0$

Table 4.4: Summary of selection rules for radiation transitions

4.7 Alkali Atoms

In the ground state the alkali atoms (Li I, Na I, K I, $\cdots$) have one (valence) electron in the outermost occupied shell in addition to one or more filled inner shells or subshells. The same fact applies to a number of ionized species (Be II, B III, C IV, Mg II, Ca II, $\cdots$). In figure 4.4 energy levels are given for some of these elements for electron configurations in which only the electron outside of filled subshells is excited. The energy levels have been marked with the $n\ell$-value of the excited electron. Energy levels are given relative to that of the ground state. From the figure we see as a general trend the increase in the energy levels with increasing principal quantum number n of the excited electron. Similarly, there is an increase in energy with increasing orbital quantum number ℓ within each subshell.

The energy levels in figure 4.4 have been ordered according to spectral term. Energy level diagrams ordered according to spectral term and with allowed (and forbidden) transitions and their corresponding wavelengths indicated are called *Grotrian diagrams*. The alkali atoms with their $n\ell$ electron configuration always form doublet states. This means that the energy levels displayed in figure 4.4, except for the S-term levels, will split as a result of the spin-orbit interaction. For practical reasons, this splitting has not been given in the figure, nor has the corresponding splitting of spectral lines been indicated. For the elements displayed in figure 4.4, the maximum splitting of 223 cm^{-1} occurs for the $4\mathrm{p}\,{}^2\mathrm{P}^o$ states of Ca II. The wavelengths given represent the "average" wavelength for the corresponding sets of individual spectral lines.

In the ground state, Li I has two electrons filling the K-shell, the last electron occupying the L-shell. The ground state is therefore $1\mathrm{s}^2 2\mathrm{s}\,{}^2\mathrm{S}_{1/2}$ (or $2\mathrm{s}\,{}^2\mathrm{S}_{1/2}$). The lowest excited states are $2\mathrm{p}\,{}^2\mathrm{P}^o_{1/2,3/2}$ Transitions between the ground state and these two excited states are both allowed. The corresponding resonance line is a doublet at $\lambda 6707$ with a doublet separation of

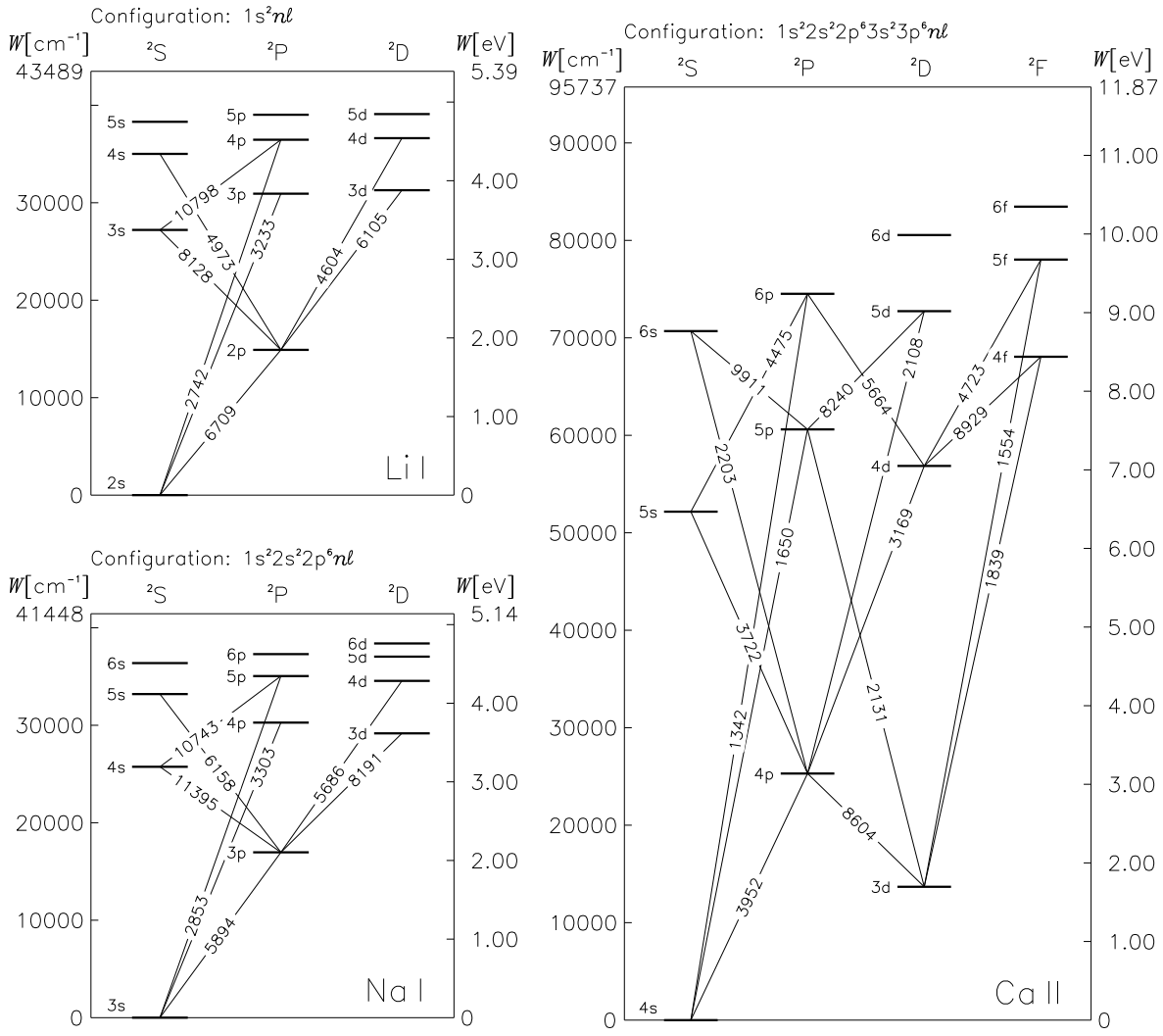

Figure 4.4: Energy levels for alkali elements

0.151Å due to the spin-orbit splitting of the two P-states.

Two isotopes of lithium are common, ^{6}Li and ^{7}Li. The resonance lines of these isotopes will be slightly different due to the different nuclear masses. For ^{6}Li the components λ6707.921 and λ6708.072 are found while for ^{7}Li the corresponding wavelengths are λ6707.761 and λ6707.912. In stellar spectra these components will not be separately resolved because of line broadening. By determining the center of the resonance line it is, however, still possible to determine the relative abundance of these two isotopes.

For Na I in the ground state the K- and L-shells are filled with the valence electron occupying the M-shell. The ground state is therefore $1s^2 2s^2 2p^6 3s\,^2S_{1/2}$ (or $3s\,^2S_{1/2}$). The resonance line results from transitions between the ground state and the lowest excited states $3p\,^2P^o_{1/2,3/2}$. The resonance line of Na I is again a doublet with wavelengths λ5895.9 (D_1) and λ5889.9 (D_2) and a doublet separation of 6 Å. The Na I resonance line is found both as emission lines in stellar spectra and as weak absorption lines due to the interstellar medium.

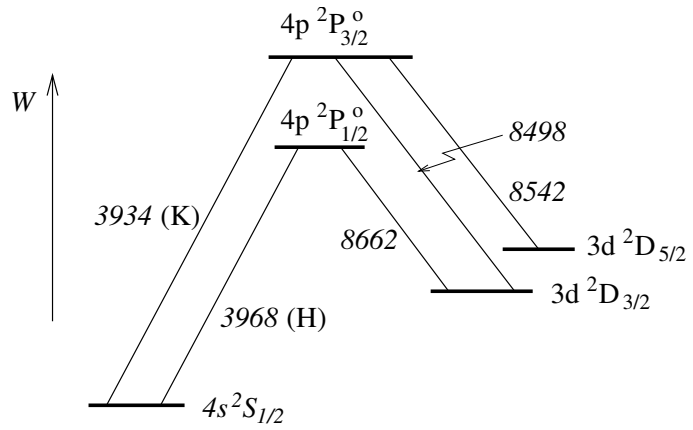

Figure 4.5: Details of the lowest energy levels for Ca II

Let us conclude with some comments on the Ca II resonance lines. These are often important components of astronomical spectra. The ground state of Ca II is $4s\,^2S_{1/2}$. Transitions to the excited states $4p\,^2P^o_{1/2,3/2}$ result in the lines λ3934 (K-line) and λ3968 (H-line). The doublet separation is thus 35 Å. Details of the transitions among the three lowest sets of energy levels of Ca II are illustrated in figure 4.5. In particular, the H- and K-lines are indicated. In addition we see that the λ8604 line in figure 4.4c in fact consists of three individual lines at λ8498, λ8542 and λ8662. For purpose of illustration, the energy separations of the $4p\,^2P$ and $3d\,^2D$ multiplets have been strongly exaggerated in the figure, the real energy separations being comparable to the line thickness. K I has similar electron configurations as Ca II, but since the nuclear charge is less, the ionization energy of K I is only 4.34 eV compared to 11.87 eV for Ca II. This means that energy levels of K I are generally different from the corresponding ones for Ca II. In particular, the relative ordering of np and (n-1)d levels in K I is opposite from that of Ca II.

4.8 Helium and the Alkaline Earths

The ground state of He I and the alkaline Earth elements Be I, Mg I, Ca I, $\cdots$ are all characterized by having a filled s-subshell as the outermost occupied subshell. The same applies for a number of ionized elements Li II, C III, N IV, $\cdots$. We will recognize similarities in energy levels of these elements.

If one of the electrons of He I occupies an excited state, the electron spins may be antiparallel, $S = 0$, or parallel, $S = 1$. This means that the excited states of He I will be singlet or triplet states. For the case that the excited electron occupies the L-shell, the different states possible are

$$1s2s\,^1S_0, \quad 1s2p\,^1P^o_1$$

or

$$1s2s\,^3S_1, \quad 1s2p\,^3P^o_{0,1,2}.$$

The singlet and triplet states of He I are also called para-helium and ortho-helium, respectively.

In figure 4.6 the energies of some of the lower excited states of He I are given relative to the energy of the ground state. Only energy levels for states with one electron remaining in the

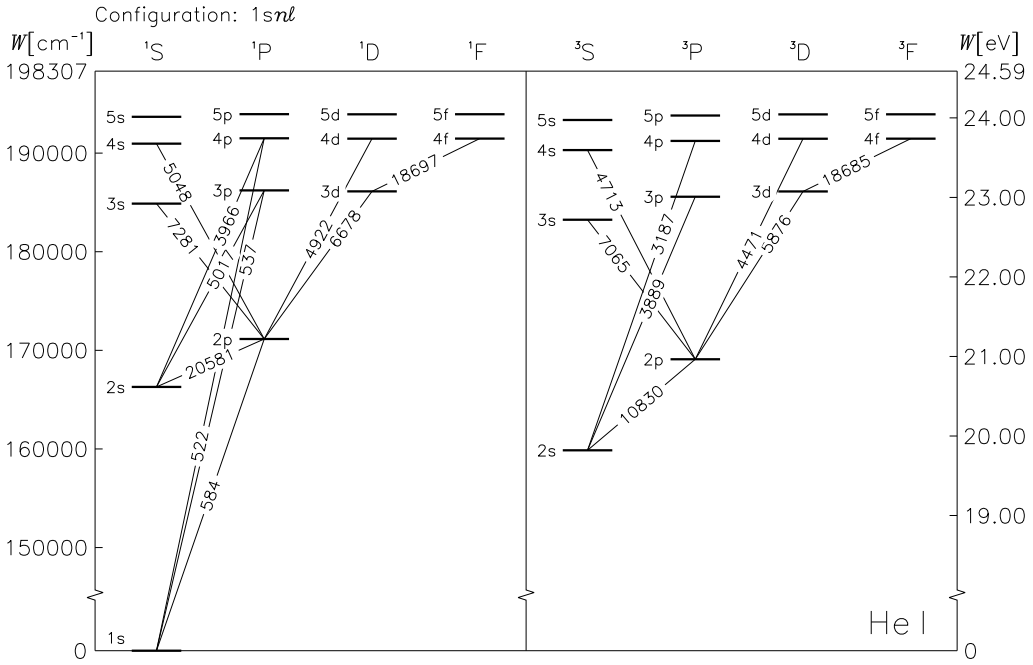

Figure 4.6: Energy levels of He I with allowed transitions.

K-shell have been included. The energy levels have accordingly been labeled by the $n\ell$-value of the excited electron. For practical reasons the splitting of the triplet energy levels due to spin-orbit interactions have not been indicated. For instance, the splitting of the 1s2p $^3\mathrm{P}^o_{0,1,2}$ multiplet is only 1 cm^{-1}, well within the line thickness in the figure.

Comparing electron states with identical principal quantum numbers n for the excited electron, the corresponding energy levels are seen to increase monotonously with increasing values of ℓ. This is in accordance with the discussion of the radial part of the one-electron eigenfunctions in the central field approximation in section 4.2. A larger value of ℓ means that the excited electron on the average will be further away from the nucleus, experiencing a better shielding of the nuclear charge due to the K-shell electron and thus being less strongly bound to the nucleus. This is in contrast to the one-electron atom where the energy levels to lowest order only depend on the principal quantum number n. The first of the two Hund rules is seen to be satisfied. The energy levels of the triplet states are all lower than their singlet counterparts. The second of the Hund rules does not apply in this case because for a given configuration and a given multiplicity, only one value of L is possible for the states considered in figure 4.6.

A transition between the ground state and the two lowest excited energy states 1s2s $^1\mathrm{S}_0$ and 1s2s $^3\mathrm{S}_1$ are both forbidden. These states are therefore meta-stable states with lifetimes of .02 s and 9000 s, respectively. The resonance line of He I at λ584.4 corresponds to the transition between the ground state and the singlet 1s2p $^1\mathrm{P}^o_1$ state.

Transitions between singlet and triplet states are not allowed, nor have such transitions been found in observed spectra. Thus, we get two separate series of spectral lines from He I, one for transitions among singlet states (para-helium) and one for transitions among triplet states (ortho-helium). In the figure, allowed transitions and corresponding wavelengths are given. In the visible part of the spectrum the most important He I lines involve transitions

between singlet or triplet 1s2p states and the corresponding higher 1sns and 1snd states.

In the examples considered in sections 4.7 and 4.8, it is one electron, namely the weakest bound electron, that has been lifted to an excited energy level. The smallest amount of added energy that is needed in order to remove this electron from the atom is per definition the ionization energy of the atom. It is also possible to lift one of the more strongly bound electrons to an excited energy level. In this case, the excited atom may well end up with an energy that exceeds the ionization energy of the atom. This situation may lead to a transition in which the more strongly bound electron state is refilled and the excess energy given to one of the more weaker bound electrons and which is thus given sufficient energy to escape the atom. This process is called an *autoionization* transition. Only a few examples of atoms with excitation energies exceeding the ionization energy can be given here. The Helium I state $2s^2\,{}^1S$ has the energy 57.870 eV while the ionization energy is 24.587 eV. The Li I state $1s2s2p\,{}^4P^o$ has energy 57.469 eV while the ionization energy is 5.392 eV. For the O I multiplet $1s^22s^22p^3({}^2P^o)3s\,{}^3P^o$ the energy is 14.124 eV while the ionization energy is 13.618 eV.

4.9 Effects of External Fields

We shall conclude our discussion of properties of many-electron atoms by considering the effects of external magnetic and electric fields on the energy levels and therefore on the spectral lines formed.

4.9.1 The Zeeman effect

We have seen that an external magnetic field gave rise to splitting of the energy levels of the one-electron atom. We find the similar effect for many-electron atoms. The interaction term of the Hamiltonian takes the identical form (3.140) in these two cases

$$H_B = \frac{e}{2m}\boldsymbol{B}\cdot(\boldsymbol{L}+2\boldsymbol{S}) = \frac{e}{2m}\boldsymbol{B}\cdot(\boldsymbol{J}+\boldsymbol{S}),$$

except that $\boldsymbol{L}$, $\boldsymbol{S}$ and $\boldsymbol{J}$ now represent the total angular momenta of all the electrons of the atom. This time we shall only consider the case of "weak" external fields, that is, we limit our discussion to situations where the effect of the external magnetic field is small compared to the splitting introduced by the spin-orbit interaction. This means that we have to evaluate the average value of H_B for eigenfunctions $|\,LSJM_J\,\rangle$. The calculation is completely analogous to that of section 3.13. Since the component of the total angular momentum $\boldsymbol{J}$ along the external magnetic field will be quantized according to (4.18), the external magnetic field will split each energy level into $2J+1$ equally spaced levels according to

$$\Delta W_B = \hbar\omega_L g M_J. \tag{4.22}$$

Here $\omega_L = eB/2m$ is the angular Larmor frequency. The Landé g-factor,

$$g = 1 + \frac{J(J+1)-L(L+1)+S(S+1)}{2J(J+1)}, \tag{4.23}$$

is now expressed in terms of the total orbital, the total spin and the total angular quantum numbers L, S and J. We note that for $J=0$ states no magnetic splitting occurs.

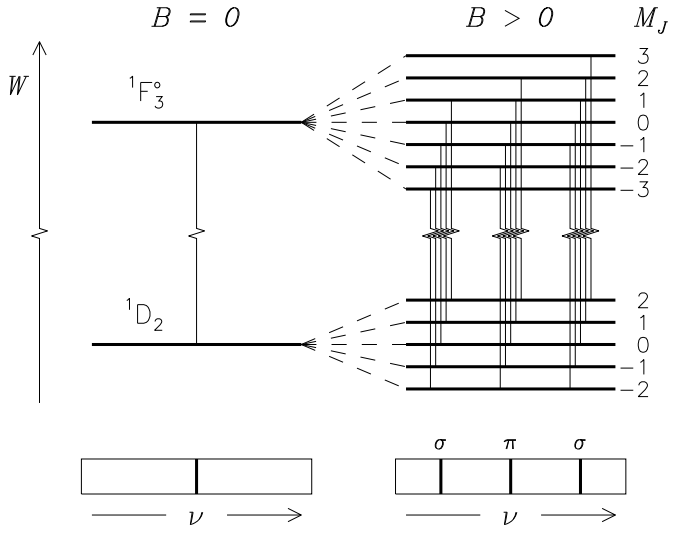

Figure 4.7: Normal Zeeman effect in the He I spectrum

One important difference relative to the one-electron atom result should be noted. Even including the effects of electron spin, the total spin of the many-electron atom may vanish. For these singlet states the Landé factor $g = 1$ and the magnetic splitting reduces to the *normal Zeeman effect*, that is, the energy difference between splitted levels is independent of the values of $J = L$.

Let us illustrate the effect with a simple example. In an external magnetic field the singlet states of He I experience normal Zeeman splitting of the energy levels. In figure 4.7 the Zeeman effect on transitions involving $^1D_2$ and $^1F^o_3$ states is illustrated. The former state splits into five equidistant levels with individual spacing $\hbar\omega_L$. The latter state splits into seven equidistant levels with the identical individual spacing. According to the selection rules of section 4.6, transitions are only allowed between states satisfying $\Delta M_J = 0, \pm 1$. This means that the former single spectral line in the absence of a magnetic field now splits into three individual lines as indicated in the lower part of figure 4.7: the plane polarized π-component at the original location corresponding to $\Delta M_J = 0$, and the two elliptical σ-components, corresponding to $\Delta M_J = \pm 1$, shifted by the Larmor frequency $\pm\omega_L$.

Non-singlet states, for which $S \neq 0$, will through the value of the Landé g-factor show different sensitivity to an external magnetic field. The more sensitive states have $g > 3$. An example is $^6D_{1/2}$ states with $g = 10/3$. Other states are insensitive to magnetic fields, for instance, for $^5F_1$ states we find $g = 0$. These facts may be used to advantage when choosing spectral lines for the study of different phenomena.

The Zeeman effect is routinely used for the determination of magnetic fields in the Solar and stellar atmospheres and in interstellar clouds. In the latter case, the 21 cm line due to the hyperfine splitting of the ground state of hydrogen will be used. From the Zeeman effect of this line, interstellar magnetic fields down into the 10^{-9} T range have been measured. For Solar and stellar applications the Zeeman splitting of spectral lines is competing with the thermal line broadening effect. For these applications magnetically sensitive lines of heavy elements like Fe, Cr or V with small thermal line widths may be used. If we instead are interested in studying flow velocities in the Solar atmosphere through spectral line Doppler

shifts produced by the flow, we would instead choose to work with magnetic insensitive lines of heavy element in order to minimize the perturbing effects of both magnetic fields and thermal line broadening. In table 4.5 selected magnetic sensitive and magnetic insensitive lines of some heavier elements are listed. Note that also some forbidden transitions have been included.

Wavelength Å	Element	Transition	Landé g-factor
Magnetic sensitive lines			
5131.5	Fe	$^5P_1 - {}^5P_1$	5/2
5247.6	Cr	$^5D_0 - {}^5P_1$	5/2
5250.2	Fe	$^5D_0 - {}^7D_1$	3
6258.6	V	$^6D_{1/2} - {}^6D_{1/2}$	10/3
6302.5	Fe	$^5P_1 - {}^5D_0$	5/2
Magnetic insensitive lines			
5123.7	Fe	$^5F_1 - {}^5F_1$	0
5434.5	Fe	$^5F_1 - {}^5D_0$	0
6613.8	Fe	$^5F_1 - {}^7F_0$	0

Table 4.5: Selected magnetic sensitive and insensitive lines from heavier elements

Quiz 4.13 : We will study the effect of an external magnetic field on the resonance line of He I. What is the resulting splitting of the energy levels involved in the formation of the resonance line? What transitions are allowed? How many individual spectral lines will be formed? What is the line spacing? Carry through the same discussion for the transition $^1P_1^o - {}^1D_2$. How do your results differ from the $^1D_2 - {}^1F_3^o$ case illustrated in figure 4.7?

Quiz 4.14 : We want to determine the strength B of the surface magnetic field of a star with radius $R = 7 \cdot 10^8$ m, rotation period $T = 2 \cdot 10^6$ s and with the rotation axis at right angle to the line of sight. For the measurements we want to make use of the Zeeman effect on singlet He I lines. What is the minimum field strength required for the spectral line splitting to exceed the Doppler broadening of each individual line due to the stellar rotation?

Quiz 4.15 : The Fe I λ6302 line is often used for determination of magnetic fields in the lower solar atmosphere. The line results from the transition $3d^64s4p\,^5P_1 - 3d^64s5s\,^5D_0$. The energy levels for these states are tabulated as 29732.73 cm^{-1} and 45595.08 cm^{-1}. What are the parities of each state? Which energy level belongs to each state? How does the given energy information fit with the corresponding listed experimental value 6302.508 Å for the wavelength λ? Explain!

Quiz 4.16 : Do you see a common feature regarding the Zeeman splitting of the magnetic sensitive lines listed in table 4.5. [Hint: How may lines are produced?]

4.9.2 The Stark effect

The potential energy term of the Hamiltonian resulting from the presence of an external electric field $\boldsymbol{E}$ are identical in form for the one-electron and the many-electron atom,

$$H_E = -\boldsymbol{m}_E \cdot \boldsymbol{E}.$$

We only have to replace the electric dipole moment $\boldsymbol{m}_E = -e\boldsymbol{r}$ of the single electron with that of $\mathcal{N}$ electrons,

$$\boldsymbol{m}_E = -e(\boldsymbol{r}_1 + \boldsymbol{r}_2 + \cdots + \boldsymbol{r}_N).$$

Yet, the resulting Stark effect is distinctly different in these two cases. For the Coulomb field of the one-electron atom, the energy levels are degenerate. For a given unperturbed energy level we find corresponding eigenfunctions with both parities. As discussed in section 3.14 this means that the linear Stark effect is the dominating one.

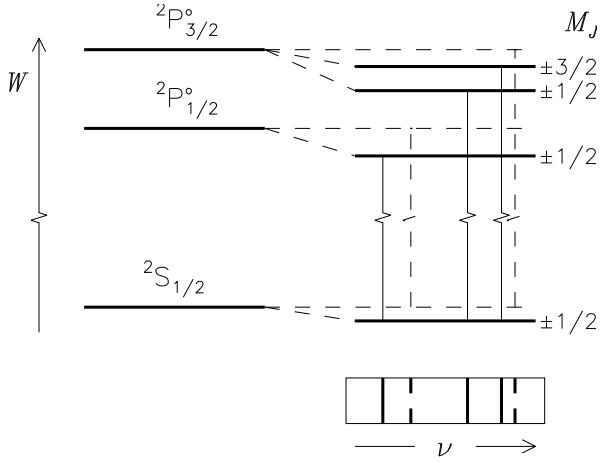

Figure 4.8: Quadratic Stark effect for the D-lines of Na I

In the many-electron atom, the effective central force potential acts to partially lift this energy degeneracy. The energy level of an electron orbital now depends, not only on the principal quantum number n, but also on the orbital quantum number ℓ. Even if the energy degeneracy is not completely removed, the eigenfunctions belonging to a given unperturbed eigenvalue now have identical parities. This means that the linear Stark term vanishes identically. In a perturbation expansion taken to second order (see table 3.9), the contribution to the energy level W_{LSJ} belonging to the state $|\, LSJM_J \,\rangle$ from the quadratic Stark effect is given as

$$\Delta W = E^2 \sum{}' \frac{|\langle\, LSJM_J \,|m_{Ez}|\, L'S'J'M'_J \,\rangle|^2}{W_{LSJ} - W_{L'S'J'}}. \tag{4.24}$$

The primed sum runs over all states $|\, L'S'J'M'_J \,\rangle$ different from $|\, LSJM_J \,\rangle$. In our notation for eigenvalues and eigenfunctions we have for convenience only included total angular quantum numbers. We notice that the numerator is non-vanishing only if $S' = S$, $M'_J = M_J$ and $J' = J, J \pm 1$. This follows from reasoning similar to that used for the derivation of selection rules in the electric dipole approximation. The final result can be shown to be of the general form

$$\Delta W = E^2(A + BM_J^2), \tag{4.25}$$

where A and B both depend on L, S and J. That is, the energy perturbation is independent of the sign of M_J. The most important contribution to the sum (4.24) will come from the nearest levels of opposite parity. For example, the perturbation of the $3p\,^2P_{1/2,3/2}$ levels of the Na I atom will be approximately twice as large as that of the ground state $3s\,^2S_{1/2}$ because the former levels are perturbed by the $3d\,^2D_{3/2,5/2}$ levels which lie closer. For the ground state the energy shift in cm^{-1} is $7.8 \cdot 10^{-5} E^2$ where E is given in V/m. The situation is illustrated in figure 4.8.

As discussed in section 3.14, one of the main effects of the Stark effect is to give rise to pressure broadening of spectral lines. The pressure broadening results from the variable electric field seen by atoms during close encounters with charged particles in ionized gases.

Chapter 5

Molecular Spectra

In the previous chapters we studied properties of single atoms, containing one positively charged nucleus and one or more surrounding electrons. A molecule is made up of two or more individual atoms. The molecule thus consists of two or more nuclei together with one or more electrons. The electrons are necessary for the stability of the molecule. In the absence of electrons, the positively charged nuclei would repel each other and no stable system can be formed.

As in the single atom, the electrons of the molecule are only allowed to occupy a specific set of energy states. However, due to the presence of two or more nuclei, new types of motion are possible in the molecule. The nuclei may participate in vibrational motion or the molecule may rotate as a whole. In the present chapter we shall see that quantum mechanics requires also these additional types of motion to be quantized with a corresponding set of allowed energy states. It turns out that these additional energy states are more closely spaced than those due to the electron motion. They will therefore give rise to easily recognizable features in molecular spectra.

The first molecules to be discovered in cold interstellar gas clouds, the CH, CH^+ and CN molecules, were identified in 1937. These molecules constitute some of the basic building blocks for the formation of more complex organic molecules and were observed as exceedingly narrow absorption lines in stellar spectra in the red and near infra-red spectral range. With the development of microwave technology and the ability to observe spectra also in the far infra-red and millimeter wavelengths, a large number of different interstellar molecules have later been identified, particularly after 1970. Some of these molecules are surprisingly large and complicated. In the following we shall for the greater part limit our discussion to the simpler diatomic molecule.

5.1 The Diatomic Molecule

The diatomic molecule may be considered formed in a process in which the two participating atoms are successively brought closer to each other. Let R be the inter-nuclear distance. When far apart, the two atoms behave independently of each other. At a certain distance the two electron clouds feel the presence of each another and therefore start to deform. This means that the electron clouds are no longer able to completely shield the charge of the nuclei. The two nuclei then start to repel each other. This repulsion increases as the two nuclei are brought closer still, eventually increasing proportional to R^{-2}, corresponding to the

electrostatic repulsion between two "naked" nuclei. At the same time, however, the electrons of one atom is attracted by the nucleus of the other and vise versa.

A stable molecule may form during this process if the potential energy takes a minimum value for some inter-nuclear distance R_e. Typical values of the inter-nuclear distance R_e are of the order $1\text{Å} = 10^{-10}$ m. We remember that this is also the typical radius of the atomic electron cloud in an individual atom. We expect that at least the outermost electrons are then no longer able to identify their original mother nuclei. Indeed, it will be necessary to consider the electrons of the molecule as belonging to one common system for which the Pauli principle must be satisfied. Only one electron may occupy each allowed electron state of the molecule as a whole.

Two types of molecular bindings can be identified. In *hetero-nuclear molecules*, where the two atoms forming the molecule are of different types, as in CO, the energy necessary to remove one valence electron from each of these atoms will generally be different. This leads to a binding in which the atom with the weakest bound electron, the most electro-positive atom, supplies the electro-negative one with parts of one or more of its own electrons. One end of the molecule becomes electrically more positive than the other. The molecule thus acquires a permanent electric dipole moment. The binding in this case is said to be *ionic* or *hetero-polar*. In *homo-nuclear molecules*, where the two atoms forming the molecule are identical, as in H_2, no permanent electric dipole moment is formed. The binding is now established through electrons with opposite spins occupying the space between the two nuclei. The binding in this case is called *covalent* or *homo-polar*. In both cases it is the deformed electron cloud that supply the necessary "glue" to keep the two nuclei together.

A quantitative discussion of the properties of the diatomic molecule starts with the identification of the proper Hamiltonian. Significant simplifications result if we assume that the combined mass of the electrons is negligible compared with that of the nuclei, that is, that the center-of-mass of the molecule coincides with the center-of-mass of the two nuclei. In analogy with the discussion of section 3.4, the Hamiltonian then reduces to

$$H = T_n + T_e + V(\boldsymbol{R}, \boldsymbol{r}_1, \cdots, \boldsymbol{r}_N) \tag{5.1}$$

where T_n and T_e are the kinetic energies of the nuclei and electrons, respectively, and V is the total potential energy of the molecule. $\boldsymbol{R}_i$ and $\boldsymbol{r}_i$ are the position vectors of the individual nuclei and electrons relative to the center-of-mass, and $\boldsymbol{R} = \boldsymbol{R}_2 - \boldsymbol{R}_1$ is the inter-nuclear vector. The geometry is illustrated in figure 5.1. In terms of the corresponding momenta $\boldsymbol{P} = \mu\dot{\boldsymbol{R}}$ and $\boldsymbol{p}_i = m\dot{\boldsymbol{r}}_i$ where $\mu = M_1M_2/(M_1 + M_2)$ is the reduced mass of the two nuclei and m the electron mass, the kinetic energies of the nuclei and the N electrons are

$$T_n = \frac{\boldsymbol{P}^2}{2\mu}$$

and

$$T_e = \sum_{i=1}^{N} \frac{\boldsymbol{p}_i^2}{2m}.$$

The electrostatic potential energy is given by

$$V(\boldsymbol{R}, \boldsymbol{r}_1, \cdots, \boldsymbol{r}_N) = -\sum_{i=1}^{N} \frac{Z_1 e^2}{4\pi\epsilon_0 \,|\, \boldsymbol{r}_i - \boldsymbol{R}_1 \,|} - \sum_{i=1}^{N} \frac{Z_2 e^2}{4\pi\epsilon_0 \,|\, \boldsymbol{r}_i - \boldsymbol{R}_2 \,|} + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{e^2}{4\pi\epsilon_0 \,|\, \boldsymbol{r}_i - \boldsymbol{r}_j \,|} + \frac{Z_1 Z_2 e^2}{4\pi\epsilon_0 R}$$

where Z_1 and Z_2 are the charge numbers of the two nuclei at positions $\boldsymbol{R}_1 = -M_2\boldsymbol{R}/(M_1+M_2)$ and $\boldsymbol{R}_2 = M_1\boldsymbol{R}/(M_1+M_2)$ relative to the center-of-mass. The Hamiltonian (5.1) only includes electrostatic contributions to the total energy. Relativistic and spin effects were neglected.

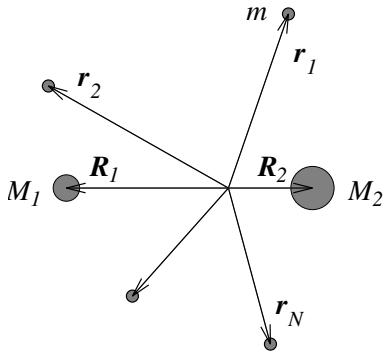

Figure 5.1: Geometry of the diatomic molecule

Solving the Schrödinger equation

$$H \,|\, \Psi \,\rangle = W \,|\, \Psi \,\rangle \tag{5.2}$$

for the simplest diatomic molecule is a demanding task even for the zero-order Hamiltonian (5.1). We therefore need to look for ways to simplify the problem. From a classical point of view, we would expect the light electrons to adjust almost immediately to any relative motion of the two nuclei. It therefore makes sense to study the *electron wave equation*

$$(T_e + V(\boldsymbol{R}, \boldsymbol{r}_1, \cdots, \boldsymbol{r}_N)) \,|\, \Phi_\sigma \,\rangle = W_\sigma(R) \,|\, \Phi_\sigma \,\rangle, \tag{5.3}$$

where the inter-nuclear vector $\boldsymbol{R}$, appearing in the eigenvalue $W_\sigma(R)$ and in the *electron wave function* $|\, \Phi_\sigma \,\rangle = |\, \Phi_\sigma(\boldsymbol{R}; \boldsymbol{r}_1, \cdots, \boldsymbol{r}_N) \,\rangle$, is considered as an external parameter. The subscript σ specifies the electron configuration of the molecule. For symmetry reasons the electron energy W_σ only depends on the inter-nuclear distance R and not on the orientation of the molecule.

Let us assume that the electron wave equation (5.3) has been solved. Let us further assume that the molecular wave function $|\, \Psi \,\rangle$ can be written as a product of the electron wave function $|\, \Phi_\sigma \,\rangle$ with another function of $\boldsymbol{R}$ only,

$$|\, \Psi \,\rangle = \mathcal{F}_\sigma(\boldsymbol{R}) \,|\, \Phi_\sigma \,\rangle. \tag{5.4}$$

The *nuclear wave function* $\mathcal{F}_\sigma(\boldsymbol{R})$ may be determined by substituting (5.4) into (5.2). Considerable simplifications result if the electron wave function $|\, \Phi_\sigma \,\rangle$ may be considered to be a

slowly varying function of $\boldsymbol{R}$ relative to that of $\mathcal{F}_\sigma(\boldsymbol{R})$. This assumption is referred to as the *Born-Oppenheimer approximation* and is normally found to be well satisfied for values of R close to the equilibrium inter-nuclear distance R_e. Under this assumption, the *nuclear wave equation*

$$\left(-\frac{\hbar^2}{2\mu}\nabla^2_{\boldsymbol{R}} + W_\sigma(R)\right)\mathcal{F}_\sigma = W\mathcal{F}_\sigma \tag{5.5}$$

is easily derived. The electron energy $W_\sigma(R)$ represents the effective bonding energy of the molecule. That is, it is the electrons that supply the "glue" for keeping the molecule together.

5.2 Molecular Vibration and Rotation

The nuclear wave equation (5.5) differs from the wave equation (3.18) for the one-electron atom only in that the free variable is now the inter-nuclear vector $\boldsymbol{R}$ and that the Coulomb potential (3.16) is replaced by the spherically symmetric "molecular" potential energy $W_\sigma(R)$, depending on the particular electron configuration σ. It follows trivially that the *nuclear orbital angular momentum* $\boldsymbol{N} = \boldsymbol{R}\times\boldsymbol{P}$ may be quantized simultaneously with the molecular energy in the familiar manner,

$$\boldsymbol{N}^2\mathcal{F}_\sigma = \hbar^2 N(N+1)\mathcal{F}_\sigma \tag{5.6}$$

$$N_z\mathcal{F}_\sigma = \hbar M_N\mathcal{F}_\sigma. \tag{5.7}$$

The *nuclear orbital quantum number* N is a non-negative integer.[1] The *nuclear azimuthal quantum number* M_N takes values in the range

$$M_N = -N, -N+1, \cdots, N.$$

The angular part of the nuclear wave function $\mathcal{F}_\sigma$ is identical in form to that of the one-electron atom. Indeed, if we substitute

$$\mathcal{F}_\sigma(\boldsymbol{R}) = \frac{1}{R}F_\sigma(R)P_N^{|M_N|}(\cos\theta)\exp(\iota M_N\phi), \tag{5.8}$$

the nuclear wave equation (5.5) reduces to

$$\left(-\frac{\hbar^2}{2\mu}\frac{\mathrm{d}^2}{\mathrm{d}R^2} + W_\sigma(R) + \frac{\hbar^2 N(N+1)}{2\mu R^2}\right)F_\sigma = WF_\sigma. \tag{5.9}$$

To proceed we need to know the molecular potential energy W_σ as a function of inter-nuclear distance R for the specified electron configuration σ. From an experimental basis it has been demonstrated that this function may often be approximated in the form

$$W_\sigma(R) = W_\sigma(R_e) + U_M(R) \tag{5.10}$$

where $U_M(R)$ is the three parameter *Morse potential*

$$U_M(R) = D_e(1-\exp(-\beta Q))^2. \tag{5.11}$$

[1] Notice that we in this section are using the symbol N in two different meanings: as the number of electrons in the molecule and as the new quantum number! Hopefully, which meaning, if not stated explicitly, should be clear from the context and such that no confusion should arise.

Here $Q = R - R_e$ is the deviation of the inter-nuclear distance from the equilibrium value R_e. The shape of the Morse potential (full curve) is given in figure 5.2. The depth D_e of the potential well is the *binding energy* of the molecule, that is, the amount of energy that must be supplied to the molecule in order to break it up into its individual parts. The parameter β determines the width of the potential energy minimum. The three parameters R_e, D_e and β are to be fitted for each molecule and for each electron configuration σ.

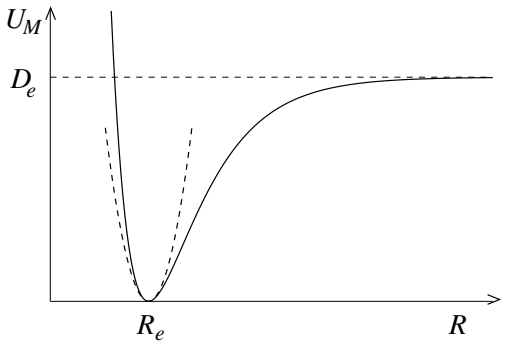

Figure 5.2: Morse potential (solid curve) and the corresponding harmonic oscillator potential (dash-dotted curve)

If the nuclei of the diatomic molecule are brought out of their equilibrium configuration, forces will arise trying to restore the equilibrium state. For small distortions, the potential energy function (5.10) may be approximated by an expansion to second order in Q,

$$W_\sigma(R) = W_\sigma(R_e) + \frac{1}{2}k_\sigma Q^2, \tag{5.12}$$

where we defined the "spring" constant $k_\sigma = U''_M(R_e) = 2\beta^2 D_e$. We recognize the second order term as the potential energy of an *harmonic oscillator*. This potential (dash-dotted curve) is compared with the Morse potential (solid curve) in figure 5.2. For a lowest order approximation we also replace R in the angular momentum term of (5.9) by the equilibrium value R_e. The latter approximation is referred to as the *rigid rotor approximation*.

It will now be seen that the total molecular energy can be written as a sum of three contributions,

$$W = W_\sigma(R_e) + W_v + W_N. \tag{5.13}$$

We will return to a discussion of the importance of the first term $W_\sigma(R_e)$ in section 5.5.

The second term of (5.13) represents the *vibrational energy* of the molecule. It is determined from the *harmonic oscillator wave equation*

$$-\frac{\hbar^2}{2\mu}\frac{\mathrm{d}^2F}{\mathrm{d}Q^2} + \frac{1}{2}kQ^2F = W_vF. \tag{5.14}$$

For simplicity we suppressed the appropriate electron configuration subscript σ for k, W_v and F. Equation (5.14) is conveniently solved through the substitution

$$F(Q) = \exp(-\frac{x^2}{2})u(x) \quad \text{with} \quad x = \sqrt{\alpha}Q \quad \text{and} \quad \alpha = \frac{\sqrt{\mu k}}{\hbar}. \tag{5.15}$$

The unknown function $u(x)$ must satisfy the *Hermite differential equation*

$$\frac{\mathrm{d}^2u}{\mathrm{d}x^2} - 2x\frac{\mathrm{d}u}{\mathrm{d}x} + 2vu = 0, \tag{5.16}$$

with the vibrational energy W_v related to the *vibrational quantum number* v through

$$W_v = \hbar\omega_0(v + \frac{1}{2}) \quad \text{with} \quad \omega_0 = \sqrt{\frac{k}{\mu}}. \tag{5.17}$$

The Hermite differential equation is known to have finite solutions only if v takes non-negative integer values (see quiz 5.3). The solutions are then the *Hermite polynomials* $u(x) = H_v(x)$. Some useful properties of these polynomials are listed in table 5.1.

Hermite polynomial — generating function

$$H_v(x) = (-1)^v \exp(x^2)\frac{\mathrm{d}^v}{\mathrm{d}x^v}\exp(-x^2)$$

Some lower order polynomials

$$H_0(x) = 1 \quad H_1(x) = 2x \quad H_2(x) = -2 + 4x^2$$
$$H_3(x) = -12x + 8x^3 \quad H_4(x) = 12 - 48x^2 + 16x^4$$

Recurrence relation

$$H_{v+1}(x) = 2xH_v(x) - 2vH_{v-1}(x)$$

Normalization and orthogonality

$$\int_{-\infty}^{\infty} \exp(-x^2)H_v(x)H_{v'}(x)\,\mathrm{d}x = 2^v v!\sqrt{\pi}\,\delta_{v,v'}$$

Table 5.1: Hermite polynomials

The energy levels W_v available to the harmonic oscillator form a series of equidistant levels. In figure 5.3 the energy levels are plotted superposed on the harmonic oscillator potential $kQ^2/2$. The harmonic oscillator wave functions $F_v(Q) = \exp(-\alpha Q^2/2)H_v(\sqrt{\alpha}Q)$ for each level are also shown. We note that $\mid F_v(Q) \mid^2$ represents the probability density function for finding the two nuclei a distance $R = R_e + Q$ apart. We also note that maxima of this probability density function will (for $v > 0$) be found near the classical turning point for a particle with energy W_v in the given potential well. The presence of the *vibrational rest energy* $W_0 = \frac{1}{2}\hbar\omega_0$ means that the *dissociation energy* of the molecule should be defined as

$$D_0 = D_e - \frac{1}{2}\hbar\omega_0.$$

Some of the energy required to break the molecule apart is already present in the form of the vibrational rest energy.

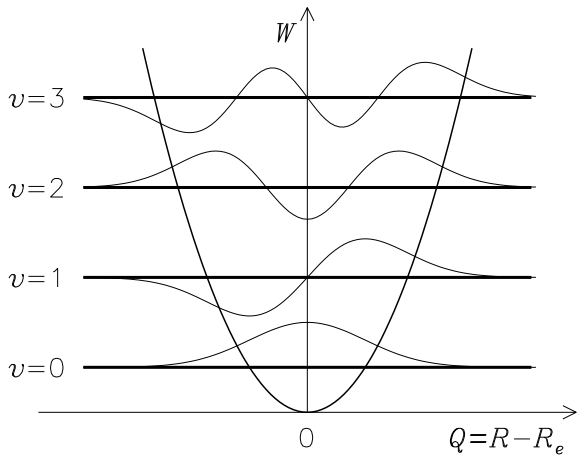

Figure 5.3: Vibrational wave functions, energies and harmonic potential

The third term of (5.13),

$$W_N = B\,N(N+1) \quad \text{with} \quad B = \frac{\hbar^2}{2I_e}, \tag{5.18}$$

is the *rotational energy* of the two nuclei. The *rotational constant* B is here expressed in terms of the *moment of inertia* $I_e = \mu R_e^2$ of the rigid rotor. The quantities I_e, B and W_N all depend on the electron configuration σ. From (5.18) it is seen that the energy difference between adjacent rotational energy levels increases linearly with $N+1$,

$$W_{N+1} - W_N = 2B(N+1). \tag{5.19}$$

Quiz 5.1 : Verify that the moment of inertia of the rotating diatomic molecule relative to an axis of rotation through the center-of-mass and at right angle to the line connecting the two masses M_1 and M_2, can be written

$$I_e = M_1R_1^2 + M_2R_2^2 = \mu R_e^2,$$

where R_1 and R_2 are the distances from the center-of-mass to the two masses, $R_e = R_1 + R_2$ and $\mu = M_1M_2/(M_1+M_2)$.

Quiz 5.2 : Show that (5.14) reduces to (5.16) in terms of the new variables x and $u(x)$ as given by (5.15).

Quiz 5.3 : Argue that solutions of (5.16) may be written as power series

$$u(x) = \cdots + a_n x^n + a_{n+2}x^{n+2} + \cdots$$

where n is even or odd. Show that the coefficients must satisfy

$$\frac{a_{n+2}}{a_n} = \frac{2(n-v)}{(n+2)(n+1)}.$$

Thus, verify that "physically acceptable" solutions of (5.16) will only be found if the quantum number v takes non-negative integer values. [*Hint:* The infinite power series with coefficients satisfying $a_{n+2}/a_n \to 2/n$ as $n \to \infty$ (resulting if v takes non-integer values) behaves asymptotically as $\exp(2x^2)$ for large x.]

5.3 Selection Rules for Vibrational-Rotational Transitions

With the vibrational and rotational energy levels determined, we proceed to study radiation transitions between these levels when the molecule is exposed to an external electromagnetic radiation field. This discussion closely resembles the previous one for radiation transitions in one-electron atoms in section 3.9. The interaction between the molecule and the external radiation field may be approximated by the extra term

$$H' = -\boldsymbol{m}_E \cdot \boldsymbol{E} \tag{5.20}$$

in the Hamiltonian. Here $\boldsymbol{m}_E = \boldsymbol{m}_E(\boldsymbol{R}, \boldsymbol{r}_1, \cdots, \boldsymbol{r}_N)$ represents the electric dipole moment resulting from all the individual charges in the molecule and $\boldsymbol{E}$ is the oscillating electric field of the wave.

In the electric dipole approximation, in which the variation of the wave electric field over the molecule is neglected,

$$\boldsymbol{E} = \frac{E_0}{2}\left(\hat{\boldsymbol{\epsilon}}\, \exp(-\iota\omega t) + \hat{\boldsymbol{\epsilon}}^* \exp(\iota\omega t)\right),$$

the transition rate for an absorption transition from an initial state $|\, i\,\rangle = |\, \sigma v N M_N\,\rangle$ to another state $|\, f\,\rangle = |\sigma' v' N' M'_N\,\rangle$ is given by

$$w^a_{fi} \sim |\langle\, f\,|\, \hat{\boldsymbol{\epsilon}} \cdot \boldsymbol{m}_E\, |\, i\,\rangle|^2\,. \tag{5.21}$$

Similar expressions are valid for the corresponding emission rates.

We restrict the discussion to transitions for which the electron configuration remains unchanged, $\sigma' = \sigma$, and introduce the corresponding average electric dipole moment

$$\overline{\boldsymbol{m}}_E(\boldsymbol{R}) = \langle\, \Phi_\sigma\,|\, \boldsymbol{m}_E\, |\, \Phi_\sigma\,\rangle.$$

The averaging is performed with respect to the electron wave function $|\, \Phi_\sigma\,\rangle$, unperturbed by the radiation field. $\overline{\boldsymbol{m}}_E$ represents the *permanent* electric dipole moment of the molecule. We notice that $\overline{\boldsymbol{m}}_E$ is a function of the inter-nuclear vector $\boldsymbol{R}$ only. From symmetry arguments we expect $\overline{\boldsymbol{m}}_E$ to be parallel with $\boldsymbol{R}$. We also expect the magnitude of $\overline{\boldsymbol{m}}_E$ to vary with inter-nuclear distance R. To first order we write

$$\overline{m}_E(R) = \overline{m}_E(R_e) + Q\, \overline{m}'_E(R_e).$$

The integrals appearing in the expression (5.21) for the transition rate now simplify to

$$\langle\, f\,|\, \hat{\boldsymbol{\epsilon}} \cdot \boldsymbol{m}_E\, |\, i\,\rangle = \iiint \mathcal{F}_{\sigma v' N' M'_N}(\boldsymbol{R})\, \hat{\boldsymbol{\epsilon}} \cdot \overline{\boldsymbol{m}}_E(\boldsymbol{R})\, \mathcal{F}_{\sigma v N M_N}(\boldsymbol{R}) R^2\, \mathrm{d}R\, \mathrm{d}\cos\theta\, \mathrm{d}\phi. \tag{5.22}$$

The nuclear wave function $\mathcal{F}_\sigma$ for the specified electron configuration σ has been indexed with the vibrational and rotational quantum numbers for the states involved. The angular integrals are identical in form to those studied for one-electron atoms in section 3.10. We may therefore immediately conclude that necessary conditions for non-vanishing transition rates are that $\Delta N \equiv N' - N = \pm 1$ and $\Delta M_N \equiv M'_N - M_N = 0, \pm 1$.

The remaining radial integral in (5.22) takes the form

$$\int \exp(-x^2) H_{v'}(x) \left(\overline{m}_E(R_e) + x \frac{\overline{m}'_E(R_e)}{\sqrt{\alpha}}\right) H_v(x)\, \mathrm{d}x. \tag{5.23}$$

From the recurrence relation and orthogonality property of Hermite polynomials listed in table 5.1, it is seen that the first term of (5.23) will be non-vanishing if $\overline{m}_E(R_e) \neq 0$ and $\Delta v \equiv v' - v = 0$. The second term will be non-vanishing if $\overline{m}'_E(R_e) \neq 0$ and $\Delta v = \pm 1$. We shall refer to transitions for which $\Delta v = 0$ as *rotational transitions*, those for which $\Delta v \neq 0$ as *vibrational-rotational transitions*.

We may sum up our findings as the *selection rules* for rotational and vibrational-rotational transitions:

A diatomic molecule, in the harmonic oscillator, rigid rotor approximation, may absorb or emit radiation in rotational or vibrational-rotational transitions if

- $\Delta N = \pm 1$
- $\Delta M_N = 0, \pm 1$.
- $\Delta v = 0$ *and the molecule has a permanent electric dipole moment* $\overline{\boldsymbol{m}}_E$.
- $\Delta v = \pm 1$ *and the molecule has a permanent electric dipole moment* $\overline{\boldsymbol{m}}_E$ *varying with inter-nuclear distance* R.

Hetero-nuclear molecules usually have permanent electric dipole moments. Homo-nuclear diatomic molecules do not have permanent electric dipole moments, and therefore do not usually exhibit radiative rotational or vibrational-rotational transitions.

On the basis of these selection rules, we may understand spectral features due to rotational and vibrational-rotational transitions of the diatomic molecule. Let us first consider purely rotational transitions. According to (5.18), the energy difference between adjacent rotational energy levels increases linearly with the angular quantum number of the higher energy state. Transitions are only allowed between energy levels corresponding to adjacent values of N. This means that rotational transitions give rise to an equidistant sequence of spectral lines with frequency. The first line in this series,

$$\frac{1}{\lambda} = \frac{2B}{hc},$$

corresponds to the $N = 0$ to $N = 1$ transition. In figure 5.4 some lower rotational energy levels, allowed rotational transitions between these levels and the corresponding spectral lines are illustrated. Rotational spectral lines are for most molecules found in the microwave spectral range.

The corresponding situation for vibrational-rotational transitions is illustrated in figure 5.5. In the figure the lower rotational energy levels for two different vibrational states, $v = 0$ and $v = 1$, are given. Notice the break in the energy scale. The energy difference between adjacent vibrational states is much larger that those between adjacent rotational states. Vibrational energy differences usually correspond to the infra-red spectral range. Transitions between rotational energy levels belonging to the two vibrational states constitute the $(0, 1)$-*molecular band*. We notice that the purely vibrational transition from $v = 0, N$ to $v' = 1, N' = N$ is forbidden. To each side of this forbidden line, traditionally called the Q-*branch*, there appears series of equidistant lines with frequency. The collection of lines corresponding to transitions $v = 0, N$ to $v' = 1, N' = N - 1$ is referred to as the P-*branch*,

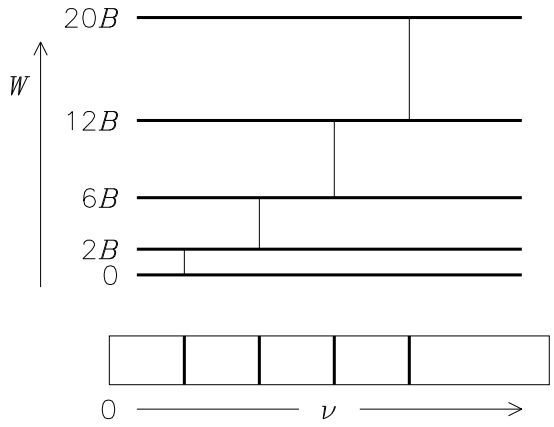

Figure 5.4: Rotational energy levels, allowed transitions and spectral lines

the collection of lines corresponding to transitions $v = 0, N$ to $v' = 1, N' = N + 1$ is called the R-*branch*.

Quiz 5.4: For the CO molecule, the inter-nuclear distance is $R_e = 0.113$ nm. The "spring" constant is tabulated as $k = 1.87 \cdot 10^3$ N/m. Determine the wavelengths λ for the first three lines of the rotational spectrum. Determine the three lowest vibrational energy levels and the corresponding wavelengths of the vibrational spectrum. Which spectral ranges do these transitions belong to?

Quiz 5.5: The HCl molecule has a potential energy curve described by dissociation energy $D_e = 36300$ cm^{-1}, shape factor $\beta = 1.90$ Å^{-1} and inter-nuclear equilibrium distance $R_e = 1.28$ Å. Determine the "spring" constant k and the energy difference between the two first vibrational levels. Estimate the number of vibrational states for the molecule.

5.4 Generalized Oscillator-Rotor Models

With increasing vibrational and rotational energies the harmonic oscillator, rigid rotor approximation discussed above needs to be improved upon.

With increasing vibrational energies, the harmonic oscillator potential deviates appreciably from the Morse potential as illustrated in figure 5.2. It will then be necessary to include higher order terms in Q (at least third and fourth order terms) in the potential expansion (5.12). The oscillator is then called an *anharmonic oscillator*. The vibrational energy levels in this case get negative correction terms. At the same time the radial parts F of the nuclear wave functions $\mathcal{F}$ are modified, on the average being more extended in the radial direction. Due to the latter effect the average moment of inertia increases with a corresponding decrease in rotational energy levels.

At high rotational quantum numbers N, we would also expect increasing centrifugal forces to stretch the inter-nuclear distance. This would also increase the moment of inertia of the molecule and therefore lead to a decrease in rotational energy levels. A model for which the

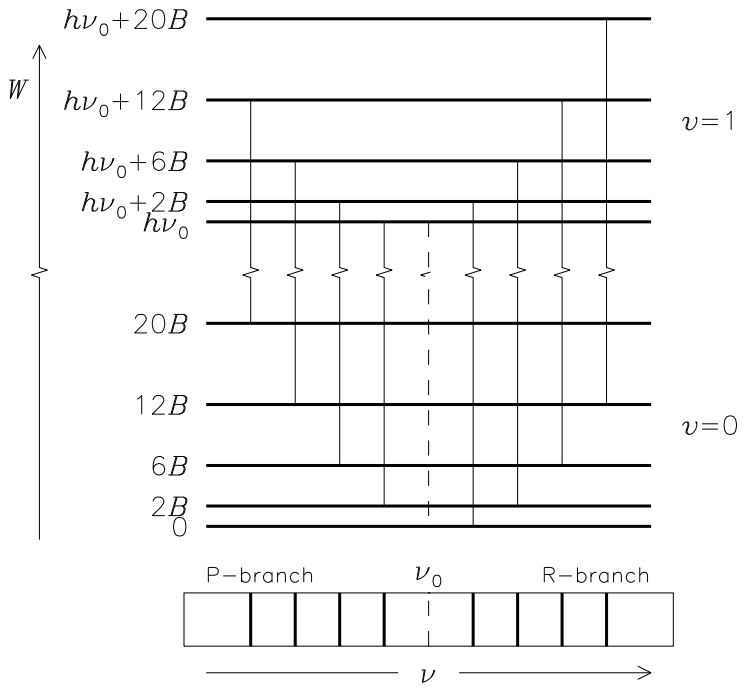

Figure 5.5: Vibrational-rotational transitions

variation in R with increasing values of N in the angular momentum term of (5.9) is taken into account, is referred to as a *non-rigid rotor* model.

Including anharmonic and stretching effects the expressions for the vibrational and rotational energies of (5.13) will be modified. A quantitative study (second order perturbation theory) shows that the lowest order correction to these expression are

$$W_v = \hbar\omega_0(v+\frac{1}{2}) - \frac{\hbar^2\omega_0^2}{4D_e}(v+\frac{1}{2})^2 \tag{5.24}$$

and

$$W_N = B_v\, N(N+1) \tag{5.25}$$

where the rotational constant B_v is now a function for the vibrational quantum number v,

$$B_v = \frac{\hbar^2}{2I_e}\left(1 - \frac{3\hbar\omega_0}{\beta R_e D_e}\left(1 - \frac{1}{\beta R_e}\right)(v+\frac{1}{2})\right). \tag{5.26}$$

The expressions (5.24) and (5.26) must in fact be considered as the first terms of power series in $v+\frac{1}{2}$ and $N(N+1)$ of a more complete analysis.

In figure 5.6 the Morse potential (solid curve), the anharmonic and harmonic oscillator potentials (dash-dotted and dashed curves) and the corrected vibrational energy levels are illustrated schematically. We note the decreasing distance between adjacent vibrational levels with increasing v-values. The corresponding wave functions are modified such that the maxima of $|F|^2$ tend to occur near the classical turning points R_v, satisfying $U_M(R_v) = W_v$. Another consequence of the anharmonic potential is that the selection rules for vibrational transitions must be generalized. Transitions corresponding to $\Delta v = \pm 1, \pm 2, \cdots$ can now be expected. Still, however, the $\Delta v = \pm 1$ transitions usually lead to the strongest lines. Because

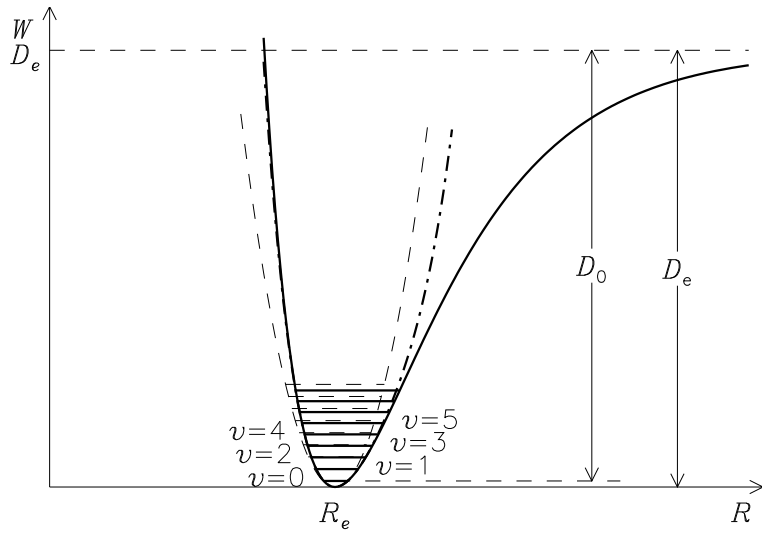

Figure 5.6: Anharmonic potential (dash-dotted curve) and energy levels of the anharmonic oscillator compared with the corresponding potential and energy levels of the harmonic oscillator (dashed)

of the decreasing inter-level spacing, the $\Delta v = \pm 1$ transitions for the anharmonic oscillator will produce a series of slightly shifted $(v, v+1)$-molecular bands. Other groups of lines result from $\Delta v = \pm 2, \pm 3, \cdots$ transitions.

The correctional term of the rotational constant B_v stems from the modification of the "spring" constant due to rotational stretching of the inter-nuclear equilibrium distance. A physical interpretation of this result is that the average inter-nuclear distance and therefore the average moment of inertia, depends on the vibrational state of the molecule. This means that the distance between the rotational energy levels for $v = 1$ in figure 5.5 will be slightly smaller than the corresponding ones for $v = 0$. This in turn means that individual lines in the P- and R-branches are no longer equidistant. In fact, for the (0,1)-molecular band the frequencies of the different lines in the P-branch ($N \to N-1$) and the R-branch ($N \to N+1$) are given by

$$\hbar\omega_P(N) = \hbar\omega_0 + B_1\,(N-1)N - B_0\,N(N+1) \tag{5.27}$$

$$\hbar\omega_R(N) = \hbar\omega_0 + B_1\,(N+1)(N+2) - B_0\,N(N+1), \tag{5.28}$$

with B_v decreasing with increasing v. The result is that there will be a gradual decrease in the frequency difference between successive vibrational-rotational lines going from the P- to the R-branch. The effect is illustrated in figure 5.7. This effect is even more pronounced in vibrational-rotational transitions also involving transitions in the electron configuration σ of the molecule. This will allow for larger differences in the values of the rotational constants involved. We will return to a further discussion of this effect in section 5.5.

Quiz 5.6 : Make use of (5.27) and (5.28) to show that

$$\hbar\omega_R(N) - \hbar\omega_P(N) = 2B_1\,(2N+1)$$
$$\hbar\omega_R(N) - \hbar\omega_P(N+2) = 2B_0\,(2N+3).$$

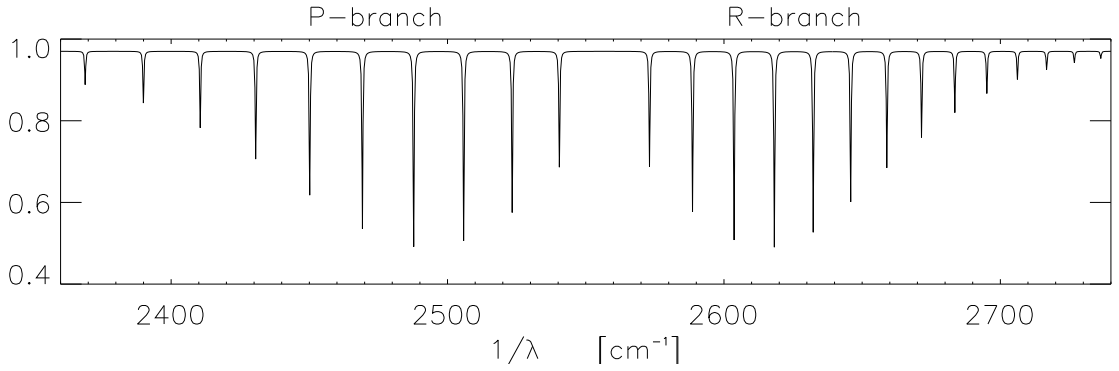

Figure 5.7: Vibrational-rotational absorption spectrum of a diatomic molecule displaying anharmonic and non-rigid rotor effects

Identify the P- and R-branches of the (0,1)-molecular band of the idealized diatomic molecule spectrum shown in figure 5.7. Determine values for B_0 and B_1 for the molecule. What is the equilibrium inter-nuclear distance R_e? What is the "spring" constant k? The data in the figure refers to the HBr molecule ($M_{\mathrm{Br}} = 79m_p$).

5.5 Electronic-Vibrational-Rotational Spectra

Molecules can absorb or emit radiation not only as a result of changes in their vibrational or rotational energies, but also as a result of changes in their electron configuration. The energy changes associated with a transition from one electron state to another are usually relatively large, corresponding to radiation in the visible or ultraviolet spectral ranges, in contrast to the infrared or microwave ranges for vibrational-rotational or rotational transitions.

As discussed in section 5.1, the electrons of the molecule may only exist in one of a discrete set of eigenstates. Each electron configuration σ will as a function of the nuclear vector $\boldsymbol{R}$ give rise to a different effective bonding potential $W_\sigma(\boldsymbol{R})$. This is illustrated in figure 5.8 where two such functions are given, one corresponding to the lowest electron energy state and one for an excited state. The bonding potentials may have different depths, different widths and have their minima at different inter-nuclear distances. For each bonding potential a set of allowed vibrational and rotational states exist. Some vibrational energy levels for the two potential curves are drawn in the figure. Due to the anharmonic potential, the distance between successive energy levels is not constant, but decreases as the vibrational excitation energy approaches the dissociation energy of the molecule.

In a transition involving changes in electron configuration, the molecule will jump from an energy level belonging to one potential energy curve to a level belonging to another curve. Due to the small electron mass, we expect the electrons to be able to complete their transition before the nuclei have time to react. This is known as the *Franck-Condon principle* and means that the transition may be represented by a vertical line in figure 5.8. We must expect the line to start at an inter-nuclear distance near a maximum of the vibrational probability density function and end up near another maximum. In analogy with the harmonic oscillator case illustrated in figure 5.3 we expect to find maxima of the probability density function $|\,F\,|^2$ near the minimum of the potential energy function for $v = 0$, but near the classical

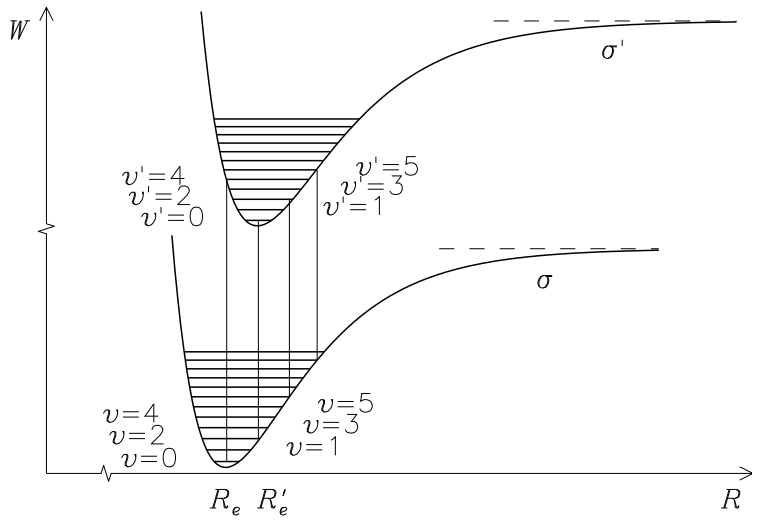

Figure 5.8: Potential energy functions and, vibrational energy levels for two different electron configurations with some probable transitions.

turning points for $v > 0$. But then we must expect that a transition from a vibrational state $v = 0$ to another $v' = 0$ state belonging to the other electron configuration will be possible if the minima of the two potential energy curves fall at approximately the same inter-nuclear distance, $R_e = R'_e$, but not if they are appreciably different. The latter case is illustrated in figure 5.8. Here we must instead expect transitions between $v = 0$ and $v' = 3$ and between $v = 2$ and $v' = 0$ to be among the more probable. In the figure we have also indicated two other probable transitions. This is an indication that we must expect a richer set of selection rules for transitions involving changes in the electron, vibrational and rotational state of the molecule. The selection rules for allowed changes in the vibrational and rotational quantum numbers v and N depend on the electron states involved in the transitions. We must expect to see $\Delta v = 0, \pm 1, \pm 2, \cdots$ while $\Delta N = 0, \pm 1$ are frequently allowed.

Considerable richness and complexity are to be expected in the interpretation of molecular spectra. Each allowed electron transition will consist of several bands, each band corresponding to particular values of the initial and final vibrational quantum numbers v and v'. Each band in turn consists of several branches, P-, Q- and R-branches corresponding to transitions from rotational quantum number N to $N' = N - 1$, N or $N + 1$, respectively. The Q-branch now represents allowed transitions. This is in contrast to our previous results for pure vibrational-rotational transitions.

With initial and final states belonging to different electron configurations, we should expect to see larger differences in the rotational constants B_v. This means that there will be differences in the distance between corresponding adjacent rotational energy levels for different vibrational states. Instead of the regular spaced spectral lines of pure vibrational-rotational transitions illustrated in figure 5.5, we expect to see rotational spectral features with variations in inter-line spacing exceeding those of the anharmonic non-rigid oscillator in figure 5.7. Indeed, it is commonly observed that, for instance, the R-branch first extends toward higher frequencies, then comes to a halt, before turning toward lower frequencies again with increasing rotational quantum numbers N. This follows from the expressions for the

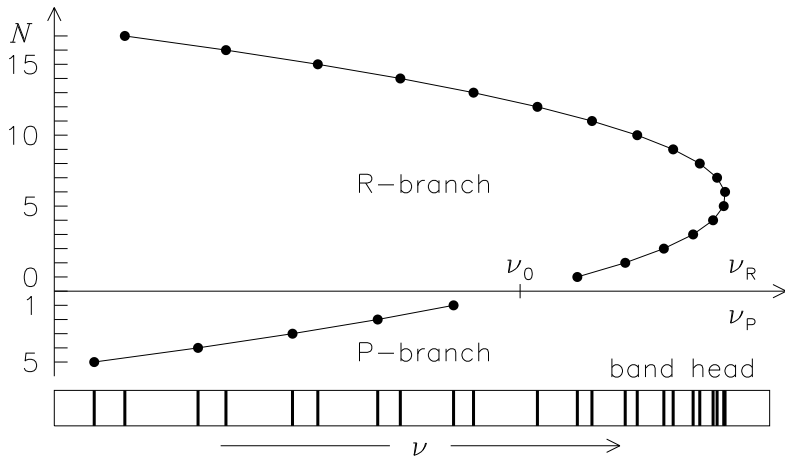

Figure 5.9: Formation of band heads

frequencies of the individual lines of the P- and R-branches as functions of N. Thus, it will be seen from (5.27) and (5.28) that if $B_1 < B_0$ and the difference in these values is large enough, both $\nu_P(N)$ and $\nu_R(N)$ will eventually both decrease with increasing values of N. This effect leads to the formation of *band heads* in the spectrum, that is, the convergence of individual rotational spectral lines toward a limiting frequency. This phenomenon is illustrated in figure 5.9 for the case $(B_1 - B_0)/B_0 = -.14$. For clarity the P- and R-branches have been plotted "back to back" along the N-axis.

5.6 Comments on Polyatomic Molecules

The harmonic oscillator model for the diatomic molecule easily extends to polyatomic molecules. Instead of two nuclear masses and one connecting "spring", the M-atomic molecule will consist of M nuclear masses connected by a larger number of "springs" of different strengths, one spring for each pair of forces acting between the nuclei. In a 3-atomic linear molecule, like HCN, the number of springs is 2. Linear here means that the molecule is configured such that the three nuclei lie on a straight line. The more general 3-atomic molecule, in which the three nuclei lie at the vertices of a triangle, like H_2O, requires three springs. Each spring corresponds to one harmonic oscillator with its own set of equidistant energy levels. Of the $3M$ coordinates necessary to specify the positions of the nuclei in the most general M-atomic molecule, 3 may be chosen to describe the center-of-mass motion of the molecule, another 3 to describe the orientation of the molecule. The remaining $3M - 6$ coordinates can always be chosen to correspond to independent harmonic oscillators. Selection rules similar to the ones derived above will apply to each of these oscillators. The result is therefore that we would expect one set of vibrational-rotational spectral lines for each oscillator. The linear HCN molecule, identified in interstellar spectra, show spectral lines at $\lambda^{-1} = 2089$ and $3312\ \text{cm}^{-1}$. For H_2O with the two HO-bonds making an angle of 105° with each other, three vibrational spectral features are found at $\lambda^{-1} = 1545, 3652$ and $3756\ \text{cm}^{-1}$. The HCN and H_2O molecules with the corresponding spring models are illustrated in figure 5.10.

The rigid rotor model for the diatomic molecule also extends to polyatomic molecules. This extension is simple in the case of "symmetric top" molecules, that is, molecules for

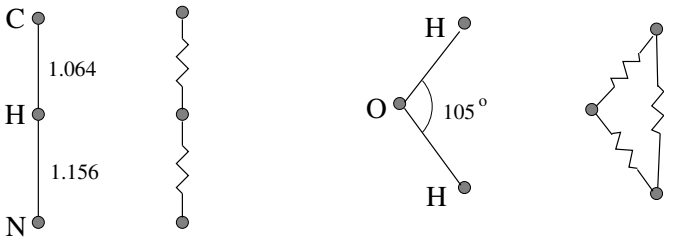

Figure 5.10: Linear and arbitrary 3-atomic molecules

which orthogonal x, y and z-axis may be found such that the moments of inertia I_x, I_y and I_z of the molecule with respect to these axis satisfy

$$I = I_x = I_y \neq I_z.$$

The classical rotational energy for such molecules takes the form

$$H = \frac{N_x^2}{2I_x} + \frac{N_y^2}{2I_y} + \frac{N_z^2}{2I_z} = \frac{1}{2I}\boldsymbol{N}^2 - (\frac{1}{2I} - \frac{1}{2I_z})N_z^2.$$

The operators $\boldsymbol{N}^2$ and N_z commute. The corresponding physical quantities are therefore simultaneously measurable and we may immediately write down the expression for the allowed rotational energy levels

$$W_{NM_N} = \frac{\hbar^2}{2I}N(N+1) - \frac{\hbar^2}{2}(\frac{1}{I} - \frac{1}{I_z})M_N^2. \tag{5.29}$$

The angular part of the nuclear wave function for the symmetric top molecule is identical to that of the rigid rotor as given in (5.19), but the z-axis is now fixed to the axis of rotational symmetry of the molecule. The extra degrees of freedom with respect to energy levels expressed by (5.29) naturally leads to additional richness as regards rotational spectral features. The CH_3CN molecule which was discovered in the interstellar medium in 1971 is an example of a symmetric top molecule. The molecule is illustrated in figure 5.11. The three H atoms are located at the vertices of an equilateral triangle, the three C-H bonds forming an angle of 109° with each other.

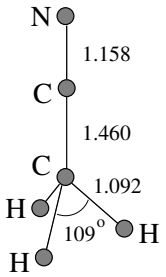

Figure 5.11: The CH_3CN molecule

5.7 Coupling of Angular Momenta in Molecules

In our discussion of the the rigid rotor any coupling with the electrons was neglected. This is clearly an oversimplification. The electrons may have both orbital and spin angular momenta. We are also from our discussion of atomic physics well acquainted with complicated ways of combining angular vectors. Below we shall only give a few comments on the corresponding angular momentum problem for molecules.

Let us this time start with the electrons. For the electrons of the molecule the situation is different from that of electrons in individual atoms. With the spherically symmetric potential of the atom, the square of the electron orbital angular momentum $\boldsymbol{L}^2$ may be quantized together with H and L_z. In the diatomic molecule the electric field of the two nuclei introduces a rotational symmetry around the molecular axis. The component of L_z along this axis may thus still be quantized together with H. But $\boldsymbol{L}^2$ is no longer a good quantum number. The quantum number corresponding to L_z is traditionally denoted Λ,

$$L_z\Phi = \pm\Lambda\Phi, \quad \Lambda = 0, 1, 2, \cdots . \tag{5.30}$$

The electron energy is independent of the choice sign in (5.30). This introduces a double energy degeneracy for $\Lambda \neq 0$.

Λ	0	1	2	3
symbol	Σ	Π	Δ	Φ

Table 5.2: Molecular spectroscopic notation

Molecular electron states are identified, according to the value of Λ, with capital Greek letters as shown in table 5.2. The notation is analogous to the situation for atoms except for the fact that it is now the azimuthal component of the orbital angular momentum Λ which is the ordering agency instead of the total orbital angular momentum L.

Electron states of the molecule corresponding to identical value of Λ are distinguished by a Roman capital letter preceding the Greek symbol for the value of Λ. The ground state is represented by the letter X, the following states by A, B, $C, \cdots$ in order of increasing energy.

The spin of the individual electrons combine to form the total electron spin angular momentum $\boldsymbol{S}$. As usual $\boldsymbol{S}^2$ may be quantized together with H with eigenvalues $\hbar^2 S(S+1)$, the spin quantum number S taking integer or half-integer values depending on the number of electrons in the molecule. Provided spin-orbit interactions are neglected there is a $2S+1$ energy degeneracy associated with S. The value $2S+1$ is the multiplicity of the electron state and appears as a subscript immediately preceding the Greek symbol for the value of Λ in molecular spectroscopic notation.

The spin angular momentum is not effected by the molecular electric field. If the molecule does not rotate and $\Lambda = 0$, the spin angular momentum does not show any preference for the molecular axis. If $\Lambda \neq 0$, however, the electron orbital motion induces an internal magnetic field in the direction of the molecular axis. This internal magnetic field causes a precession of $\boldsymbol{S}$ around this axis, the axial component taking values $\hbar\Sigma$. Σ is the azimuthal spin quantum number and takes values $-S, -S+1, \cdots, S$. The azimuthal spin quantum number Σ should not be confused with the Greek symbol Σ for electron states in table 5.2.

The electron orbital and spin angular momenta $\boldsymbol{L}$ and $\boldsymbol{S}$ may couple with the nuclear angular momentum $\boldsymbol{N}$ at right angles with the molecular axis to form the total angular momentum

$$\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S} + \boldsymbol{N}$$

of the molecule in different ways. The *Hund rules* specify which combinations are possible. Two common cases are shown in figure 5.12.

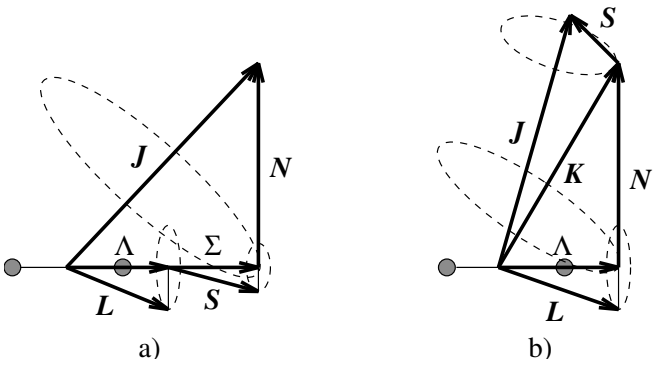

Figure 5.12: Coupling of angular momenta according to Hund's rules (a) and (b)

Hund's rule (a) applies where a strong nuclear electric field causes a rapid precession of $\boldsymbol{L}$ around the molecular axis, while a strong induced magnetic field causes a similar rapid precession of $\boldsymbol{S}$ around the molecular axis. The effective values of $\boldsymbol{L}$ and $\boldsymbol{S}$ are then both collinear with the molecular axis, giving rise to an effective azimuthal quantum number

$$\Omega = |\,\Lambda + \Sigma\,|\,.$$

The collinear and strongly coupled effective $\boldsymbol{L}$ and $\boldsymbol{S}$ vectors then couple with the perpendicular $\boldsymbol{N}$ vector to give total angular quantum numbers of the molecule as

$$J = \Omega, \Omega + 1, \Omega + 2, \cdots\,.$$

Hund's rule (b) applies where the coupling between $\boldsymbol{S}$ and the molecular axis is weak. This decoupling takes place for Σ states ($\Lambda = 0$). The orbital angular momentum $\boldsymbol{L}$ may still be strongly coupled with the molecular axis, making Λ a good quantum number. The effective $\boldsymbol{L}$ and $\boldsymbol{N}$ then couple to form a resultant angular vector $\boldsymbol{K}$ with integer quantum numbers

$$K = \Lambda, \Lambda + 1, \Lambda + 2, \cdots\,.$$

The total angular momentum $\boldsymbol{J}$ is finally obtained by coupling $\boldsymbol{K}$ and $\boldsymbol{S}$ with quantum numbers

$$J = |\,K - S\,|, |\,K - S\,| + 1, \cdots, K + S.$$

Chapter 6

Thermal and Statistical Physics

Thermal physics is the branch of physics that deals with the properties of systems in thermal equilibrium. It describes the state of systems that have been brought in contact with each other and been allowed to exchange energy or particles. It describes types of processes that systems may undertake when some parameter of the system is changed. Characteristic of thermal physics is that very general and far-reaching conclusions on the properties of systems can be achieved on the basis of very simple statistical assumptions.

6.1 Probability Concepts

Before embarking on our discussion of thermal physics, it will be useful to review some basic concepts and definitions from the theory of probability. We introduce these through the discussion of a simple experiment: throwing darts toward a target on the wall. Let the center of the target be located at the origin of a Cartesian xy-coordinate system. Let X and Y represent the "coordinates of the dart on target". X and Y are examples of *stochastic variables*. Before a throw the values of X and Y are not known. We will indicate the result of a particular throw i by coordinates x_i and y_i. In statistical jargon, x_i and y_i represent a *realization* of the stochastic variables X and Y. In the next throw, X and Y will almost certainly find another realization. In principle, x_i and y_i may both take any real value. The *state space* of our particular experiment is therefore two-dimensional, real-valued and continuous.

Now consider the target divided into rectangular cells with sides $\mathrm{d}x$ and $\mathrm{d}y$. Each cell is recognized by for instance its center coordinates (x, y). After throwing a large number N of darts, a player will find $\mathrm{d}N(x, y)$ hits in the cell with center at (x, y). She may therefore assign the number $\mathrm{d}\mathcal{N}(x, y)/N$ as the *probability* of hitting this particular cell. If a new series of N throws were made, a different result would be expected. One would argue that as the number N increases, the difference in the results of two consecutive series of N throws for a given player would decrease. This being the case, we could assign a *probability density function* $f_{XY}(x, y)$ to the particular experiment such that $N f_{XY}(x, y)\,\mathrm{d}x\,\mathrm{d}y$ represents the *expected* number of hits in the given cell after N throws. For convenience, we shall in the following refer to probability density functions also as *distribution functions*.

The distribution function $f_{XY}(x, y)$ must satisfy two constraints to be acceptable. First, it will have to be *positive definite*,

$$f_{XY}(x, y) \geq 0. \tag{6.1}$$

We cannot think of a negative number of hits in any cell. Secondly, any valid throw will have to hit somewhere, that is, summing over all cells, all throws should be accounted for. Mathematically, this means that the distribution function must be normalized,

$$\int f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1. \tag{6.2}$$

It is now useful to introduce some additional concepts related to our distribution function $f_{XY}(x,y)$. We could devise a new experiment in which we were only interested in studying one of the coordinates of the dart on target, for instance the X variable. The distribution function for the new experiment can be derived from the old one by summing over all possible outcomes of the now non-interesting Y variable,

$$f_X(x) = \int f_{XY}(x,y)\,\mathrm{d}y. \tag{6.3}$$

The distribution function $f_X(x)$ is an example of a *marginal* probability density function. It will automatically satisfy the requirements of being positive definite and normalized. In figure 6.1 a distribution function $f_{XY}(x,y)$ is illustrated by equi-contours. The two marginal distributions $f_X(x)$ and $f_Y(y)$ result through integrating $f_{XY}(x,y)$ along the y- and x-axis, respectively.

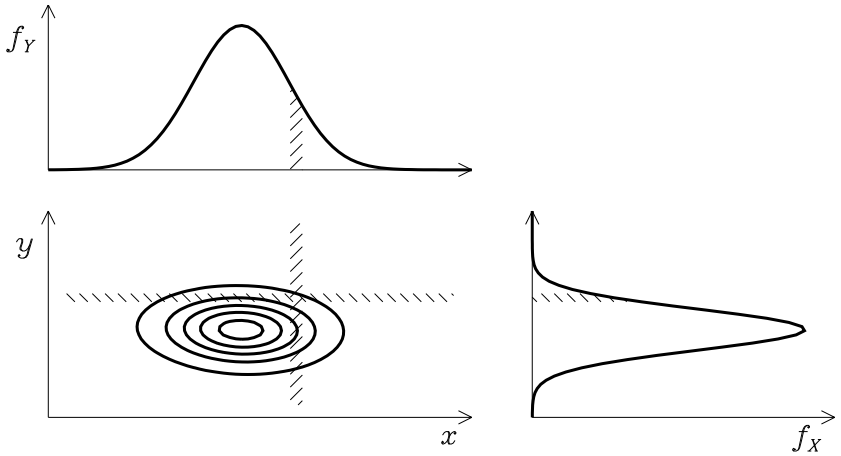

Figure 6.1: A distribution and the corresponding marginal distributions

A given function $g(X,Y)$ of the two stochastic variables X and Y will take different values in each realization. A measure for the expected value of g for a particular realization can be found by weighting the different outcomes $g(x,y)$ with the probability for that particular outcome, that is $f_{XY}(x,y)\,\mathrm{d}x\mathrm{d}y$, and then summing over all possible outcomes. This procedure defines the *mean value* of the given function,

$$\langle\, g(X,Y)\,\rangle = \int g(x,y) f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y. \tag{6.4}$$

For the simple choice $g(X,Y) = X$ we thus have

$$\langle\, X\,\rangle = \int x f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int x f_X(x)\,\mathrm{d}x.$$

The uncertainty in the measurement of X is often represented by the *standard deviation* σ_X defined by

$$\sigma_X^2 \equiv \langle\, (X - \langle\, X \,\rangle)^2 \,\rangle = \int (x - \langle\, X \,\rangle)^2 f_X(x)\, \mathrm{d}x.$$

The quantity σ_X^2 is called the *variance* of the stochastic variable X. The standard deviation σ_X is a measure of the typical width of the distribution function $f_X(x)$ to either side of the mean value $\langle\, X \,\rangle$.

The stochastic variables X and Y will be called *statistically independent* if

$$f_{XY}(x, y) = f_X(x)\, f_Y(y). \tag{6.5}$$

Statistical independence of two stochastic variables means that we would not be better off in predicting the outcome of the experiment in one variable even if we were given information as to the corresponding outcome of the other variable. Similarly, two stochastic variables X and Y are called *statistically uncorrelated* if

$$\langle\, XY \,\rangle = \langle\, X \,\rangle \langle\, Y \,\rangle. \tag{6.6}$$

Two stochastic variables that are statistically independent will also be statistically uncorrelated. Two stochastic variables that are statistically uncorrelated need *not* be statistically independent. In figure 6.2 examples of two collections of realizations (scatters plots) for two stochastic variables X and Y are illustrated. For case a) it would look like we would not be better off in predicting the value of Y in a given realization if we were informed on the corresponding value of X. This indicates that X and Y are uncorrelated variables and even that they might possibly be statistically independent. For case b) for which small values of X tend to be accompanied by corresponding small values of Y, this is not true.

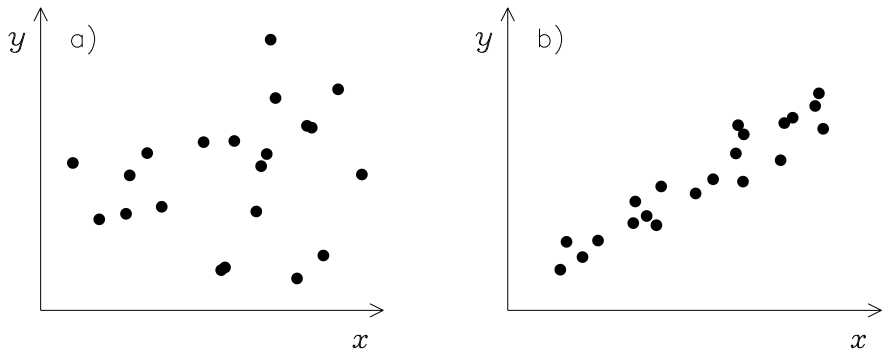

Figure 6.2: Scatter plots

It is sometimes convenient to work with distribution functions that are not normalized to unity as required by (6.2), but to another positive constant value N,

$$\int f_{XY}(x, y)\, \mathrm{d}x\, \mathrm{d}y = N. \tag{6.7}$$

In this case some of the formulas above need to be modified. In particular, the formula for the mean value (6.4), will now read

$$\langle\, g(X,Y)\,\rangle = \frac{\int g(x,y) f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y}{\int f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y}, \tag{6.8}$$

with similar modifications for $\langle\, X\,\rangle$ and σ_X^2.

In the above discussion the particular realizations of the stochastic variables were assumed to be real-valued – possibly within a limited range. That is, we assumed the state space of the experiment to be real-valued and continuous. In a different version of the experiment we may be interested in knowing which of a predetermined set of cells is hit and not the particular coordinates x and y within the cell. Naming the cells by an integer index i the state space of the modified experiment is integer-valued and thus discrete. The probability $\mathcal{P}_i$ of hitting cell i may again be found as the limiting ratio of the number of hits N_i in cell i and the total number of throws N, that is, $\mathcal{P}_i = N_i/N$. The discrete probabilities $\mathcal{P}_i$ must satisfy constraints similar to those of the continuous probability density function discussed above. The discrete probabilities must be positive definite,

$$\mathcal{P}_i \geq 0 \quad \text{for all } i. \tag{6.9}$$

The discrete probabilities must also be normalized,

$$\sum_i \mathcal{P}_i = 1, \tag{6.10}$$

where the summation is carried out over all possible outcomes. The mean value of any discrete function g_i defined over the set of cells is again found by weighting the different realization with the probability for that particular outcome

$$\langle\, g\,\rangle = \sum_i g_i \mathcal{P}_i. \tag{6.11}$$

Quiz 6.1 : In the measurement of the quantity $Z = f(X,Y)$, the quantities X and Y are measured separately and independently. Due to different measurement errors, each time a measurement is performed different results are found. We therefore consider X, Y and Z as stochastic variables.

If we consider the "true" value and the "typical" measurement error for X to be $\mu_X = \langle\, X\,\rangle$ and σ_X, and similarly for Y, what are the corresponding values for $Z = X + Y$?

If instead $Z = XY$ verify that $\mu_Z = \mu_X \mu_Y$ and $\sigma_Z^2 = \mu_Y^2\sigma_X^2 + \mu_X^2\sigma_Y^2 + \sigma_X^2\sigma_Y^2$. Thus, if the relative errors σ_X/μ_X and σ_Y/μ_Y are both small, show that the square of relative error of $Z = XY$ is the *sum* of the squares of the relative errors of X and Y, that is, the factor with the largest relative error dominates the relative error of the product.

6.2 Entropy and Temperature

Thermal physics adopts the view that at the microscopic level any physical system is only allowed to occupy one of a discrete set of quantum states – in accordance with quantum mechanics. As systems we may consider a volume of gas, a single atom or even a single electron orbital. A *closed system* is one for which energy U and particle number N are given and for which all external parameters, for instance volume, are held constant. We will in the following denote all such parameters by the symbol V. An *accessible state* for the system is one quantum state that complies with the constrains imposed on the system.

At the basis of thermal physics lies the *fundamental statistical assumption* that *any accessible state of a given closed system is equally probable*. The assumption of equal probability for any accessible state is the only realistic assumption that can be made if specific information to the contrary does not exist.

The number of accessible states for a given system Σ_i with given values of N_i, U_i and V_i will be denoted the *multiplicity* $g_i(N_i, U_i, V_i)$ of the system. The multiplicity g_n of a single atom with energy W_n equals the number of quantum states with energy W_n, that is, multiplicity is identical to the quantum mechanical degeneracy of the given energy state. For single atoms the multiplicities involved are typically small number, for the agregates of atoms we are now dealing with (a volume of gas or liquid or a solid block) the corresponding multiplicities are truely huge numbers!

Let the closed system Σ consist of two closed subsystems Σ_1 and Σ_2 with given individual values of particle numbers, energies and external parameters and with individual multiplicities $g_1(N_1, U_1, V_1)$ and $g_2(N_2, U_2, V_2)$. The situation is illustrated in the left-hand part of figure 6.3. The multiplicity of the total system Σ is given as

$$g(N_1, N_2, U_1, U_2, V_1, V_2) = g_1(N_1, U_1, V_1)\, g_2(N_2, U_2, V_2). \tag{6.12}$$

This result implies that the two subsystems Σ_1 and Σ_2 are at most only *weakly interacting*, that is, the Pauli principle does not pose significant additional constraints on the combined system, contraints in addition to those existing for each subsystem taken isolated. The two subsystems are now brought in *thermal contact*, that is, the two subsystems are allowed to exchange energy while maintaining the total energy $U = U_1 + U_2$, as illustrated in the right-hand part of figure 6.3. The multiplicity of system Σ must now be expressed as the sum of multiplicities for each allowed distribution of the energy between the two subsystems

$$g(N_1, N_2, U, V_1, V_2) = \sum_{U_1' + U_2' = U} g_1(N_1, U_1', V_1)\, g_2(N_2, U_2', V_2). \tag{6.13}$$

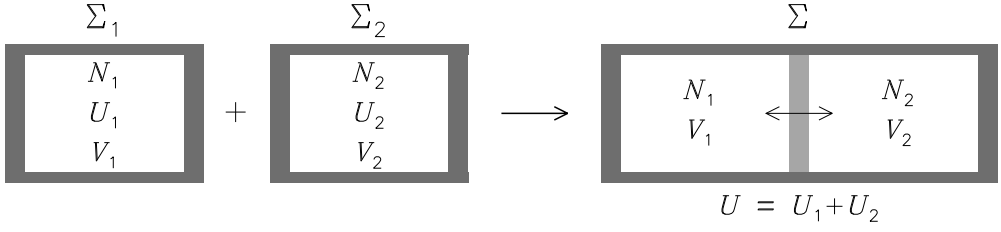

Figure 6.3: Systems in thermal contact

The *most probable* distribution of the total energy U between the two subsystems in thermal contact is the distribution $U_1' = U_1^*$, $U_2' = U_2^*$ corresponding to the maximum contribution to the total multiplicity g of the combined system. This particular distribution is determined by the relation

$$\frac{\partial g_1}{\partial U_1} g_2 \, \mathrm{d}U_1 + g_1 \frac{\partial g_2}{\partial U_2} \, \mathrm{d}U_2 = 0 \quad \text{with} \quad \mathrm{d}U_1 + \mathrm{d}U_2 = 0,$$

or equivalently,

$$\frac{\partial \ln g_1}{\partial U_1} = \frac{\partial \ln g_2}{\partial U_2}. \tag{6.14}$$

The number of accessible states for typical systems to be encountered are very large numbers. In particular, the number of accessible states (6.13) for system Σ is normally very much larger that the product of the numbers of accessible states (6.12) for the original systems Σ_1 and Σ_2. It is now a most fortunate fact that of the different terms appearing on the right hand side of (6.13), it is normally one term, namely that corresponding to the most probable distribution of energy between subsystems, that completely dominates the sum of the other terms. This property has been verified for systems for which detailed analytical results are available. We shall make use of this fact in the following, that is, we shall assume that the total multiplicity of the closed system Σ consisting of the subsystems Σ_1 and Σ_2 in thermal contact can be approximated as

$$g(N_1, N_2, U, V_1, V_2) \approx \max_{U_1' + U_2' = U} g_1(N_1, U_1', V_1) \, g_2(N_2, U_2', V_2) = g_1(N_1, U_1^*, V_1) \, g_2(N_2, U_2^*, V_2). \tag{6.15}$$

At this stage it is useful to introduce two important quantities. The *entropy* $\mathcal{S}$ of a system Σ with multiplicity g is defined as

$$\mathcal{S}(N, U, V) \equiv \ln g(N, U, V). \tag{6.16}$$

The entropy $\mathcal{S}$ is a dimensionless quantity. From (6.15) we note in particular that the entropy of the closed system Σ consisting of the subsystems Σ_1 and Σ_2 in thermal contact is given as the sum of the entropies of the two subsystems

$$\mathcal{S}(N_1, N_2, U, V_1, V_2) \approx \mathcal{S}_1(N_1, U_1^*, V_1) + \mathcal{S}_2(N_2, U_2^*, V_2). \tag{6.17}$$

The *fundamental temperature* $\mathcal{T}$ of a the system Σ with entropy $\mathcal{S} = \mathcal{S}(N, U, V)$ is defined by

$$\frac{1}{\mathcal{T}} \equiv \frac{\partial \mathcal{S}}{\partial U} \Big|_{N,V} . \tag{6.18}$$

Alternatively, if the energy of the system is expressed in terms of particle number, entropy and external parameters, $U = U(N, \mathcal{S}, V)$, the fundamental temperature will be given by

$$\mathcal{T} = \frac{\partial U}{\partial \mathcal{S}} \Big|_{N,V} . \tag{6.19}$$

The subscripts $|_{N,V}$ in (6.18) and (6.19) indicate which variables are to be kept constant while evaluating the partial derivatives. Indirectly, this notation is a reminder that the entropy $\mathcal{S}$ in the definition (6.18) is expected to be expressed as an explicit function of N, U and V before

the differentiation is performed. The similar comment applies to the dependent variable U in the alternative definition (6.19). From its definition we see that fundamental temperature $\mathcal{T}$ has energy as physical dimension.

Temperature is traditionally measured in degrees Kelvin. Traditional temperature T is related to the fundamental temperature $\mathcal{T}$ through

$$\mathcal{T} \equiv \kappa T,$$

where κ is the *Boltzmann constant*, $\kappa = 1.38054 \cdot 10^{-23}$ J/K. We note that the constant κ has no physical significance except for its role in connecting the Kelvin temperature scale with the unit of energy. When making use of the traditional definition of temperature it is common to introduce the corresponding redefined entropy

$$S \equiv \kappa \mathcal{S}.$$

Fundamental and traditional temperature and entropy will be used freely in the following discussion. They will both be referred to simply as temperature or entropy.

With these definitions at hand let us return to our discussion of the most probable energy distribution of the two subsystems in thermal contact. In terms of entropy and temperature the condition (6.14) reduces to

$$\frac{1}{\mathcal{T}_1} = \frac{\partial \mathcal{S}_1}{\partial U_1} = \frac{\partial \mathcal{S}_2}{\partial U_2} = \frac{1}{\mathcal{T}_2}. \tag{6.20}$$

The most probable energy distribution between two closed subsystems in thermal contact is thus the one for which the temperatures of the subsystems are identical, $\mathcal{T}_1 = \mathcal{T}_2$. We shall refer to this result as the condition for *thermal equilibrium* of the two subsystems in thermal contact.

Quiz 6.2 : On the basis of the assumption (6.13) argue that the entropy will in general increase as two subsystems Σ_1 and Σ_2 are brought in thermal contact.

6.3 The Boltzmann Distribution

With the definitions of entropy, temperature and thermal equilibrium at hand, we are ready for one of the most fundamental results of thermal physics. Let us consider a closed system with energy U_0, consisting of a system Σ in thermal contact with a large reservoir $\mathcal{R}$ at temperature $\mathcal{T}$, as illustrated in figure 6.4. We assume that the energy of the reservoir is much larger than that of the system Σ. System Σ may for instance be a single atom. We seek an explicit expression for the probability $\mathcal{P}_s(W_s)$ that the system Σ will be found in a particular quantum state s corresponding to the energy W_s.

With the state s of the system Σ specified, the multiplicity of the total system[12] equals the multiplicity $g_{\mathcal{R}}(U_0 - W_s)$ of the reservoir $\mathcal{R}$. The probability that the system Σ will be

[2]We note in particular that the assumption behind (6.15) is not needed here.

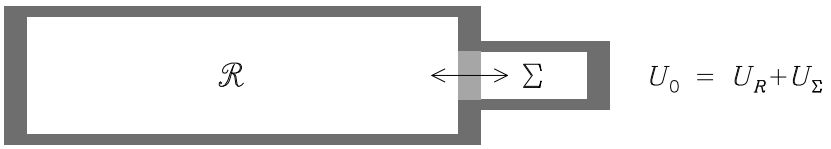

Figure 6.4: Simple system in thermal contact with reservoir

found in the given state s must therefore satisfy

$$\begin{aligned}\mathcal{P}_s(W_s) \sim g_{\mathcal{R}}(U_0 - W_s) &= \exp\left(\mathcal{S}_{\mathcal{R}}(U_0 - W_s)\right) \\ &\approx \exp\left(\mathcal{S}_{\mathcal{R}}(U_0) - \frac{\partial \mathcal{S}_{\mathcal{R}}}{\partial U_0} W_s\right) \sim \exp\left(-\frac{W_s}{\mathcal{T}}\right).\end{aligned}$$

The Taylor series expansion of the entropy expression is valid as long as the energy W_s is small enough compared with the total energy U_0. Taking the normalization requirement into account, this result can be written as

$$\mathcal{P}_s(W_s) = \frac{1}{Z} \exp\left(-\frac{W_s}{\mathcal{T}}\right), \tag{6.21}$$

where the normalization factor

$$Z = \sum_s \exp\left(-\frac{W_s}{\mathcal{T}}\right) \tag{6.22}$$

is the *partition function* (die Zustandsumme) for system Σ. The summation in (6.22) extends over all states s available to the system. The factor $\exp(-W_s/\mathcal{T})$ will be referred to as the *Boltzmann factor*. The discrete probability distribution function (6.21) is the *Boltzmann distribution*.

With the distribution function given, the average energy of the system Σ may now be calculated in accordance with (6.11) as

$$U = \sum_s W_s \mathcal{P}_s(W_s) = \mathcal{T}^2 \frac{\partial \ln Z}{\partial \mathcal{T}}\Big|_V. \tag{6.23}$$

This result exhibits the intimate relationship between the partition function Z and thermodynamic state variables of the system.

Quiz 6.3 : What is the probability $\mathcal{P}_{W_s}(W_s)$ for finding the system in energy state W_s, that is, in any state s corresponding to the specified energy W_s?

6.4 Particles in a Box

At this point let us pause for some simple, yet important examples. We start with the simplest of them all: one particle in a box.

Consider a free particle with mass m and without any internal structure confined to a box with sides of length L. The potential energy of the particle vanishes inside the box and is infinite outside. The properties of the particle in the box are determined by the solution of the Schrödinger equation

$$-\frac{\hbar^2}{2m}\nabla^2\Psi = W\Psi,$$

with the condition $\Psi = 0$ imposed at the boundary of the box. The solution, which we shall also refer to as an *orbital*, is

$$\Psi(\boldsymbol{r}) = \sqrt{\frac{8}{L^3}} \sin\frac{\pi q_x x}{L} \sin\frac{\pi q_y y}{L} \sin\frac{\pi q_z z}{L},$$

with the corresponding energy

$$W_{\boldsymbol{q}} = \frac{\pi^2\hbar^2}{2mL^2}(q_x^2 + q_y^2 + q_z^2) = \frac{\pi^2\hbar^2}{2mL^2}\,\boldsymbol{q}^2.$$

The quantum numbers q_x, q_y and q_z are all positive integers. The square of the particle momentum is a measurable quantity with the value

$$\boldsymbol{p}^2 = \frac{\pi^2\hbar^2}{L^2}(q_x^2 + q_y^2 + q_z^2) = \frac{h^2}{4L^2}\,\boldsymbol{q}^2. \tag{6.24}$$

To evaluate the partition function Z_1 for this one-particle system, the Boltzmann factor needs to be summed over all available states. For the case that the energy difference between neighboring states remains small compared to $\mathcal{T}$, this sum can be evaluated as an integral

$$\sum_s \cdot = \sum_{q_x=1}^{\infty}\sum_{q_y=1}^{\infty}\sum_{q_z=1}^{\infty} \cdot \to \frac{1}{8}\int_0^\infty 4\pi q^2\,\mathrm{d}q\,\cdot = \frac{V}{h^3}\int_0^\infty 4\pi p^2\,\mathrm{d}p\,\cdot\,. \tag{6.25}$$

Here we made use of the "spherical symmetry" of the Boltzmann factor in the quantum number $\boldsymbol{q}$-space, and that the density of modes in $\boldsymbol{q}$-space is unity, that is, the number of states contained within a spherical shell of radius q and thickness $\mathrm{d}q$ is $4\pi q^2/8$. The factor $1/8$ is introduced since the three quantum numbers can only take positive values. The right hand expression in (6.25) follows from (6.24) with $V = L^3$. The partition function is now easily evaluated as

$$Z_1 = \frac{V}{h^3}\int_0^\infty \exp(-\frac{p^2}{2m\mathcal{T}})\,4\pi p^2\,\mathrm{d}p = V n_{\mathrm{Q}} \tag{6.26}$$

where

$$n_{\mathrm{Q}} \equiv \left(\frac{2\pi m\mathcal{T}}{h^2}\right)^{3/2} \tag{6.27}$$

is called the *quantum concentration*.

Let us now generalize our system to include N non-interacting, identical particles. The Boltzmann distribution (6.21) is valid also for the N particle system, but the specification of each quantum state s available to the system must now include the specification of three quantum numbers for each particle. The partition function Z_N for the N particle system may still be easily evaluated. In fact, the sum over all available quantum states will consist of a product of sums over available states for each particle,

$$Z_N = \frac{1}{N!}\,Z_1^N. \tag{6.28}$$

The extra factor $N!$ in (6.28) has to be included because the particles were assumed to be identical. If two given particles for a given quantum state s interchanged names (or

numbers), we would not be able to recognize the difference. The new state should therefore not be counted as an independent state of the system. And N identical particles can be interchanged in $N!$ different ways. For large values of N the extra factor can be expressed in terms of the Stirling approximation

$$\ln N! \approx N \ln N - N. \tag{6.29}$$

Making use of (6.26) and (6.29) leads to

$$\ln Z_N \approx N \left(\ln(\frac{V}{N} n_{\mathrm{Q}}) + 1 \right). \tag{6.30}$$

As a first application of this result, the average energy U for the system may be calculated by making use of (6.23). A simple calculation leads to

$$U = \frac{3}{2} N \mathcal{T}. \tag{6.31}$$

Each particle in the gas has an average kinetic energy $U/N = 3\mathcal{T}/2$, that is, apart from the factor $3/2$ the temperature $\mathcal{T}$ is a measure of the kinetic energy in the "thermal" motions of an average gas particle.

6.5 The Maxwell Velocity Distribution

The Boltzmann distribution is a discrete probability distribution function over the quantum states of the free particle. It can be transformed into a continuous probability density function by summing the probability contributions from all quantum states corresponding to a certain phase space element, for example, over all quantum numbers $\boldsymbol{q}$ such that the corresponding absolute value of the particle momentum $\mid \boldsymbol{p} \mid$ falls in the interval $(p, p + \mathrm{d}p)$. The same reasoning underlying (6.25) leads to

$$f_p(p)\,\mathrm{d}p = \sum_{s \in (p, p+\mathrm{d}p)} \frac{1}{Z_1} \exp\left(-\frac{W_s}{\mathcal{T}}\right) = \frac{1}{V} (2\pi m \mathcal{T})^{-3/2}\, 4\pi p^2 \exp\left(-\frac{p^2}{2m\mathcal{T}}\right) V\,\mathrm{d}p.$$

Here $f_p(p)\,\mathrm{d}p$ represents the probability that the particle will be found (within the the total volume V) with absolute momentum in the interval $(p, p + \mathrm{d}p)$. We may thus also interpret

$$f_{\boldsymbol{r}p}(p) = n (2\pi m \mathcal{T})^{-3/2}\, 4\pi p^2 \exp\left(-\frac{p^2}{2m\mathcal{T}}\right) \tag{6.32}$$

as the probability density function for finding particles in $(\boldsymbol{r}, p)$-space. In this interpretation we implicitly assumed the probability density function to be *homogeneous* in $\boldsymbol{r}$-space, that is, we assumed the probability density function to be independent of $\boldsymbol{r}$. This is in accordance with the form of the Boltzmann distribution (6.21). In (6.32) we also introduced the *particle density* $n \equiv N/V$. This means that we have chosen to normalize the distribution function, this time not to unity, but to the total number of particles in the volume V,

$$\int_0^\infty \int_V f_{\boldsymbol{r}p}\,\mathrm{d}^3\boldsymbol{r}\,\mathrm{d}p = N.$$

This is a convenient and usual choice.

From (6.32), the probability density function $f_{\boldsymbol{rp}}$ in $(\boldsymbol{r},\boldsymbol{p})$-space may be derived if the momentum space is assumed to be *isotropic*. This assumption again conforms with the form of the Boltzmann distribution (6.21) and means that there is no preferred direction for the particle momentum $\boldsymbol{p}$. The probability density function in $(\boldsymbol{r},\boldsymbol{p})$-space must then be a function of p only, that is, $f_{\boldsymbol{rp}} = f_{\boldsymbol{rp}}(p)$. Since f_{rp} is uniquely determined as a marginal distribution of $f_{\boldsymbol{rp}}$,

$$f_{rp} = \iint f_{\boldsymbol{rp}}(p)\, p^2 \,\mathrm{d}\cos\theta \,\mathrm{d}\varphi = 4\pi p^2 f_{\boldsymbol{rp}}(p), \tag{6.33}$$

if follows that

$$f_{\boldsymbol{rp}}(p) = n(2\pi m\mathcal{T})^{-3/2} \exp\left(-\frac{p^2}{2m\mathcal{T}}\right). \tag{6.34}$$

For non-relativistic particles, velocity $\boldsymbol{v}$ and momentum $\boldsymbol{p}$ are related through $\boldsymbol{p} = m\boldsymbol{v}$ where mass m is a constant. With this relation, the probability density function in $(\boldsymbol{r},\boldsymbol{v})$-space may also be derived. With $f_{\boldsymbol{rv}}\,\mathrm{d}^3\boldsymbol{r}\,\mathrm{d}^3\boldsymbol{v} = f_{\boldsymbol{rp}}\,\mathrm{d}^3\boldsymbol{r}\,\mathrm{d}^3\boldsymbol{p}$ and $\mathrm{d}^3\boldsymbol{p} = m^3\,\mathrm{d}^3\boldsymbol{v}$, the result is

$$f_{\boldsymbol{rv}}(v) = n\left(\frac{m}{2\pi\mathcal{T}}\right)^{3/2} \exp\left(-\frac{mv^2}{2\mathcal{T}}\right). \tag{6.35}$$

The distribution functions (6.34) and (6.35) will be referred to as the *Maxwell momentum and velocity distributions*, respectively.

Some comments on notation may be useful at this point. The probability density function, for instance the Maxwell velocity distribution $f_{\boldsymbol{rv}}$, in combination with a state space element $\mathrm{d}^3\boldsymbol{r}\,\mathrm{d}^3\boldsymbol{v}$, that is $f_{\boldsymbol{rv}}(\boldsymbol{r},\boldsymbol{v})\,\mathrm{d}^3\boldsymbol{r}\,\mathrm{d}^3\boldsymbol{v}$, is interpreted as the expected number of particle in the given state space element at location $(\boldsymbol{r},\boldsymbol{v})$. The state space location is indicated by the argument of the density function, while the actual choice of state space, and therefore the physical dimension of the density function, is indicated by the function subscripts. As the function arguments may be used to indicate a particular symmetry property, for instance homogeneity or isotropy, the arguments and the subscripts may differ. In (6.35) the argument list has been modified to indicate that in thermal equilibrium we expect the density function to be homogeneous in $\boldsymbol{r}$-space and isotropic in $\boldsymbol{v}$-space. In situations where no confusion can arise as to which probability density function is involved, the function subscripts may safely be suppressed. In the literature this is often the case even if the choice of state space may not always be readily evident from the context.

Several useful marginal probability density functions can be derived from (6.35). Two examples are the distribution functions for each component of the velocity $\boldsymbol{v}$ and the speed v. An example of the former is given by

$$f_{\boldsymbol{r}v_x}(v_x) = \iint f_{\boldsymbol{rv}}(v_x, v_y, v_z)\,\mathrm{d}v_y\,\mathrm{d}v_z = n\sqrt{\frac{m}{2\pi\mathcal{T}}}\exp\left(-\frac{mv_x^2}{2\mathcal{T}}\right). \tag{6.36}$$

The latter is found by introducing spherical coordinates (v,θ,φ) in velocity space and integrating over the angular variables,

$$f_{\boldsymbol{r}v}(v) = \iint f_{\boldsymbol{rv}}(v)\, v^2 \,\mathrm{d}\cos\theta \,\mathrm{d}\varphi = n\left(\frac{m}{2\pi\mathcal{T}}\right)^{3/2} 4\pi v^2 \exp\left(-\frac{mv^2}{2\mathcal{T}}\right). \tag{6.37}$$

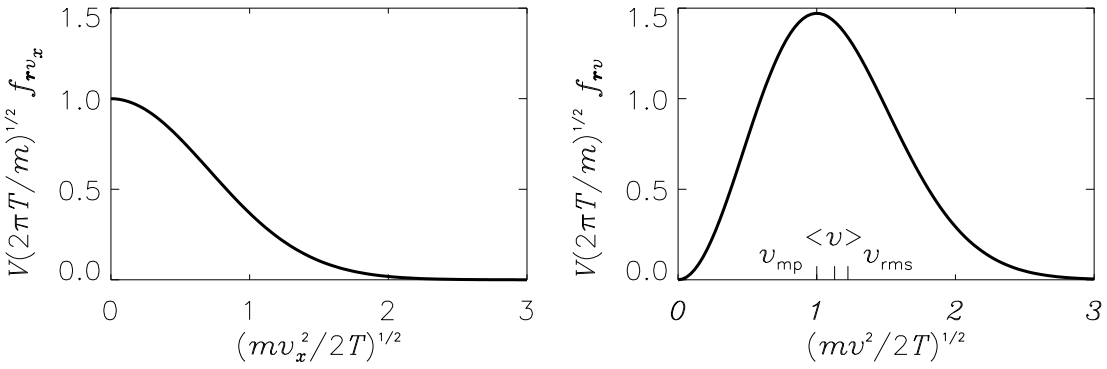

Figure 6.5: Velocity distribution functions

These functions in normalized form are plotted in figure 6.5. Since $f_{\boldsymbol{r}v_x}(v_x)$ is an even function of v_x, it has only been plotted for positive values of v_x.

With the velocity distribution function known, different average values may be calculated. For instance, $\langle\, v^2 \,\rangle$ may be calculated by making use of (6.37) and the general mean value definition (6.8) as

$$\langle\, v^2 \,\rangle = \frac{1}{N}\int_0^\infty \int_V v^2 f_{\boldsymbol{r}v}(v)\, \mathrm{d}^3\boldsymbol{r}\, \mathrm{d}v = \frac{3\mathcal{T}}{m}. \tag{6.38}$$

The factor $1/N$ appears because we have chosen to normalize our velocity distribution functions to the total particle number N in the volume V. In figure 6.5 the *mean* speed $\langle\, v \,\rangle$ and the *root mean square* speed $v_{\rm rms} \equiv \sqrt{\langle\, v^2 \,\rangle} = \sqrt{3\mathcal{T}/m}$ have been indicated. In the figure we have also indicated the *most probable* speed $v_{\rm mp}$, defined as the particle speed most often encountered, that is, the speed satisfying $\partial f_{\boldsymbol{r}v}(v)/\partial v = 0$. It is important to note that these three different ways of defining "expected" speed all differ!

Quiz 6.4 : What are the physical dimensions of $f_{\boldsymbol{rv}}$, $f_{\boldsymbol{r}v}$ and $f_{\boldsymbol{rp}}$?

Quiz 6.5 : Show that

$$\langle\, v \,\rangle = \sqrt{\frac{8\mathcal{T}}{\pi m}} \quad \text{and} \quad v_{\rm mp} = \sqrt{\frac{2\mathcal{T}}{m}}.$$

Quiz 6.6 : Calculate $\langle\, v_x \,\rangle$ and $\sqrt{\langle\, v_x^2 \,\rangle}$.

Quiz 6.7 : Derive $f_{\boldsymbol{r}w}(w)$ where $w = \frac{1}{2}mv^2$. Find $\langle\, w \,\rangle$ and $w_{\rm mp}$.

Quiz 6.8 : A gas consisting of atoms with mass m and emitting at electromagnetic radiation at frequency ν_0, has temperature $\mathcal{T}$. Because of thermal motions the spectral line at ν_0 has finite width. The normalized line profile $\phi(\nu)$ is determined by the relative number of atoms that is emitting in the interval $(\nu, \nu + \mathrm{d}\nu)$ as seen by a stationary observer. Express $\phi(\nu)$ in terms of ν_0, ν and $\Delta\nu_D = \nu_0\sqrt{\mathcal{T}/m}/c$. Evaluate the width of the spectral line as given by the standard deviation σ_ν. Can you think of alternative definitions of line width? What is the ratio of the thermal line widths for Fe and H lines in the same frequency range?

6.6 The Ideal Gas

Let us next consider a gas of identical particles without internal structure in a box. The particles will be reflected when hitting the walls of the box. Through this process the particles impart impulse to the wall. The sum of individual impulses per unit time and per unit wall area is equivalent to the gas *pressure* P. The number d^6N of particles with velocities in the cell $\mathrm{d}^3\boldsymbol{v}$ centered at $\boldsymbol{v}$ that will hit the wall element $\mathrm{d}^2\boldsymbol{\mathcal{A}} = \mathrm{d}^2\mathcal{A}\,\hat{\boldsymbol{n}}$ at position $\boldsymbol{r}$ during the time interval $\mathrm{d}t$ equals the number of such particles that are contained within a cylindrical volume with base $\mathrm{d}^2\boldsymbol{\mathcal{A}}$ and side $\boldsymbol{v}\,\mathrm{d}t$, that is,

$$\mathrm{d}^6N = f_{\boldsymbol{rv}}(\boldsymbol{r},\boldsymbol{v})\ |\boldsymbol{v}\cdot\mathrm{d}^2\boldsymbol{\mathcal{A}}|\ \mathrm{d}t\,\mathrm{d}^3\boldsymbol{v}.$$

The geometry is illustrated in figure 6.6a. We remember that the distribution function $f_{\boldsymbol{rv}}$ has been normalized to the particle number N. Each particle experiences during a reflection a change in momentum of magnitude

$$|\Delta\boldsymbol{p}| = 2\,|\boldsymbol{p}\cdot\hat{\boldsymbol{n}}|\,.$$

As seen from figure 6.6b, the momentum change $\Delta\boldsymbol{p}$ is directed normal to the wall. The contribution to the pressure from these particles is therefore

$$\mathrm{d}^3P = |\Delta\boldsymbol{p}|\,\frac{\mathrm{d}^6N}{\mathrm{d}^2\mathcal{A}\,\mathrm{d}t} = 2p_xv_xf_{\boldsymbol{rv}}(v)\,\mathrm{d}^3\boldsymbol{v}, \tag{6.39}$$

where the x-axis was chosen to be parallel to $\hat{\boldsymbol{n}}$.

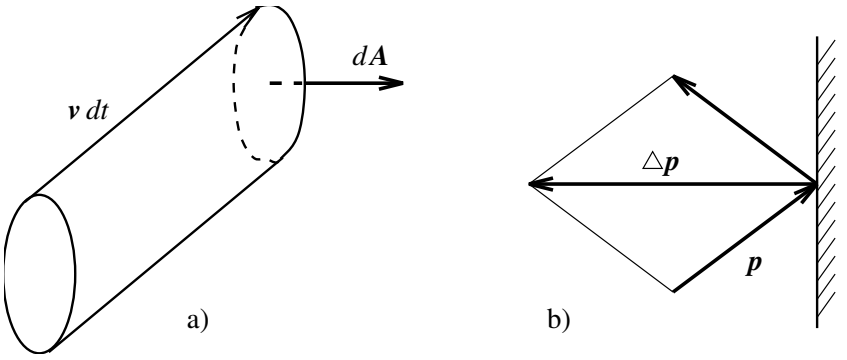

Figure 6.6: Geometry of impinging particles

The total pressure on the surface element is found by summing the pressure contributions from all approaching particles, that is, integrating over all values of v_y and v_z and all positive values of v_x. Since p_xv_x and $f_{\boldsymbol{rv}}(v)$ are both even functions of v_x, the velocity integrals can be extended over all velocity space by omitting the factor 2 in (6.39). Finally, again making use of the isotropy of the distribution function in velocity space, we may write

$$P = \int p_xv_xf_{\boldsymbol{rv}}(v)\,\mathrm{d}^3\boldsymbol{v} = \frac{1}{3}\int \boldsymbol{p}\cdot\boldsymbol{v}\,f_{\boldsymbol{rv}}(v)\,\mathrm{d}^3\boldsymbol{v}. \tag{6.40}$$

With $\boldsymbol{p} = m\boldsymbol{v}$ and $f_{\boldsymbol{rv}}(v)$ given by (6.35), the resulting pressure is

$$P = n\mathcal{T}. \tag{6.41}$$

The results, (6.41) and (6.31) are recognized as the *equation of state* and the *internal energy* of the *ideal gas*.

If several particle species are present in the system, these species contribute independently to the total pressure and the energy of the gas. In thermal equilibrium, the total result will be

$$P = \sum_i n_i \mathcal{T} \quad \text{and} \quad U = \frac{3}{2} V \sum_i n_i \mathcal{T},$$

where n_i is the particle density for species i and the summation is extended over all particle species. We note that each particle in the gas in thermal equilibrium contributes equally to pressure and energy, irrespective of particle mass m_i. An electron or an iron atom "at the same temperature" will contribute equally to the total pressure or internal thermal energy of the gas mixture.

The formulas for pressure and energy are sometimes expressed in terms of the *mass density* ρ_m and the *mean molecular weight* $\overline{\mu}$, defined by

$$\rho_m = \sum_i m_i n_i = \overline{\mu} m_{\rm H} \sum_i n_i, \tag{6.42}$$

where $m_{\rm H}$ is the mass of the hydrogen atom. In terms of these variables, (6.41) and (6.31) can be written

$$P = \frac{\rho_m}{\overline{\mu} m_{\rm H}} \mathcal{T} \quad \text{and} \quad U = \frac{3}{2} \frac{\rho_m V}{\overline{\mu} m_{\rm H}} \mathcal{T}. \tag{6.43}$$

Some comments on the validity of the ideal gas results may be appropriate at this point. The equation of state (6.41) and the energy equation (6.31) were both derived on the basis of a number of simplifying assumptions. For the energy equation (6.31) we assumed the particles to be point particles without internal structure. From our quantum mechanical discussion, we know that for atoms and molecules this assumption is not strictly valid. In addition to kinetic energy associated with their motion, atoms contribute to the total energy of the gas through their internal energy states. Molecules also contribute through their rotational and vibrational degrees of freedom. We will return to these questions in the next section.

Our derivation neglected the presence of inter-particle forces. These forces also contribute toward the total energy of the gas. In a neutral gas the inter-molecular forces have a short effective range, often falling off with inter-molecular distance proportional to $r^{-\lambda}$ with $\lambda = 6$. In an ionized gas (or plasma) the forces between individual charged particles have much longer range, $\lambda = 2$. For dilute gases and even dilute plasmas, the neglect of the inter-molecular forces will be acceptable at high enough temperatures. With increasing density and therefore smaller average inter-molecular distances, however, these forces eventually lead to phase transitions from the gas phase to the liquid or solid phases. For these situations the ideal gas results are definitely no longer valid. With these limitations in mind, the ideal gas laws (6.41) and (6.31) often represent good approximations to real gases and even plasmas. We shall make repeated use of these laws in the following.

Quiz 6.9 : Show that the average kinetic energy of the ideal gas equals the internal energy U as given by (6.31), that is,

$$\int \frac{1}{2} m \boldsymbol{v}^2 f_{\boldsymbol{rv}}(v)\, \mathrm{d}^3\boldsymbol{r}\, \mathrm{d}^3\boldsymbol{v} = U. \tag{6.44}$$

Also show that the pressure P and the kinetic energy U of any non-relativistic gas will satisfy the relation

$$P = \frac{2}{3}\frac{U}{V}, \tag{6.45}$$

whatever the detailed form of $f_{\boldsymbol{rv}}$.

Quiz 6.10 : Determine the mean molecular weight $\overline{\mu}$ for the following cases:
a) a fully ionized hydrogen gas,
b) a neutral gas consisting of 80 % H and 20 % He particles,
c) a neutral gas consisting of 80 % by mass of H and the rest He.
d) a fully ionized gas consisting of 80 % by mass of H and the rest He.

Quiz 6.11 : In astrophysics we often encounter fully ionized gases in which hydrogen and helium are by far the dominant species. Show that for these situations

$$\overline{\mu} \approx \left(2X + \frac{3}{4}Y + \frac{1}{2}Z\right)^{-1}, \tag{6.46}$$

where X, Y and Z are the mass fractions of hydrogen, helium and all the heavier elements (normally called the "metals"). The mass fractions satisfy $X + Y + Z = 1$. [Hint: For the derivation you may want to make use of the mean atomic number A of the "metals". The "metals" may be assumed to contain $A/2$ electrons on the average.]

6.7 Particles with Internal Energy States

As the next example, consider an atom or a molecule that, in addition to different translational states, may exist in different internal quantum states. The total energy consists of the sum of translational and internal energies,

$$W_s = W_{\boldsymbol{q}} + W_i,$$

where $\boldsymbol{q}$ and i represent the three translational quantum numbers and any suitable set of quantum numbers i describing the internal state, as discussed in chapters 3, 4 and 5. The Boltzmann distribution (6.21) is valid also for the present problem. Since any combination of translational and internal states is possible, the sum over all available quantum states in the partition function (6.22) reduces to the product of the sums over translational and internal quantum states. The translational part of this problem was discussed in section 6.4. Here we focus on the internal part and therefore define the "internal" Boltzmann distribution and partition function

$$\mathcal{P}_i(W_i) = \frac{1}{Z^{\rm int}} \exp\left(-\frac{W_i}{\mathcal{T}}\right) \tag{6.47}$$

$$Z^{\rm int} = \sum_i \exp\left(-\frac{W_i}{\mathcal{T}}\right). \tag{6.48}$$

The summation in (6.48) is to be carried out over all *internal quantum states* i. We note that splitting the Boltzmann distribution into a product of one translational and one internal part

is equivalent to the assumption of statistical independence between translational and internal degrees of freedom.

We often encounter systems with energy degeneracy, that is, systems where several quantum states correspond to the same energy. In these cases it may be convenient to group together quantum states corresponding to the same energy and sum over the different energy states W_i instead. With this procedure the Boltzmann distribution and partition function take slightly modified forms,

$$\mathcal{P}_{W_i}(W_i) = \frac{g_i}{Z^{\text{int}}} \exp\left(-\frac{W_i}{\mathcal{T}}\right) \tag{6.49}$$

$$Z^{\text{int}} = \sum_{W_i} g_i \exp\left(-\frac{W_i}{\mathcal{T}}\right), \tag{6.50}$$

where g_i is the multiplicity of energy state W_i and the summation should be carried out over all *internal energy levels* W_i.

The total partition function for the particle in box with internal energy states is trivially seen to be

$$Z_1^{\text{total}} = Z_1 \cdot Z^{\text{int}} \tag{6.51}$$

with Z_1 given by (6.26). In the following we shall make repeated use of the *normalized partition function* Z' defined by

$$Z^{\text{int}} \equiv Z' \exp(-\frac{W_0}{\mathcal{T}}) \tag{6.52}$$

where W_0 is the ground state energy. It is important to note that Z^{int} and Z' are both functions of $\mathcal{T}$ only.

The generalization of the above results to a system consisting of N non-interacting and identical particles with internal energy states is straightforward. In particular, the number of particles that we expect to find in any given atomic state i is given by

$$N_i = N\mathcal{P}_i(W_i).$$

The ratio of expected numbers of particles in states i and j is therefore

$$\frac{N_i}{N_j} = \exp\left(-\frac{W_i - W_j}{\mathcal{T}}\right). \tag{6.53}$$

The corresponding ratio of the expected numbers of particles in energy states W_i and W_j is

$$\frac{N_{W_i}}{N_{W_j}} = \frac{g_i}{g_j} \exp\left(-\frac{W_i - W_j}{\mathcal{T}}\right). \tag{6.54}$$

It is important to note that the given ratios of particle numbers may be calculated without knowledge of the actual value of the internal partition function Z^{int}. The results (6.53) and (6.54) are referred to as the *Boltzmann relation.*

The total partition function for the N particle system is

$$Z_N^{\text{total}} = \frac{1}{N!}(V n_{\text{Q}} Z^{\text{int}})^N,$$

or approximately

$$\ln Z_N^{\text{total}} \approx N\left(\ln\left(\frac{V}{N}\, n_{\text{Q}}\, Z^{\text{int}}\right) + 1\right). \tag{6.55}$$

In the evaluation of the internal part of the partition function, (6.48) or (6.50), the summation should be carried out over all quantum states or all energy levels available to the system, respectively. If the number of states is infinite and we for convenience assume that the numbering of states is such that the energy is monotonically increasing with state number, we see that we will face a problem if the energy levels W_i do not tend toward infinity with increasing i. This is actually the case with the energy levels of any atom or molecule. The problem is resolved only by taking into account that, close to the continuum limit, the atoms can no longer be assumed to act as independent particles. Charged particles in a gas establish electric micro-fields which broaden the energy levels of atoms or ions, particularly the upper levels. The upper levels start to overlap and the ionization energy ΔW_I appears to be lowered by an amount δW which depends on the electric micro-field intensity and therefore on the electron density. In a neutral hydrogen gas we expect similar effects for principal quantum numbers n for which the classical electron orbit radius $r = n^2 a_B$ compares with the average distance between atoms $d = n_H^{-1/3}$, that is $n^{-2} \approx a_B n_H^{1/3}$. The corresponding lowering of the ionization energy is

$$\delta W \approx 13.6\, a_B n_H^{1/3} \quad [\text{eV}].$$

For dilute gases or plasmas the effect is small. In the partition function summation only the discrete energy levels below the reduced ionization energy $\Delta W_I^* = \Delta W_I - \delta W$ are to be included.

A detailed discussion of this problem falls outside our scope. We note, however, that in situations where the two lowest energy levels are widely separated compared to the thermal energy, $W_1 - W_0 \gg \mathcal{T}$, it may be enough to include only the lowest energy state (the ground state) in the sum. For this case the normalized partition function reduces to the multiplicity of the ground state,

$$Z' \approx g_0. \tag{6.56}$$

This approximation is good for hydrogen and helium and many other element, but poor for the alkali metals. In table 6.1 the ground state multiplicity g_0 together with the value of the normalized partition function for two different temperatures $T = 5040$ K and 10080 K are given for the neutral and the singly ionized states of some selected elements. For later convenience, the table also includes the corresponding ionization energies, ΔW_I.

Normally we expect Z' to be equal to or slightly larger than the multiplicity of the ground state. The fact that Z' for some elements, for instance O I, is less than the tabulated g_0 is a result of fine structure splitting of the ground state spectral term. Thus, for O I the 5 $^3P_2$ states contribute fully to Z' while the contributions from the 4 $^3P_{1,0}$ states are slightly reduced by the exponential factor.

Quiz 6.12 : In the definition of the ground state multiplicities in table 6.1, the fine structure splitting for different J-values have not been taken into account. With the help of table 4.3, explain the values of the ground state multiplicities g_0 listed in table 6.1. For what elements does (6.56) represent a valid approximation for the temperatures listed?

			Z'_I		ΔW_I		Z'_{II}		ΔW_{II}
Z	Element	g_0	5040 K	10080 K	eV	g_0	5040 K	10080 K	eV
1	H	2	2.00	2.00	13.60	1	1.00	1.00	
2	He	1	1.00	1.00	24.59	2	2.00	2.00	54.42
3	Li	2	2.09	3.09	5.39	1	1.00	1.00	75.64
4	Be	1	1.02	1.35	9.32	2	2.00	2.00	18.21
5	B	6	6.03	6.03	8.30	1	1.00	1.00	25.16
6	C	9	9.33	10.00	11.26	6	6.03	6.03	24.38
7	N	4	4.07	4.57	14.53	9	8.91	9.33	29.60
8	O	9	8.71	9.33	13.62	4	4.00	4.07	35.12
9	F	6	5.62	5.89	17.42	9	8.32	8.71	34.97
10	Ne	1	1.00	1.00	21.56	6	5.37	5.62	40.96
11	Na	2	2.04	2.88	5.14	1	1.00	1.00	47.29
12	Mg	1	1.03	1.38	7.65	2	2.04	2.04	15.04
13	Al	6	5.89	6.17	5.99	1	1.00	1.00	18.83
14	Si	9	9.55	10.96	8.15	6	5.75	5.89	16.35
15	P	4	4.47	6.17	10.49	9	8.13	8.71	19.73
16	S	9	8.13	8.71	10.36	4	4.17	5.25	23.33
17	Cl	6	5.25	5.62	12.97	9	7.76	8.32	23.81
18	Ar	1	1.00	1.00	15.76	6	4.90	5.13	27.63
19	K	2	2.19	3.98	4.34	1	1.00	1.00	31.63
20	Ca	1	1.17	3.55	6.11	2	2.19	3.47	11.87
21	Sc	10	12.02	30.90	6.54	15	22.91	33.11	12.80
22	Ti	21	30.20	75.86	6.82	28	56.23	83.18	13.58
23	V	28	41.69	107.15	6.74	25	43.65	77.62	14.65
24	Cr	7	10.47	32.36	6.77	6	7.24	13.49	16.50
25	Mn	6	6.46	14.45	7.44	7	7.76	13.49	15.64
26	Fe	25	26.92	54.95	7.87	30	42.66	63.10	16.16
27	Co	28	33.11	57.54	7.86	21	28.84	45.71	17.06
28	Ni	21	29.51	39.81	7.64	10	10.47	19.05	18.17
29	Cu	2	2.29	3.80	7.73	1	1.02	1.51	20.29
30	Zn	1	1.00	1.07	9.39	2	2.00	2.00	17.96

Table 6.1: Multiplicity, partition function and ionization energy for selected elements

Quiz 6.13 : In figure 4.4 the lower energy levels for Ca II have been plotted. Make use of (6.50) and (6.52) and the three lower energy levels for Ca II to verify the tabulated value (see table 6.1) of the normalized partition function Z' for Ca II for the temperature $T = 10080$ K.

Quiz 6.14 : Are you able to explain that for some elements the exact value of the normalized partition functions are smaller than the corresponding value of g_0?

Quiz 6.15 : The inter-nuclear distance of the CO molecule is $R_e = 0.113$ nm. CO gas is kept at temperature $T = 500$ K. Determine the ratio n_N/n_0 of number densities n of molecules with nuclear orbital quantum number N and $N = 0$. Give an order of magnitude estimate of the number of rotational absorption lines that you would expect for the gas.

6.8 Reversible Processes

Let us now return to the general formalism of thermal physics. According to the discussion of section 6.2, a system in thermal equilibrium with a reservoir will acquire a temperature $\mathcal{T}$ equal to that of the reservoir. The temperature was expressed in (6.19) as the partial derivative of the energy $U = U(\mathcal{S}, V)$ with respect to entropy $\mathcal{S}$ while keeping all external parameters, here the system volume V, constant.

Now let the system participate in a process in which entropy and volume change by small amounts $d\mathcal{S}$ and dV, but such that *at each stage of the process the system remains infinitesimally close to a thermal equilibrium state.* Such a process will be called a *reversible process.* At any stage of a reversible process, the process may be arrested, reversed and the system brought back to its original state. Any process that does not satisfy this requirement is an *irreversible process.*

The energy change $\mathrm{d}U$ associated with the reversible process in which the system is taken through a continuous succession of thermal equilibrium states from a state A to another state B corresponding to infinitesimal total changes $\mathrm{d}\mathcal{S}$ and $\mathrm{d}V$ in entropy and volume (see figure 6.7a) is

$$\mathrm{d}U = \frac{\partial U}{\partial \mathcal{S}}|_V \, \mathrm{d}\mathcal{S} + \frac{\partial U}{\partial V}|_{\mathcal{S}} \, \mathrm{d}V. \tag{6.57}$$

The first partial derivative on the right hand side equals the temperature $\mathcal{T}$. We interpret the first term as the amount of energy added to the system during the process in the form of *heat*,

$$\delta\mathcal{Q} = \mathcal{T} \, \mathrm{d}\mathcal{S}. \tag{6.58}$$

The second derivative is interpreted as the negative of the system pressure

$$P = -\frac{\partial U}{\partial V}|_{\mathcal{S}} \, . \tag{6.59}$$

The negative of the second term is interpreted as the *work* done by the system on its surroundings during the process,

$$\delta\mathcal{W} = P \, \mathrm{d}V. \tag{6.60}$$

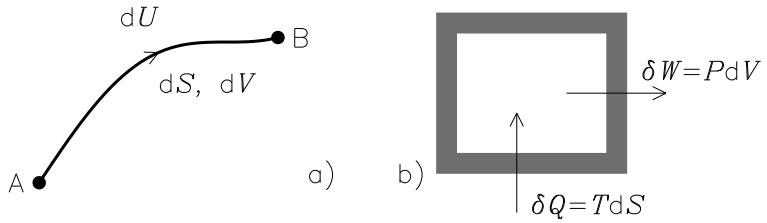

Figure 6.7: Schematics of a reversible process

We shall see below that the thermodynamic definition (6.59) of pressure is identical to the previous kinetic definition (6.40).

With this notation the mathematical identity (6.57) represents a statement of the *first law of thermodynamics*: *In any reversible process the change in system energy equals the difference between the heat added to the system and the work done by the system on its surroundings*,

$$\mathrm{d}U = \delta\mathcal{Q} - \delta\mathcal{W} = \mathcal{T}\,\mathrm{d}\mathcal{S} - P\,\mathrm{d}V. \tag{6.61}$$

The situation is illustrated in figure 6.7b. The choice of notation in (6.61) is important. The system energy U is a function of the state variables $\mathcal{S}$ and V, $\mathrm{d}U$ is the total differential of this function with respect to small changes in these variables. The quantity of heat added $\delta\mathcal{Q}$ and the work done $\delta\mathcal{W}$ are, however, *not* each of them total differentials. This means that while the change in energy U only depends on the initial and final states of the system, the latter quantities both depend on the details of the process chosen, that is, the way in which the system is brought from its initial state to the final state.

In (6.61) the energy U is assumed to be given as a function of entropy and volume $U = U(\mathcal{S}, V)$. According to (6.18), this also applies to the temperature $\mathcal{T} = \mathcal{T}(\mathcal{S}, V)$. In principle, this relation can be solved with respect to entropy giving $\mathcal{S} = \mathcal{S}(\mathcal{T}, V)$ and therefore also $U = U\left(\mathcal{S}(\mathcal{T}, V), V\right)$. But this means that the total differential of energy may also be expressed in the form

$$\mathrm{d}U = \frac{\partial U}{\partial \mathcal{T}}\Big|_V \,\mathrm{d}\mathcal{T} + \frac{\partial U}{\partial V}\Big|_\tau \,\mathrm{d}V. \tag{6.62}$$

When combining (6.62) and (6.61), an expression for the total differential of the entropy results

$$\mathrm{d}\mathcal{S} = \frac{1}{\mathcal{T}}\frac{\partial U}{\partial \mathcal{T}}\Big|_V \,\mathrm{d}\mathcal{T} + \frac{1}{\mathcal{T}}\left(P + \frac{\partial U}{\partial V}\Big|_\tau\right)\mathrm{d}V \tag{6.63}$$

and therefore

$$\frac{\partial \mathcal{S}}{\partial \mathcal{T}}\Big|_V = \frac{1}{\mathcal{T}}\frac{\partial U}{\partial \mathcal{T}}\Big|_V \quad \text{and} \quad \frac{\partial \mathcal{S}}{\partial V}\Big|_\tau = \frac{1}{\mathcal{T}}\left(P + \frac{\partial U}{\partial V}\Big|_\tau\right).$$

Since $\partial^2\mathcal{S}/\partial V\partial\mathcal{T} = \partial^2\mathcal{S}/\partial\mathcal{T}\partial V$, we also find

$$\mathcal{T}\frac{\partial P}{\partial \mathcal{T}}\Big|_V = P + \frac{\partial U}{\partial V}\Big|_\tau. \tag{6.64}$$

The above examples are but a small number of relations between thermodynamic state variables and their partial derivatives that can be derived on the basis of the first law of thermodynamics. Other quantities are defined through similar partial derivatives. Thus, the heat capacities C_V and C_P at constant volume and at constant pressure are given as the

quantity of heat needed to increase the temperature by one unit (fundamentally one energy unit, traditionally one degree Kelvin) under the two stated conditions. With

$$\delta\mathcal{Q} = \mathrm{d}U + P\,\mathrm{d}V = \frac{\partial U}{\partial \mathcal{T}}|_V\,\mathrm{d}\mathcal{T} + \left(P + \frac{\partial U}{\partial V}|_\tau\right)\mathrm{d}V \tag{6.65}$$

we find

$$C_V \equiv \frac{\delta\mathcal{Q}}{\delta\mathcal{T}}|_V = \frac{\partial U}{\partial \mathcal{T}}|_V. \tag{6.66}$$

If V and U are instead expressed in terms of P and $\mathcal{T}$, $V = V(P,\mathcal{T})$ and $U\,(\mathcal{T}, V(P,\mathcal{T}))$, it follows from (6.65) that

$$\delta\mathcal{Q} = \left(\frac{\partial U}{\partial \mathcal{T}}|_V + \left(P + \frac{\partial U}{\partial V}|_\tau\right)\frac{\partial V}{\partial \mathcal{T}}|_P\right)\mathrm{d}\mathcal{T} + \left(P + \frac{\partial U}{\partial V}|_\tau\right)\frac{\partial V}{\partial P}|_\tau\,\mathrm{d}P$$

and therefore

$$C_P \equiv \frac{\delta\mathcal{Q}}{\delta\mathcal{T}}|_P = \frac{\partial U}{\partial \mathcal{T}}|_V + \left(P + \frac{\partial U}{\partial V}|_\tau\right)\frac{\partial V}{\partial \mathcal{T}}|_P.$$

Making use of (6.66) and (6.64), the latter result may also be written as

$$C_P = C_V + \mathcal{T}\frac{\partial P}{\partial \mathcal{T}}|_V\,\frac{\partial V}{\partial \mathcal{T}}|_P. \tag{6.67}$$

Thermodynamic variables can generally be divided into two classes. Consider the following thought experiment. If a diaphragm suddenly divides a gas box in thermal equilibrium with its surroundings in two halves, each of the two halves will still be in a state of thermal equilibrium. The energies and the volumes of the two halves will be reduced by a factor two relative to the original values. Pressure and temperature will remain unchanged. The former type of variables are called *extensive* variables, the latter type *intensive* variables. By dividing an extensive variable by the particle number of the system, that variable is turned into an intensive variable. Thus, the energy per particle or the *specific energy* $u = U/N$ and the *specific volume* $v = V/N$ are both intensive variables. It is evident that any product of two intensive variables will itself be an intensive variable, while a product of an extensive variable with an intensive one will result in an extensive variable.

Quiz 6.16 : Demonstrate that (6.64) is satisfied for an ideal gas.

Quiz 6.17 : Show that for the ideal gas at pressure P the work $\delta\mathcal{W}$ performed on the surroundings during an expansion $\mathrm{d}V$ of its volume due to the action of pressure forces is given by

$$\delta\mathcal{W} = P\,\mathrm{d}V. \tag{6.68}$$

Quiz 6.18 : For the ideal gas evaluate the heat capacities at constant volume and constant pressure, C_V and C_P. Show that the ratio of these two quantities is

$$\gamma = \frac{C_P}{C_V} = \frac{5}{3}. \tag{6.69}$$

Quiz 6.19 : The coefficients of thermal expansion and compressibility are given by

$$\beta_V \equiv \frac{1}{V}\frac{\partial V}{\partial \mathcal{T}}\Big|_P \quad \text{and} \quad \kappa_P \equiv -\frac{1}{V}\frac{\partial V}{\partial P}\Big|_{\mathcal{T}}. \tag{6.70}$$

On the basis of the mathematical definitions (6.70), give a precise description of the physical contents of these coefficients and prescribe processes which can be used to measure them. Calculate these coefficients for the ideal gas.

Quiz 6.20 : Make use of (6.63) to show that in a reversible and *adiabatic* process, that is, in a process in which there is no heat exchange with the surroundings, $\delta\mathcal{Q} = 0$, the ideal gas must satisfy the following three relations

$$\mathcal{T}V^{\gamma-1} = \text{constant}, \tag{6.71}$$

$$PV^{\gamma} = \text{constant} \tag{6.72}$$

and

$$P^{1-\gamma}\mathcal{T}^{\gamma} = \text{constant}. \tag{6.73}$$

where $\gamma = C_P/C_V$.

Quiz 6.21 : For an ideal gas show that (6.63) reduces to

$$\mathrm{d}\mathcal{S} = \frac{3N}{2\mathcal{T}}\,\mathrm{d}\mathcal{T} + \frac{N}{V}\,\mathrm{d}V.$$

Show that this expression may be integrated to give

$$\mathcal{S} = \mathcal{S}_0 + N\ln\left(V\mathcal{T}^{3/2}\right),$$

where $\mathcal{S}_0$ is an integration constant. Taking into account the extensive/intensive properties of the thermodynamic variables involved, are you able to specify the N dependence of $\mathcal{S}_0$?

6.9 The Helmholtz Free Energy

The choice of entropy $\mathcal{S}$ and volume V as independent variables (6.61) is sometimes inconvenient. In the previous section it was demonstrated that by a formal inversion of the temperature definition (6.18), a change from $\mathcal{S}$ and V to $\mathcal{T}$ and V is possible. An alternative and powerful approach for effectuating a change of independent variables is to make use of a *Legendre transformation*. In the Legendre transformation the change of independent variables is induced through a change of dependent variable. Thus, if we prefer to work with $\mathcal{T}$ and V as independent variables instead of $\mathcal{S}$ and V, this is easily achieved by replacing the energy U as dependent variable with a function

$$F \equiv U(\mathcal{S}, V) - \mathcal{T}\mathcal{S}. \tag{6.74}$$

The new dependent variable F is the *Helmholtz free energy*. By making use of the expression for the differential of U in terms of $\mathcal{S}$ and V, (6.57), we find

$$\mathrm{d}F = -\mathcal{S}\,\mathrm{d}\mathcal{T} - P\,\mathrm{d}V. \tag{6.75}$$

This means that we may consider the Helmholtz free energy as a "natural" function of $\mathcal{T}$ and V, $F = F(\mathcal{T}, V)$. In equilibrium at constant $\mathcal{T}$ and V the Helmholtz free energy F will take an extremum value, $\mathrm{d}F = 0$. It can be shown that this extremum is a minimum. From (6.75) we also immediately identify alternative expressions for entropy and pressure as partial derivatives of the Helmholz free energy in terms of temperature $\mathcal{T}$ and volume V,

$$\mathcal{S} = -\frac{\partial F}{\partial \mathcal{T}}|_V \quad \text{and} \quad P = -\frac{\partial F}{\partial V}|_\tau \ . \tag{6.76}$$

With the expression (6.76) for entropy substituted back into (6.74), the Helmholtz free energy can be expressed in terms of energy U

$$\mathcal{T}^2 \frac{\partial}{\partial \mathcal{T}} \left(\frac{F}{\mathcal{T}} \right) |_V = -F + \mathcal{T} \frac{\partial F}{\partial \mathcal{T}}|_V = -U.$$

Together with the expression (6.23) for the average energy U, this leads to an explicit expression for the average Helmholtz free energy in terms of temperature and partition function

$$F = -\mathcal{T} \ln Z. \tag{6.77}$$

With this general result at hand we may now return to the case of an ideal gas consisting of N non-interacting and identical particles in the volume V. With the partition function given by (6.30) we find from (6.76) the following results for pressure and entropy in thermal equilibrium,

$$P = \frac{N}{V}\mathcal{T}$$

and

$$\mathcal{S} = N \left(\frac{5}{2} + \ln(\frac{V}{N} n_{\mathrm{Q}}) \right), \tag{6.78}$$

with the quantum concentration n_{Q} defined by (6.27). The expression for the pressure is identical to the previous kinetic result (6.41). The expression (6.78) for the entropy is called the *Sackur-Tetrode formula* and is seen to be in accordance with the result of quiz 6.21.

6.10 Irreversible Processes

Most real processes are irreversible. It is, strictly speaking, usually not possible to reverse a process and bring the system back to the original state. Processes involving any kind of dissipation are good examples. During such processes the thermodynamic variables are not defined. Only when the system has found a new equilibrium state can these variables again be evaluated.

Still, something can be said about such processes. The *second law of thermodynamics* focuses on these questions: *If a closed system at a given time t is in a configuration which is not an equilibrium state, the most probable development is that the entropy of the system increases monotonically with time*,

$$\frac{\mathrm{d}\mathcal{S}}{\mathrm{d}t} \geq 0. \tag{6.79}$$

This means that we expect the entropy to take a maximum value as the system approaches a thermal equilibrium state. This is equivalent to the expectation that the Helmholtz free energy will take a minimum value in thermal equilibrium.

Quiz 6.22 : Two volumes V_1 and V_2, separated by an ideal membrane, contain two different ideal gases at identical pressure P and temperature $\mathcal{T}$. Show that in a process in which the membrane is removed, the entropy of the system increases by an amount

$$\Delta\mathcal{S} = \frac{PV_1}{\mathcal{T}} \ln \frac{V_1 + V_2}{V_1} + \frac{PV_2}{\mathcal{T}} \ln \frac{V_1 + V_2}{V_2}.$$

Is the process reversible or irreversible? What would the corresponding results be if the two volumes contained identical gas?

6.11 The Chemical Potential

In our discussion of thermal physics we have so far assumed the particle number to remain constant. We now want to investigate what changes are needed if this restriction is relaxed. Consider two systems Σ_1 and Σ_2 containing identical particles both in thermal contact with a common heat reservoir at temperature $\mathcal{T}$. In equilibrium the Helmholtz free energy of both systems will take minimum values. Let the two systems now also be allowed to exchange particles, as indicated in figure 6.8. The two systems will be said to be in thermal as well as *diffusive contact*. The equilibrium that will be established will be called a *thermal and diffusive equilibrium state*.

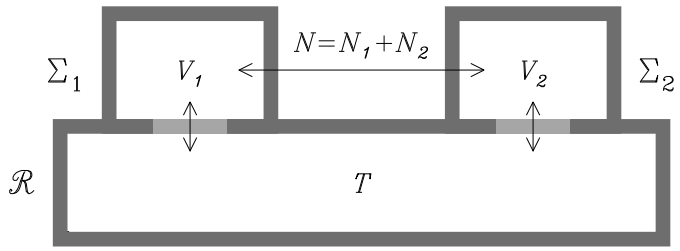

Figure 6.8: Systems in thermal and diffusive equilibrium

For processes in which volume and temperature are kept constant, the system is conveniently described in terms of the Helmholtz free energy. The total Helmholtz free energy of the two systems in diffusive contact is equal to the sum of the corresponding quantities for the two individual systems,

$$F = F_1(N_1, \mathcal{T}, V_1) + F_2(N_2, \mathcal{T}, V_2).$$

With diffusive contact between the two systems the particle numbers N_1 and N_2 must now be considered as independent variables, subject only to the constraint of a constant total particle number, $N = N_1 + N_2$. In diffusive equilibrium (with $\mathcal{T}$, V_1 and V_2 kept constant) the total Helmholtz free energy must find a minimum value under the new constraint,

$$\mathrm{d}F = \frac{\partial F_1}{\partial N_1}\Big|_{V_1\mathcal{T}} \mathrm{d}N_1 + \frac{\partial F_2}{\partial N_2}\Big|_{V_2\mathcal{T}} \mathrm{d}N_2 = 0 \quad \text{with} \quad \mathrm{d}N_1 + \mathrm{d}N_2 = 0.$$

Diffusive equilibrium between the two systems is seen to be characterized by the *chemical potential*, defined by

$$\mu \equiv \frac{\partial F}{\partial N}\Big|_{V\mathcal{T}}, \tag{6.80}$$

taking identical values for both systems,

$$\mu_1 = \mu_2.$$

This result is analogous to the requirement of identical temperatures for systems in thermal equilibrium.

With N considered as an independent variable, the differential of the Helmholtz free energy $F = F(N, \mathcal{T}, V)$ for an arbitrary reversible process must now be generalized as

$$\mathrm{d}F = -\mathcal{S}\,\mathrm{d}\mathcal{T} - P\,\mathrm{d}V + \mu\,\mathrm{d}N \tag{6.81}$$

and similarly for the energy $U = U(N, \mathcal{S}, V)$,

$$\mathrm{d}U = \mathcal{T}\,\mathrm{d}\mathcal{S} - P\,\mathrm{d}V + \mu\,\mathrm{d}N \tag{6.82}$$

The latter statement allows for an alternative definition of the chemical potential μ,

$$\mu \equiv -\mathcal{T}\frac{\partial \mathcal{S}}{\partial N}|_{UV}\ . \tag{6.83}$$

If several particle species are present in the volume V, the term $\mu\,\mathrm{d}N$ needs to be replaced by a sum of such contributions from each species.

The chemical potential μ for an ideal gas consisting of non-interacting identical particles with internal energy levels is readily calculated from the definition (6.80). For this we make use of the explicit expression for the Helmholz free energy (6.77) in terms of the logarithm of the total partition function (6.55). Since Z^{int} is a function of temperature $\mathcal{T}$ only, a simple calculation leads to

$$\mu = -\mathcal{T}\ln\left(\frac{V}{N}\,n_{\mathrm{Q}}\,Z^{\mathrm{int}}\right). \tag{6.84}$$

With several particle species present, the chemical potential for each species will be given by (6.84) with the proper particle mass, particle number and internal partition function Z^{int} substituted.

Quiz 6.23 : Is the chemical potential μ an extensive or an intensive variable?

6.12 The Law of Mass Action

In gas mixtures, chemical reactions, dissociation or ionization processes can occur in which the number of particles of different types change. Simple examples are

$$\begin{aligned} 2\mathrm{H_2O} &\rightleftharpoons 2\mathrm{H_2} + \mathrm{O_2} \\ \mathrm{H_2} &\rightleftharpoons 2\mathrm{H} \\ \mathrm{H} &\rightleftharpoons \mathrm{H^+} + \mathrm{e^-}. \end{aligned} \tag{6.85}$$

When thermal equilibrium is reached, a certain balance between the different reaction products will be established. We want to study what factors determine the equilibrium balance.

The processes mentioned may all be described by a reaction equation of the form

$$\sum_i \nu_i \mathrm{A}_i = 0, \tag{6.86}$$

where integers ν_i indicate the number of particles of type A_i that are involved in an elementary reaction. If the number N_i of particles of type i is changed by $\mathrm{d}N_i$ for a given reaction, related changes will also occur in the particle numbers of the other species taking part in the reaction. In fact, simple considerations show that the changes in particle numbers induced by the reaction equation (6.86) must satisfy

$$\mathrm{d}N_i = k\nu_i, \tag{6.87}$$

where k is a constant.

The Helmholtz free energy F takes a minimum value when approaching a thermal and diffusive equilibrium state for any process proceeding at constant temperature and volume. This means that the total differential of F must vanish when evaluated at equilibrium

$$\mathrm{d}F\,|_{\mathcal{T}V} = \sum_i \mu_i \,\mathrm{d}N_i = 0.$$

Taking the constraint (6.87) into account, this result is equivalent to the requirement

$$\sum_i \mu_i \nu_i = 0,$$

or alternatively,

$$\exp\left(-\frac{\sum_i \mu_i \nu_i}{\mathcal{T}}\right) = 1. \tag{6.88}$$

In terms of the normalized internal partition functions Z_i' and particle densities $n_i = N_i/V$, and making use of the relation (6.84) valid for ideal gases, this condition can be rewritten in the form

$$\prod_i n_i^{\nu_i} = \prod_i \left(Z_i'\, n_{\mathrm{Q}i}\, \exp\left(-\frac{W_{i,0}}{\mathcal{T}}\right)\right)^{\nu_i}. \tag{6.89}$$

We notice that the right hand side expression is a function of temperature $\mathcal{T}$ only. The result (6.89) is known as the *law of mass action* and forms the basis for an understanding of chemical reactions, dissociation and ionization processes at thermal equilibrium. It also represents one key to an understanding of spectral line formation in stellar atmospheres.

Let us now apply the law of mass action (6.89) to the general ionization process

$$\mathrm{X}_J \rightleftharpoons \mathrm{X}_{J+1} + \mathrm{e}^-,$$

where X_J represents an arbitrary element X in the ionization state J. The reaction equation coefficients may be chosen as $\nu_{\mathrm{X}_{J+1}} = \nu_e = 1$ and $\nu_{\mathrm{X}_J} = -1$. The electron is a particle with two spin states, but otherwise no internal structure. This means that $W_e = 0$, $Z_e' = 2$ and therefore that the ionization balance in the gas is determined by

$$\frac{n_{\mathrm{X}_{J+1}} n_e}{n_{\mathrm{X}_J}} = 2\frac{Z'_{\mathrm{X}_{J+1}}}{Z'_{\mathrm{X}_J}}\, n_{\mathrm{Q}_e}\, \exp\left(-\frac{\Delta W_{\mathrm{X}_J}}{\mathcal{T}}\right). \tag{6.90}$$

Here

$$\Delta W_{\mathrm{X}_J} = W_{\mathrm{X}_{J+1},0} - W_{\mathrm{X}_J,0}$$

is the ionization energy for element X_J. In (6.90) we have made use of the fact that the quantum concentrations for the X_J and X_{J+1} elements are for all practical purposes equal. Values of the ionization energies for selected elements are listed in table 6.1 together with the corresponding values of the "normalized" partition functions Z'. The result (6.90) is known as the *Saha equation* and applies to any ionization process in ideal gases at thermal equilibrium.

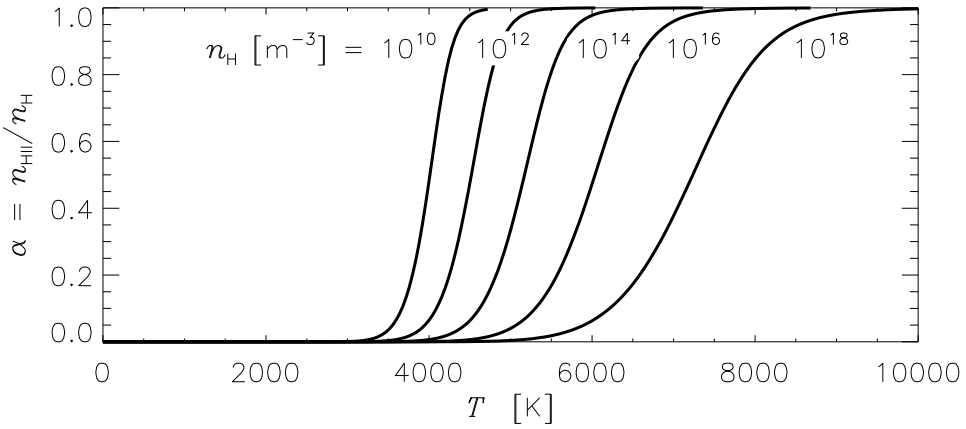

Figure 6.9: Degree of ionization in hydrogen as a function of temperature for different densities

In a pure hydrogen gas, charge neutrality requires $n_e = n_{\mathrm{H\,II}}$. The Saha equation (6.90) may then also be expressed in terms of the *degree of ionization* of the gas

$$\alpha \equiv \frac{n_{\mathrm{H\,II}}}{n_{\mathrm{H}}},$$

where $n_{\mathrm{H}} = n_{\mathrm{H\,I}} + n_{\mathrm{H\,II}}$ represents the total density of neutral and ionized hydrogen. We find

$$\frac{\alpha^2}{1-\alpha} n_{\mathrm{H}} = n_{\mathrm{Q}e} \exp\left(\frac{-\Delta W_{\mathrm{H\,I}}}{\mathcal{T}}\right), \tag{6.91}$$

where $\Delta W_{\mathrm{H\,I}}$ is the ionization energy of hydrogen. In figure 6.9 the degree of ionization α is plotted as a function of temperature T for different total hydrogen densities n_{H}. Due to the presence of the exponential function, the degree of ionization switches from almost zero to near unity over a relatively small temperature interval. Above this interval the ionized gas will not produce any hydrogen absorption lines because there is essentially no neutral hydrogen left to take part in any absorption process. With increasing total hydrogen density, the ionization balance is shifted toward a decreasing degree of ionization. This will also happen if other particle species are added to the gas. In this case the contribution to the electron density from the other species will increase the n_e factor in (6.90) while the right hand side expression remains constant. The ratio $n_{\mathrm{H\,II}}/n_{\mathrm{H\,I}}$ must therefore decrease.

If several different ionization processes may take place simultaneously, an equation of type (6.90) will be valid for each reaction. In particular, if a given particle specie participates in

several reactions, the *total* particle density of this specie will appear in each reaction equation. Pure helium gas represents a simple example. In this case the two ionization processes

$$\mathrm{He} \rightleftharpoons \mathrm{He}^{+} + \mathrm{e}^{-}$$
$$\mathrm{He}^{+} \rightleftharpoons \mathrm{He}^{++} + \mathrm{e}^{-}$$

give rise to the equations

$$\frac{n_{\mathrm{He\,II}}\, n_e}{n_{\mathrm{He\,I}}} = 2\frac{Z'_{\mathrm{He\,II}}}{Z'_{\mathrm{He\,I}}}\, n_{\mathrm{Q}e}\, \exp\left(-\frac{\Delta W_{\mathrm{He\,I}}}{\mathcal{T}}\right) \tag{6.92}$$

$$\frac{n_{\mathrm{He\,III}}\, n_e}{n_{\mathrm{He\,II}}} = 2\frac{Z'_{\mathrm{He\,III}}}{Z'_{\mathrm{He\,II}}}\, n_{\mathrm{Q}e}\, \exp\left(-\frac{\Delta W_{\mathrm{He\,II}}}{\mathcal{T}}\right). \tag{6.93}$$

Charge neutrality this time requires

$$n_e = n_{\mathrm{He\,II}} + 2n_{\mathrm{He\,III}}, \tag{6.94}$$

while the total helium density is

$$n_{\mathrm{He}} = n_{\mathrm{He\,I}} + n_{\mathrm{He\,II}} + n_{\mathrm{He\,III}}. \tag{6.95}$$

For given n_{He} and $\mathcal{T}$ the equations (6.92)–(6.95) represent four equations for the four unknowns n_e, $n_{\mathrm{He\,I}}$, $n_{\mathrm{He\,II}}$ and $n_{\mathrm{He\,III}}$.

In figure 6.10 solutions of these equations for two different values of the total helium density n_{He} are given. Below 9000 K the helium gas for $n_{\mathrm{He}} = 10^{16}$ m^{-3} consists of He I atoms. Between 11000 K and 22000 K the gas is dominated by singly ionized He II ions. Above 25000 K the gas is completely ionized. By varying n_{He} the result will be modified. With $n_{\mathrm{He}} = 10^{12}$ m^{-3} the pure helium gas will be fully ionized already at $T = 18000$ K. For $n_{\mathrm{He}} = 10^{22}$ m^{-3} the temperature will have to be raised to 60000 K to achieve the same effect. Keeping the amount of helium constant while adding other species will also lead to changes in the result. This happens because the electron density appearing in the Saha equations for helium is the total electron density in the gas. Through the addition of other species the electron density increases and the ionization curves for helium shift toward higher temperatures.

In section 6.7 comments were made on the effect of electric micro-fields in a plasma in reducing the effective ionization energies. This effect has importance for the evaluation of the partition function. It has also importance for the calculation of the ionization state. The proper ionization energy to use in the Saha equation (6.90) is the reduced ionization energy value $\Delta W_I^* = \Delta W_I - \delta W$.

An atom may loose an electron either because of a collision with another particle or as a result of an interaction with a photon. Similarly, the inverse process may be a collisional recombination or a radiative recombination. The former involves a three-body collision of the ion, the electron to be absorbed and a third particle necessary to satisfy the requirement of momentum conservation. In the latter process, this requirement is taken care of by the emitted photon. Since both particles and radiation field are involved in the ionization process, the Saha equation (6.90) will only be valid for gases in thermal equilibrium with the radiation field (see section 6.15). In the outer, low density part of stellar atmospheres this condition will not be satisfied. The gas temperature may here be substantially higher than the equivalent radiation temperature and predictions based on the Saha equation will therefore be in error.

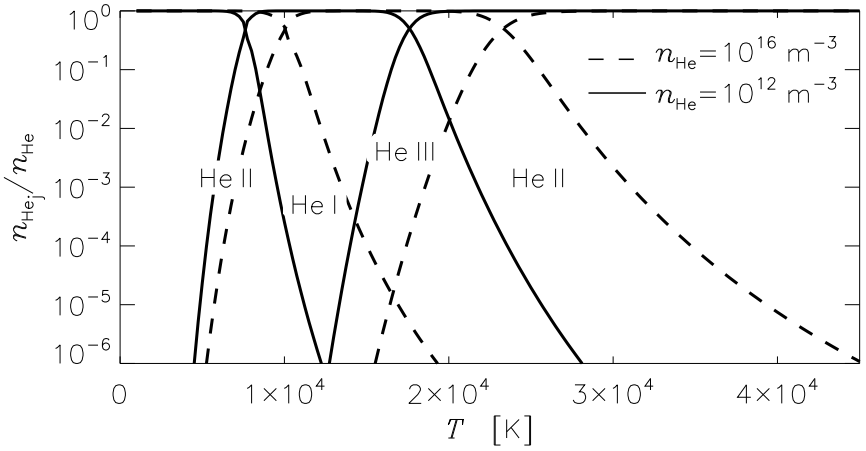

Figure 6.10: Ionization of helium as a function of temperature for different densities

Quiz 6.24 : Identify appropriate coefficients ν_i for the three reactions listed in (6.85).

Quiz 6.25 : For the dissociative process $H_2 \rightleftharpoons 2H$ show that the equilibrium balance is determined by

$$\frac{n_{\rm H}^2}{n_{\rm H_2}} = \frac{Z_{\rm H}'^2}{Z'_{\rm H_2}} \left(\frac{\pi m_{\rm H} \mathcal{T}}{h^2}\right)^{3/2} \exp\left(-\frac{\Delta W_{\rm H_2}}{\mathcal{T}}\right),$$

where

$$\Delta W_{\rm H_2} = 2W_{\rm H,0} - W_{\rm H_2,0}$$

is the dissociation energy of the H_2 molecule. Define the degree of dissociation in the gas and discuss qualitatively how this quantity varies with gas density and temperature.

Quiz 6.26 : The intensities of the absorption lines in the Balmer series are proportional to the number density $n_{\rm HI,1}$ of neutral hydrogen atoms in the lowest excited state. For a pure hydrogen gas, derive expressions for the ratios

$$\frac{n_{\rm H\,I,1}}{n_{\rm H\,I,0}}, \quad \frac{n_{\rm H\,I,1}}{n_{\rm H\,I}} \quad \text{and} \quad \frac{n_{\rm H\,I,1}}{n_{\rm H\,I} + n_{\rm H\,II}}.$$

Plot these ratios (qualitatively) as a function of temperature T for a total hydrogen density $n_{\rm H} = n_{\rm H\,I} + n_{\rm H\,II} = 10^{16}$ m^{-3}. For what temperature range would you expect the formation of the Balmer series lines to be an important process?

Quiz 6.27 : Carbon is a minority species in the Solar atmosphere. The resonance line of C IV is observed as an emission line that forms in a narrow temperature range around $T \sim 10^5$ K. The ionization energies for C III and C IV are $\Delta W_{\rm C\,III} = 47.9$ eV and $\Delta W_{\rm C\,IV} = 64.5$ eV. If you assume local thermal equilibrium to exist throughout the atmosphere, establish criteria for in what temperature range you expect C IV lines to form. For a typical total electron density $n_e = 10^{15}$ m^{-3}, what is your temperature range? Can you explain the discrepancy between your estimate and the stated temperature range?

6.13 The Gibbs Distribution

For a system, in which the particle number is allowed to vary, the fundamental Boltzmann distribution (6.21) needs to be generalized. Thus, consider a system Σ in thermal and diffusive contact with a large reservoir $\mathcal{R}$. The number of particles N_0 and energy U_0 of the total system $\mathcal{R}+\Sigma$ are held constant. If Σ is specified to contain N particles and be in quantum state $s(N)$ corresponding to energy $W_{s(N)}$, then the reservoir $\mathcal{R}$ has $N_0 - N$ particles and energy $U_0 - W_{s(N)}$.

In analogy with the discussion of section 6.3, if the state $s(N)$ of system Σ is specified, the multiplicity of the total system $\mathcal{R}+\Sigma$ equals the multiplicity $g_{\mathcal{R}}(N_0 - N, U_0 - W_{s(N)})$ of the reservoir $\mathcal{R}$. The probability that the system Σ in thermal and diffusive equilibrium with the reservoir $\mathcal{R}$ will be found in the given state $s(N)$ must then according to the fundamental statistical assumption satisfy

$$\begin{aligned}\mathcal{P}_{Ns}(N, W_{s(N)}) &\sim g_{\mathcal{R}}(N_0 - N, U_0 - W_{s(N)}) \\ &= \exp\left(\mathcal{S}_{\mathcal{R}}(N_0 - N, U_0 - W_{s(N)})\right) \\ &\approx \exp\left(\mathcal{S}_{\mathcal{R}}(N_0, U_0) - \frac{\partial \mathcal{S}_{\mathcal{R}}}{\partial N_0} N - \frac{\partial \mathcal{S}_{\mathcal{R}}}{\partial U_0} W_{s(N)}\right) \\ &\sim \exp\left(\frac{N\mu - W_{s(N)}}{\mathcal{T}}\right).\end{aligned}$$

Here we made use of (6.83) and the fact that temperature and chemical potential will both be identical for system and reservoir. In normalized form the result may be written

$$\mathcal{P}_{Ns}(N, W_{s(N)}) = \frac{1}{\mathcal{Z}} \exp\left(\frac{N\mu - W_{s(N)}}{\mathcal{T}}\right), \tag{6.96}$$

where the partition function $\mathcal{Z}$ now includes summations over particle numbers N and quantum states $s(N)$,

$$\mathcal{Z} = \sum_{N=0}^{\infty} \sum_{s(N)} \exp\left(\frac{N\mu - W_{s(N)}}{\mathcal{T}}\right). \tag{6.97}$$

We shall refer to (6.96) as the *Gibbs distribution*. The factor $\exp\left((N\mu - W_{s(N)})/\mathcal{T}\right)$ is called the *Gibbs factor*. The partition function $\mathcal{Z}$ is also called the *Gibbs* or *grand sum*. We note in particular that the average or expected number of particles in the system Σ is given as

$$\langle\, N \,\rangle = \sum_{N=0}^{\infty} \sum_{s(N)} N \mathcal{P}_{Ns}(N, W_{s(N)}) = \mathcal{T} \frac{\partial \ln \mathcal{Z}}{\partial \mu}\Big|_{\mathcal{T}V}. \tag{6.98}$$

6.14 The Degenerate Electron Gas

In the discussion on the validity of the ideal gas law we noted that short-range inter-molecular forces would under certain conditions lead to significant deviations, the phenomenon of phase transitions from gas to liquid or solid state representing extreme examples. There is, however, also another reason for deviations from the ideal gas law at large enough densities or low enough temperatures, even without any phase transition taking place. These deviations are related to quantum mechanical effects and will be evident for the free (conduction) electron

gas in metals and in the interiors of high density type stars (neutron stars, white dwarfs, etc.). The effect is occurring for odd half spin particles (particularly electrons, but under extreme conditions also for protons and neutrons). These particles are called *fermions* and obey the Pauli principle, that is, there can be maximum one such particle in any given quantum state in any interacting system.

Let us therefore return to the discussion of a gas of free particles in a volume V in section 6.6. Specifically, let the free particles be electrons. As long as the number of available quantum states in the relevant energy range is much larger than the number of electrons N, the probability that two electrons distributed at random should occupy the identical orbital is negligible. This means that the constraints laid by the Pauli principle is unimportant. In this low density limit the previous discussion of sections 6.4 and 6.6 is valid.

In the high density limit, the Pauli principle must be taken into account explicitly. To this end, consider as system an electron orbital with quantum numbers $\boldsymbol{q} = (q_x, q_y, q_z)$ as described in section 6.4. The given orbital may be empty or contain exactly one electron. With a variable number of particles, the system is described by the Gibbs distribution (6.96). The partition function takes the simple form

$$\mathcal{Z}_{\boldsymbol{q}} = 1 + \exp\left(\frac{\mu - W_{\boldsymbol{q}}}{\mathcal{T}}\right).$$

The average number of electrons in the orbital is according to (6.98) given by

$$\langle\, N_{\boldsymbol{q}} \,\rangle = \frac{1}{\mathcal{Z}_{\boldsymbol{q}}} \exp\left(\frac{\mu - W_{\boldsymbol{q}}}{\mathcal{T}}\right) = \frac{1}{\exp\left(\frac{W_{\boldsymbol{q}} - \mu}{\mathcal{T}}\right) + 1}.$$

In analogy with the discussion of section 6.4, the average number of electrons in the momentum range $(p, p + \mathrm{d}p)$ is found by summing the contributions from all orbitals corresponding to this range

$$\sum_{\boldsymbol{q} \in (p, p+\mathrm{d}p)} \langle\, N_{\boldsymbol{q}} \,\rangle = \frac{1}{\exp\left(\frac{W - \mu}{\mathcal{T}}\right) + 1} \frac{V}{h^3}\, 4\pi p^2\, \mathrm{d}p.$$

This means that the probability density function for electrons in $(\boldsymbol{r}, \boldsymbol{p})$-space, taking the Pauli principle into account, is given by

$$f_{\boldsymbol{rp}}(p) = f^{FD}_{\boldsymbol{rp}}(p) \equiv \frac{2}{h^3} \frac{1}{\exp\left(\frac{W - \mu}{\mathcal{T}}\right) + 1}. \tag{6.99}$$

The extra factor 2 was introduced because of the spin multiplicity of the electron. The result is known as the *Fermi-Dirac momentum distribution.* In this result $W = p^2/2m$ and the chemical potential $\mu(\mathcal{T})$ is related to the total electron density n through the normalization condition

$$\int_0^\infty f_{\boldsymbol{rp}}(p)\, 4\pi p^2\, \mathrm{d}p = n. \tag{6.100}$$

In the zero temperature limit we write $\mu(\mathcal{T} = 0) \equiv p_F^2/2m$ where p_F is the *Fermi momentum.* In this limit the Fermi-Dirac momentum distribution (6.99) reduces to the distribution function for the *fully degenerate* electron gas

$$f^{D}_{\boldsymbol{rp}}(p) = \begin{cases} \dfrac{2}{h^3} & p \le p_F \\[2ex] 0 & p > p_F \end{cases}. \tag{6.101}$$

The Fermi momentum p_F is related to the electron density n through the normalization condition (6.100),

$$p_F = \left(\frac{3h^3}{8\pi}\, n\right)^{1/3}. \tag{6.102}$$

The fully degenerate result (6.101) has a simple physical interpretation. The Pauli principle sets an upper limit to the electron number in any phase space element. In every element $d^3\boldsymbol{r}d^3\boldsymbol{p}$ of size h^3 there may at most be 2 electrons, one spin up and one spin down. If N electrons are to be placed in a volume V, the momentum space will be filled with this maximum phase space density out to a radius p_F. Thus, the momentum distribution function for the free electron gas will have a finite width, *even at zero temperature*. This means that even at $\mathcal{T} = 0$ the electron gas will possess a rest kinetic energy U_0 and therefore also maintain a non-vanishing rest pressure P_0. We find

$$U_0 = V \int_0^{p_F} \frac{p^2}{2m} f^D_{\boldsymbol{rp}}(p)\, 4\pi p^2\, \mathrm{d}p = \frac{4\pi V}{5h^3 m} p_F^5 = \frac{3V}{40}\left(\frac{3}{\pi}\right)^{2/3} \frac{h^2}{m} n^{5/3} \tag{6.103}$$

and, making use of (6.40),

$$P_0 = \frac{1}{3}\int_0^{p_F} pv f^D_{\boldsymbol{rp}}(p)\, 4\pi p^2\, \mathrm{d}p = \frac{2U_0}{3V} = \frac{1}{20}\left(\frac{3}{\pi}\right)^{2/3} \frac{h^2}{m} n^{5/3}. \tag{6.104}$$

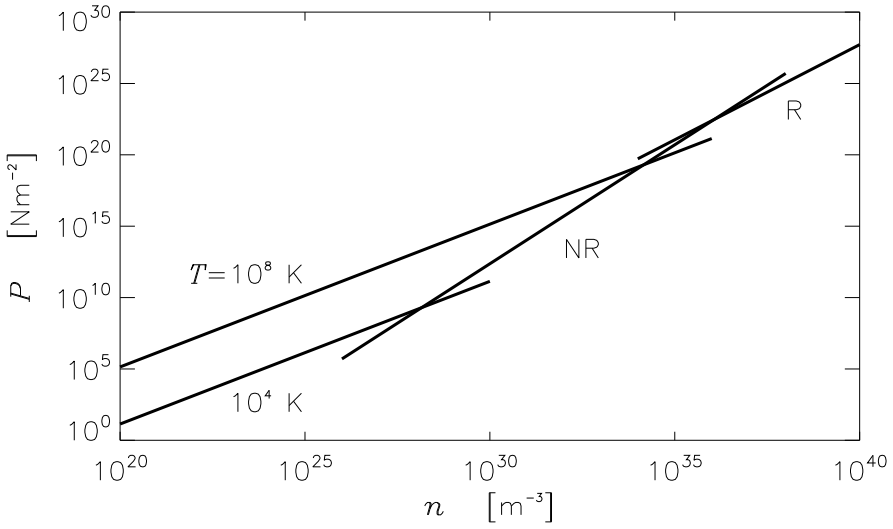

Figure 6.11: Electron pressure as a function of electron density

In (6.103) and (6.104) the non-relativistic expressions for kinetic energy and particle speed were used. These relations therefore represent the internal energy and equation of state for the *fully degenerate non-relativistic* electron gas. With increasing particle densities, the Fermi momentum p_F eventually becomes large enough that relativistic formulas need to be applied, even at $\mathcal{T} = 0$. For the equation of state (6.104), this means that we have to substitute $v = pc/(m^2c^2 + p^2)^{1/2}$. In the limiting case, $p_F \gg mc$, we have $v \approx c$ and therefore

$$P_0 \approx \frac{8\pi c}{3h^3}\int_0^{p_F} p^3\, \mathrm{d}p = \frac{hc}{8}\left(\frac{3}{\pi}\right)^{1/3} n^{4/3}. \tag{6.105}$$

This is the equation of state for the *fully degenerate relativistic* electron gas. In figure 6.11 the electron pressure as a function of electron density is plotted for the non-relativistic degenerate gas (NR) and the limiting case of the fully relativistic degenerate gas (R). In reality there will be a smooth transition from the one case to the other. In the figure the ideal gas result for temperatures $T = 10^4$ K and 10^8 K have also been plotted.

The normalization integral (6.100) for $\mathcal{T} > 0$ is best suited for numerical evaluations. In figure 6.12 the Fermi-Dirac momentum distribution function has been plotted (solid lines) for different ratios of electron density n to the quantum concentration n_{Q} as given by (6.27). For a given density, the ratio n/n_{Q} decreases with increasing temperature $\mathcal{T}$. As seen from the figure, the Pauli principle effectively sets an upper limit of $2/h^3$ to the maximum value of the momentum distribution function. This can also be seen directly from (6.99). The maximum value is achieved in the case when the exponential function in the denominator can be neglected compared with unity, that is, when $p < p_F$ and $\mathcal{T} \to 0$. At a finite temperature and large enough p on the other hand, the unity in the denominator of (6.99) can be neglected compared with the exponential function. In this limit the Maxwell momentum distribution function (6.34) is recovered as expected.

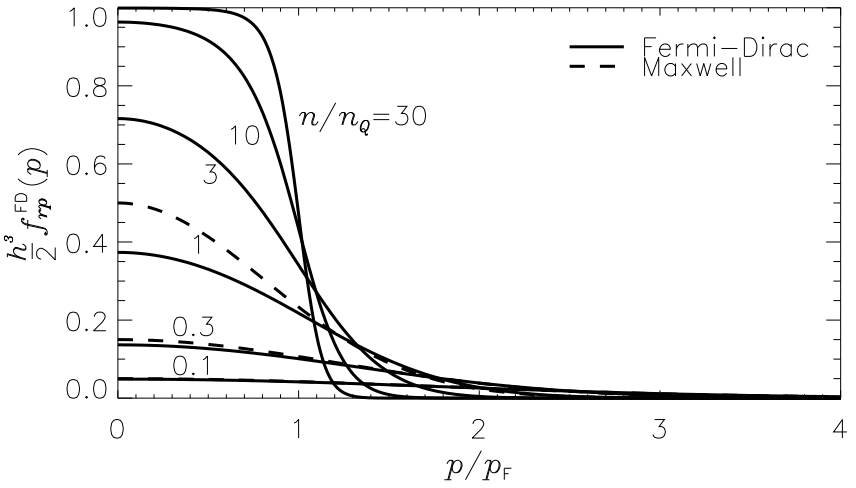

Figure 6.12: The Fermi-Dirac distribution function

These arguments may also be turned around to set limits to the validity of the Maxwell momentum distribution function. The maximum value of $f^M_{\boldsymbol{rp}}$ is found for $p = 0$. When this value exceeds the maximum value allowed by the Pauli principle, that is, when

$$n\left(\frac{1}{2\pi m\mathcal{T}}\right)^{3/2} = \frac{2}{h^3},$$

electron degeneracy will already have set in. Apart from the factor 2, this means when the electron density n equals the quantum concentration n_{Q}. The quantum concentration n_{Q} is therefore seen to be the critical density when to expect effects of quantum degeneracy to set in. In figure 6.12 also the Maxwell momentum distribution function $f^M_{\boldsymbol{rp}}(p)$ for $n/n_{\mathrm{Q}} = 0.1, 0.3$ and have been plotted (dotted lines) for comparison with the corresponding Fermi-Dirac

distribution. For $n/n_{\rm Q} = 0.1$ any difference can hardly be seen, for $n/n_{\rm Q} = 1$ significant differences exist.

Quiz 6.28 : If $f_{\boldsymbol{rp}} = f^D_{\boldsymbol{rp}}$ for the fully degenerate electron gas, what is the corresponding energy distribution function $f_{\boldsymbol{r}w}$ where $w = p^2/2m$?

Quiz 6.29 : Show that the average kinetic energy per particle in the fully degenerate gas is given by $\overline{w} = \frac{3}{5} w_F$ where w_F is the energy corresponding to the Fermi momentum p_F.

Quiz 6.30 : Derive a condition for beginning quantum degeneracy in an electron gas. In a $\log n$ – $\log T$ diagram delimit regions where the electron gas will behave as an ideal gas and where it will show effects of quantum degeneracy. Indicate in the same diagram where the quantum degeneracy will be relativistic. Finally, in the same diagram, plot the locations of the free electron gas in a) metallic Cu at room temperature, b) in the interior of the Sun ($\rho_m = 1.5 \cdot 10^5$ kg/m^3, $T = 1.5 \cdot 10^7$ K), and c) in the interior of a white dwarf ($\rho_m = 10^{10}$ kg/m^3, $T = 10^8$ K).

Quiz 6.31 : If we define "effective" temperature $T_{\rm eff}$ for the fully degenerate electron gas through $\overline{w} = \frac{3}{2}\mathcal{T}_{\rm eff}$, what is $T_{\rm eff}$ for the electron gas in metallic silver? [Assume each atom to contribute one free electron (conduction electron) to the gas.] What is $T_{\rm eff}$ for the interior of the Sun ($\rho_m = 1.5 \cdot 10^5$ kg/m^{-3}, $T = 1.5 \cdot 10^7$ K) and in a white dwarf ($\rho_m = 10^{10}$ kg/m^{-3}, $T = 10^8$ K)?

Quiz 6.32 : Protons and neutrons are also subject to quantum degeneracy. Why is the quantum degeneracy effect for these particles often negligible compared to that of the electrons?

Quiz 6.33 : Show that the distribution function $f^D_{\boldsymbol{rp}}(p)$ for the fully degenerate electron gas in a volume V at $\mathcal{T} = 0$ can be re-derived by assuming that in every phase space element $V\, d^3\boldsymbol{p}$ of size h^3 there can at most be 2 electrons (2 because of the spin multiplicity of the electron).

6.15 The Photon Gas

The photon is the elementary quantum of the electromagnetic radiation field. A photon of angular frequency ω has energy $W = \hbar\omega$ and momentum $p = \hbar\omega/c$. Even if the photon has only two different polarization or spin states, the photon is a spin $s = 1$ particle and therefore behaves as a *boson*. For bosons the Pauli principle does not apply. This means that in any given quantum state of the radiation field there may be an arbitrary number of photons. In the following we shall also make use of the fact that the chemical potential μ of the photon is known to vanish.

6.15.1 Equilibrium Radiation Field in Vacuum

To derive the photon distribution function we consider the simple case of a radiation field enclosed in a perfectly conducting box with sides of length L. The radiation field is determined

by the Maxwell equations (1.1)-(1.4). In vacuum, the electric field must obey the wave equation

$$\nabla^2 \boldsymbol{E} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\boldsymbol{E} = 0. \tag{6.106}$$

The boundary conditions require the electric field to be perpendicular to the perfectly conducting wall. For $x = 0, L$ the electric field must satisfy the conditions $E_y = E_z = 0$, and similarly at the other walls. It is easy to check that

$$\begin{aligned} E_x &= E_{0x}\cos(\omega t)\cos\frac{\pi q_x x}{L}\,\sin\frac{\pi q_y y}{L}\,\sin\frac{\pi q_z z}{L} \\ E_y &= E_{0y}\cos(\omega t)\sin\frac{\pi q_x x}{L}\,\cos\frac{\pi q_y y}{L}\,\sin\frac{\pi q_z z}{L} \\ E_z &= E_{0z}\cos(\omega t)\sin\frac{\pi q_x x}{L}\,\sin\frac{\pi q_y y}{L}\,\cos\frac{\pi q_z z}{L} \end{aligned} \tag{6.107}$$

represents a solution of (6.106) satisfying the imposed boundary conditions. The mode (or quantum) numbers $\boldsymbol{q} = (q_x, q_y, q_z)$ are all positive integers. The angular frequency ω is related to these mode numbers through the dispersion relation

$$\omega^2 = \frac{\pi^2 c^2}{L^2}(q_x^2 + q_y^2 + q_z^2) = \frac{\pi^2 c^2}{L^2}\,\boldsymbol{q}^2. \tag{6.108}$$

For each choice of mode numbers $\boldsymbol{q}$, the polarization of the radiation field must be chosen such as to satisfy the divergence condition on the electric field, $\nabla \cdot \boldsymbol{E} = 0$, that is,

$$E_{0x}q_x + E_{0y}q_y + E_{0z}q_z = 0.$$

Two different choices are always possible.

With photons acting as bosons, an arbitrary number of photons may occupy any given mode $\boldsymbol{q}$. Let us now consider as our system one such mode in thermal equilibrium at temperature $\mathcal{T}$. With a vanishing chemical potential for the photons, the partition function (6.97) is easily evaluated,

$$\mathcal{Z}_{\boldsymbol{q}} = \sum_{N=0}^{\infty} \exp\left(-\frac{N\hbar\omega_{\boldsymbol{q}}}{\mathcal{T}}\right) = \frac{1}{1 - \exp\left(-\frac{\hbar\omega_{\boldsymbol{q}}}{\mathcal{T}}\right)}. \tag{6.109}$$

The average number of photons in the given mode is then

$$\langle\, N_{\boldsymbol{q}} \,\rangle = \frac{1}{\mathcal{Z}_{\boldsymbol{q}}}\sum_{N=0}^{\infty} N\exp\left(-\frac{N\hbar\omega_{\boldsymbol{q}}}{\mathcal{T}}\right) = -\frac{1}{\mathcal{Z}_{\boldsymbol{q}}}\frac{\mathrm{d}\mathcal{Z}_{\boldsymbol{q}}}{\mathrm{d}\left(\frac{\hbar\omega_{\boldsymbol{q}}}{\mathcal{T}}\right)} = \frac{1}{\exp\left(\frac{\hbar\omega_{\boldsymbol{q}}}{\mathcal{T}}\right) - 1}. \tag{6.110}$$

The average number of photons in the volume V with angular frequency in the interval $(\omega, \omega + \mathrm{d}\omega)$ is found by summing over all modes with frequencies belonging to this interval. From (6.108) we have

$$\sum_{\boldsymbol{q}\in(\omega,\omega+\mathrm{d}\omega)} \cdot \to \frac{1}{8}4\pi q^2\,\mathrm{d}q\,\cdot = \frac{\omega^2}{2\pi^2 c^3}\,V\mathrm{d}\omega\,\cdot\,.$$

The average photon number in the specified frequency range and in a given polarization is therefore

$$\langle\, N_{\boldsymbol{q}} \,\rangle\,\frac{\omega^2}{2\pi^2 c^3}\,V\mathrm{d}\omega = \frac{1}{2\pi^2 c^3}\frac{\omega^2}{\exp\left(\frac{\hbar\omega}{\mathcal{T}}\right) - 1}\,V\mathrm{d}\omega,$$

from which the angular *frequency probability density* for the photons of one polarization is identified as

$$f_{\boldsymbol{r}\omega}(\omega) = \frac{1}{2\pi^2 c^3} \frac{\omega^2}{\exp\left(\frac{\hbar\omega}{\mathcal{T}}\right) - 1}. \tag{6.111}$$

With the added assumption of isotropy of the radiation field the corresponding directional frequency probability density and the *wave vector probability density* for one polarization is readily seen to be (see quiz 6.34)

$$f_{\boldsymbol{r}\omega\hat{\boldsymbol{\Omega}}} = \frac{1}{4\pi} f_{\boldsymbol{r}\omega} = \frac{1}{(2\pi)^3 c^3} \frac{\omega^2}{\exp\left(\frac{\hbar\omega}{\mathcal{T}}\right) - 1} \tag{6.112}$$

and

$$f_{\boldsymbol{r}\boldsymbol{k}}(\omega) = \frac{1}{(2\pi)^3} \frac{1}{\exp\left(\frac{\hbar\omega}{\mathcal{T}}\right) - 1}, \tag{6.113}$$

respectively.

From the photon probability densities (6.111)-(6.113) a number of important quantities can be derived. The spesific intensity represents the energy transported per time unit per solid angle and is therefore given by

$$\mathcal{I}_\omega = \hbar\omega\, c\, f_{\boldsymbol{r}\omega\hat{\boldsymbol{\Omega}}} = B_\omega(\mathcal{T}), \tag{6.114}$$

where

$$B_\omega(\mathcal{T}) \equiv \frac{\hbar}{(2\pi)^3 c^2} \frac{\omega^3}{\exp\left(\frac{\hbar\omega}{\mathcal{T}}\right) - 1} \tag{6.115}$$

is the *Planck radiation function* for one polarization. The associated radiation function in terms of frequency ν, $B_\nu(\mathcal{T}) = 2\pi B_\omega(\mathcal{T})$, is plotted in figure 6.13 for different temperatures T. The shape of the different curves are identical, but the curves shift to higher frequencies and higher intensities as the temperature increases. We note again that the function plotted refers to one polarization. To include both polarizations of the radiation field an additional factor 2 is needed.

The spectral energy density in the radiation field including both polarizations, that is, the energy in the radiation field per unit volume and per unit angular frequency for both polarizations is expressed as

$$u_\omega = 2\hbar\omega f_{\boldsymbol{r}\omega} = \frac{8\pi}{c} B_\omega(\mathcal{T}). \tag{6.116}$$

By integrating the spectral energy density u_ν over all frequencies, the total energy density u in the radiation field in thermal equilibrium at temperature $\mathcal{T}$ is found as

$$u = \int_0^\infty u_\omega \,\mathrm{d}\omega = \frac{8\pi\mathcal{T}^4}{h^3c^3} \int_0^\infty \frac{x^3\,\mathrm{d}x}{\exp(x) - 1} = \frac{8\pi^5}{15h^3c^3}\mathcal{T}^4 = a\mathcal{T}^4, \tag{6.117}$$

with the constant $a = 2.0821732 \cdot 10^{76}\mathrm{Jm}^{-3}\mathrm{J}^{-4}$ (or $a = 7.5659122 \cdot 10^{-16}\ \mathrm{Jm}^{-3}\mathrm{K}^{-4}$).

The amount of energy that is transported by the equilibrium radiation field per unit frequency interval and per unit time through a unit area with a given orientation $\hat{\boldsymbol{z}}$ is now easily determined. For this the specific intensity multiplied with the directional factor $\cos\theta$

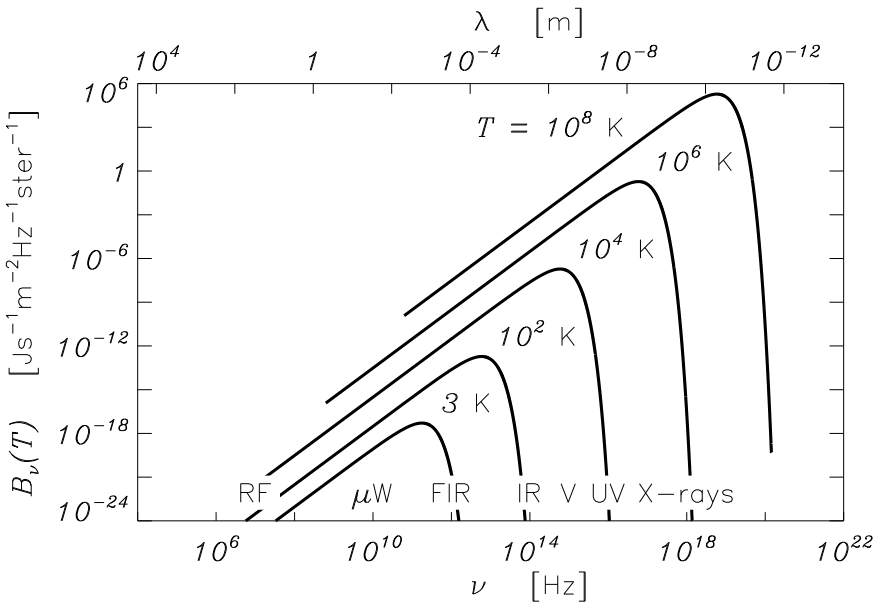

Figure 6.13: The Planck radiation function for different temperatures T

(see figure 1.8) is integrated over solid angles satisfying $\hat{\boldsymbol{z}} \cdot \boldsymbol{\Omega} > 0$. We refer to this quantity as the *unidirectional energy flux* in the direction $\hat{\boldsymbol{z}}$,

$$\mathcal{F}_\omega^+ = \int_0^1 \int_0^{2\pi} \cos\theta \, \mathcal{I}_\omega \, \mathrm{d}\cos\theta \, \mathrm{d}\varphi = \pi B_\omega(\mathcal{T}). \tag{6.118}$$

The corresponding total unidirectional flux is found by summing the contribution from all frequencies

$$F^+ = \int_0^\infty \mathcal{F}_\omega^+ \, \mathrm{d}\omega = \frac{2\pi^5}{15h^3c^2}\mathcal{T}^4 = \sigma\mathcal{T}^4 \tag{6.119}$$

where $\sigma = 1.5605495 \cdot 10^{84}$ $\mathrm{Wm^{-2}J^{-4}}$ (or $\sigma = 5.6705085 \cdot 10^{-8}$ $\mathrm{Wm^{-2}K^{-4}}$) is the *Stefan-Boltzmann constant*. In thermal equilibrium equal amounts of energy are transported through the given area in the opposite directions, $\mathcal{F}_\omega^- = \mathcal{F}_\omega^+$ and $F^- = F^+$.

Like material particles, the photons carry momentum. When reflected at a wall a photon will impart impulse to the wall. A photon gas therefore exerts a pressure on a wall just like a gas of particles does. The photon pressure can be found from (6.40) by replacing $f_{\boldsymbol{r}v}\,\mathrm{d}v$ with $2f_{\boldsymbol{r}\omega}\,\mathrm{d}\omega$, substituting $p = \hbar\omega/c$ and $v = c$ and summing over all frequencies. The extra factor 2 is again needed in order to take the contributions from both polarizations into account. The result is

$$P_r = 2\,\frac{1}{3}\int_0^\infty \frac{\hbar\omega}{c} c\, f_{\boldsymbol{r}\omega}\,\mathrm{d}\omega = \frac{1}{3}u. \tag{6.120}$$

6.15.2 Equilibrium Radiation Field in a Plasma

The results above all refer to electromagnetic radiation in thermodynamic equilibrium at temperature $\mathcal{T}$ in vacuum. The presence of a medium will lead to some modifications.

We take the wave vector probability density $f_{\boldsymbol{rk}}$ as given in (6.113) as our starting point. The photon distribution function $f_{\boldsymbol{r}\omega\hat{\boldsymbol{\Omega}}}$ can now be derived by making use of the expression

(1.117) for the $\mathrm{d}^3\boldsymbol{k}$ element in terms of angular frequency ω and the solid angle element $\mathrm{d}^2\Omega$ centered on the group velocity $\boldsymbol{v}_g$,

$$\mathrm{d}^3\boldsymbol{k} = \frac{\omega^2}{c^2}\frac{\mathcal{N}_r^2}{v_g}\,\mathrm{d}\omega\mathrm{d}^2\boldsymbol{\Omega}$$

with the ray refractive index $\mathcal{N}_r$ for the medium given by (1.118). From

$$f_{\boldsymbol{r}\boldsymbol{k}}\,\mathrm{d}^3\boldsymbol{k} = f_{\boldsymbol{r}\omega\hat{\boldsymbol{\Omega}}}\,\mathrm{d}\omega\mathrm{d}^2\Omega$$

we readily identify

$$f_{\boldsymbol{r}\omega\hat{\boldsymbol{\Omega}}} = \frac{\mathcal{N}_r^2}{(2\pi)^3c^2v_g}\,\frac{\omega^2}{\exp(\frac{\hbar\omega}{\mathcal{T}})-1}. \tag{6.121}$$

Evanescent modes for which the refractive index $\mathcal{N}^2$ and therefore also $\mathcal{N}_r^2$ are negative should be disregarded. We remember that for isotropic media the ray refractive index $\mathcal{N}_r$ reduces to the refractive index $\mathcal{N} = ck/\omega$.

With the photon distribution functions at hand derived quantities are easily calculated. At thermal equilibrium the spectral energy density in the plasma for a given propagating mode is

$$u_\omega = \int \hbar\omega f_{\boldsymbol{r}\omega\hat{\boldsymbol{\Omega}}}\,\mathrm{d}\Omega = \frac{\hbar}{(2\pi)^3c^2}\,\frac{\omega^3}{\exp(\frac{\hbar\omega}{\mathcal{T}})-1}\int\frac{\mathcal{N}_r^2}{v_g}\,\mathrm{d}^2\Omega \tag{6.122}$$

where the solid angle integral should ony be taken over directions for which the mode is propagating. The specific intensity for a given polarization in the plasma is given as

$$\mathcal{I}_\omega = \hbar\omega\,v_g\,f_{\boldsymbol{r}\omega\hat{\boldsymbol{\Omega}}} = \frac{\hbar\mathcal{N}_r^2}{(2\pi)^3c^2}\,\frac{\omega^3}{\exp(\frac{\hbar\omega}{\mathcal{T}})-1} = \mathcal{N}_r^2 B_\omega(\mathcal{T}). \tag{6.123}$$

This important result shows that in thermal equilibrium the ratio $\mathcal{I}_\omega/\mathcal{N}_r^2$ for any mode only depends on temperature, being independent of the presence of a material medium or not as long as the mode is propagating. The ratio will be equal to the Planck radiation function in vacuum.

In a volume containing both plasma and radiation field, the pressure contributions from the plasma (6.41) and the radiation field (6.120) must be added. It is important to note the widely different temperature dependence of these pressure contributions. At the high temperatures of the interior of stars, the radiation field may be a major contributor to the total pressure, near the cool stellar surface the radiation pressure is insignificant in comparison.

Quiz 6.34 : With $\mathrm{d}^3\boldsymbol{k} = k^2\,(\mathrm{d}\omega/\mathrm{d}k)^{-1}\,\mathrm{d}\omega\,\mathrm{d}^2\Omega_k$ where $\mathrm{d}^2\Omega_k$ is the infinitesimal solid angle element centered on the $\boldsymbol{k}$ direction, rederive the frequency probability density $f_{\boldsymbol{r}\omega}$ (6.111) in vacuum as a marginal probability density function of $f_{\boldsymbol{r}\boldsymbol{k}}$ (6.113).

Quiz 6.35 : The small frequency approximation, $\hbar\omega/\mathcal{T} \ll 1$, of the Planck radiation function $B_\omega(\mathcal{T})$ is called the Rayleigh-Jeans approximation. Give the explicit expression for this approximation and show that if this approximation is taken as generally valid, it will result in an "ultra-violet catastrophe" – the total energy in the radiation field becomes infinite.

Quiz 6.36 : What is the expression for the Planck radiation function B_λ as a function for wavelength λ, that is, the intensity per unit wavelength interval?

Quiz 6.37 : The Wien displacement law gives the location of maximum of the Planck radiation function as a function of wavelength λ for a given temperature T. Derive this law.

Quiz 6.38 : Carry through the arguments leading to (6.119). Verify that the explicit expression for the Stefan-Boltzmann constant σ and thus show that $\sigma = \frac{1}{4}ac$ where the constant a is defined in (6.117). Express the energy density u and the radiation pressure P_r in terms of the unidirectional energy flux $\mathcal{F}^+$.

Quiz 6.39 : Compare the contributions from electrons, ions and the radiation field to the total pressure in the interior of the Sun ($\rho_m = 1.5 \cdot 10^5$ kg/m^{-3}, $T = 1.5 \cdot 10^7$ K) and in a white dwarf ($\rho_m = 10^{10}$ kg/m^{-3}, $T = 10^8$ K)? With a surface temperature $T = 10^4$ K, what will the corresponding comparison near the surface of the star be?

Quiz 6.40 : In a photon gas for which the energy and pressure are given as $U = Vu(\mathcal{T})$ and $P = \frac{1}{3}u(\mathcal{T})$, show that the relation (6.64) requires that $u(\mathcal{T}) \sim \mathcal{T}^4$.

Chapter 7

Fluid Mechanics

In this chapter we shall establish the equations of motion for fluids and study some of their basic properties. The discussion will be based on an intuitive, macroscopic approach. This approach will normally lead to equations of motions suitable for large-scale properties of fluids systems while some microscopic phenomena are being overlooked.

By a *fluid* we shall mean any continuous medium in which the atoms or molecules are allowed to move freely with respect to each other. Fluids thus include liquids, gases and plasmas. The fluid will be described in terms of quantities like mass density ρ_m, flow velocity $\boldsymbol{v}$, and pressure P. These quantities will be functions of position $\boldsymbol{r}$ and time t, that is, the fluid will not be assumed to be in any global thermodynamic equilibrium state. The fluid will, however, be assumed to satisfy *local thermal equilibrium*. This means that we will assume a local equation of state for the fluid to exist.

7.1 The Continuity Equation

Consider an arbitrary volume V, fixed with respect to the chosen inertial system, and bounded by the surface $\mathcal{A}$. Let $\mathrm{d}^2\boldsymbol{\mathcal{A}} = \mathrm{d}^2\mathcal{A}\,\hat{\boldsymbol{n}}$ be an infinitesimal outward pointing element of this surface at position $\boldsymbol{r}$ in a fluid where the mass density is ρ_m and the flow velocity $\boldsymbol{v}$. Due to the flow velocity, mass will be transported across the surface A. The amount of mass transported through the surface element $\mathrm{d}^2\boldsymbol{\mathcal{A}}$ during the time interval $\mathrm{d}t$ is equal to the mass $\rho_m \boldsymbol{v}\,\mathrm{d}t\cdot\mathrm{d}^2\boldsymbol{\mathcal{A}}$ contained within a cylinder of base $\mathrm{d}^2\boldsymbol{\mathcal{A}}$ and side $\boldsymbol{v}\,\mathrm{d}t$. The geometry is illustrated in figure 7.1. Integrating over the closed surface A, the total amount of mass transported out of the volume V per unit time is given by $\oint_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}}\cdot\rho_m\boldsymbol{v}$. If mass is not created or destroyed in the fluid, the mass transport rate must be balanced by a corresponding rate of change of the total mass residing within V,

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_V \mathrm{d}^3\boldsymbol{r}\,\rho_m + \oint_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}}\cdot\rho_m\boldsymbol{v} = 0. \tag{7.1}$$

This is the *integral form* of the *mass continuity equation*.

With the integration volume V fixed with respect to the inertial system, the time derivative in the first term of (7.1) may be taken inside the volume integral. The second term can be rewritten as a volume integral by making use of Gauss integral theorem (A.40),

$$\oint_A \mathrm{d}^2\boldsymbol{\mathcal{A}}\cdot\rho_m\boldsymbol{v} = \int_V \mathrm{d}^3\boldsymbol{r}\,\nabla\cdot(\rho_m\boldsymbol{v}).$$

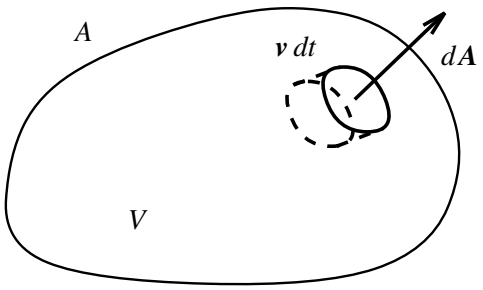

Figure 7.1: Fluid transport across a closed surface.

Equation (7.1) then takes the form

$$\int_V \mathrm{d}^3\boldsymbol{r}\left(\frac{\partial\rho_m}{\partial t}+\nabla\cdot(\rho_m\boldsymbol{v})\right)=0,$$

a relation which must be satisfied for any choice of the integration volume V. This can be achieved only if the integrand vanishes identically,

$$\frac{\partial\rho_m}{\partial t}+\nabla\cdot(\rho_m\boldsymbol{v})=0. \tag{7.2}$$

Equation (7.2) is the *differential form* of the *mass continuity equation.*

The individual terms in (7.2) are easily interpreted. The first term describes the local rate of change of mass per unit volume. From the definition of the divergence operator (1.5), the second term represents the outward directed mass flux per unit volume. For local mass conservation, the sum of these two terms must vanish.

The continuity equation determines the local mass density ρ_m in the fluid as a function of the flow velocity $\boldsymbol{v}$. In particular, if the flow velocity $\boldsymbol{v}(\boldsymbol{r},t)$ is given and the mass density ρ_m is specified at time $t=0$, then the continuity equation (7.2) determines $\rho_m(\boldsymbol{r},t)$ at all later times.

By expanding the divergence term, the continuity equation (7.2) may be given in the alternative form

$$\frac{\mathrm{D}\rho_m}{\mathrm{D}t}\equiv\left(\frac{\partial}{\partial t}+\boldsymbol{v}\cdot\nabla\right)\rho_m=-\rho_m\nabla\cdot\boldsymbol{v}. \tag{7.3}$$

The operator $\mathrm{D}/\mathrm{D}t$ defined in (7.3) is called the *convective derivative* and will be met repeatedly in the following. It has an intuitive physical interpretation. The first (local) part of the operator describes the rate of change as seen by an observer at a fixed location. The second (directional derivative) part multiplied by $\mathrm{d}t$, describes the change seen by an observer being subject to an "instantaneous displacement" $\boldsymbol{v}\,\mathrm{d}t$. Taken together, the two terms describe the rate of change seen by an observer convected with the local fluid velocity $\boldsymbol{v}$.

A flow for which the mass density ρ_m remains constant as seen by a co-moving observer, $\mathrm{D}\rho_m/\mathrm{D}t=0$, is called an *incompressible flow.* From (7.3) it follows that such a flow will satisfy $\nabla\cdot\boldsymbol{v}=0$. It should be noted that a fluid participating in an incompressible flow does not have to be an *incompressible fluid.*

Quiz 7.1 : Can you describe a situation for which the continuity equation (7.2) is *not* strictly valid.

Quiz 7.2 : Decide if any of the following velocity fields represent an incompressible flow:

$$\text{i) } \boldsymbol{v} = \frac{A}{r^n}\hat{\boldsymbol{r}} \quad \text{for} \quad n = \frac{1}{2}, 1 \text{ or } 2 \quad \text{ii) } \boldsymbol{v} = \boldsymbol{\Omega} \times \boldsymbol{r},$$

where A and $\boldsymbol{\Omega}$ are constants.

Quiz 7.3 : Interstellar gas falls radially in towards a star with mass M_* with velocity $v_r(r)$ given by

$$\frac{1}{2}v_r^2 \approx \frac{GM_*}{r}.$$

The increase in the stellar mass with time is negligible. Determine the mass density $\rho_m(r)$ in the gas for a stationary flow situation. The analysis should be carried through with the help of the continuity equation once in the form (7.2), and once in the form (7.1). [Hint: in the latter case, choose a volume in the form of a spherical shell with inner and outer radii r_1 and r_2.]

Quiz 7.4 : Equation (7.2) is the conservation law for mass density and mass flux. What is the analogous conservation law for electric charge density ρ and electric current density $\boldsymbol{j}$?

Quiz 7.5 : A fluid has flow velocity $\boldsymbol{v}(\boldsymbol{r})$. A volume V bounded by the surface $\mathcal{A}$ is flowing with the fluid. Show that the rate of change of the volume V is given by

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \oint_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}} \cdot \boldsymbol{v} = \int_V \mathrm{d}^3\boldsymbol{r}\, \nabla \cdot \boldsymbol{v}.$$

Thus argue that

$$\lim_{V\to 0} \frac{1}{V}\frac{\mathrm{d}V}{\mathrm{d}t} = \nabla \cdot \boldsymbol{v},$$

that is, $\nabla \cdot \boldsymbol{v}$ represents the local relative change in the volume of any fluid element per unit time.

7.2 The Momentum Equation

From the continuity equation, the density ρ_m may be determined once the flow velocity $\boldsymbol{v}$ is known. To determine the flow velocity, however, another equation is required. Such an equation can be derived by equating the rate of change of the momentum of a given fluid element to the total force acting on the element.

Consider again an arbitrary volume V, fixed with respect to the inertial system, and bounded by the closed surface $\mathcal{A}$. A force $\boldsymbol{F}$ is acting on the fluid contained within this volume. The momentum $\int_V \mathrm{d}^3\boldsymbol{r}\, \rho_m \boldsymbol{v}$ of the fluid within V will change with time due to the action of the force $\boldsymbol{F}$, but also because the fluid is transporting momentum by flowing through the chosen volume.

In complete analogy with the discussion of mass transport across the surface $\mathcal{A}$ in section 7.1, the corresponding momentum transport across $\mathcal{A}$ is given by $\oint_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}} \cdot \rho_m \boldsymbol{v}\,\boldsymbol{v}$. Momentum balance for the fluid within the volume V is therefore given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_V \mathrm{d}^3\boldsymbol{r}\,\rho_m\boldsymbol{v} + \oint_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}}\cdot\rho_m\boldsymbol{v}\boldsymbol{v} = \boldsymbol{F}. \tag{7.4}$$

It is again convenient to transform the surface integral into a volume integral. With the help of the Gauss integral theorem (A.40) and the continuity equation (7.2), the x-component of the left hand side of (7.4) is easily transformed into

$$\begin{aligned}
&\int_V \mathrm{d}^3\boldsymbol{r}\,\left\{\frac{\partial\rho_m v_x}{\partial t} + \nabla\cdot(\rho_m\boldsymbol{v}v_x)\right\} = \\
&\int_V \mathrm{d}^3\boldsymbol{r}\,\left\{\left(\frac{\partial\rho_m}{\partial t} + \nabla\cdot(\rho_m\boldsymbol{v})\right)v_x + \rho_m\left(\frac{\partial v_x}{\partial t} + \boldsymbol{v}\cdot\nabla v_x\right)\right\} = \\
&\int_V \mathrm{d}^3\boldsymbol{r}\,\rho_m\frac{\mathrm{D}v_x}{\mathrm{D}t}.
\end{aligned}$$

The y- and z-components transform accordingly. Equation (7.4) therefore reduces to

$$\int_V \mathrm{d}^3\boldsymbol{r}\,\rho_m\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = \boldsymbol{F}. \tag{7.5}$$

With the proper expression for the total force $\boldsymbol{F}$ acting on the fluid instantaneously contained within the volume V, equation (7.5) (or equivalently (7.4)) is the *integral form* of the *momentum equation* for the fluid.

Quiz 7.6 : Are you able to re-derive (7.5) from the "Newtonian" statement

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{V(t)} \mathrm{d}^3\boldsymbol{r}\,\rho_m\boldsymbol{v} = \boldsymbol{F}$$

where $V(t)$ is a volume element *flowing* with the fluid itself?

7.2.1 The ideal fluid

Forces acting on the fluid element may be of two types, a *volume force* $\boldsymbol{F}_V$ acting throughout the volume, or a *surface force* $\boldsymbol{F}_S$ describing forces acting upon the given fluid element from the surrounding fluid elements.

The gravitational force is an example of a volume force,

$$\boldsymbol{F}_V = -\int_V \mathrm{d}^3\boldsymbol{r}\,\rho_m\nabla\Phi_g. \tag{7.6}$$

Here Φ_g is the *gravitational potential* with the corresponding gravitational acceleration field $\boldsymbol{g} = -\nabla\Phi_g$. If the contribution to the total gravitational field from the mass of the system being studied is negligible, the gravitational potential Φ_g may be considered as an externally given potential. A self-gravitating system, where the gravitational potential Φ_g is completely

determined by the mass distribution $\rho_m(\boldsymbol{r})$ itself, represents the opposite extreme. For this case the gravitational potential is given by

$$\Phi_g(\boldsymbol{r}) = -G \int \mathrm{d}^3\boldsymbol{r}' \, \frac{\rho_m(\boldsymbol{r}')}{|\,\boldsymbol{r} - \boldsymbol{r}'\,|}, \tag{7.7}$$

where $G = 6.672 \cdot 10^{-11}$ Nm^2/kg^2 is the *gravitational constant*. Relation (7.7) represents an explicit expression for the gravitational potential Φ_g in terms of the mass density ρ_m. Even so, it is often convenient to make use of the equivalent differential relationship

$$\nabla^2\Phi_g = 4\pi G\rho_m. \tag{7.8}$$

This result is simply derived by applying the ∇^2-operator to the integral in (7.7) and making use of the properties of Dirac's δ-function as given in (A.70)-(A.75).

The pressure force is an example of a surface force,

$$\boldsymbol{F}_S = -\oint_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}}\, P = -\int_V \mathrm{d}^3\boldsymbol{r}\, \nabla P. \tag{7.9}$$

Here P represents the pressure at the boundary A. The negative sign is introduced because the surface element $\mathrm{d}^2\boldsymbol{\mathcal{A}}$ is pointing outward while the pressure force from the surrounding fluid is acting in the opposite direction. The expression of the surface force as a volume integral over the pressure gradient ∇P is a simple application of the generalized Gauss integral theorem (A.50).

A fluid that in addition to the pressure force (7.9) is only subject to forces per unit mass derivable from a potential, as exemplified by the gravitational force (7.6), is referred to as an *ideal fluid*. The momentum equation (7.5) for the ideal fluid can now be written

$$\int_V \mathrm{d}^3\boldsymbol{r}\, \rho_m \frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = -\int_V \mathrm{d}^3\boldsymbol{r}\, (\nabla P + \rho_m \nabla\Phi_g).$$

This equation must be valid for any choice of integration volume V and may therefore be reduced to

$$\rho_m \frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = -\nabla P - \rho_m \nabla\Phi_g. \tag{7.10}$$

This is the *differential form* of the *momentum equation* for an ideal fluid.

Real fluids may often to a good approximation be represented as an ideal fluid. Other times additional forces acting on the fluid have to be taken into account. One example is the magnetic force acting on electrically conducting fluids. We shall return to a discussion of this force and its consequences in subsequent sections. Here we shall introduce still another additional force, the *viscous force*. This is a force which appears when two adjacent fluid layers with a relative velocity are exchanging momentum.

Quiz 7.7 : Verify the identity

$$\oint_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}}\, P = \int_V \mathrm{d}^3\boldsymbol{r}\, \nabla P$$

by direct integration. [Hint: Calculate the x-component of the surface force for a volume V in the form of a regular cube with sides parallel to the coordinate planes.]

Quiz 7.8 : Make use of (7.2) to show that the LHS of (7.10) may be rewritten in the form

$$\rho_m \frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = \frac{\partial \rho_m \boldsymbol{v}}{\partial t} + \nabla \cdot (\rho_m \boldsymbol{v}\boldsymbol{v}) \tag{7.11}$$

7.2.2 The viscous force

Even if neighboring atoms or molecules may move relative to each other in fluids, this does not mean that a steady state velocity shear may be maintained in the fluid without the action of external forces. Here we shall see how an empirical expression for the *viscous force* acting between neighboring fluid layers can be found.

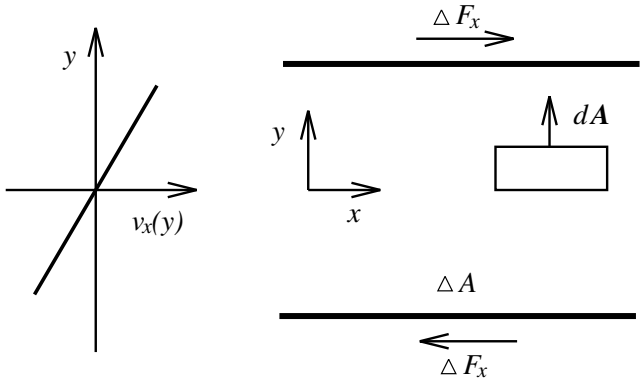

Figure 7.2: Simple slab velocity shear geometry

For this purpose consider a uniform fluid slab in the xz-plane and thickness Δy in the y-direction as illustrated in figure 7.2. A velocity profile $\boldsymbol{v} = v_x(y)\hat{\boldsymbol{x}}$ with constant velocity shear $\mathrm{d}v_x(y)/\mathrm{d}y$ can be maintained by pulling the two bounding planes of area $\Delta\mathcal{A}$ at $y = \pm\Delta y/2$ with equal and opposite forces $\pm\Delta F_x$ in the x-direction. Experimentally, it is found that the force per unit area, $\Delta F_x/\Delta\mathcal{A}$, necessary to maintain a stationary flow, increases linearly with the velocity shear,

$$\frac{\Delta F_x}{\Delta \mathcal{A}} = \eta \frac{\mathrm{d}v_x}{\mathrm{d}y}. \tag{7.12}$$

The constant of proportionality η is called the *coefficient of viscosity* and is a constant depending on the particular fluid and its temperature. The coefficient of viscosity η is also called the *shear viscosity* or the *molecular viscosity*.

A force per unit area as given by (7.12), must appear at any plane in the fluid parallel to the two bounding planes. The faster moving fluid on one side of this plane will be trying to pull the slower fluid on the other side. The slower fluid is at the same time trying to brake the faster fluid. The expression $\Delta F_x/\Delta A$ has physical dimensions equal to that of pressure. In contrast to the pressure force acting *perpendicular* to any bounding surface, the present force is directed *in* this plane.

The experimental relation (7.12) may be used to express the viscous surface force acting on any fluid element. For the instantaneous rectangular shaped element indicated in figure 7.2, the force $\mathrm{d}^2\boldsymbol{F}_\eta$ acting on the top side surface element $\mathrm{d}^2\boldsymbol{\mathcal{A}} = \mathrm{d}^2\mathcal{A}\,\hat{\boldsymbol{y}}$ is given as

$$\mathrm{d}^2\boldsymbol{F}_\eta = \eta\,\mathrm{d}^2\mathcal{A}\,\hat{\boldsymbol{y}} \cdot \nabla v_x(y)\hat{\boldsymbol{x}} = \eta\,\mathrm{d}^2\boldsymbol{\mathcal{A}} \cdot \nabla \boldsymbol{v}. \tag{7.13}$$

The right hand expression will be valid also for the bottom side as long as $\mathrm{d}^2\boldsymbol{\mathcal{A}}$ is the corresponding outward pointing surface element.

When arguing for the expression (7.13), a special velocity profile was assumed. The question is now if this expression is valid also for an arbitrary velocity field. The case of a rigidly rotating fluid mass shows that this is *not* so. For the rigidly rotating fluid mass,

$$\boldsymbol{v} = \boldsymbol{\Omega} \times \boldsymbol{r}, \tag{7.14}$$

where the *angular velocity* $\boldsymbol{\Omega}$ is any constant vector, there is no relative motion between neighboring fluid elements. The viscous shear force on any surface element must therefore necessarily vanish in this case. But if (7.14) is substituted in (7.13), a non-vanishing result is found.

The proper expression for the viscous surface force can instead be shown to be

$$\mathrm{d}^2\boldsymbol{F}_\eta = \eta\,(\mathrm{d}^2\boldsymbol{\mathcal{A}}\cdot\nabla\boldsymbol{v} + \nabla\boldsymbol{v}\cdot\mathrm{d}^2\boldsymbol{\mathcal{A}} - \frac{2}{3}\nabla\cdot\boldsymbol{v}\,\mathrm{d}^2\boldsymbol{\mathcal{A}}) + \zeta\nabla\cdot\boldsymbol{v}\,\mathrm{d}^2\boldsymbol{\mathcal{A}}. \tag{7.15}$$

The extra coefficient ζ in (7.15) is known as the *bulk viscosity*. η and ζ are both non-negative quantities. The two terms of (7.15) represent the contributions to friction forces from straining and dilatational motions in the fluid respectively. We note in particular that for incompressional fluid motions the bulk viscosity term vanishes. The particular combination of terms given in (7.15) reduces to (7.13) for the simple slab geometry has in addition the following properties: it vanishes for the rigidly rotating fluid, conserves angular momentum, and has the form necessary to make sure that the entropy production associated with the irreversible viscous heating in the fluid always remains non-negative.

The total viscous surface force acting on a given fluid element from the neighboring fluid, is found by summing individual contributions over the surface $\mathcal{A}$. If we assume η ζ to be constants and make use of Gauss integral theorem (A.40), the viscous force can be expressed as a volume force,

$$\boldsymbol{F}_\eta = \oint_A \left[\eta\left(\mathrm{d}^2\boldsymbol{\mathcal{A}}\cdot\nabla\boldsymbol{v} + \nabla\boldsymbol{v}\cdot\mathrm{d}^2\boldsymbol{\mathcal{A}} - \frac{2}{3}\nabla\cdot\boldsymbol{v}\,\mathrm{d}^2\boldsymbol{\mathcal{A}}\right) + \zeta\nabla\cdot\boldsymbol{v}\,\mathrm{d}^2\boldsymbol{\mathcal{A}}\right] = \int_V \mathrm{d}^3\boldsymbol{r}\,\boldsymbol{\mathcal{F}}_\eta, \tag{7.16}$$

where

$$\boldsymbol{\mathcal{F}}_\eta = \eta\,(\nabla^2\boldsymbol{v} + \frac{1}{3}\nabla\nabla\cdot\boldsymbol{v}) + \zeta\nabla\nabla\cdot\boldsymbol{v} \tag{7.17}$$

is the *viscous force* per unit volume.

The viscous force per unit volume $\boldsymbol{\mathcal{F}}_\eta$ should be added to the right hand side of the momentum equation (7.10). The momentum equation with the viscous force included,

$$\rho_m\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = -\nabla P - \rho_m\nabla\Phi_g + \boldsymbol{\mathcal{F}}_\eta, \tag{7.18}$$

is known as the *Navier-Stoke equation*.

Quiz 7.9 : Verify that (7.15) reduces to (7.13) for the simple slab geometry and vanishes for the case of the rigid rotor velocity field (7.14).

Quiz 7.10 : Show that (7.17) follows from (7.15).

7.3 The Energy Equation

The next task is the derivation of an equation describing the energy balance in the fluid. In addition to the kinetic energy $\boldsymbol{v}^2/2$ per unit fluid mass associated with the macroscopic motion of the fluid, the fluid will also possess internal energy. We shall refer to the internal energy per unit fluid mass as the *specific internal energy* u. The specific internal energy may consist of several contributions, thermal energy associated with the thermal motions of individual atoms in the fluid will always be one. For the moment the detailed form of u is not needed.

7.3.1 The ideal fluid

We first consider the case of an ideal fluid for which effects due to viscous forces are absent. We shall, however, allow for energy transport through heat conduction in the fluid. Consider again a fixed volume element V. The time rate of change of kinetic and internal energy in this volume, will in addition to the direct transport of such energy out of the volume through the motion of the fluid itself, be determined by the work done by the fluid against the pressure forces at the boundary, the work done by volume forces inside the fluid element, and the heat conduction across the boundary. In analogy with the previous derivation of the conservation laws for mass and momentum, and with the heat flux given by $-\lambda\nabla\mathcal{T}$, where λ is the *coefficient of heat conduction*, we may immediately write down the *integral form* of the *energy equation* as

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int_V \mathrm{d}^3\boldsymbol{r}\,\left(\frac{1}{2}\rho_m\boldsymbol{v}^2+\rho_m u\right)+\oint_{\mathcal{A}}\mathrm{d}^2\boldsymbol{\mathcal{A}}\cdot\boldsymbol{v}\left(\frac{1}{2}\rho_m\boldsymbol{v}^2+\rho_m u\right)\\ =-\oint_{\mathcal{A}}\mathrm{d}^2\boldsymbol{\mathcal{A}}\cdot\boldsymbol{v}P-\int_V\mathrm{d}^3\boldsymbol{r}\,\rho_m\boldsymbol{v}\cdot\nabla\Phi_g+\oint\mathrm{d}^2\boldsymbol{\mathcal{A}}\cdot\lambda\nabla\mathcal{T}.\end{aligned}\tag{7.19}$$

The surface integrals may be transformed into volume integrals with the help of Gauss integral theorem (A.40). Then, since (7.19) must be valid for any choice of the volume V, it may also be written

$$\begin{aligned}\frac{\partial}{\partial t}\left(\frac{1}{2}\rho_m\boldsymbol{v}^2+\rho_m u\right)+\nabla\cdot\left(\frac{1}{2}\rho_m\boldsymbol{v}^2\boldsymbol{v}+\rho_m u\boldsymbol{v}\right)\\ =-\boldsymbol{v}\cdot\nabla P-P\nabla\cdot\boldsymbol{v}-\rho_m\boldsymbol{v}\cdot\nabla\Phi_g+\nabla\cdot(\lambda\nabla\mathcal{T}).\end{aligned}$$

Making use of (7.2) and (7.10), most terms cancel and we are left with the *differential form* of the *energy equation*,

$$\rho_m\frac{\mathrm{D}u}{\mathrm{D}t}=-P\nabla\cdot\boldsymbol{v}+\nabla\cdot(\lambda\nabla\mathcal{T}).\tag{7.20}$$

The $P\nabla\cdot\boldsymbol{v}$ term on the right hand side represents the work done by an expanding unit volume fluid element against the pressure force from the surrounding fluid. This follows easily from the interpretation of $\nabla\cdot\boldsymbol{v}$ as the rate of increase of the unit volume (see quiz 7.5). The second term describes the effect of heat conduction, that is, how internal energy is transported within the fluid. Heat conduction may increase or decrease the specific internal energy of a given fluid element, depending on the local temperature profile. We notice that the gravitational force does *not* contribute to the internal energy balance in the fluid.

7.3.2 Effects of viscous forces

The viscous force $\mathrm{d}^2\boldsymbol{F}_\eta$ acting at the surface element $\mathrm{d}^2\boldsymbol{\mathcal{A}}$, as given by (7.15), will contribute with one extra term $\oint_{\mathcal{A}} \mathrm{d}^2\boldsymbol{F}_\eta \cdot \boldsymbol{v}$ on the right hand side of the energy equation (7.19). If the coefficients of shear and bulk viscosities η and ζ may be considered constant throughout the fluid, this extra term may be written

$$\begin{aligned}\oint_{\mathcal{A}} \mathrm{d}^2\boldsymbol{F}_\eta \cdot \boldsymbol{v} &= \int_V \mathrm{d}^3\boldsymbol{r}\ \left[\eta\nabla\cdot\left(\nabla\boldsymbol{v}\cdot\boldsymbol{v}+\boldsymbol{v}\cdot\nabla\boldsymbol{v}-\frac{2}{3}\boldsymbol{v}\nabla\cdot\boldsymbol{v}\right)+\zeta\nabla\cdot(\boldsymbol{v}\nabla\cdot\boldsymbol{v})\right]\\ &= \int_V \mathrm{d}^3\boldsymbol{r}\ \left[\eta\left((\nabla^2\boldsymbol{v}+\frac{1}{3}\nabla\nabla\cdot\boldsymbol{v})\cdot\boldsymbol{v}+\nabla\boldsymbol{v}:\widetilde{\nabla\boldsymbol{v}}+\nabla\boldsymbol{v}:\nabla\boldsymbol{v}-\frac{2}{3}(\nabla\cdot\boldsymbol{v})^2\right)+\zeta(\nabla\cdot\boldsymbol{v})^2\right]\\ &= \int_V \mathrm{d}^3\boldsymbol{r}\ \left(\boldsymbol{\mathcal{F}}_\eta\cdot\boldsymbol{v}+\mathcal{D}\right).\end{aligned}$$

The viscous force per unit volume $\mathcal{F}_\eta$ was defined in (7.17), while the *viscous dissipation rate*

$$\mathcal{D} = \eta\left(2\boldsymbol{\Phi}:\boldsymbol{\Phi}-\frac{2}{3}(\nabla\cdot\boldsymbol{v})^2\right)+\zeta(\nabla\cdot\boldsymbol{v})^2 \tag{7.21}$$

is given in terms of the *rate of strain increase*

$$\boldsymbol{\Phi} = \frac{1}{2}(\nabla\boldsymbol{v}+\widetilde{\nabla\boldsymbol{v}}). \tag{7.22}$$

The rate of strain increase is a symmetric second order tensor. The transpose and the double contraction operators "$\widetilde{\ }$" and "$:$" are defined in appendix A. The viscous dissipation rate is a non-negative quantity. This can be seen by noting that $\mathcal{D}$ in Cartesian coordinates may be written as a sum of explicit non-negative terms

$$\mathcal{D} = 4\eta\left(\Phi_{xy}^2+\Phi_{yz}^2+\Phi_{zx}^2\right)+\frac{2}{3}\eta\left((\Phi_{xx}-\Phi_{yy})^2+(\Phi_{yy}-\Phi_{zz})^2+(\Phi_{zz}-\Phi_{xx})^2\right)+\zeta(\nabla\cdot\boldsymbol{v})^2.$$

For the differential form of the energy equation, only the viscous dissipation rate term $\mathcal{D}$ will survive. That is, when taking the effects of viscous forces into account, the energy equation (7.20) must be replaced with

$$\rho_m\frac{\mathrm{D}u}{\mathrm{D}t} = -P\nabla\cdot\boldsymbol{v}+\nabla\cdot(\lambda\nabla\mathcal{T})+\mathcal{D}. \tag{7.23}$$

The last term describes how viscous effects transform macroscopic kinetic energy in the fluid into internal energy. Because of the non-negative property of the dissipation rate $\mathcal{D}$, viscous effects will always act to increase the internal energy. In this way viscous effects always lead to irreversible processes.

7.4 The Closure Problem

The continuity and momentum equations (7.2) and (7.10) (or (7.18)) represent four scalar equations for five scalar quantities, the density ρ_m, three components of the flow velocity $\boldsymbol{v}$ and the pressure P. The system of equations is not yet complete. We shall need at least one more equation in order to have a self-consistent description of the fluid behavior. The problem is not solved by adding the energy equation (7.20). This increases the number of

equations by one, but also adds the internal energy per unit mass u to the list of dependent variables. To find a way out of this problem, normally referred to as the *closure problem*, we will make use of our previous discussion of thermal physics. We assume that the thermal physics results are still valid locally in the fluid. As thermodynamic system, we consider any moving fluid element interacting with its neighboring fluid elements.

With this point of view, we may predict the existence of an equation of state for the fluid, that is, the existence of a functional relationship between the local values of pressure P, density ρ_m and temperature $\mathcal{T}$,

$$P = P(\rho_m, \mathcal{T}). \tag{7.24}$$

If a general expression for the specific internal energy u in terms of density and temperature can be provided,

$$u = u(\rho_m, \mathcal{T}), \tag{7.25}$$

then a complete set of equations of motion for the fluid have been secured.

In the following we shall consider the case of an ideal mono-atomic gas with equation of state and specific internal energy given by

$$P = \frac{\rho_m \mathcal{T}}{\mu m_{\mathrm{H}}} \quad \text{and} \quad u = \frac{3\mathcal{T}}{2\mu m_{\mathrm{H}}}. \tag{7.26}$$

Here μ is the mean molecular weight of the particles in the gas. Elimination of temperature leads to

$$u = \frac{P}{(\gamma - 1)\rho_m} \tag{7.27}$$

in terms of the ratio $\gamma = 5/3$ of the specific heat at constant pressure and constant volume in the mono-atomic gas. With (7.27) as closure procedure we will now consider some limiting cases.

7.4.1 Adiabatic process

In an adiabatic process, the heat exchange between a fluid element and its surroundings vanishes, $\lambda = 0$. If the process is also reversible, then the viscous dissipation rate vanishes, $\mathcal{D} = 0$. Substitution of (7.27) into (7.20) and making use of (7.3) to eliminate $\nabla \cdot \boldsymbol{v}$ on the right hand side of (7.20) leads to

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(P\rho_m^{-\gamma}\right) = 0. \tag{7.28}$$

We shall refer to (7.28) as the *adiabatic law*. The result may be compared with the previous result (6.72) for reversible adiabatic processes. The constancy of PV^γ in an adiabatic process for a moving fluid element with infinitesimal volume varying as $V \sim 1/\rho_m$ reduces to the requirement that the convective derivative of $P\rho_m^{-\gamma}$ vanishes. This should not be a surprising result.

The adiabatic law (7.28) often describes physical processes to a fair degree of accuracy if these occur over short enough time scales such that heat conduction during the process can be neglected. This will normally apply to most wave phenomena, but also to some quasi-steady situations. The adiabatic equation of state, together with the continuity and momentum equations, will be sufficient in order to secure a complete set of hydrodynamic equations. That is, with suitable boundary conditions added, the equations (7.2), (7.10) and (7.28) will allow for the determination of ρ_m, $\boldsymbol{v}$ and P as functions of space and time.

7.4.2 Isothermal process

The isothermal process represents the opposite extreme. The process now has to proceed slowly enough that heat exchange between fluid elements is able to maintain constant temperature throughout the system. A constant temperature $\mathcal{T}$ is the trivial solution of (7.20) in the limit $\lambda \to \infty$.

For a fluid behaving as an ideal gas, a complete set of hydrodynamic equations is obtained by adding the equation of state

$$P = \frac{\rho_m \mathcal{T}}{\mu m_{\rm H}}, \tag{7.29}$$

to the continuity and momentum equations, (7.2) and (7.10). We note for the following discussion that with temperature $\mathcal{T}$ considered constant, (7.29) can in differential form be written

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(P\rho_m^{-1}\right) = 0. \tag{7.30}$$

7.4.3 Polytropic processes

The adiabatic and isothermal processes represent the two limiting cases of vanishing and very large heat conduction. The real situation will normally lie somewhere in between. For this situation the full energy equation (7.20) will strictly have to be solved together with the continuity and momentum equations (7.2) and (7.10), and the proper closure procedure (7.27).

Instead of proceeding with the full energy equation it is often possible to make use of an ad hoc procedure. If (7.28) and (7.30) represent the two limiting cases, one would expect

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(P\rho_m^{-n}\right) = 0, \tag{7.31}$$

where n is an adjustable constant taking values in the range $1 < n < \gamma$, to approximate the combined effects of the energy equation (7.20) and the closure procedure (7.27). A fluid satisfying (7.31) for some choice of n, is referred to as a *polytrop*. The polytropic model has been widely used, for instance, in the modeling of stellar interiors.

7.5 Hydrostatic Equilibrium

We now turn to a discussion of some of the consequences of the fluid equations introduced above. Consider the simple case of a fluid in *hydrostatic equilibrium*. This is a state in which the fluid is everywhere at rest $\boldsymbol{v} = \mathbf{0}$, and where density and pressure are both constant with respect to time t. For this case, the continuity equation (7.2) is trivially satisfied. Viscous forces generally vanish, and the momentum equation (7.10) reduces to

$$\nabla P = -\rho_m \nabla \Phi_g. \tag{7.32}$$

For a solution of (7.32) to exist, the curl of the right hand side must vanish,

$$\nabla \rho_m \times \nabla \Phi_g = \mathbf{0}. \tag{7.33}$$

In hydrostatic equilibrium, the gradient of the mass density must therefore everywhere be parallel to the gravitational force field. This is equivalent to the requirement that any constant density surface must everywhere coincide with a corresponding constant gravitational potential surface.

We shall study consequences of (7.32) for two different situations, one for which the magnitude of the gravitational acceleration can be considered constant, and one for which the variation of the gravitational acceleration is important. The former case applies to discussions of the properties of planetary atmospheres confined to the immediate neighborhood of the planet. The latter case is relevant to the question of the internal structure of stars and extended stellar atmospheres.

7.5.1 The barometric formula

For a fluid in a constant gravitational field $\nabla\Phi_g = g\hat{\boldsymbol{z}}$, the result (7.33) requires the mass density ρ_m to be a function of z only. It follows from the pressure balance condition (7.32) that this must also be the case for the pressure P,

$$\frac{\mathrm{d}}{\mathrm{d}z}P(z) = -\rho_m(z)g. \tag{7.34}$$

To proceed, we invoke the equation of state for an ideal gas in the form (7.29). Equation (7.34) then reduces to an equation for the pressure alone,

$$\frac{\mathrm{d}P}{\mathrm{d}z} = -\frac{\mu m_{\mathrm{H}} g}{\mathcal{T}}P. \tag{7.35}$$

For an isothermal atmosphere in which $\mathcal{T}$ is constant, the solution is

$$P(z) = P_0 \exp\left(-\frac{\mu m_{\mathrm{H}} g}{\mathcal{T}}z\right). \tag{7.36}$$

P_0 is the pressure at some reference level $z = 0$. The result, known as the *barometric formula*, is illustrated in figure 7.3. The pressure in the isothermal atmosphere decreases exponentially with height with scale height $H = \mathcal{T}/\mu m_{\mathrm{H}} g$. From the physical point of view, the pressure at any point in the atmosphere adjusts to withstand the total weight per unit area of the atmosphere above that point.

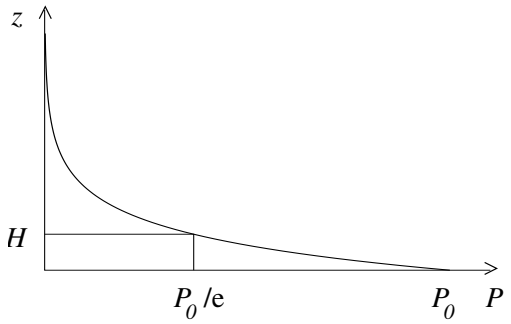

Figure 7.3: Pressure profile of an isothermal atmosphere

The barometric formula (7.36) is easily generalized to a situation where heat sources or other energy deposition mechanisms maintain a stationary temperature profile in the atmosphere, $\mathcal{T} = \mathcal{T}(z)$. The mean molecular weight mass may also be allowed to vary with height. The solution of the pressure balance equation (7.35) is this time

$$P(z) = P_0 \exp\left(- \int_0^z d\zeta \, \frac{\mu(\zeta)\, m_{\rm H} g}{\mathcal{T}(\zeta)} \right). \tag{7.37}$$

An explicit solution requires the height profile of temperature and mean molecular weight to be known. Any such profile does not, however, give rise to a physically relevant solution. As we shall see below, formal solutions of the form (7.37) may be unstable to perturbations and therefore not physically realistic.

Quiz 7.11 : What is the scale height H for the lower part of the Earths atmosphere?

Quiz 7.12 : From (7.34) show that the pressure in an hydrostatic atmosphere everywhere equals the total weight of the overlying gas per unit area.

Quiz 7.13 : Show that the barometric formula for a spherically symmetric, local atmosphere, in which the gravitational acceleration is radially directed, but of constant magnitude, is identical to the plane parallel case with r replacing z.

Quiz 7.14 : We consider the possibility of an extended static and isothermal atmosphere around a star with mass M_* and radius R_*. Let the atmospheric mass density near the stellar surface be $\rho_{m*} = \rho_m(R_*)$. Neglecting self-gravitational effects, the gravitational potential is given by $\Phi_g = -GM_*/r$. Show that

$$\rho_m(r) = \rho_{m*} \exp\left(-\frac{GM_* \mu m_{\rm H}}{\mathcal{T} R_*} \left(1 - \frac{R_*}{r} \right) \right). \tag{7.38}$$

What is the limiting density at infinity? Would you consider this type of solution to be realistic?

Quiz 7.15 : The coefficient of heat conduction λ for an ionized gas is a strongly varying function of gas temperature $\mathcal{T}$, $\lambda \sim \mathcal{T}^{5/2}$. Make use of the energy equation (7.23) (neglecting viscous effects) to show that in a static stellar atmosphere, where energy transport is dominated by heat conduction, the temperature will vary with radial distance r as $\mathcal{T} \sim r^{-2/7}$. What is the corresponding mass density $\rho_m(r)$ in such an atmosphere?

7.5.2 Static stellar models

Let us now consider the corresponding situation inside a spherically symmetric star of radius R_* and mass M_*. The gravitational field inside the star is determined by the mass distribution inside the star itself. At radius r the gravitational acceleration is given by

$$\boldsymbol{g}(r) = -\frac{GM(r)}{r^2}\, \hat{\boldsymbol{r}}, \tag{7.39}$$

where

$$M(r) = 4\pi \int_0^r \rho_m(r')\, r'^2 \, \mathrm{d}r' \tag{7.40}$$

is the total stellar mass inside radius r.

With the help of the pressure balance condition (7.32), the required pressure at the center of the star may now be expressed in terms of the density profile $\rho_m(r)$,

$$P_c = G \int_0^{R_*} \frac{M(r)\rho_m(r)}{r^2} \, \mathrm{d}r. \tag{7.41}$$

To evaluate P_c it is necessary to know the mass density profile of the star,

$$\rho_m(r) = \rho_{mc} f(r/R_*)$$

where ρ_{mc} is the central density and the function $f(\xi)$ decreases from one to zero as ξ increases from zero to one. The detailed form of $f(\xi)$ depends on the equation of state for the stellar interior and also the temperature profile of the star. The latter in turn depends on the stellar heat source and the way energy is transported in the star. A discussion of these questions falls outside the scope of our presentation. We shall see, however, that important scaling conclusions can be reached even without this detailed knowledge if we assume the normalized mass density profile to remain largely constant for each main class of stars, for instance, among the main sequence stars and for the different classes of compact stars (white dwarfs, neutron stars).

For a given density profile, the central density and pressure can be expressed in terms of the stellar mass and radius,

$$M_* = 4\pi\alpha\rho_{mc}R_*^3 \tag{7.42}$$

$$P_c = 4\pi G\beta\rho_{mc}^2 R_*^2, \tag{7.43}$$

where α and β are approximate constants for each class of stars, expressed in terms of integrals over the normalized density profile $f(\xi)$,

$$\alpha = \int_0^1 \mathrm{d}\xi\, \xi^2 f(\xi) \quad \text{and} \quad \beta = \int_0^1 \mathrm{d}\xi\, \frac{f(\xi)}{\xi^2} \int_0^\xi \mathrm{d}\eta\, \eta^2 f(\eta).$$

For static equilibrium to exist, it is necessary that the combined gas and radiation pressure at the center of the star can be maintained at the value required by (7.41),

$$P_c = P_e + P_i + P_{rad}. \tag{7.44}$$

The evaluation of P_e, P_i and P_{rad} normally requires the central particle densities for electrons and ions and the temperature to be known.

Main sequence stars

Let us first consider main sequence stars. We know that the gas in the center of such stars can be approximated as an ideal gas. Let the combined central gas pressure constitute a fraction δ of the necessary central pressure P_c. We may then write

$$P_e + P_i = \frac{\rho_{mc}}{\mu m_{\mathrm{H}}} \mathcal{T}_c = \delta P_c$$

and

$$P_{rad} = \frac{1}{3} a \mathcal{T}_c^4 = (1-\delta) P_c,$$

where the constant a is defined in (6.117). Elimination of the central temperature $\mathcal{T}_c$ between these relations gives

$$P_c = \left(\frac{1-\delta}{\delta^4}\right)^{1/3} \left(\frac{3}{a}\right)^{1/3} \left(\frac{1}{\mu m_{\mathrm{H}}}\right)^{4/3} \rho_{mc}^{4/3}.$$

Together with (7.42) and (7.43) this relation in turn leads to

$$M_*^2 \frac{\delta^4}{1-\delta} = \frac{3\alpha^2}{4\pi G^3 \beta^3 a} \left(\frac{1}{\mu m_{\mathrm{H}}}\right)^4.$$

According to our assumptions, the right hand side will remain an approximate constant among the main sequence stars. It follows that for stars of this type to maintain hydrostatic equilibrium, the radiation pressure must necessarily play an increasingly important role as the mass of the star increases. As stars tend to go unstable when the radiation pressure gets comparable to the gas pressure, the result also limits the maximum mass of any star to less than about $100 M_\odot$ where $M_\odot$ is the solar mass. This is in accordance with observations.

White dwarfs

Next consider white dwarf stars. At the high central densities of these stars, the electron gas becomes degenerate. The electron pressure as given by (6.104) may then be the dominating contributor to the central pressure, and what is even more, the electron pressure may be calculated without the knowledge of the central temperature,

$$P_e = \frac{1}{20} \left(\frac{3}{\pi}\right)^{2/3} \frac{h^2}{m_e} \left(\frac{\rho_{mc}}{\mu_e m_{\mathrm{H}}}\right)^{5/3}.$$

The quantity μ_e is defined such that $\rho_{mc}/\mu_e m_{\mathrm{H}}$ represents the electron density n_e at the center of the star.

Expressing the central density ρ_{mc} in terms of the stellar mass and radius according to (7.42), the pressure balance requirement (7.44) now leads to the surprising result

$$R_* M_*^{1/3} = \frac{1}{20} \left(\frac{3}{\pi}\right)^{2/3} \frac{h^2}{4\pi G m_e} \frac{1}{(\mu_e m_{\mathrm{H}})^{5/3}} \frac{(4\pi\alpha)^{1/3}}{\beta}. \tag{7.45}$$

With the right hand side of (7.45) remaining approximately constant, it is seen that with increasing stellar mass M_*, the radius R_* of the compact star must decrease in order to maintain hydrostatic equilibrium conditions.

With increasing stellar mass the central density of the white dwarf will eventually be high enough that the electron gas approaches the relativistic degenerate state. In this limit the electron pressure increases more slowly with increasing density. From (6.105) the electron pressure approaches the asymptotic form

$$P_e = \frac{hc}{8} \left(\frac{3}{\pi}\right)^{1/3} \left(\frac{\rho_{mc}}{\mu_e m_{\mathrm{H}}}\right)^{4/3}.$$

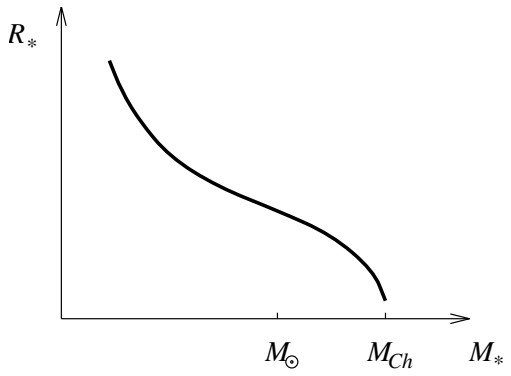

Figure 7.4: Mass - radius relation for white dwarfs

This means that the radius R_* of the star must decrease even faster with increasing mass M_* than expressed by (7.45). Eventually, no further equilibrium state can be found. This happens at a critical mass, the Chandrasekhar mass $M_{Ch} \approx 1.4 M_\odot$. The resulting stellar mass - radius relationship is illustrated schematically in figure 7.4.

Neutron stars

The collapse of the star will only stop when the stellar density reaches such extreme values that electrons are absorbed by the atomic nuclei, transforming protons into neutron with the emission of escaping neutrinos and dissolving atoms into a gas of free neutrons. The star has transformed into a neutron star. Like the free electrons, the free neutrons are fermions and must therefore satisfy the Pauli principle. At high enough densities also the neutron gas will be degenerate with a neutron pressure

$$P_n = \frac{1}{20}\left(\frac{3}{\pi}\right)^{2/3} \frac{h^2}{m_n}\left(\frac{\rho_{mc}}{m_{\rm H}}\right)^{5/3},$$

where m_n is the mass of the neutron. A new class of static equilibria may then be found. The gravitational field around the star is now strong enough to require a general relativity treatment. The main conclusions for the white dwarf case are, however, still valid. The radius R_* of the neutron star will decrease with increasing mass M_*. Eventually, the neutron gas transform into a relativistic degenerate neutron gas and a new critical mass, the Oppenheimer-Volkof mass, M_{OV} in the range 1.5 - 2.0 $M_\odot$, is reached beyond which no static equilibrium can be found. The neutron star then transforms into a black hole.

Quiz 7.16 : Verify (7.39)-(7.40) by making use of the differential relationship (7.8) between the gravitational potential Φ_g and the mass density ρ_m, the assumed spherical symmetry and the proper boundary condition for $d\Phi_g/dr$ at $r = 0$.

Quiz 7.17 : Determine the pressure in the Earth as a function of radius r assuming the interior as an incompressible fluid with constant density ρ_m in hydrostatic equilibrium. Would the pressure at the center increase or decrease if the mass density increased towards the center?

7.6 Stability of Static Equilibria

It is now time to return to the important question of the stability properties of the hydrostatic equilibrium solutions studied above. A simple mechanical analog will clarify the matter involved. With some care, a ball may be placed on top of a sphere in a constant gravitational field. This is an equilibrium state for the ball. However, only minor perturbations will make the ball fall off. With even the slightest displacement of the ball from the top point, the gravitational force will act to move the ball further away from its equilibrium position. An equilibrium state with this property is called unstable. Similar situations also exist in the description of fluids. In particular, under certain conditions this applies to the hydrostatic equilibrium studied above.

Consider the case of a fluid heated from below, like the air above a hot plate or the interior of a star outside the fusion burning core. To remove the generated heat by radiation transport or heat conduction, a decreasing temperature profile with height will result. In such an environment consider a fluid element at local thermal equilibrium with the surrounding fluid at some height z. Let the fluid element due to some disturbing action, be moved to another height $z + \delta z$ while expanding adiabatically in order to accommodate the changing surrounding pressure. This means that the mass density of the fluid element will change by an amount $\delta^{ad}\rho_m$. The process is illustrated in figure 7.5.

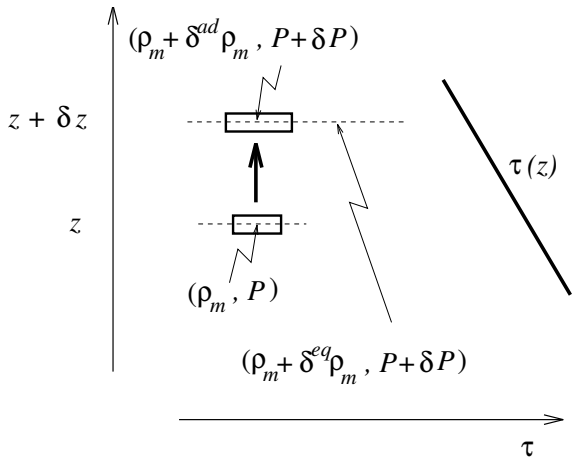

Figure 7.5: The thermal instability

The relative change in the mass density suffered by the fluid element is easily calculated from the adiabatic law (7.28),

$$\frac{\delta^{ad}\rho_m}{\rho_m} = \frac{\delta P}{\gamma P}. \tag{7.46}$$

The corresponding variation in the mass density $\delta^{eq}\rho_m$ of the equilibrium atmosphere in going from height z to $z + \delta z$ follows from the equation of state (7.26),

$$\frac{\delta^{eq}\rho_m}{\rho_m} = \frac{\delta\left(\mu\frac{P}{\mathcal{T}}\right)}{\left(\mu\frac{P}{\mathcal{T}}\right)} = \frac{\delta P}{P} - \frac{\delta\mathcal{T}}{\mathcal{T}}. \tag{7.47}$$

The latter result requires that the variation of the mean molecular weight μ with height z can be neglected. If $\delta z > 0$ and the density of the transported fluid element $\rho_m + \delta^{ad}\rho_m$ is less than the density of the surrounding fluid $\rho_m + \delta^{eq}\rho_m$, the fluid element will experience a buoyancy force that will be acting to drive it further away from its original location. The equilibrium state is then unstable.

The adiabatic process considered above represents the most critical case. Indeed, if the fluid element during its motion exchanges heat with its surroundings, the resulting difference in mass density will be less. As nature generally has the ability of seeking out the most unstable modes, the condition for the onset of the instability can therefore be formulated as

$$\frac{\delta P}{P} - \frac{\delta \mathcal{T}}{\mathcal{T}} = \frac{\delta^{eq}\rho_m}{\rho_m} > \frac{\delta^{ad}\rho_m}{\rho_m} = \frac{\delta P}{\gamma P}.$$

The instability is called the *thermal instability*. In differential form, and taking the barometric formula (7.37) into account, the criterion for the onset of the instability may be written

$$\frac{\mathrm{d}\mathcal{T}}{\mathrm{d}z} < \frac{\gamma - 1}{\gamma}\frac{\mathcal{T}}{P}\frac{\mathrm{d}P}{\mathrm{d}z} = -\frac{\gamma - 1}{\gamma}\mu m_{\mathrm{H}} g. \tag{7.48}$$

The results of our simple discussion may be verified through a more rigorous stability analysis. The instability gives rise to the development of a Bérnard convection cell pattern, leading to an effective convective transport of the excess heat from below. It is seen from (7.48) that the condition for instability is easier satisfied, the smaller the mean molecular weight of the fluid. The fluid that is most unstable to the thermal instability is thus the ionized hydrogen gas. This has important consequences for the internal structure of the Sun and other stars. Indeed, the thermal instability gives rise to the formation of convection zones in most stars.

Quiz 7.18 : Derive (7.46) and (7.47) from (7.28) and (7.26).

Quiz 7.19 : From the tabulated values of radius, mass and central temperature, make an order of magnitude estimate to see whether the thermal instability may be acting inside the Sun.

7.7 Fluid Flows

Consider the flow velocity $\boldsymbol{v}(\boldsymbol{r}, t)$ of the fluid to be given as a function of space and time. In section 7.1 it was demonstrated that the divergence of the flow velocity $\nabla \cdot \boldsymbol{v}$ can be interpreted geometrically as the local relative expansion of the fluid per unit time. A similar geometric interpretation can be given to the curl or the *vorticity* $\boldsymbol{\omega} \equiv \nabla \times \boldsymbol{v}$ of the flow field. Consider any *simply connected* curve C in the fluid. A closed curve C is simply connected if it can be shrunk to a point without engaging any physical obstacle or singularity. The *circulation* Γ along the curve C is defined as

$$\Gamma \equiv \oint_C \mathrm{d}\boldsymbol{\ell} \cdot \boldsymbol{v} = \int_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}} \cdot \boldsymbol{\omega}. \tag{7.49}$$

The right hand side expression follows from Stokes integral theorem (A.41). The circulation is seen to indicate the tendency of the flow field to follow the direction of the curve C, that is,

the tendency of the flow to "curl back" on itself. The vorticity $\boldsymbol{\omega}$ is consequently interpreted as this "curling back" tendency of the flow field per unit area. For a simply connected region for which the flow velocity vanishes at the boundary, knowledge of the divergence and the vorticity of the flow field uniquely determines the flow velocity $\boldsymbol{v}$ itself according to the Helmholtz formula (A.56).

A given fluid particle will as a function of time follow a certain trajectory in space. This trajectory is called a *path-line*. In a laboratory experiment, path-lines can be made visible by adding colored particles to the flow – assuming that the added particles are able to follow in the motion of the neighboring fluid particles. A *streamline*, on the other hand, is a curve that is everywhere parallel to the flow velocity $\boldsymbol{v}$ at a particular time t. For stationary flows, flows that at any given point do not change with time, streamlines and path-lines coincide. For unsteady flows this does not hold. Streamlines evolve with time and the two types of lines generally differ. It will also be convenient to introduce the definition of a *vortex line*. A vortex line is a curve that is everywhere parallel to the vorticity $\boldsymbol{\omega}$ of the fluid at a given time t.

The evolution of the flow velocity in the fluid is determined by the momentum equation (7.10). For a discussion of the physical content of this equation it is often convenient to make use of a slightly rearranged equation. Thus, dividing through by ρ_m and making use of the identity

$$\boldsymbol{v} \cdot \nabla \boldsymbol{v} = \nabla \frac{\boldsymbol{v}^2}{2} - \boldsymbol{v} \times (\nabla \times \boldsymbol{v}), \tag{7.50}$$

the momentum equation may be rewritten in the form

$$\frac{\partial \boldsymbol{v}}{\partial t} - \boldsymbol{v} \times \nabla \times \boldsymbol{v} = -\frac{\nabla P}{\rho_m} - \nabla \left(\frac{\boldsymbol{v}^2}{2} + \Phi_g \right). \tag{7.51}$$

Before turning to a discussion of some consequences of the flow equations, it is convenient to introduce yet another definition. A fluid for which $\nabla \times (\nabla P/\rho_m)$ vanishes identically, will be called *barytropic*. This means that $\nabla P/\rho_m$ can be written as the gradient of a scalar function h,

$$\frac{\nabla P}{\rho_m} = \nabla h. \tag{7.52}$$

A fluid for which this condition is *not* satisfied everywhere, is called *baryclinic*. A necessary condition for the fluid to be barytropic is that constant pressure and constant density surfaces coincide everywhere in the fluid and therefore that the pressure can be expressed as a function of density alone, $P = P(\rho_m)$. This also means that the scalar function h may be written explicitly as

$$h = \int_0^P \frac{\mathrm{d}P}{\rho_m}. \tag{7.53}$$

In particular, for an adiabatic flow we have $P \sim \rho_m^\gamma$. This means that the adiabatic flow is barytropic and we find

$$h = \frac{\gamma}{\gamma - 1} \frac{P}{\rho_m}. \tag{7.54}$$

Once the flow velocity $\boldsymbol{v}$ is known the vorticity $\boldsymbol{\omega}$ can be calculated. In the absence of analytic solutions for $\boldsymbol{v}$ it is useful to establish a separate equation for the evolution of the

vorticity. Such an equation is easily found by applying the curl operator to each term of the momentum equation in the form (7.51). The result is

$$\frac{\partial \boldsymbol{\omega}}{\partial t} - \nabla \times (\boldsymbol{v} \times \boldsymbol{\omega}) = -\nabla \times \left(\frac{\nabla P}{\rho_m} \right). \tag{7.55}$$

From (7.55) it may now be concluded that if in a barytropic flow the vorticity vanishes identically initially, then the vorticity will continue to be identically zero in the flow. This conclusion explains why fluid flows are traditionally divided into two separate classes. The flow is classified as *potential (irrotational) flow* if the vorticity $\boldsymbol{\omega}$ vanishes everywhere in the fluid. In this case the velocity field can be derived from a scalar *velocity potential* Ψ, $\boldsymbol{v} = \nabla\Psi$. If on the other hand, the vorticity at any point and time is non-vanishing, the flow is classified as a *vortex flow*. The former case leads to considerable simplifications in the theory, but the latter case is often met with in reality.

Quiz 7.20 : In a stationary flow the fluid velocity is given by

$$\boldsymbol{v} = \begin{cases} \boldsymbol{\Omega} \times \boldsymbol{r} & \text{for} \quad r_\perp < a \\ a^2 \boldsymbol{\Omega} \times \boldsymbol{r}/r_\perp^2 & \text{else,} \end{cases}$$

where $\boldsymbol{\Omega}$ is constant and $r_\perp = |\boldsymbol{\Omega} \times \boldsymbol{r}| \, /\Omega$. Find the vorticity $\boldsymbol{\omega}$ for the flow and draw some examples of streamlines and vortex lines.

Quiz 7.21 : For the stationary flow

$$\boldsymbol{v} = a^2 \boldsymbol{\Omega} \times \boldsymbol{r}/r_\perp^2$$

where $\boldsymbol{\Omega}$ is constant and $r_\perp = |\boldsymbol{\Omega} \times \boldsymbol{r}| \, /\Omega$, calculate the circulation Γ along a circle centered at the origin and lying in the plane perpendicular to $\boldsymbol{\Omega}$. How does the result fit with the value of the vorticity $\boldsymbol{\omega}$? Explain!

Quiz 7.22 : Show that the function h defined in (7.53) for the case of an isothermal flow takes the form

$$h = \frac{\mathcal{T}}{m} \ln \frac{\rho_m}{\rho_{m0}}$$

where ρ_{m0} is an arbitrary reference density. What is the corresponding result for an incompressible flow?

7.8 The Bernoulli Theorem

For *stationary flows* the time derivative term in (7.51) vanishes. If the flow is barytropic, we see that

$$\boldsymbol{v} \cdot \nabla \left(\frac{\boldsymbol{v}^2}{2} + h + \Phi_g \right) = 0.$$

For stationary flows we remember that a streamline is also the trajectory that a fluid element will follow during its motion. The last result may therefore be formulated in the following

way:
For a stationary, barytropic flow in an ideal fluid

$$\frac{\boldsymbol{v}^2}{2} + h + \Phi_g = C, \tag{7.56}$$

where C is constant along any given streamline. If the flow is also irrotational, C will be constant for the whole fluid. This result is known as the *Bernoulli theorem.*

Compared to the hydrostatic equilibrium case as represented by the barometric formula (7.36), Bernoulli's theorem prescribes a completely different relationship between the pressure and the gravitational field. The pressure, entering through the the function h, now also depends on the flow velocity $\boldsymbol{v}$. High flow velocities are associated with decreased values of the pressure. What this means is that the fluid by increasing its flow velocity is transforming some of its thermal energy in the form of random thermal motion into a more systematic and directed flow pattern associated with the flow velocity $\boldsymbol{v}$.

7.8.1 Stellar winds

As an application of Bernoulli's theorem, let us consider the problem of the stationary expansion of hot coronal gas into interstellar space. We assume the expansion to be spherically symmetric and therefore $\boldsymbol{v} = v_r\,\hat{\boldsymbol{r}}$. This means that the flow is irrotational and that $\boldsymbol{v}^2/2 + h + \Phi_g$ will take the same value everywhere. The gravitational potential for a star with mass M_* is $\Phi_g = -GM_*/r$. In differential form Bernoulli's theorem (7.56) may be written

$$v_r\,\mathrm{d}v_r + C_s^2\frac{\mathrm{d}\rho_m}{\rho_m} + GM_*\frac{\mathrm{d}r}{r^2} = 0, \tag{7.57}$$

where we introduced $C_s^2 \equiv \mathrm{d}P/d\rho_m$. C_s will later be shown to represent the *sound speed* in the fluid. The equation of continuity (7.2) for a stationary and spherically symmetric flow requires $\rho_m v_r r^2$ to remain constant, or

$$\frac{\mathrm{d}\rho_m}{\rho_m} + \frac{\mathrm{d}v_r}{v_r} + \frac{2\,\mathrm{d}r}{r} = 0. \tag{7.58}$$

Elimination of $d\rho_m/\rho_m$ between (7.57) and (7.58) leads to

$$(\mathcal{M}^2 - 1)\frac{\mathrm{d}v_r}{v_r} = (1 - \frac{1}{\mathcal{R}})\frac{2\,\mathrm{d}r}{r}, \tag{7.59}$$

where $\mathcal{M} \equiv v_r/C_s$ is the *Mach number* and $\mathcal{R} = 2C_s^2 r/GM_*$ a normalized radius. The sound speed C_s appearing in the definitions of $\mathcal{M}$ and $\mathcal{R}$, is itself generally a function of r.

The qualitative properties of (7.59) can be understood by considering the normalized $\mathcal{R}$-$\mathcal{M}$ diagram of figure 7.6. Four regions may be identified. For $\mathcal{R} < 1$ and $\mathcal{M} < 1$ and for $\mathcal{R} > 1$ and $\mathcal{M} > 1$ it is seen from (7.59) that v_r increases with increasing r. The opposite result applies for $\mathcal{R} < 1$ and $\mathcal{M} > 1$ and for $\mathcal{R} > 1$ and $\mathcal{M} < 1$. These conclusions are illustrated in figure 7.6 by the direction of the arrows. We see that $\mathrm{d}v_r/\mathrm{d}r$ vanishes at $\mathcal{R} = 1$ while it tends to infinity as $\mathcal{M} \to 1$. The point $(\mathcal{R} = 1, \mathcal{M} = 1)$ is singular in the respect that the coefficients of both $\mathrm{d}v_r$ and $\mathrm{d}r$ in (7.59) vanish here.

For the hot stellar corona to exist, it is necessary that $v_r \to 0$ for small r. At the same time a stellar wind type solution will require that v_r increases or at least remains finite for

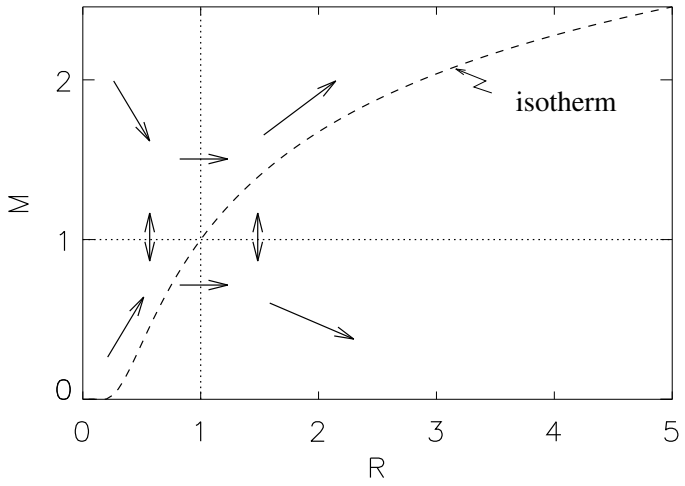

Figure 7.6: Stellar wind R-M plane, rising (falling) arrows indicates $\mathrm{d}v_r/\mathrm{d}r > (<) 0$

large r. The latter condition means that $v_r^2 r$ and therefore also $\mathcal{M}^2\mathcal{R}$ tend to infinity with increasing r, irrespective of the particular functional form of C_s. The only possible path for a solution starting in the $(\mathcal{R} < 1, \mathcal{M} < 1)$ region and ending in the $(\mathcal{R} > 1, \mathcal{M} > 1)$ region, however, goes through the critical point.

For an isothermal flow C_s is a constant, $C_s^2 = \mathcal{T}/\mu m_{\mathrm{H}}$ where μ is the mean molecular weight of the particles in the gas. For this case (7.59) may be integrated to give

$$\frac{1}{2}\mathcal{M}^2 - \ln \mathcal{M} = \frac{2}{\mathcal{R}} + \ln \mathcal{R}^2 - \frac{3}{2}. \tag{7.60}$$

The solution is plotted in figure 7.6. The flow makes a *subsonic* to *supersonic* transition at the critical point, increasing asymptotically as $\mathcal{M} \sim 2(\ln \mathcal{R})^{1/2}$ for large $\mathcal{R}$.

The isothermal flow assumption is an extreme one, requiring in fact an infinite capacity heat regulator, for example through heat conduction, able to maintain constant temperature in the flow. Let us therefore proceed to the opposite case and consider the possibility of an adiabatic wind solution. For this case C_s decreases with increasing r. This follows from the fact that $P \sim \rho_m^\gamma$ and therefore $C_s^2 = \mathrm{d}P/\mathrm{d}\rho_m \sim \rho_m^{\gamma-1} \sim (v_r r^2)^{1-\gamma}$. For the last step we made use of the equation of continuity (7.2). This means that $\mathcal{R} \sim C_s^2 r \sim r^{3-2\gamma} v_r^{1-\gamma}$, and therefore that $\mathcal{R}$ *decreases* as r increases if $\gamma > 3/2$ and v_r is non-decreasing. This is in contradiction with boundary conditions for a wind solution. For a plasma $\gamma > 3/2$. A solution with $\mathrm{d}v_r/\mathrm{d}r > 0$ and satisfying $v_r \to 0$ as $r \to R_*$ would have to remain within the $\mathcal{R} < 1, \mathcal{M} < 1$ domain in the $\mathcal{R}$-$\mathcal{M}$ plane. This solution would thus not be able to satisfy the boundary condition $\mathcal{M}^2\mathcal{R} \to \infty$ as $r \to \infty$. We must conclude that an adiabatic equation of state is not reconcilable with the existence of a stellar wind. To drive a wind, energy must be supplied to the outer corona by some irreversible process. This could, in accordance with the discussion of section 7.4, be heat conduction, viscous dissipation, or more likely some kind of wave dissipation. The latter possibility will be treated in section 7.16.

Quiz 7.23 : Consider an isothermal stellar wind. Argue that the mass loss rate of the

star can be expressed as

$$\frac{\mathrm{d}M}{\mathrm{d}t} = 4\pi r_c^2 \rho_{mc} v_{rc}$$

where subscript c indicates quantities evaluated at the critical point. Write down the expressions for r_c and v_{rc}. Make use of Bernoulli's theorem (7.56) and the boundary conditions $v_r \to 0$ and $\rho_m \to \rho_{m*}$ as $r \to R_*$ to show that

$$\rho_{mc} = \rho_{m*} \exp\left(-\frac{GM_*\mu m_{\mathrm{H}}}{\mathcal{T} R_*}\left(1 - \frac{R_*}{r_c}\right) - \frac{1}{2}\right). \tag{7.61}$$

Can you estimate a characteristic lifetime of the Sun due to mass loss in the wind? By comparing (7.61) with (7.38) argue that the isothermal wind corona is "almost" static out to the critical point. How does the limiting density in the isothermal wind solution at large distances compare with the isothermal static corona model?

7.9 The Kelvin Circulation Theorem

Let us now return to the more general flow situation for which the vorticity $\boldsymbol{\omega}$ does not vanish identically. The left hand side operator of the vorticity equation (7.55) allows for a simple geometric interpretation. Consider an arbitrary simply connected *material* curve $C(t)$ in the fluid at time t, bounding the open surface $A(t)$. A material curve flows and deforms with the fluid and may be thought of as being made up of the same fluid particles at all times. The following mathematical identity is valid

$$\begin{aligned}\int_{A(t)} \mathrm{d}^2\boldsymbol{\mathcal{A}} \cdot \left[\frac{\partial \boldsymbol{\omega}}{\partial t} - \nabla \times (\boldsymbol{v} \times \boldsymbol{\omega})\right] &= \int_{A(t)} \mathrm{d}^2\boldsymbol{\mathcal{A}} \cdot \frac{\partial \boldsymbol{\omega}}{\partial t} + \oint_{C(t)} \boldsymbol{\omega} \cdot (\boldsymbol{v} \times \mathrm{d}\boldsymbol{\ell}) \\ &= \frac{\mathrm{d}}{\mathrm{d}t} \int_{A(t)} \mathrm{d}^2\boldsymbol{\mathcal{A}} \cdot \boldsymbol{\omega}. \end{aligned} \tag{7.62}$$

The second term of the middle expression results from the application of Stoke's integral theorem (A.41). As may be seen from figure 7.7, the factor $(\boldsymbol{v}\,\mathrm{d}t \times \mathrm{d}\boldsymbol{\ell})/\mathrm{d}t$ equals the local rate of change of the integration area. The two terms of the middle expression therefore represent the total time derivative of the surface integral of the vorticity $\boldsymbol{\omega}$ over a time dependent integration domain, as stated by the right hand side expression. For the subsequent discussion of magnetic fields in electrically conducting fluids in section 7.14 it is important to note that nowhere in this derivation did we make use of the particular relationship between the fluid velocity $\boldsymbol{v}$ and the vorticity $\boldsymbol{\omega}$.

In terms of the circulation Γ defined in (7.49) and the identity (7.62), the vorticity equation (7.55) may be rewritten as

$$\frac{\mathrm{d}}{\mathrm{d}t}\Gamma = -\oint_C \mathrm{d}\boldsymbol{\ell} \cdot \frac{\nabla P}{\rho_m}. \tag{7.63}$$

For a barytropic fluid the right hand side of (7.55) vanishes identically, as does the line integral in (7.63). This result forms the basis for the *Kelvin circulation theorem*:
If along the simply connected closed material curve C the fluid is ideal and barytropic, the circulation Γ around C remains constant,

$$\frac{d}{dt}\Gamma = 0. \tag{7.64}$$

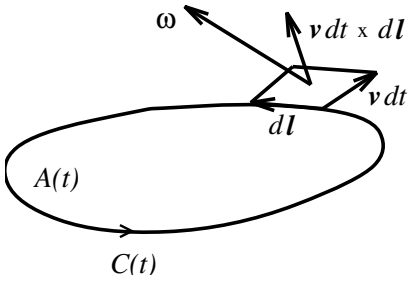

Figure 7.7: Surface element traced by material curve

Several important conclusions follow from Kelvin's theorem. For an ideal, barytropic fluid the conditions of the theorem will be satisfied for *any* simply connected material curve. This means that:
If the flow field in an ideal, barytropic fluid is irrotational initially, it will remain so.

Now, consider an arbitrary *vortex tube* in a vortex flow. A vortex tube is a cylindrical volume whose side surface is made up entirely from vortex lines. Let C be any simply connected closed material curve bounding a surface A lying on the vortex tube at time t. The geometry is illustrated in figure 7.8a. By definition, $\boldsymbol{\omega} \cdot \mathrm{d}\boldsymbol{\mathcal{A}}$ vanishes over A at time t. But from Kelvin's theorem for a ideal, barytropic fluid, the circulation around the curve C remains constant and therefore vanishes for all times. This means that the material curve C will continue to lie on a vortex tube and therefore also that the vortex tube is itself convected by the flow. Going to the limit of an infinitely thin vortex tube, this is equivalent to:
Any vortex line in an ideal, barytropic fluid is convected by the flow.

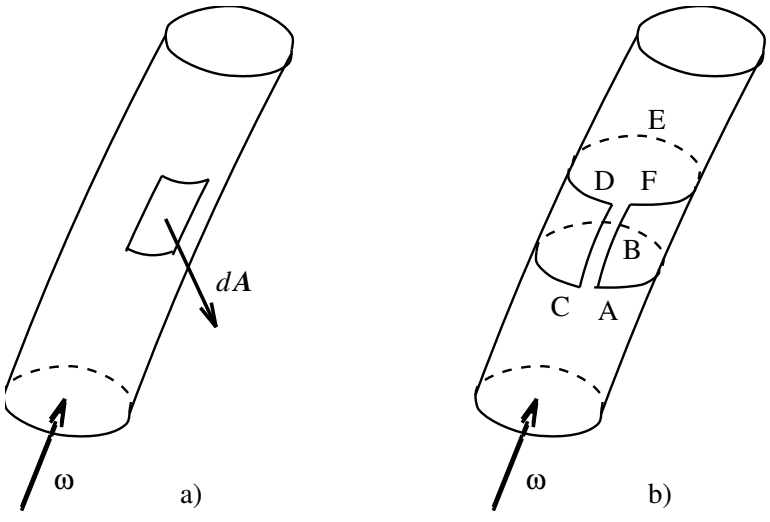

Figure 7.8: Vortex tubes with material curves

Next, we return to the vortex tube of finite thickness containing the particular material curve $ABCDEF$ of figure 7.8b. Let the sides AF and CD coincide. Our material curve is

then equivalent to the sum of two material curves ABC and DEF encircling the vortex tube at different positions along the tube. From Kelvin's theorem, it follows that the circulation around the two material curves ABC and FED must remain identical:
In an ideal, barytropic fluid the circulation around any vortex tube is constant along the tube.

The latter conclusion means that a vortex tube is not allowed to start or end inside the fluid. Any vortex tube in an ideal, barytropic fluid either ends at a boundary or at a discontinuity in the fluid, or closes on itself.

Quiz 7.24 : By expanding the second term on the left hand side of (7.55) and making use of the continuity equation (7.3) show that an alternative form of the vorticity equation is

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(\frac{\boldsymbol{\omega}}{\rho_m}\right) = \frac{\boldsymbol{\omega}}{\rho_m}\cdot\nabla\boldsymbol{v} - \frac{1}{\rho_m}\nabla\times\left(\frac{\nabla P}{\rho_m}\right). \tag{7.65}$$

Quiz 7.25 : For any scalar field f prove that

$$\nabla\frac{\mathrm{D}}{\mathrm{D}t}f = \frac{\mathrm{D}}{\mathrm{D}t}\nabla f + \nabla\boldsymbol{v}\cdot\nabla f.$$

Making use of this identity, show through a scalar multiplication of (7.65) with the gradient of any scalar quantity f conserved by the flow,

$$\frac{\mathrm{D}f}{\mathrm{D}t} = 0,$$

that the vorticity equation (7.55) can be written in the alternative form

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(\frac{\boldsymbol{\omega}}{\rho_m}\cdot\nabla f\right) = -\frac{1}{\rho_m}\nabla\times\left(\frac{\nabla P}{\rho_m}\right)\cdot\nabla f. \tag{7.66}$$

The quantity in the left hand side parenthesis is traditionally referred to as "potential vorticity".

7.10 Rotating Coordinate Frames

The derivation of the equations of motions in the previous paragraphs assumed that all relevant quantities were measured relative to an inertial frame. In a geophysical or astrophysical context it is often convenient to make use of rotating frames of reference, for example rotating with a celestial body as illustrated in figure 7.9a. The natural phenomena are of course unaffected by our particular choice of reference frame. We shall, however, see that our description of the phenomena will be affected by the choice. To an observer in a rotating frame, an object fixed in an inertial frame will appear to rotate and be accelerating. To this observer, the object thus seems to be subject to forces not present in the inertial frame. We therefore next identify the modified equations of motion when written entirely in terms of quantities directly observed from the rotating frame.

Consider a frame rotating with angular velocity $\boldsymbol{\Omega}$ with respect to an inertial frame. Let $\boldsymbol{V}$ be an arbitrary vector. Observers in the two frames of reference will agree on the

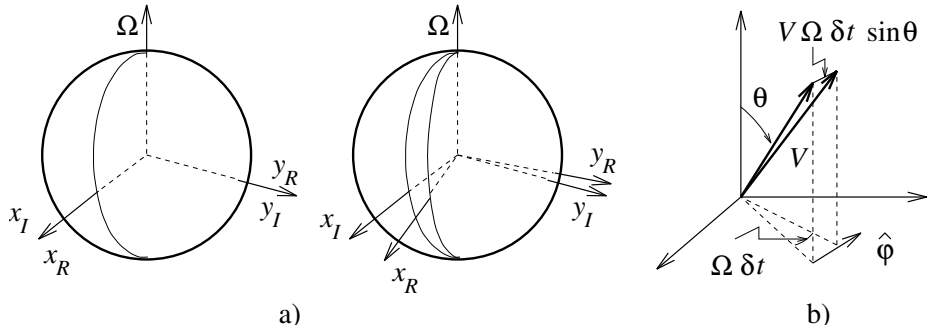

Figure 7.9: Geometry of rotating frames

instantaneous magnitude and direction of $\boldsymbol{V}$. If $\boldsymbol{V}$ is fixed with respect to the rotating frame, the observer in the rotating frame will report no rate of change of $\boldsymbol{V}$. To the observer in the inertial frame, however, $\boldsymbol{V}$ will change by an amount

$$\delta \boldsymbol{V}_I = V\Omega \sin\theta\, \hat{\boldsymbol{\varphi}}\, \delta t = \boldsymbol{\Omega} \times \boldsymbol{V}\, \delta t$$

during the time interval δt. The geometry, as seen from the inertial frame, is illustrated in figure 7.9b. The result is independent of the position of the vector $\boldsymbol{V}$ relative to the axis of rotation. If $\boldsymbol{V}$ is allowed to change with time also in the rotating frame, changing with an amount $\delta \boldsymbol{V}_R = \mathrm{d}\boldsymbol{V}/dt\,|_R\, \delta t$ during the same time interval δt, the observer in the inertial frame will report $\delta \boldsymbol{V}_R$ as an additional change in $\boldsymbol{\mathcal{A}}$. We did here neglect terms of order $(\delta t)^2$. The total rate of change of $\boldsymbol{V}$ in the inertial frame is therefore

$$\frac{\mathrm{d}\boldsymbol{V}}{\mathrm{d}t}\,|_I = \frac{\mathrm{d}\boldsymbol{V}}{\mathrm{d}t}\,|_R + \boldsymbol{\Omega} \times \boldsymbol{V}. \tag{7.67}$$

The notations $|_I$ and $|_R$ refer to the the reference frames in which the evaluation is performed.

With $\boldsymbol{V}$ representing the changing position vector $\boldsymbol{r}$ of a given fluid element, the transformation law (7.67) leads to the result

$$\boldsymbol{v}_I = \boldsymbol{v}_R + \boldsymbol{\Omega} \times \boldsymbol{r}, \tag{7.68}$$

where $\boldsymbol{v}_I$ and $\boldsymbol{v}_R$ are the velocities of the given fluid element as seen from the two frames. Application of the transformation law (7.67) to the vector $\boldsymbol{V} = \boldsymbol{v}_I$ in turn leads to

$$\begin{aligned} \frac{\mathrm{d}\boldsymbol{v}_I}{\mathrm{d}t}\,|_I &= \frac{\mathrm{d}}{\mathrm{d}t}\,(\boldsymbol{v}_R + \boldsymbol{\Omega} \times \boldsymbol{r})\,|_R + \boldsymbol{\Omega} \times (\boldsymbol{v}_R + \boldsymbol{\Omega} \times \boldsymbol{r}) \\ &= \frac{\mathrm{d}\boldsymbol{v}_R}{\mathrm{d}t}\,|_R + 2\boldsymbol{\Omega} \times \boldsymbol{v}_R + \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \boldsymbol{r})\,. \end{aligned} \tag{7.69}$$

Here we assumed $\boldsymbol{\Omega}$ to be constant. We identify (7.69) as the relationship connecting accelerations in the two frames. It is important to note that the three right hand side terms of (7.69) are expressed entirely in terms of quantities relating to the rotating frame. The first term is the acceleration as measured in the rotating frame. The second term is the *Coriolis acceleration*, directed perpendicular to both $\boldsymbol{\Omega}$ and $\boldsymbol{v}_R$. The third term represents the *centripetal acceleration*, directed radially inward towards the axis of rotation.

The two time derivative terms in (7.69) represent the total rate of change of the velocity vector $\boldsymbol{v}(t)$ of a given fluid element, as seen from the two frames. If we replace $\boldsymbol{v}(t)$ with the

velocity fields $\boldsymbol{v}(\boldsymbol{r},t)$ as measured in the two frames, the time derivative operators have to be replaced by the corresponding convective derivatives $\mathrm{D}\boldsymbol{v}(\boldsymbol{r},t)/\mathrm{D}t$. This means that the relation (7.69), when applied to the velocity fields $\boldsymbol{v}(\boldsymbol{r},t)$, takes the form

$$\frac{\mathrm{D}\boldsymbol{v}(\boldsymbol{r},t)}{\mathrm{D}t}\,|_{\,I} = \left(\frac{\mathrm{D}\boldsymbol{v}(\boldsymbol{r},t)}{\mathrm{D}t} + 2\boldsymbol{\Omega}\times\boldsymbol{v}(\boldsymbol{r},t) + \boldsymbol{\Omega}\times(\boldsymbol{\Omega}\times\boldsymbol{r})\right)|_{\,R}. \tag{7.70}$$

We notice that while $\boldsymbol{r}=\boldsymbol{r}(t)$ in (7.69) represents the time varying position of a given fluid element, $\boldsymbol{r}$ in (7.70) is a general position vector, independent of time t. The origins for both types of position vectors are located on the rotational axis. It is useful to note that the centripetal acceleration term in (7.70) can be expressed as the gradient of a scalar *centripetal potential* Φ_c,

$$\boldsymbol{\Omega}\times(\boldsymbol{\Omega}\times\boldsymbol{r}) = \nabla\Phi_c \quad \text{with} \quad \Phi_c = -\frac{1}{2}(\boldsymbol{\Omega}\times\boldsymbol{r})^2. \tag{7.71}$$

Scalar fields like the mass density ρ_m, the pressure P or the gravitational potential Φ_g at any given time instance take identical values in both frames, their partial derivatives with respect to space coordinates will also be identical, for example $\nabla\rho_m\,|_{\,I} = \nabla\rho_m\,|_{\,R}$. With the help of (7.70), the momentum equation (7.10) in the rotating frame can therefore be written

$$\rho_m\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = -\nabla P - \rho_m\nabla\Phi + 2\rho_m\boldsymbol{v}\times\boldsymbol{\Omega}. \tag{7.72}$$

Here the potential Φ, in addition to the gravitational potential Φ_g, also includes the centripetal potential Φ_c,

$$\Phi = \Phi_g + \Phi_c. \tag{7.73}$$

By moving the Coriolis acceleration term of (7.70) to the right hand side of the momentum equation (7.72), it appears to the observer in the rotating frame as an additional force acting per unit volume, the *Coriolis force*

$$\boldsymbol{\mathcal{F}}_c = 2\rho_m\boldsymbol{v}\times\boldsymbol{\Omega}. \tag{7.74}$$

This force is always perpendicular to the velocity and therefore does no work on the fluid. For an observer aligned with $\boldsymbol{\Omega}$, for instance for an observer located in the Northern hemisphere of a rotating body, the Coriolis force will tend to deflect moving fluid elements to the right. In the Southern hemisphere the deflection will be opposite.

We have already claimed that any scalar field $s(\boldsymbol{r},t)$ will take identical instantaneous values in the two frames of reference. However, while the spatial derivatives of $s(\boldsymbol{r},t)$ also take identical values in the two frames, this does not apply to the time derivatives $\partial s(\boldsymbol{r},t)/\partial t$. The notation $\partial s(\boldsymbol{r},t)/\partial t$ means that the position vector $\boldsymbol{r}$ should be kept constant during the time derivation. But keeping $\boldsymbol{r}$ constant means different things in the two frames. A point fixed in the rotating frame will be moving in the inertial frame. This leads to the relation

$$\frac{\partial s}{\partial t}\,|_{\,I} = \frac{\partial s}{\partial t}\,|_{\,R} - (\boldsymbol{\Omega}\times\boldsymbol{r})\cdot\nabla s \tag{7.75}$$

and therefore

$$\frac{\mathrm{D}s}{\mathrm{D}t}\,|_{\,I} = \frac{\mathrm{D}s}{\mathrm{D}t}\,|_{\,R}. \tag{7.76}$$

If use is also made of the fact that

$$\nabla\cdot\boldsymbol{v}\,|_{\,I} = \nabla\cdot\boldsymbol{v}\,|_{\,R}, \tag{7.77}$$

it follows that the continuity equation (7.3) and, for instance, the adiabatic law (7.28) both transform without change of form in going from the inertial frame to the rotating frame.

Quiz 7.26 : Verify (7.71).

Quiz 7.27 : Make use of the relation (7.75) to prove (7.76).

Quiz 7.28 : Verify (7.77).

Quiz 7.29 : Show that in the presence of the Coriolis force, the vorticity equation (7.55) may be rewritten in the form

$$\frac{\partial \tilde{\boldsymbol{\omega}}}{\partial t} - \nabla \times (\boldsymbol{v} \times \tilde{\boldsymbol{\omega}}) = -\nabla \times \left(\frac{\nabla P}{\rho_m} \right)$$

where $\tilde{\boldsymbol{\omega}} = \boldsymbol{\omega} + 2\boldsymbol{\Omega}$.

Quiz 7.30 : In the presence of the Coriolis force, show that the rate of change of the circulation Γ around a simply connected closed material curve C (7.63) must be generalized to

$$\frac{\mathrm{d}\Gamma}{\mathrm{d}t} = -\oint_C \mathrm{d}\boldsymbol{\ell} \cdot \frac{\nabla P}{\rho_m} - 2\boldsymbol{\Omega} \cdot \oint_C \boldsymbol{v} \times \mathrm{d}\boldsymbol{\ell}.$$

Further show that the last integral in this result represents the rate of change of the vector area $\boldsymbol{\mathcal{A}}(t)$ enclosed by the moving curve C,

$$\oint_C \boldsymbol{v} \times \mathrm{d}\boldsymbol{\ell} = \frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{\mathcal{A}}(t).$$

Quiz 7.31 : At the North pole of a planet with an ideal atmosphere, a high pressure region is formed. Show that a stationary, strictly horizontal wind pattern, with streamlines forming circles centered at the pole and the wind blowing in the clockwise direction, can be maintained. What is the corresponding flow pattern if the high pressure region is replaced with a low pressure region? What are the modification to these solutions if the high or low pressure regions are maintained at other latitudes?

7.11 Magneto-hydrodynamics

In our discussion of fluid mechanics, we have so far assumed the fluid to be electrically neutral. That is, we did not allow the fluid to become electrically charged, or to be able to carry electric currents. In astrophysics we often find fluids for which these assumptions are not met. This is the case for the liquid cores of planets and for the ionized gases of stellar interiors, stellar and planetary atmospheres and interstellar space. We therefore next turn to the question of how the macroscopic fluid model will have to be modified in order to include effects due to the interaction between conducting fluids and electric and magnetic fields.

The additional force acting per unit volume on a fluid having charge density ρ_e and carrying a current density $\boldsymbol{j}$ at a location where the electric and magnetic fields are $\boldsymbol{E}$ and $\boldsymbol{B}$, is given by the *Lorentz force*

$$\boldsymbol{\mathcal{F}}_{em} = \rho_e \boldsymbol{E} + \boldsymbol{j} \times \boldsymbol{B}. \tag{7.78}$$

The electric charge and current densities are related through the equation of charge continuity

$$\frac{\partial \rho_e}{\partial t} + \nabla \cdot \boldsymbol{j} = 0. \tag{7.79}$$

But ρ_e and $\boldsymbol{j}$ also act as sources for the electric and magnetic fields through Maxwell's equations

$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \tag{7.80}$$

$$\nabla \times \boldsymbol{B} = \mu_0 \left(\boldsymbol{j} + \epsilon_0 \frac{\partial \boldsymbol{E}}{\partial t} \right) \tag{7.81}$$

$$\nabla \cdot \boldsymbol{B} = 0 \tag{7.82}$$

$$\nabla \cdot \boldsymbol{E} = \frac{\rho_e}{\epsilon_0} \tag{7.83}$$

The electric current density $\boldsymbol{j}$ in the moving frame is often determined by the *electric conductivity* σ of the fluid and the effective electric field seen by an observer moving with the fluid,

$$\boldsymbol{j} = \sigma(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}). \tag{7.84}$$

We shall refer to this relation as the *Ohm law*.

With the addition of Maxwell's equations, we are left with a more complicated set of equations for the conducting fluid. Instead of dealing with five hydrodynamic variables ρ_m, $\boldsymbol{v}$ and P, we are now also facing a larger number of electromagnetic variables. Important simplifications are, however, possible under certain conditions. First of all, we will demonstrate that the first term of the Lorentz force (7.78) may be neglected in most situations of interest.

To this end consider the simple case of a homogeneous fluid at rest. When substituting (7.84) and (7.83) into (7.79), we find

$$\frac{\partial \rho_e}{\partial t} = -\sigma \nabla \cdot \boldsymbol{E} = -\frac{\sigma}{\epsilon_0} \rho_e.$$

This result shows that the charge density in the fluid will decay exponentially with time, with a typical decay time $\tau_\rho = \epsilon_0/\sigma$. For good conductors this decay time is very short. Values of the decay time for some relevant examples are listed in table 7.1.

Conducting fluids thus have the ability to react very quickly to any charge unbalance that may arise. In most situations of practical interest we may therefore safely assume the conducting fluid to maintain local charge neutrality to a high degree of accuracy at all times. This property of the conducting fluid is referred to as *quasi-neutrality*. This is also the

System	Conductivity σ [mho m^{-1}]	Decay time τ_ρ [s]
Earth's core	10^5	10^{-16}
Ionosphere	10	10^{-12}
Solar atmosphere	10^3	10^{-14}
Solar corona	10^6	10^{-17}

Table 7.1: Typical conductivities and charge decay times

argument needed for neglecting the electric field term of the Lorentz force. We shall later discuss the decay of electric current density in conducting fluids. Contrary to the case of charge density, it will be found that the decay times for electric currents in astrophysically interesting situations can be very long. There is thus no similar argument to neglect also the magnetic field part of the Lorentz force.

With the assumption of quasi-neutrality and the neglect of the charge density in the fluid, we shall at the same time have to disregard the equation of charge continuity (7.79) and Gauss law (7.83). Another approximation will also act to simplify our conducting fluid model. The $\epsilon_0 \partial \boldsymbol{E}/\partial t$ term of Maxwell's law (7.81) is called the *displacement current*. This term is necessary for any discussion of electromagnetic wave phenomena. At low enough frequencies, however, the displacement current can be neglected compared to the conduction current $\boldsymbol{j}$. When discussing typical fluid behavior, this approximation is usually well satisfied. Thus, for the conducting fluid model we shall replace Maxwell's law (7.81) by the corresponding Ampére's law

$$\nabla \times \boldsymbol{B} = \mu_0 \boldsymbol{j}. \tag{7.85}$$

With these simplifying assumptions, the equations of motion for a conducting fluid and its interaction with electric and magnetic fields are reduced to

$$\frac{\mathrm{D}\rho_m}{\mathrm{D}t} = -\rho_m \nabla \cdot \boldsymbol{v} \tag{7.86}$$

$$\rho_m \frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = -\nabla P - \rho_m \nabla \Phi_g + \boldsymbol{j} \times \boldsymbol{B} \tag{7.87}$$

$$\boldsymbol{j} = \sigma(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}) \tag{7.88}$$

$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \tag{7.89}$$

$$\nabla \times \boldsymbol{B} = \mu_0 \boldsymbol{j} \tag{7.90}$$

$$\nabla \cdot \boldsymbol{B} = 0. \tag{7.91}$$

To these equations we again have to add a proper closure procedure. For an adiabatic process this would be the adiabatic law

$$\frac{\mathrm{D}}{\mathrm{D}t}(P\rho_m^{-\gamma}) = 0. \tag{7.92}$$

The use of the adiabatic law (7.92) restricts the processes to be studied to take place rapidly enough to allow heat conduction effects to be neglected. Still, the process must be slow enough so that electromagnetic effects are unimportant. In the more general case, this equation would be replaced with the proper equation of state together with a generalized version of the energy equation in which also Joule heating effects are taken into account.

The set of equations (7.86) - (7.92) is known as the *magneto-hydrodynamic* (MHD) model. The model constitutes a relationship between 14 different physical quantities ρ_m, $\boldsymbol{v}$, P, $\boldsymbol{B}$, $\boldsymbol{E}$ and $\boldsymbol{j}$. Some of these quantities may easily be eliminated, for instance $\boldsymbol{j}$ and $\boldsymbol{E}$. Due to the approximations made, the model suffers from minor inconsistencies. Still, the MHD model has been used with success in a large number of situations.

Viscous forces and heat conduction were neglected ($\boldsymbol{\mathcal{F}}_\eta = 0$ and $\lambda = 0$) in (7.86) - (7.92). If in addition the electric conductivity of the fluid tends to infinity ($\sigma = \infty$), the model reduces to the *ideal magneto-hydrodynamic* model. In this approximation Ohm's law (7.88)

reduces to the requirement that the electric field seen by an observer flowing with the fluid vanishes,

$$\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B} = 0. \tag{7.93}$$

We shall now turn to a discussion of some of the consequences of the MHD model.

7.12 Magnetic Pressure and Tension Forces

In (7.87) the magnetic force was expressed in terms of the electric current density and the magnetic field. These two quantities are not unrelated. As seen from (7.90) the current density acts as source for the magnetic field. Often a spatially varying magnetic field indicates the presence of a local current density.

It is sometimes convenient to express the magnetic force in terms of the magnetic field alone. This is easily achieved by the substitution of Ampére's law (7.90) into the force expression

$$\boldsymbol{j} \times \boldsymbol{B} = \frac{1}{\mu_0}(\nabla \times \boldsymbol{B}) \times \boldsymbol{B} = \frac{1}{\mu_0}\boldsymbol{B} \cdot \nabla \boldsymbol{B} - \nabla \frac{\boldsymbol{B}^2}{2\mu_0}. \tag{7.94}$$

The last term is of the same form as the pressure force. The expression $P_m = \boldsymbol{B}^2/2\mu_0$ is therefore called the *magnetic pressure*. The first term can be rearranged and given a precise physical interpretation. Introducing $\boldsymbol{B} = B\,\hat{\boldsymbol{b}}$ where $\hat{\boldsymbol{b}}$ is a unit vector along $\boldsymbol{B}$, we may write

$$\frac{1}{\mu_0}\boldsymbol{B} \cdot \nabla \boldsymbol{B} = \frac{B^2}{\mu_0}\,\hat{\boldsymbol{b}} \cdot \nabla \hat{\boldsymbol{b}} + \hat{\boldsymbol{b}}\hat{\boldsymbol{b}} \cdot \nabla \frac{B^2}{2\mu_0}. \tag{7.95}$$

But, as seen from figure 7.10, we have

$$\hat{\boldsymbol{b}} \cdot \nabla \hat{\boldsymbol{b}} = \frac{\hat{\boldsymbol{n}}}{R}, \tag{7.96}$$

where R is the local *radius of curvature* of the magnetic field line and $\hat{\boldsymbol{n}}$ is a unit vector pointing toward the local center of curvature. In particular, $\hat{\boldsymbol{n}}$ and $\hat{\boldsymbol{b}}$ are orthogonal vectors.

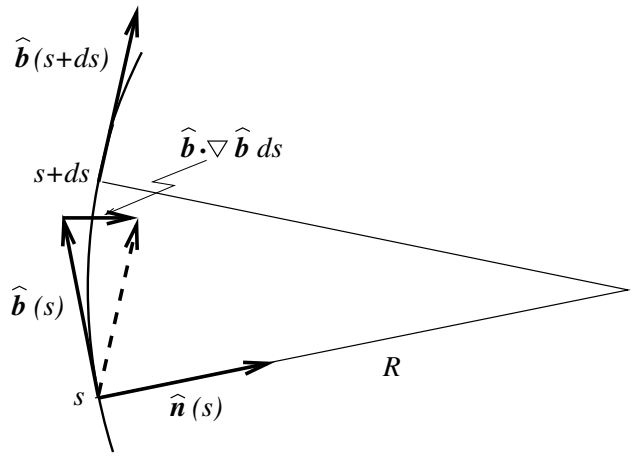

Figure 7.10: Field line geometry

Combining these expressions, we have

$$\boldsymbol{j} \times \boldsymbol{B} = \frac{B^2}{\mu_0}\, \hat{\boldsymbol{b}} \cdot \nabla \hat{\boldsymbol{b}} - \nabla_\perp \frac{B^2}{2\mu_0}, \tag{7.97}$$

where the subscript $\perp$ indicates that only the components of the gradient operator perpendicular to $\boldsymbol{B}$ are to be retained. The expression (7.97) shows that the magnetic force may be considered made up from two parts. The last term is the *magnetic pressure force* always acting perpendicular to the magnetic field. The first term represents a *magnetic tension force*, also acting perpendicular to the magnetic field. The magnetic tension force may be visualized by considering magnetic field lines as stretched elastic strings. In this picture the conducting fluid is pushed sideways by the curved magnetic field line tending to straighten up, like an arrow on the bowstring. The contents of these different force concepts will be illustrated further in the subsequent discussion.

7.13 Magneto-hydrostatic Equilibria

We have previously studied static equilibrium states in non-conducting fluids. The magnetic force leads to the possibility of new classes of static equilibrium states in the conducting fluid. Neglecting effects of gravitational forces, the momentum equation (7.87) reduces to

$$\nabla P = \boldsymbol{j} \times \boldsymbol{B}. \tag{7.98}$$

With the proper choice of magnetic field geometry, a pressure maximum may be maintained in the fluid. In particular, constant pressure surfaces will be spanned by a web of crossing current and magnetic field lines as depicted in figure 7.11,

$$\boldsymbol{j} \cdot \nabla P = 0 \quad \text{and} \quad \boldsymbol{B} \cdot \nabla P = 0.$$

In terms of the magnetic pressure and tension forces, the equilibrium condition (7.98) takes the form

$$\nabla_\perp \left(P + \frac{\boldsymbol{B}^2}{2\mu_0} \right) = \frac{\boldsymbol{B}^2}{\mu_0}\, \hat{\boldsymbol{b}} \cdot \nabla \hat{\boldsymbol{b}}.$$

In regions with straight magnetic field lines, the magnetic tension force vanishes, $\hat{\boldsymbol{b}} \cdot \nabla \hat{\boldsymbol{b}} = \boldsymbol{0}$, and the sum of fluid and magnetic pressures will be constant. In regions with curved magnetic fields the sum of the two pressures will *not* be constant.

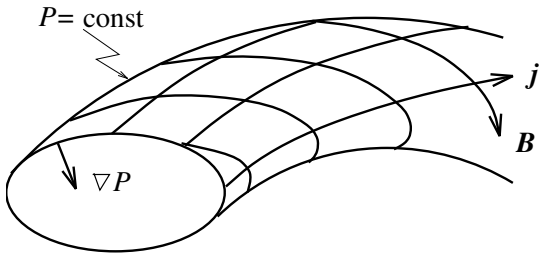

Figure 7.11: P-surface spanned by $\boldsymbol{j}$ and $\boldsymbol{B}$ lines

Let us consider a simple example. In an otherwise homogeneous fluid a circular cross section current channel is maintained along a constant external magnetic field $\boldsymbol{B}_0$. We introduce a coordinate system with the z axis along the center of the current channel. Let the current channel have constant current density $\boldsymbol{j} = j_0\hat{\boldsymbol{z}}$ inside the radius a. The situation is illustrated in figure 7.12. The current will produce an additional azimuthal magnetic field

$$\boldsymbol{B}_\varphi = \begin{cases} \frac{\mu_o j_0}{2}\,\rho\,\hat{\boldsymbol{\varphi}} & \text{for} \quad \rho < a \\ \frac{\mu_o j_0 a^2}{2\rho}\,\hat{\boldsymbol{\varphi}} & \text{for} \quad \rho > a. \end{cases}$$

Combined with the original current $\boldsymbol{j}$ this field generates a magnetic force

$$\boldsymbol{j} \times \boldsymbol{B}_\varphi = -\frac{\mu_0 j_0^2}{2}\rho\hat{\boldsymbol{\rho}}$$

acting back on the fluid for $\rho < a$. To establish a static equilibrium state, the force balance condition (7.98) requires the pressure in the fluid to vary according to

$$P(\rho) = P_0 + \frac{\mu_0 j_0^2}{4}(a^2 - \rho^2),$$

where P_0 is the fluid pressure outside the current channel. The fluid pressure and the azimuthal magnetic field strength are both plotted in the figure 7.12.

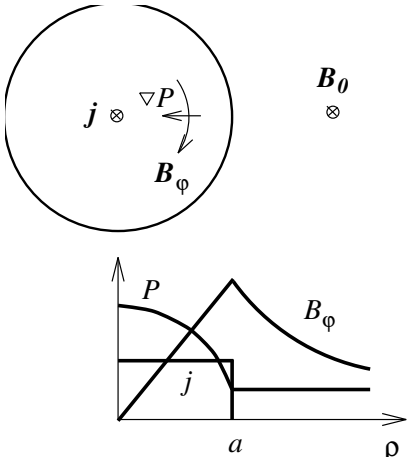

Figure 7.12: Linear Pinch

In this example the magnetic force inside the current channel is divided equally between the magnetic pressure force and the magnetic tension force,

$$-\nabla_\perp \frac{B_\varphi^2}{2\mu_0} = \frac{B_\varphi^2}{\mu_0}\hat{\boldsymbol{b}} \cdot \nabla\hat{\boldsymbol{b}} = -\frac{\mu_0 j_0^2}{4}\rho\hat{\boldsymbol{\rho}}.$$

Outside the current channel the total magnetic force vanishes. Here the magnetic pressure and tension forces are equal but of opposite direction. The tendency of the curved magnetic field line to straighten up is counteracted by an outward decreasing magnetic pressure.

The static MHD equilibria are of the greatest interest in the controlled fusion program. In this application, the nested pressure surfaces are bent into toroidal shape in the TOKAMAK device. One of the major problems in this research has been to find stable equilibria. In the stellar interiors and atmospheres, current carrying magnetic loops play important roles in the theory of magnetic flux-tube buoyancy, sun-spot formation, solar prominences and flares, and in dynamo theories for the generation and maintenance of stellar and planetary magnetic fields.

Quiz 7.32 : In a static equilibrium state a current sheath maintains the magnetic field

$$\boldsymbol{B} = \begin{cases} B_0\,\hat{\boldsymbol{z}}, & x < -d \\ \frac{1}{2d}(B_0(d-x) + B_1(d+x))\,\hat{\boldsymbol{z}}, & -d < x < d \\ B_1\,\hat{\boldsymbol{z}}, & x > d \end{cases}$$

where $B_1 > B_0$. For $x < -d$ the fluid pressure is $P = P_0$. Determine the current density $\boldsymbol{j}$ and the pressure P as functions of x. What are the magnetic pressure and tension forces acting on the fluid in this case?

7.14 Frozen Fields and Field Diffusion

The elimination of $\boldsymbol{E}$ from (7.88) and (7.89) leads to an equation for the evolution of the magnetic field in the form

$$\frac{\partial \boldsymbol{B}}{\partial t} - \nabla \times (\boldsymbol{v} \times \boldsymbol{B}) = -\nabla \times \frac{\boldsymbol{j}}{\sigma}. \tag{7.99}$$

The left hand side expression is seen to be identical in form to that of the vorticity equation (7.55). To go from (7.55) to (7.99) we only have to replace the vorticity $\boldsymbol{\omega}$ by the magnetic field $\boldsymbol{B}$. Even the right hand sides of these equations are of similar form. We may therefore copy our previous discussion of Kelvin's theorem for fluid flows in section 7.9 to the present magnetic field problem in electrically conducting fluids. Instead of (7.63) giving the time development of the circulation Γ, we here find a relation for the time variation of the magnetic flux

$$\Phi = \int_{A(t)} \mathrm{d}^2\boldsymbol{\mathcal{A}} \cdot \boldsymbol{B}$$

through the open surface $A(t)$ bounded by a simply connected material curve $C(t)$ in the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi = -\oint_{C(t)} \mathrm{d}\boldsymbol{\ell} \cdot \frac{\boldsymbol{j}}{\sigma}. \tag{7.100}$$

In analogy with the discussion of Kelvin's theorem and its consequences, we may now formulate *Alfvén's frozen field theorem*:
In a perfectly conducting fluid ($\sigma = \infty$), the magnetic flux contained within any simply connected material curve C remains constant.

In complete analogy with the discussion following Kelvin's theorem in section 7.9, if follows from Alfvén's frozen field theorem that magnetic flux tubes are convected by the perfectly conducting fluid. A *flux tube* is a cylindrical volume in the fluid whose side surface is made

entirely from magnetic field lines at any given time t. Fluid particles that at any instance lie on the walls of a flux tube will continue to form the wall of a magnetic flux tube. This means that the magnetic field lines may be visualized as frozen to the fluid. Fluid elements lying along a magnetic field line at a given time will continue to lie along a magnetic field line at later times. Furthermore, the magnetic flux Φ contained within any given flux tube is constant along the tube.

In a fluid with finite electrical conductivity σ, the magnetic field and the fluid tend to diffuse with respect to each other. This is readily seen from (7.100). Any flux tube with cross section A with a larger magnetic flux density than the surrounding region requires an azimuthal current to be flowing at its bounding surface. With the surface A directed along the magnetic field, this current density $\boldsymbol{j}$ will be directed along the positive direction of the material curve C. The situation is illustrated in figure 7.13. The line integral on the right hand side of (7.100) will thus be positive, leading to a decrease in the magnetic flux within the same curve. Any local enhancement in the magnetic field thus tends to be destroyed through the diffusion of magnetic field lines relative to the conducting fluid. For a local magnetic depletion, the current density will be reversed. Through magnetic diffusion local minima in magnetic flux density tend to be filled.

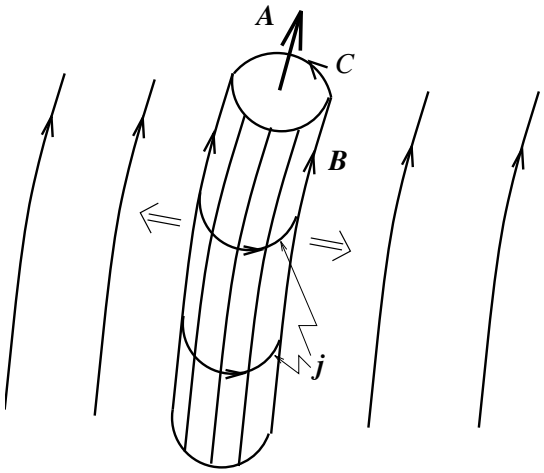

Figure 7.13: Magnetic field diffusion

Quiz 7.33 : A magnetic field directed along the z axis has a local enhancement near the axis. The field is given by $\boldsymbol{B}(\rho) = (B_0 + B_1(\rho))\hat{\boldsymbol{z}}$ where $\mathrm{d}B_1/\mathrm{d}\rho < 0$ and ρ represents the distance from the axis. What is the current density $\boldsymbol{j}$ necessary to produce the enhancement? Show that for a contour C encircling the z axis in the positive direction, the right hand side of (7.100) is negative.

7.15 The Virial Theorem

Kelvin's circulation theorem and Alfvén's frozen field theorem are both examples of far-reaching conclusions being arrived at without actually having to solve the equations of motion

explicitely. The *virial theorem* represents a similar case. For its derivation it is advantageous to rewrite the momentum equation (7.87) for an ideal and conducting fluid in the form

$$\frac{\partial}{\partial t}(\rho_m \boldsymbol{v}) = \nabla \cdot \boldsymbol{T} - \rho_m \nabla \Phi_g \tag{7.101}$$

where the stress tensor $\boldsymbol{T}$ includes kinetic, thermal and magnetic effects

$$\boldsymbol{T} = -\rho_m \boldsymbol{v}\boldsymbol{v} - (P + \frac{\boldsymbol{B}^2}{2\mu_0})\boldsymbol{I} + \frac{1}{\mu_0}\boldsymbol{B}\boldsymbol{B}. \tag{7.102}$$

We have here made use of (7.11) and (7.94). We note that $\boldsymbol{T}$ is a symmetric tensor.

Let us now consider an isolated fluid system, that is a fluid system bounded by a (fixed) surface $\mathcal{A}$ where the mass density ρ_m, and therefore also the pressure P, vanishes identically. For this system we define the scalar moment of inertia, $I = \int \mathrm{d}^3\boldsymbol{r}\, \rho_m \boldsymbol{r}^2$. The corresponding time derivative is

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \int \mathrm{d}^3\boldsymbol{r}\, \frac{\partial \rho_m}{\partial t}\boldsymbol{r}^2 = -\int \mathrm{d}^3\boldsymbol{r}\, \nabla \cdot (\rho_m \boldsymbol{v})\boldsymbol{r}^2 = 2\int \mathrm{d}^3\boldsymbol{r}\, \rho_m \boldsymbol{v} \cdot \boldsymbol{r}. \tag{7.103}$$

Thus, we also find

$$\frac{1}{2}\frac{\mathrm{d}^2 I}{\mathrm{d}t^2} = \int \mathrm{d}^3\boldsymbol{r}\, \frac{\partial}{\partial t}(\rho_m \boldsymbol{v}) \cdot \boldsymbol{r} = \int \mathrm{d}^3\boldsymbol{r}\, \left((\nabla \cdot \boldsymbol{T}) \cdot \boldsymbol{r} - \rho_m \nabla \Phi_g \cdot \boldsymbol{r}\right). \tag{7.104}$$

Through a partial integration, and assuming a surface integral to vanish, the first term on the right hand side of (7.104) is given by

$$\int \mathrm{d}^3\boldsymbol{r}\, (\nabla \cdot \boldsymbol{T}) \cdot \boldsymbol{r} = -\int \mathrm{d}^3\boldsymbol{r}\, \boldsymbol{T} : \boldsymbol{I} = 2(K + \Theta) + M \tag{7.105}$$

where

$$K = \int \mathrm{d}^3\boldsymbol{r}\, \frac{1}{2}\rho_m \boldsymbol{v}^2, \quad \Theta = \frac{3}{2}\int \mathrm{d}^3\boldsymbol{r}\, P \quad \text{and} \quad M = \frac{1}{2\mu_0}\int \mathrm{d}^3\boldsymbol{r}\, \boldsymbol{B}^2$$

are the kinetic, the thermal and the magnetic energy of the system. We note that K, Φ and M are all positive definite quantities.

With the help of (7.7) the last term on the right hand side of (7.104) can be written in the form

$$\begin{aligned}\int \mathrm{d}^3\boldsymbol{r}\, \rho_m \nabla \Phi_g \cdot \boldsymbol{r} &= G\int \mathrm{d}^3\boldsymbol{r}\, \mathrm{d}^3\boldsymbol{r}'\, \rho_m(\boldsymbol{r})\rho_m(\boldsymbol{r}')\, \frac{(\boldsymbol{r}-\boldsymbol{r}')\cdot\boldsymbol{r} + (\boldsymbol{r}'-\boldsymbol{r})\cdot\boldsymbol{r}'}{2\,|\,\boldsymbol{r}-\boldsymbol{r}'\,|^3} \\ &= \frac{G}{2}\int \mathrm{d}^3\boldsymbol{r}\, \mathrm{d}^3\boldsymbol{r}'\, \frac{\rho_m(\boldsymbol{r})\rho_m(\boldsymbol{r}')}{|\,\boldsymbol{r}-\boldsymbol{r}'\,|} = -\frac{1}{2}\int \mathrm{d}^3\boldsymbol{r}\, \rho_m(\boldsymbol{r})\Phi_g(\boldsymbol{r}) = -V_g.\end{aligned} \tag{7.106}$$

Here V_g represents the self-gravitational potential energy of the system and is a negative definite quantity.

The virial theorem can now be written in the simple form

$$\frac{1}{2}\frac{\mathrm{d}^2 I}{\mathrm{d}t^2} = 2K + 2\Theta + M + V_g. \tag{7.107}$$

The virial theorem allows for important physical insight. First of all, without a self-gravitational field the right hand side of (7.107) is positive definite. That means that $\mathrm{d}^2 I/\mathrm{d}t^2$ is positive

and therefore that $\mathrm{d}I/\mathrm{d}t$ will be steadily increasing with time. At least in the long run this means that also I will be increasing with time. A stable self-confining magnetohydrodynamic static equilibrium is therefore not possible. This conclusion will only be reversed if the self-gravitational field of the system is strong enough that the negative contribution from V_g in (7.107) balances the positive contributions from K, Θ and M. In particular, for a non-conducting fluid, and therefore no coupling to any magnetic field, we recover the well-known result that in a static equilibrium of a gravitational system the sum of kinetic and thermal energies equals one half the absolute value of the gravitational energy.

7.16 Linear Waves

An important property of all continuous media is their ability to support waves of different kinds. A wave is a particular perturbation in the properties of the medium that is allowed to propagate away from its point of origin. The wave carries energy with it. If the wave is generated in one part of the medium and absorbed in another, the wave represent a mechanism in which energy is redistributed in the medium. Under some conditions waves damp out as they propagate. Under other situations they may grow in amplitude, eventually leading to shock phenomena or to a generally turbulent state of the medium.

As an introduction to the theory of waves and their properties, we shall consider small amplitude waves in ideal fluids. We consider the simplest possible case in which the fluid before the introduction of the wave is in a uniform and static equilibrium state. This means that we are neglecting effects of any gravitational field on the wave. We shall consider two cases: a non-conducting fluid and a perfectly conducting fluid as described by the ideal MHD model. The conducting fluid will be assumed permeated by a uniform magnetic field. Since the validity of then MHD model is limited to situations where the displacement current can be neglected in comparison to conduction currents, the results we are to derive for the latter case will apply only for low-frequency waves.

7.16.1 Linear Sound Waves

For the analysis of waves in the non-conducting medium we assume small perturbations ρ_{m1}, $\boldsymbol{v}_1$ and P_1 in mass density, velocity and pressure around the corresponding constant equilibrium values, ρ_{m0}, $\boldsymbol{v}_0 = 0$ and P_0 to exist. Substituting

$$\rho_m = \rho_{m0} + \rho_{m1} \tag{7.108}$$

$$\boldsymbol{v} = \boldsymbol{v}_1 \tag{7.109}$$

$$P = P_0 + P_1 \tag{7.110}$$

into the mass and momentum conservation equations (7.2) and (7.10) and neglecting terms of second order in small quantities, a set of *linearized equations* results

$$\frac{\partial \rho_{m1}}{\partial t} = -\rho_{m0}\,\nabla \cdot \boldsymbol{v}_1 \tag{7.111}$$

$$\rho_{m0}\frac{\partial \boldsymbol{v}_1}{\partial t} = -\nabla P_1 = -\frac{\partial P}{\partial \rho_m}|_u\,\nabla \rho_{m1} - \frac{\partial P}{\partial u}|_{\rho_m}\,\nabla u_1. \tag{7.112}$$

In (7.112) we have assumed a general equation of state

$$P = P(\rho_m, u). \tag{7.113}$$

All partial derivatives in (7.112) are evaluated for the equilibrium values. In the subsequent discussion the 0-index on equilibrium quantities will be suppressed for convenience.

Next, the wave dynamics will be assumed to represent an adiabatic prosess, that is, wave motions are assumed to be rapid enough that any heat exchange between neighboring fluid elements can be neglected. From the first law of thermodynamics this means that

$$0 = T\mathrm{d}s = \mathrm{d}u + P\,\mathrm{d}\frac{1}{\rho_m} = \mathrm{d}u - \frac{P}{\rho_m^2}\,\mathrm{d}\rho_m. \tag{7.114}$$

Substituting $u_1 = P/\rho_m^2\,\rho_{m1}$ and eliminating $\boldsymbol{v}_1$ between (7.111) and (7.112), the result is

$$\frac{\partial^2}{\partial t^2}\rho_{m1} = C_s^2\,\nabla^2\rho_{m1} \tag{7.115}$$

with

$$C_s^2 = \frac{\partial P}{\partial \rho_m}|_u + \frac{P}{\rho_m^2}\,\frac{\partial P}{\partial u}|_{\rho_m}. \tag{7.116}$$

The quantity C_s will be refered to as the *sound speed* in the medium. In particular, for a medium satisfying the ideal gas equation of state (see (7.27))

$$P = (\gamma - 1)\rho_m u \tag{7.117}$$

we find

$$C_s^2 = \gamma\,\frac{P}{\rho_m}. \tag{7.118}$$

We recognize (7.115) as the wave equation for the mass density perturbations. Any perturbation in mass density imposed at time $t = 0$ will propagate away with speed C_s. A general solution depending only on one spatial coordinate, for instance z, is given by $\rho_{m1}(z,t) = f(z \pm C_s t)$ where f is any twice differentiable function. This means that an arbitrarily specified density pulse will propagate in the given direction with the sound speed C_s without changing its shape. With ρ_{m1} given the associated perturbation $\boldsymbol{v}_1$ in the fluid velocity can be found from (7.112).

In the following we shall be satisfied by looking for solutions of the linearized equations in the form of plane waves, that is, perturbations in ρ_m, $\boldsymbol{v}$ and P varying in space and time as

$$\cos(\boldsymbol{k}\cdot\boldsymbol{r} - \omega t) = \mathrm{Re}\exp(\imath\boldsymbol{k}\cdot\boldsymbol{r} - \imath\omega t). \tag{7.119}$$

This type of solution is possible since (7.111)-(7.112) is a set of linear, homogeneous partial differential equations with constant coefficients.

Allowing for a possible internal phase difference between the perturbed quantities, we include a complex amplitude factor $\delta\rho_m$ and write for the complex density perturbation

$$\boldsymbol{\rho}_{m1} = \delta\rho_m\,\exp(\imath\boldsymbol{k}\cdot\boldsymbol{r} - \imath\omega t), \tag{7.120}$$

with similar expressions for the other quantities. The complex notation for the perturbations is convenient in that the operators $\partial/\partial t$ and ∇ appearing in the linearized equations may be replaced with the algebraic factors $-\imath\omega$ and $\imath\boldsymbol{k}$. We remember that on the basis of the Fourier transform theory outlined in table 1.1, any wave form may be synthesized from plane wave solutions.

In particular for the present case, the wave equation (7.115) reduces to

$$(\omega^2 - C_s^2 \boldsymbol{k}^2)\, \rho_{m1} = 0.$$

Non-trivial plane wave solutions are thus only possible if the dispersion relation

$$\omega^2 = C_s^2\, k^2 \tag{7.121}$$

is satisfied. This is a general property of plane wave solutions, for a given wave vector $\boldsymbol{k}$, non-trivial solutions are only possible for particular values of the angular frequency ω. The form of the dispersion relation will differ from one type of wave to another, but there will always be dispersion relation. The phase speed of the present em sound or *acoustic wave*

$$v_{ph} = \omega/k = C_s \tag{7.122}$$

is seen to be independent of the wave vector $\boldsymbol{k}$. This means that plane waves that propagate in the same direction with different wave numbers k and therefore different ω all propagate together. This is also a necessary condition for the acoustic pulse to retain its shape during the propagation. Waves with this property are said to be *dispersion-less*.

The energy in the wave is transported with group velocity

$$\boldsymbol{v}_{gr} = \partial\omega/\partial\boldsymbol{k} = C_s\, \hat{\boldsymbol{k}}. \tag{7.123}$$

For the acoustic wave the group velocity is identical to the phase velocity $v_{ph}\hat{\boldsymbol{k}}$. This is *not* a general property of linear waves, as will be seen below. Finally, with ρ_{m1} given the corresponding velocity perturbation can be found from (7.112)

$$\boldsymbol{v}_1 = \frac{C_s}{\rho_m}\rho_{m1}\, \hat{\boldsymbol{k}}. \tag{7.124}$$

The material motion in the fluid induced by the wave is thus in the direction of the wave propagation. This means that $\boldsymbol{k} \cdot \boldsymbol{v}_1 \neq 0$ and therefore that $\nabla \cdot \boldsymbol{v}_1 \neq 0$. The fluid flow associated with the acoustic wave is thus a compressible flow. The wave is thus called a *compressible wave*.

7.16.2 Linear MHD Waves

For the perfectly conducting, magnetized fluid we must in additions to perturbations in mass density, velocity and pressure (7.111)-(7.112) also include perturbations in the magnetic field

$$\boldsymbol{B} = \boldsymbol{B}_0 + \boldsymbol{B}_1. \tag{7.125}$$

Substituting in the ideal MHD equations (7.86)-(7.87) and (7.99), the latter with $\sigma = \infty$, and linearizing we find this time

$$\frac{\partial \rho_{m1}}{\partial t} = -\rho_{m0}\nabla \cdot \boldsymbol{v}_1 \tag{7.126}$$

$$\rho_0 \frac{\partial \boldsymbol{v}_1}{\partial t} = -\nabla P_1 + \frac{1}{\mu_0}\,(\nabla \times \boldsymbol{B}_1) \times \boldsymbol{B}_0 \tag{7.127}$$

$$\frac{\partial \boldsymbol{B}_1}{\partial t} = \boldsymbol{B}_0 \cdot \nabla \boldsymbol{v}_1 - \nabla \cdot \boldsymbol{v}_1\, \boldsymbol{B}_0. \tag{7.128}$$

For an adiabatic process we may in accordance with discussion in section 7.16.1 make use of the replacement $P_1 = C_s^2\,\rho_{m1}$ where C_s is the sound speed.

Before turning to a full discussion of the linearized MHD equations (7.126)-(7.128), let us first demonstrate that an *incompressible wave* solution does exist, that is a solution for which $\nabla \cdot \boldsymbol{v}_1 = 0$ or $\boldsymbol{k} \cdot \boldsymbol{v}_1 = 0$. From (7.126) and (7.128) it follows that

$$\rho_{m1} = P_1 = 0 \tag{7.129}$$

$$\boldsymbol{B}_1 = -\frac{\boldsymbol{B}_0 \cdot \boldsymbol{k}}{\omega}\boldsymbol{v}_1. \tag{7.130}$$

The linearized momentum equation (7.112) then reduces to

$$\omega^2 \boldsymbol{v}_1 = \frac{\boldsymbol{B}_0 \cdot \boldsymbol{k}}{\mu_0 \rho_0}(\boldsymbol{k} \times \boldsymbol{v}_1) \times \boldsymbol{B}_0. \tag{7.131}$$

The right hand side vector is perpendicular to $\boldsymbol{B}_0$. This must also be the case for the left hand side vector, leading to the requirement on the velocity amplitude vector

$$\hat{\boldsymbol{b}} \cdot \delta\boldsymbol{v} = 0. \tag{7.132}$$

The incompressibility condition is equivalent with the requirement

$$\boldsymbol{k} \cdot \delta\boldsymbol{v} = 0. \tag{7.133}$$

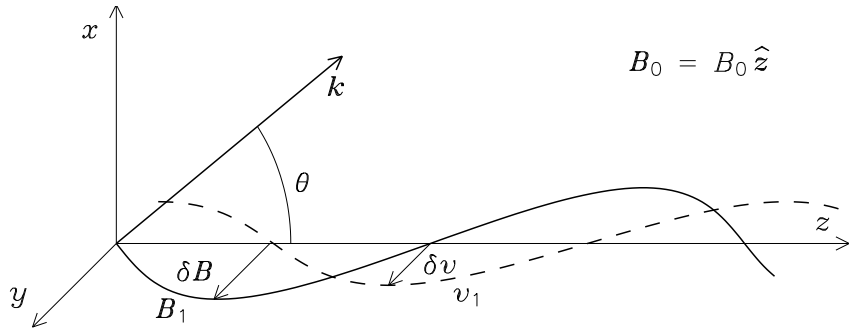

Figure 7.14: The Alfvén wave

With the latter two conditions satisfied, equation (7.131) will have non-trivial solutions provided the angular frequency ω and the wave vector $\boldsymbol{k}$ satisfy the dispersion relation

$$\omega^2 = V_A^2(\hat{\boldsymbol{b}} \cdot \boldsymbol{k})^2 = V_A^2 k^2 \cos^2\theta, \tag{7.134}$$

where $V_A = B_0/\sqrt{\mu_0\rho_0}$ is the *Alfvén speed* and θ is the angle between $\boldsymbol{B}_0$ and $\boldsymbol{k}$. The solution found is called an *Alfvén wave* (or sometimes the shear Alfvén wave). The polarization of the wave, here defined as the direction of amplitude vector $\delta\boldsymbol{v}$ of the velocity perturbation, is perpendicular to both $\boldsymbol{B}_0$ and $\boldsymbol{k}$. The wave appears as a propagating periodic buckling of the magnetic field. This is depicted in figure 7.14. The wave vector is lying in the xz-plane, the magnetic field perturbation and the corresponding fluid velocity perturbation in the yz-plane.

We note that the relation between the magnetic field and fluid velocity perturbations (7.130) is such as to satisfy the frozen field theorem. The fluid and the magnetic field move together.

From the dispersion relation (7.134), it is seen that for a given angle θ the frequency of the wave increases linearly with the wave number k. This means that the phase speed $v_{ph} = \omega/k$ for a given direction of propagation is independent of the wave number k. The Alfvén wave is therefore a dispersion-less wave. For this case the dispersion relation (7.134) can conveniently be represented in graphical form by plotting the phase speed as a function of the direction of the wave vector $\boldsymbol{k}$. Such a plot is called a *phase velocity surface*. The corresponding phase velocity surface is a surface of revolution with the equilibrium magnetic field direction as the axis of symmetry. For the rotationally symmetric case it is sufficient to plot one section through this surface. The result for the Alfvén wave is given in figure 7.15, curves marked A. Part a) of the figure corresponds to the case where $V_A > C_s$, part b) to the case $V_a < C_s$. In the figure, the polarization direction of the mode is also indicated. As already remarked, $\delta\boldsymbol{v}$ is perpendicular to the plane containing $\boldsymbol{k}$ and $\boldsymbol{B}_0$.

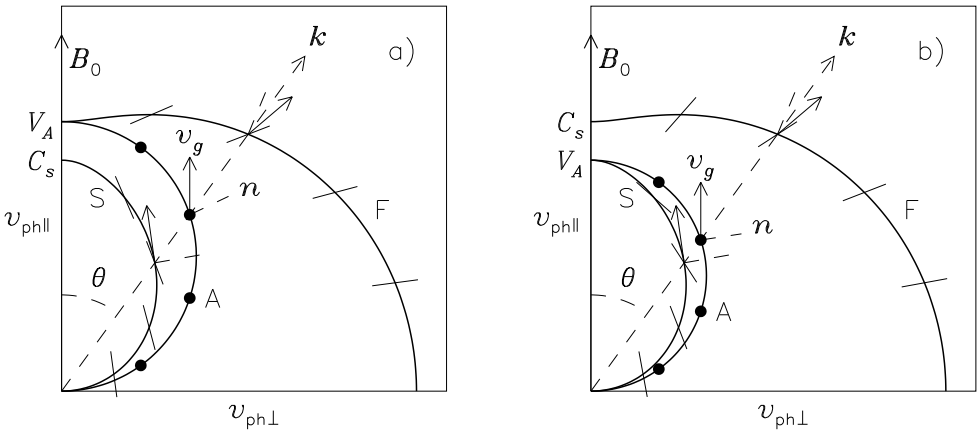

Figure 7.15: Section through the phase velocity surface for the Alfvén, the fast and the slow magneto-sonic waves for a) $V_a > C_s$ and b) $V_A < C_s$. Only one quadrant of the section is plotted.

Waves transport energy through the medium. The velocity of energy propagation is given by the group velocity

$$\boldsymbol{v}_{gr} = \frac{\partial}{\partial \boldsymbol{k}}\omega(\boldsymbol{k}) = \hat{\boldsymbol{k}}\frac{\partial\omega(k,\theta)}{\partial k} + \hat{\boldsymbol{\theta}}\frac{1}{k}\frac{\partial\omega(k,\theta)}{\partial\theta}. \tag{7.135}$$

For the Alfvén wave, the $\partial\omega/\partial\theta$ term is generally non-vanishing. The wave therefore carries energy in a direction different from the direction of the phase velocity as given by the wave vector $\boldsymbol{k}$. Indeed, the group velocity is always directed along the equilibrium magnetic field $\boldsymbol{B}_0$.

The Alfvén wave is but one of three different solutions of the linearized equations (7.126) - (7.128). To find the remaining wave modes we must abandon the simplifying assumption

$\nabla \cdot \boldsymbol{v}_1 = 0$. For the general solution equations (7.126) and (7.128) read

$$\rho_{m1} = \frac{\rho_0}{\omega} \boldsymbol{k} \cdot \boldsymbol{v}_1, \tag{7.136}$$

$$\boldsymbol{B}_1 = \frac{1}{\omega}(\boldsymbol{B}_0 \boldsymbol{k} \cdot \boldsymbol{v}_1 - \boldsymbol{B}_0 \cdot \boldsymbol{k} \boldsymbol{v}_1). \tag{7.137}$$

When substituted in the corresponding linearized momentum equation (7.112) an equation for the velocity perturbation $\delta \boldsymbol{v}$ results

$$\omega^2 \delta \boldsymbol{v} = (C_s^2 + V_A^2) \boldsymbol{k}\, \boldsymbol{k} \cdot \delta \boldsymbol{v} + V_A^2 \left[(\boldsymbol{k} \cdot \hat{\boldsymbol{b}})^2 \delta \boldsymbol{v} - \hat{\boldsymbol{b}}\, \boldsymbol{k} \cdot \hat{\boldsymbol{b}}\, \boldsymbol{k} \cdot \delta \boldsymbol{v} - \boldsymbol{k}\, \boldsymbol{k} \cdot \hat{\boldsymbol{b}}\, \hat{\boldsymbol{b}} \cdot \delta \boldsymbol{v} \right]. \tag{7.138}$$

Equation (7.138) is a linear and homogeneous system of equations in the three components of $\delta \boldsymbol{v}$. If the z axis is chosen parallel to $\hat{\boldsymbol{b}}$ and in addition the x axis is oriented such that the xz-plane contains the $\boldsymbol{k}$ vector, the equation can be written in matrix form as

$$A \cdot \delta \boldsymbol{v} = 0$$

with

$$A = \begin{bmatrix} \omega^2 - V_A^2 k^2 - C_s^2 k^2 \sin^2\theta & 0 & -C_s^2 k^2 \sin\theta \cos\theta \\ 0 & \omega^2 - V_A^2 k^2 \cos^2\theta & 0 \\ -C_s^2 k^2 \sin\theta \cos\theta & 0 & \omega^2 - C_s^2 k^2 \cos^2\theta \end{bmatrix}. \tag{7.139}$$

A necessary and sufficient condition for the existence of non-trivial solutions is that the determinant of the system matrix A vanishes. This leads to three different solutions for ω and $\delta \boldsymbol{v}$, for each choice of the wave vector $\boldsymbol{k}$, namely

$$\frac{\omega^2}{k^2} = V_A^2 \cos^2\theta \tag{7.140}$$

with

$$\delta \boldsymbol{v} \sim (0, 1, 0), \tag{7.141}$$

or

$$\frac{\omega^2}{k^2} = \frac{1}{2}(C_s^2 + V_A^2) \pm \frac{1}{2}\sqrt{(C_s^2 + V_A^2)^2 - 4 C_s^2 V_A^2 \cos^2\theta} \tag{7.142}$$

with

$$\delta \boldsymbol{v} \sim (\sin\theta \cos\theta, 0, \frac{1}{2}(1 - \frac{V_A^2}{C_s^2}) - \sin^2\theta \pm \frac{1}{2}\sqrt{(1 + \frac{V_A^2}{C_s^2})^2 - 4\frac{V_A^2}{C_s^2}\cos^2\theta}\,). \tag{7.143}$$

The first of these wave modes, (7.140) - (7.141), is the previously studied Alfvén wave. The other two modes, (7.142) - (7.143), generally have a component of $\delta \boldsymbol{v}$ along $\boldsymbol{k}$ and are thus examples of compressive waves, $\nabla \cdot \boldsymbol{v}_1 \neq 0$. They are called the *fast* and the *slow magneto-sonic modes*. Also these modes are dispersion-less. The phase velocities and corresponding polarization of these modes have already been given in figure 7.15. Note that the result will be different depending on which of the velocities C_s and V_A is the larger one. For propagation perpendicular to $\hat{\boldsymbol{b}}$, the phase velocity of the fast mode is always given by $\sqrt{C_s^2 + V_A^2}$, while for parallel propagation the velocity is $\max(C_s, V_a)$. Generally we find

$$v_{ph}^S \leq \min(C_s, V_A) \leq \max(C_s, V_A) \leq v_{ph}^F.$$

Also for the fast and slow modes, the direction of the group velocity $\boldsymbol{v}_{gr}$ will generally be different from the direction of the corresponding phase velocity. In fact, the wave vector $\boldsymbol{k}$ divides the angle between the group velocity and the normal to the phase velocity surface. The construction is illustrated in figure 7.15. Note that the construction only gives the direction of the group velocity, not its magnitude.

An alternative to plotting the phase velocity surface for displaying the properties of the different wave modes is to plot the corresponding *refractive index surface*, $\mathcal{N} = c/v_{ph}$ as a function of wave vector direction. For the refractive index surface the direction of the group velocity coincides with the surface normal.

Quiz 7.34 : From the total differential

$$\mathrm{d}P = \frac{\partial P}{\partial \rho_m}|_u \,\mathrm{d}\rho_m + \frac{\partial P}{\partial u}|_{\rho_m} \,\mathrm{d}u$$

show that

$$\frac{\partial P}{\partial \rho_m}|_u + \frac{\partial P}{\partial u}|_{\rho_m} \frac{\partial u}{\partial \rho_m}|_P = 0 \tag{7.144}$$

and therefore that an alternative general expression for C_s^2 is

$$C_s^2 = \left(\frac{P}{\rho_m^2} - \frac{\partial u}{\partial \rho_m}|_P \right) / \frac{\partial u}{\partial P}|_{\rho_m}. \tag{7.145}$$

Verify that for the ideal gas equation of state that (7.118) is still valid.

Quiz 7.35 : For the Alfvén wave show that:
i) $v_{ph}(\theta)$ is represented by a half circle in the xz-plane with center at $x = 0$, $z = V_A/2$ and radius $V_A/2$,
ii) the magnetic field perturbation associated with the Alfvén mode is always perpendicular to the equilibrium magnetic field, $\delta\boldsymbol{B} \cdot \hat{\boldsymbol{b}} = 0$,
iii) the group velocity is directed along the magnetic field, and
iv) the wave vector $\boldsymbol{k}$ cuts the angle between the group velocity and the normal to the phase velocity surface in two equal halves.

Quiz 7.36 : In the absence of an equilibrium magnetic field $\boldsymbol{B}_0$, show that the linearized equations (7.111) - (7.125) allow for only one wave mode, a *sound wave* satisfying $\delta\boldsymbol{v} \parallel \boldsymbol{k}$ and $\omega^2/k^2 = C_s^2$. What is the shape of the phase velocity surface? What is the group velocity? With $\boldsymbol{k} = k\hat{\boldsymbol{z}}$ plot ρ_1, v_{1z} and P_1 as functions of z for a given time t. Is the sound wave an incompressible or compressible wave?

Quiz 7.37 : A spectral line is formed in a local non-magnetized region in which a sound wave is propagating towards the observer. The line producing region is stationary with respect to the observer, yet a slighted blue-shifted spectral line is reported. What is the relationship between the wave induced flow velocity and the density perturbation in the medium? How will the local contribution to the spectral line depend on the local density perturbation? Are you able to give a qualitative explanation of the reported blue-shift?

Quiz 7.38 : Show that the polarization of the fast and slow modes are always orthogonal, $\delta\boldsymbol{v}^F \cdot \delta\boldsymbol{v}^S = 0$ for any given direction of $\boldsymbol{k}$.

Quiz 7.39 : Sketch the shape of the refractive index surfaces for the three MHD wave modes for the case $V_A > C_s$. By comparing with figure 7.15, does the group velocity $\boldsymbol{v}_g$ for a given wave vector $\boldsymbol{k}$ seem to coincide with the corresponding refractive index surface normal $\hat{\boldsymbol{n}}$? Are you able to prove that this is actually correct?

7.17 Shocks

The equations of motion for fluid systems are in the form of partial differential equations. We normally expect the solutions of these equations to be continuous funtions in space and time. The equations do, however, also allow for discontinuous solutions. We shall refer to these discontinuities as shocks. The equations of motion pose constraints as to the allowed form of these discontinuities and their velocity. These constraints are known as the *Rankine-Hugoniot relations*. We next derive and discuss these constraints.

Consider a general conservation law for a field quantity Q and its corresponding flux $\boldsymbol{F}$ written in conservative form

$$\frac{\partial Q}{\partial t} + \nabla \cdot \boldsymbol{F} = 0. \tag{7.146}$$

The continuity equation, the momentum equation and the energy conservation equation for an ideal non-conducting fluid can all be written in this form,

$$\frac{\partial \rho_m}{\partial t} + \nabla \cdot (\rho_m \boldsymbol{v}) = 0 \tag{7.147}$$

$$\frac{\partial}{\partial t}(\rho_m \boldsymbol{v}) + \nabla \cdot (\rho_m \boldsymbol{v}\boldsymbol{v} + P\mathbf{1}) = 0 \tag{7.148}$$

$$\frac{\partial}{\partial t}\left(\rho_m(\frac{1}{2}\boldsymbol{v}^2 + u)\right) + \nabla \cdot \left(\rho_m(\frac{1}{2}\boldsymbol{v}^2 + u)\boldsymbol{v} + P\boldsymbol{v}\right) = 0. \tag{7.149}$$

Let us now assume a discontinuity in Q to exist, moving with local speed D in the direction $\hat{\boldsymbol{n}}$. Let the discontinuity locally be enclosed in a coin-shaped volume with area $\mathcal{A}$ and thickness h, oriented with and moving with the discontinuity as indicated in figure 7.16. Integrating each term of (7.146) over the coin-shaped volume and taking the limit as the coin thickness h tends to zero, the integral form of the conservation law (7.146) reduces to

$$-D\,[Q] + \hat{\boldsymbol{n}} \cdot [\boldsymbol{F}] = 0. \tag{7.150}$$

Here $[Q] \equiv Q_1 - Q_0$ and $Q_{0,1}$ represent the limiting values of the field quantity Q on the two sides of the discontinuity. We see that (7.150) represents a relation between allowed steps in the quantity Q and its corresponding flux $\boldsymbol{F}$ across the discontinuity moving with speed D in direction $\hat{\boldsymbol{n}}$.

Using the result (7.150) the continuity equation (7.147) leads to the relation

$$[\rho_m(v_n - D)] = 0, \tag{7.151}$$

where $v_n = \hat{\boldsymbol{n}} \cdot \boldsymbol{v}$ is the normal component of the velocity $\boldsymbol{v}$.

The momentum equation (7.148) leads to

$$[\rho_m(v_n - D)\,\boldsymbol{v}] + [P]\,\hat{\boldsymbol{n}} = 0.$$

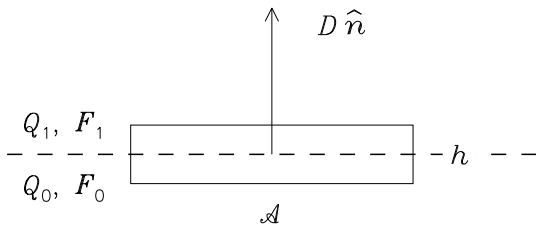

Figure 7.16: Shock discontinuity (dashed line) moving with velocity D and co-moving coin-shaped volume with side $\mathcal{A}$ and height h

This means that the tangential part of the velocity must remain continuous across the discontinuity,

$$[\boldsymbol{v}_t] = 0. \tag{7.152}$$

In addition we find

$$[P] = -[\rho_m(v_n - D)v_n] = -[\rho_m(v_n - D)^2]. \tag{7.153}$$

From the energy conservation equation (7.149) we derive the constraint

$$[\rho_m(v_n - D)(\frac{1}{2}\boldsymbol{v}^2 + u)] + [Pv_n] = 0$$

or

$$[u] = -\frac{1}{2}[v_n^2] - \frac{[Pv_n]}{\rho_{m0}(v_{n0} - D)}.) \tag{7.154}$$

By noting that

$$\frac{1}{2}[v_n^2] = \frac{1}{2}[(v_n - D)^2] + D\,[v_n - D]$$

and

$$\begin{aligned}\frac{[Pv_n]}{\rho_{m0}(v_{n0} - D)} &= \frac{P_1 v_{n1}}{\rho_{m1}(v_{n1} - D)} - \frac{P_0 v_{n0}}{\rho_{m0}(v_{n0} - D)} \\ &= \frac{P_1}{\rho_{m1}} - \frac{P_0}{\rho_{m0}} + D\frac{[P]}{\rho_{m0}(v_{n0} - D)} = [\frac{P}{\rho_m}] - D\,[v_n - D]\end{aligned}$$

the result may be written in the form

$$[u + \frac{P}{\rho_m} + \frac{1}{2}(v_n - D)^2] = 0. \tag{7.155}$$

The quantity $h \equiv u + P/\rho_m$ is the *specific enthalpy* and is equivalent to the total heat content per unit mass of the system.

The constraints (7.151), (7.153) and (7.155), together with (7.152) are known as the *Rankine-Hugoniot relations*. Together with the proper equation of state, these relations determine allowed steps in ρ_m, $\boldsymbol{v}$, P and u across any hydrodynamic shock as functions of the shock speed D.

An alternative and often useful form of the Rankine-Hugoniot relations are expressed in terms of the *spesific volume* $\mathcal{V} \equiv \rho_m^{-1}$. From (7.151) we note that

$$D\,[\rho_m] = [\rho_m v_n] = v_{n0}\,[\rho_m] + \rho_{m1}\,[v_n]$$

and therefore

$$[v_n] = \rho_{m0}(D - v_{n0})\frac{[\rho_m]}{\rho_{m1}\rho_{m0}} = -\rho_{m0}(D - v_{n0})\,[\mathcal{V}], \tag{7.156}$$

and

$$[P] = \rho_{m0}(D - v_{n0})\,[v_n] = -\rho_{m0}^2(D - v_{n0})^2\,[\mathcal{V}] \tag{7.157}$$

We also note from (7.156) and (7.153) that

$$[v_n]^2 = -[P][\mathcal{V}] \tag{7.158}$$

and thus

$$D = v_{n0} \pm \mathcal{V}_0\sqrt{-\frac{[P]}{[\mathcal{V}]}}. \tag{7.159}$$

Finally, with the help of (7.162) and (7.163) the energy equation (7.154) may be rewritten

$$\begin{aligned}[u] &= -v_{n0}\,[v_n] - \frac{1}{2}[v_n]^2 + \frac{v_{n0}\,[P] + P_1[v_n]}{\rho_{m0}(D - v_{n0})} \\ &= \frac{1}{2}[P][\mathcal{V}] - P_1[\mathcal{V}] = -\frac{1}{2}(P_0 + P_1)\,[\mathcal{V}].\end{aligned} \tag{7.160}$$

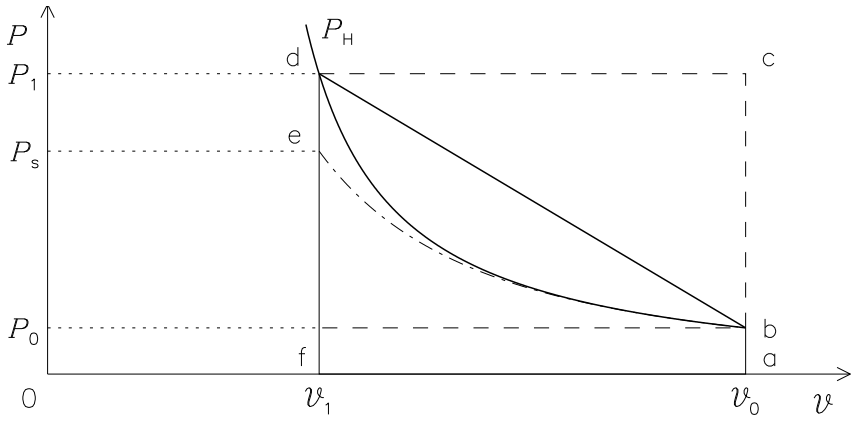

Figure 7.17: Adiabat and Hugoniot in the $P\mathcal{V}$-plane for an ideal gas

Let us now consider the case of a simple shock in an ideal gas. From (7.157) we notice that there is a linear relationship between P_1 and $\mathcal{V}_1$ for given values of P_0 and $\mathcal{V}_0$. For the ideal gas the internal energy can be expressed as $u = P\mathcal{V}/(\gamma - 1)$, valid on both sides of the shock front. Substituting this relationship for u_0 and u_1 in (7.160) gives

$$[u] = \frac{1}{\gamma - 1}(P_1\mathcal{V}_1 - P_0\mathcal{V}_0) = \frac{1}{2}(P_1 + P_0)(\mathcal{V}_0 - \mathcal{V}_1),$$

or

$$S \equiv \frac{P_1}{P_0} = \frac{(\gamma + 1)\mathcal{V}_0 - (\gamma - 1)\mathcal{V}_1}{(\gamma + 1)\mathcal{V}_1 - (\gamma - 1)\mathcal{V}_0}. \tag{7.161}$$

The quantity S is the *shock strength*. The relation (7.161) is the *Hugoniot* in the $P\mathcal{V}$-plane. It is plotted in figure 7.17 (solid curved line bd) together with the relation (7.157) (solid straight

line bd). In the figure the adiabat, that is the $P, \mathcal{V}$ states reachable from $P_0, \mathcal{V}_0$ through an (isentropic) adiabatic compression, is also plotted (dash-dotted curve).

Figure 7.17 allows for a simple physical interpretation. For simplicity let the gas ahead of the shock be at rest, $\boldsymbol{v}_0 = 0$. The specific kinetic energy acquired in the shock compression is then according to (7.158) given by

$$\frac{1}{2}v_1^2 = \frac{1}{2}(P_1 - P_0)(\mathcal{V}_0 - \mathcal{V}_1)$$

and corresponds to the triangular area bcd in the figure.

The corresponding total specific energy acquired is

$$\frac{1}{2}v_1^2 + [u] = \frac{1}{2}(P_1 - P_0)(\mathcal{V}_0 - \mathcal{V}_1) + \frac{1}{2}(P_1 + P_0)(\mathcal{V}_0 - \mathcal{V}_1) = P_1(\mathcal{V}_0 - \mathcal{V}_1).$$

Geometrically, this is represented by the rectangular area acdf. Thus, the area of the polygon abdf equals the increase in the specific internal energy of the gas during the shock compression.

Now consider the isentropic compression from state $P_0, \mathcal{V}_0$ to state $P_s, \mathcal{V}_1$. The work done in this process equals the area abef. To reach the final shocked state d the material must be heated. The necessary heat needed equals the area bde. This area also determine the entropy increase of the material due to the shock compression.

For a conducting fluid described by the ideal MHD equations some additional terms will appear in the conservation laws and the corresponding Rankine-Hugoniot relations will take a slightly more complex form. Let us here consider the simpler hydrodynamic model with an ideal gas equation of state.

Quiz 7.40 : Show that (7.150) follows from (7.146) by integrating over the coin-shaped volume indicated in figure 7.16 and letting $h \to 0$.

Quiz 7.41 : With $[Q] \equiv Q_1 - Q_0$ show that

$$[AB] = A_0\,[B] + B_1\,[A] \tag{7.162}$$

$$[A^2] = 2A_0\,[A] + [A]^2. \tag{7.163}$$

Quiz 7.42 : With $\boldsymbol{v}_0 = 0$ show that

$$\frac{\rho_{m1}}{\rho_{m0}} = \frac{S(\gamma+1) + (\gamma-1)}{S(\gamma-1) + (\gamma+1)} \tag{7.164}$$

$$v_{n1} = (S-1)\sqrt{\frac{2P_0\mathcal{V}_0}{S(\gamma+1) + (\gamma-1)}} \tag{7.165}$$

$$D = \sqrt{\frac{P_0\mathcal{V}_0}{2}(S(\gamma+1) + (\gamma-1))} \tag{7.166}$$

$$c_1 = c_0\sqrt{\frac{S(S(\gamma-1) + (\gamma+1))}{S(\gamma+1) + (\gamma-1)}} \tag{7.167}$$

where S is the shock strength and $c = \sqrt{\gamma P\mathcal{V}}$ is the sound speed. What is the maximum compression ratio ρ_{m1}/ρ_{m0} attainable in any simple shock?

7.18 Characteristic Numbers

The hydrodynamic or magneto-hydrodynamic set of equations are from a mathematical point of view rather complicated models. There is an eminent need to work with simplified models. This is also often justified. The equations may in given situations contain terms which remain small, but even so complicates an analysis considerably. What is therefore needed is a simple order-of-magnitude method for judging the importance of the different terms in the equations. To this end we introduce the notion of *characteristic scales* and *characteristic numbers*.

A characteristic length L for a variable $\phi(\boldsymbol{r}, t)$ is a distance over which ϕ varies significantly. More precisely, we define the scale length as a typical value of $\mid \phi/\nabla\phi \mid$ and express this statement mathematically as

$$L \sim |\, \phi \,/\, \nabla\phi \,| \,. \tag{7.168}$$

The characteristic length scales for two different variables may generally be different. A characteristic time scale τ is defined similarly,

$$\tau \sim |\, \phi \,/\, \partial\phi/\partial t \,| \,. \tag{7.169}$$

A number of characteristic numbers may be formed as dimensionless ratios of different terms of the model equations. In table 7.2 the more common characteristic numbers have been listed. In the right hand expressions for some of these numbers the characteristic scale lengths for different quantities have been assumed equal.

The first five numbers relate to the properties of the momentum equation. The Reynolds number represents the ratio of the spatial, non-linear part of the inertial term and the viscous forces. The Alfvén number is the ratio of the magnetic force and the non-linear part of the inertial term. This number reduces to the square of the ratio of the Alfvén and the flow speed. The Froude number is the ratio of the non-linear part of the inertial term and the gravitational force, while the Strouhal number is the ratio of the temporal and the spatial part of the inertial term.

The magnetic Reynolds number determines whether the magnetic field will be frozen to or diffuse through the conducting fluid.

The Prantl number is the ratio of the relative rate at which macroscopic kinetic energy is transformed into internal energy in the fluid and the rate at which internal energy is transported within the fluid. The explicit expression for $\mathcal{D}$ is given in (7.159) and C_V is the specific heat capacity at constant volume defined in (6.66).

The Richardson number applies for stratified, inhomogeneous flows and measures the ratio of the gravitational stability of the fluid to the free energy in the fluid associated with sheared flow available to drive the Kelvin-Helmholtz instability.

Reynolds number:

$$Re \sim \frac{|\rho_m \boldsymbol{v} \cdot \nabla \boldsymbol{v}|}{|\eta \nabla^2 \boldsymbol{v}|} \sim \frac{\rho_m v L}{\eta}$$

Alfvén number:

$$A \sim \frac{|\boldsymbol{j} \times \boldsymbol{B}|}{|\rho_m \boldsymbol{v} \cdot \nabla \boldsymbol{v}|} \sim \frac{B^2}{\mu_0 \rho_m v^2}$$

Hartmann number:

$$H \sim \frac{|\boldsymbol{j} \times \boldsymbol{B}|}{|\eta \nabla^2 \boldsymbol{v}|} \sim \frac{B^2 L}{\mu_0 \eta v}$$

Froude number:

$$Fr \sim \frac{|\rho_m \boldsymbol{v} \cdot \nabla \boldsymbol{v}|}{|\rho_m \boldsymbol{g}|} \sim \frac{v^2}{Lg}$$

Strouhal number:

$$Sh \sim \frac{|\rho_m \partial \boldsymbol{v}/\partial t|}{|\rho_m \boldsymbol{v} \cdot \nabla \boldsymbol{v}|} \sim \frac{L}{vT}$$

Magnetic Reynolds number:

$$Rm \sim \frac{|\nabla \times (\boldsymbol{v} \times \boldsymbol{B})|}{|\nabla \times (\boldsymbol{j}/\sigma))|} \sim \mu_0 \sigma v L$$

Prantl number:

$$Pr \sim \frac{|\mathcal{D}/(\rho_m \boldsymbol{v}^2/2)|}{|\nabla \cdot (\nabla \mathcal{T})/u|} \sim \eta C_V$$

Richardson number:

$$Ri \sim \frac{|g\, \partial \rho_m/\partial z|}{|\rho_m (\partial U/\partial z)^2|} \sim \frac{gL}{v^2}$$

Table 7.2: Common characteristic numbers

Chapter 8

Radiation Transport

In chapter 1 basic definitions of quantities needed for a discussion of radiation transport through matter were introduced. This included in particular the definition of the specific intensity $\mathcal{I}_\omega$ (1.72) and the refractive indix $\mathcal{N}$ (1.84) or more generally the ray refractive index $\mathcal{N}_r$ (1.118). We also formulated the radiation transport equation (1.124) in terms of macroscopic phenomenological emissivity j_ω and extinction coefficient α_ω. In chapter 6 we investigated properties of the specific intensity for systems in thermal equilibrium (TE). In chapters 3, 4 and 5 the interaction of a radiation field and individual atoms or molecules were studied from a quantum mechanical point of view (bound-bound and bound-free transitions), while interactions between a radiation field and free charged particles (free-free transitions) were treated classically in chapter 2. It is now time to tie these individual pieces together.

8.1 Emissivity and Extinction Coefficient

Thermal equilibrium conditions, implying requirements of homogeneity, isotropy and stationarity, are not strictly met with in astrophysics. Strict thermal equilibrium between matter and radiation field requires that the net radiation flux vanishes, a requirement not satisfied in any star where significant radiation energy is transported out from the fusion burning central part. Even so, thermal equilibrium requirements often form a useful reference basis. Let us therefore first discuss implications of thermal equilibrium requirements for the emissivity and the extinction coefficient.

8.1.1 Radiation and Matter in Thermal Equilibrium

In chapter 6 it was shown (6.114) that the ratio of the spectral intensity $\mathcal{I}_\omega$ for any given mode (polarization) of the radiation field and the square of the ray refractive index $\mathcal{N}_r$ for the given mode in the medium is given by the Planck radiation function B_ω,

$$\frac{\mathcal{I}_\omega}{\mathcal{N}_r^2} = B_\omega(\mathcal{T}) \equiv \frac{\hbar}{(2\pi)^3 c^2} \frac{\omega^3}{\exp(\frac{\hbar\omega}{\mathcal{T}}) - 1}.$$

Under TE conditions the specific intensity for any given frequency only depends on the temperature and, for instance, not on position s along any given ray. The radiative transfer equation (1.124) therefore reduces to

$$\mathcal{N}_r^2 \frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{\mathcal{I}_\omega}{\mathcal{N}_r^2}\right) = j_\omega - \alpha_\omega \mathcal{I}_\omega = 0,$$

or

$$\frac{j_\omega^{\rm TE}}{\mathcal{N}_r^2\alpha_\omega^{\rm TE}} \equiv S_\omega^{\rm TE} = B_\omega(\mathcal{T}).$$

For TE conditions there thus exists a simple functional relationship between the emissivity j_ω and the extinction coefficient α_ω,

$$j_\omega^{\rm TE} = \mathcal{N}_r^2\alpha_\omega^{\rm TE}\, B_\omega(\mathcal{T}), \tag{8.1}$$

valid for any frequency ω, be it at a line center or in the continuous part of the spectrum.

We remember again that this expression may be simplified. In an isotropic medium the ray refractive index $\mathcal{N}_r$ reduces to the refractive index $\mathcal{N}$, while for frequencies $\omega \gg \max(\omega_p, \omega_b)$, where ω_p and ω_b are the electron plasma and gyro frequencies for the medium, both indices approach unity, $(\mathcal{N}_r, \mathcal{N}) \to 1$.

8.1.2 Local Thermal Equilibrium (LTE)

For TE it is required that matter and radiation field are at mutual equilibrium under one constant temperature. A necessary consequence is that the radiation field can not be responsible for any net energy transport in the medium. For an astrophysical setting this is an unacceptable constraint. A less strict requirement is to assume that matter through the action of collisions between the individual particles of the medium is maintained at equilibrium conditions corresponding to a local, but slowly varying temperature $\mathcal{T}(\boldsymbol{r})$. This assumption will be referred to as *local thermal equilibrium* (LTE) and implies that the Maxwell, Boltzmann and Saha equations (6.35), (6.21) and (6.89), evaluated at the local kinetic temperature, are still assumed to be good approximations to velocity, excitation and ionization distributions in the medium. It seems natural that for this assumption to be valid the mean free collision lengths for particles in the medium must be short compared to the typical scale lengths for temperature variations. This means that individual electrons, ions and atoms through the action of mutual collisions remain spatially localized and thus may also be allowed to establish equilibrium distributions corresponding to the local temperature.

The radiative energy distribution under LTE is, however, allowed to deviate from TE conditions. Even with particle mean free paths short compared with typical scale lengths of the medium, the specific intensity will be allowed deviate significantly from isotropy. This will in turn allow for net radiative energy flux in stellar interiors. Near the stellar surface the deviations from isotropy will be more severe, the specific intensity taking a near hemi-spherical form with essentially negligible intensity directed back into the star. In these regions the LTE assumption may be questionable. This leads to the need to look for non-LTE solutions where the state of matter and radiation must be solved for in a mutually consistent manner. A discussion of these topics is clearly outside our present scope.

With these introductory remarks let us then approach the question of giving physical content to the emissivity and extinction coefficient.

8.1.3 Contributions to Emissivity and Extinction Coefficient

The contributions from microscopic effects to the extinction coefficient may conveniently be divided into two different types, contributions of destructive type and contributions of scattering type. In the first case photons are removed from the beam by being absorbed by atoms, later possibly to be reemitted in arbitrary directions. In the latter case individual

photons are lost by being scattered out of the beam. The two types of contributions will be denoted with superscripts a and s respectively, $\alpha_\omega = \alpha^a_\omega + \alpha^s_\omega$. The similar division will be made for the emissivity, $j_\omega = j^a_\omega + j^s_\omega$. In figure 8.1 the individual processes contributing to α^a_ω, α^s_ω, j^a_ω and j^s_ω are illustrated schematically. Note in particular that stimulated emission is classified as an absorption process even if this process actually contributes to an increase in the specific intensity of the beam.

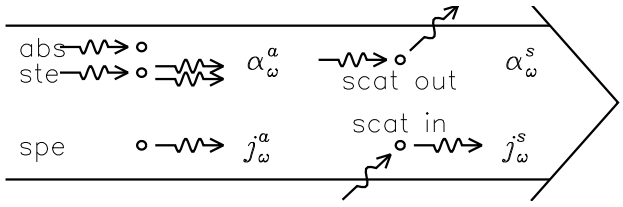

Figure 8.1: Contributors to a- and s-processes

In section 3.9 transition rates for the microscopic radiative absorption and emission processes between a radiation field and a single atom was derived. The absorption and stimulated emission transition rates, (3.97) and (3.98), both depend on the specific intensity of the radiation field. The form of the loss term in the radiative transport equation $\alpha_\omega \mathcal{I}_\omega$, is therefore well suited to incorporate the combined effect of these two processes. What is needed is to multiply the expressions for the individual transition rates with the number of atoms per unit volume available in the relevant atomic states, the number of atoms in the lower energy state n_l for the absorption process, the number of atoms in the upper energy state n_u for the stimulated emission process. In the LTE approximation these quantities are determined through the Boltzmann distribution (6.21). The expressions for the transition rates (3.97) and (3.98) both refer to the total transition rate within a given spectral line. To incorporate also the different line broadening effects (natural, Doppler, Stark, pressure) it is also necessary to include an additional line profile factor $\psi(\omega - \omega_0)$ where ω_0 indicates the line center. The line profile factor will be normalized, $\int \psi(\omega - \omega_0)\,\mathrm{d}\omega = 1$. The contribution to the loss term in (1.128) from a given bound-bound transition is thus

$$\alpha^a_\omega \mathcal{I}_\omega = \hbar\omega\,(n_l\, w^{\mathrm{abs}}_{ul} - n_u\, w^{\mathrm{st.e}}_{lu})\,\psi(\omega - \omega_0) = \hbar\omega\, n_l\, w^{\mathrm{abs}}_{ul}\,\psi(\omega - \omega_0)\,(1 - \exp(-\hbar\omega/\mathcal{T})).$$

The corresponding spontaneous emission rate (3.102) similarly contributes to the emissivity term,

$$j^a_\omega = \hbar\omega\, n_u\, w^{\mathrm{sp.e}}_{lu}\,\psi(\omega - \omega_0).$$

Here the factor $\hbar\omega$ is the energy supplied to the radiation field from each spontaneous transition. In these expressions the line profile factor was assumed to be identical for all three contributions to α^a_ω and j^a_ω. This is referred to as the *complete redistribution* assumption.

The bound-bound transitions contribute to more or less narrow spectral line features. Radiative bound-free transitions represent another type of contribution to the emissivity and extinction coefficient. These processes include photo-ionization and photo-recombination, the former being a process where a photon removes an electron in a certain energy state from the atom. These processes have a frequency dependence distinctly different from bound-bound processes. For the ionization a minimum threshold frequency is required. There is, however,

no sharp upper limit for the photon frequency, excess energy will appear as kinetic energy for the ejected electron. Typically it is found that the transition rate for bound-free transitions decreases as ω^{-3} for frequencies above the threshold value. A large number of such bound-free transitions will contribute to the continuous emissivity and extinction coefficient in stellar atmospheres. For contribution from H I we thus talk about the Lyman jump, the Balmer jump and so on, depending on from what energy level the ionization takes place. Bound-free transition in the negative hydrogen ion H^- is a significant contributor to the continuous extinction coefficients in the visual and infrared spectral ranges in stellar atmospheres with negligible hydrogen ionization.

For the processes discussed above the contributions to the extinction coefficient and emissivity are intimately related to the properties of matter through the distributions of velocities, excitation and ionization. In the LTE approximation we therefore expect the corresponding source function to be well approximated by the local Planck function,

$$S_\omega^a = \frac{j_\omega^a}{\mathcal{N}_r^2 \alpha_\omega^a} \approx B_\omega(\mathcal{T}(\boldsymbol{r})). \tag{8.2}$$

This will not be so for scattering contributions.

One of the most important scattering processes and the only one to be mentioned here is Thomson scattering. This is a free-free radiative process in which electrons are accelerated by the electric fields of the incident radiation. The accelerated electrons subsequently act as sources of radiation in accordance with the discussion of chapter 2, drawing energy from the incident radiation field. The contribution from Thomson scattering to the extinction coefficient is independent of frequency and will be the dominating source of continuous extinction in the atmospheres of hot stars where hydrogen is ionized.

A photon that is scattered out of a given beam is not lost. It will reappear, if we disregard Doppler effects, with the same frequency but now traveling in a different direction. And photons scattered out of other beams may be redirected into the beam considered. The photon supply available for the scattering emissivity is the local mean intensity $\mathcal{J}_\omega$. We thus expect a relation between the scattering parts of the emissivity and extinction coefficient of the form

$$j_\omega^s = \mathcal{N}_r^2 \alpha_\omega^s \mathcal{J}_\omega. \tag{8.3}$$

With these approximations the combined source function can now be written in the form

$$S_\omega = \frac{j_\omega^a + j_\omega^s}{\mathcal{N}_r^2(\alpha_\omega^a + \alpha_\omega^s)} = (1 - \epsilon_\omega)\mathcal{J}_\omega + \epsilon_\omega B_\omega, \tag{8.4}$$

where we introduced the *photon destruction probability per extinction*

$$\epsilon_\omega \equiv \frac{\alpha_\omega^a}{\alpha_\omega^a + \alpha_\omega^s}. \tag{8.5}$$

We notice that ϵ_ω will take values in the range $(0, 1)$. The presence of the scattering contribution in (8.4) makes an explicit solution of the radiative transport equation (1.128) more complicated since the source function is now a functional of the specific intensity $\mathcal{I}_\omega$ itself, through $\mathcal{J}_\omega = \int \mathcal{I}_\omega \, \mathrm{d}^2\Omega/(4\pi)$. We note that simplifications result if the destructive processes dominate, $\epsilon_\omega \approx 1$, or the mean intensity is maintained close to the Planck function, $\mathcal{J}_\omega \approx B_\omega$.

With this introduction to the connections between microscopic processes and the emissivity and extinction coefficients in mind, let us return to study additional consequences of the radiative transport equation (1.128).

Quiz 8.1 : How do you transform from S_ω, τ_ω, j_ω and α_ω to S_ν, τ_ν, j_ν and α_ν?

8.2 Plane-parallel Medium

In realistic situations the assumption of a homogeneous medium must often be abandoned. It is therefore useful next to consider an inhomogeneous, but plane-parallel medium, where $\mathcal{N}_r^2$, j_ω and α_ω are only allowed to vary in one direction, the z-direction. This model will be relevant for radiative transport in stars and stellar atmospheres given that the linear extinction coefficient is large enough that optical thickness is acquired over distances much shorter than stellar dimensions. For this model it is customary to perform the analysis in terms of optical depth measured along a given ray.

Optical depth τ_ω was defined in (1.127). Optical depth is measured from an observer, normally located outside the plane-parallel medium itself and looking *backward* along an emerging ray. In particular, the *radial optical depth* $\tau'_\omega(z)$ refers to a ray directed along the positive z-axis. The geometry is illustrated in figure 8.2. The radial optical depth at a certain level z in the medium is defined as

$$\tau'_\omega(z) \equiv \int_z^\infty \alpha_\omega(z')\,\mathrm{d}z'.$$

In this expression we have assumed the observer to be located outside the medium where $\alpha_\omega = 0$ and therefore that we are allowed to put the upper integration limit to infinity. The radial optical depth is related to the *optical depth* $\tau_{\omega\mu}(z)$ for an oblique ray, propagating at an angle θ with the z-axis. Due to the longer path length in the medium the optical depth at level z for the oblique ray is given by

$$\tau_{\omega\mu}(z) = \tau'_\omega(z)/\mu \tag{8.6}$$

where $\mu = \cos\theta$. For simplicity we have here assume the refractive index to be unity and therefore that ray paths are straight lines!

For the following discussion we shall make a number of simplifying assumptions. We have already assumed that the refractive index is close to unity and therefore that each individual ray travel along straight lines. We further assume that the intensity distribution is independent of the azimuthal angle φ, that is, the radiation field will be assumed to exhibit axisymmetry around the z-axis. The source function at a given z will formally be considered a function of z through the radial optical depth $\tau'_\omega(z)$. We shall furthermore find it useful to distinguish the spectral intensity for rays propagating with a component of the wave vector along the positive z-axis from that of rays propagating with a component of the wave vector along the negative z-axis,

$$\mathcal{I}_\omega(\tau'_\omega(z),\mu) = \begin{cases} \mathcal{I}^+_\omega(\tau'_\omega(z),\mu) & \text{for } \mu > 0 \\ \mathcal{I}^-_\omega(\tau'_\omega(z),\mu) & \text{for } \mu < 0. \end{cases}$$

As boundary conditions we assume that $\mathcal{I}^-_\omega(\tau'_\omega = 0,\mu) = 0$ and that $\mathcal{I}^+_\omega(t,\mu)\exp(-(t - \tau'_\omega(z))/\mu) \to 0$ as $t \to \infty$. The boundary conditions imply that there is no radiation falling on

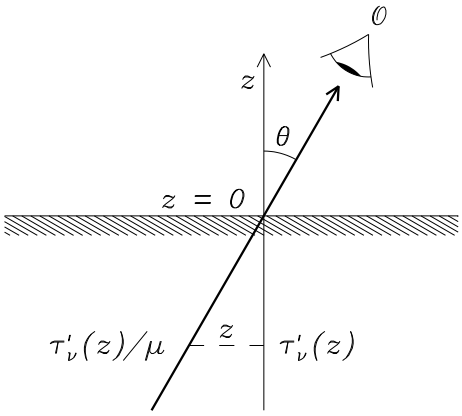

Figure 8.2: Geometry for straight rays in plane-parallel medium

the medium from the outside along the negative z-axis, while the medium is optically thick such that no radiation is "shining through" in the positive z-direction.

With these assumptions the formal solution (1.129) reduces to

$$\mathcal{I}_\omega^+(\tau'_\omega(z),\mu) = \int_{\tau'_\omega(z)}^{\infty} S_\omega(t')\,\exp(-\frac{t'-\tau'_\omega}{\mu})\,\frac{\mathrm{d}t'}{\mu} \quad \text{for} \quad \mu > 0 \tag{8.7}$$

$$\mathcal{I}_\omega^-(\tau'_\omega(z),\mu) = \int_0^{\tau'_\omega(z)} S_\omega(t')\,\exp(-\frac{t'-\tau'_\omega}{\mu})\,\frac{\mathrm{d}t'}{|\,\mu\,|} \quad \text{for} \quad \mu < 0. \tag{8.8}$$

The plane-parallel medium solutions (8.7) and (8.8) are only formal solutions. There is still an integration over the source function to be performed. In particular, we notice that a source function being a functional of the desired solution itself represents a mathematical challenge, instead of evaluating an integral for a known integrand, it is a question of solving an integral equation.

8.2.1 Interior approximation

Deep inside the plane-parallel medium the mean free path lengths of photons are expected to be short compared to the typical scale lengths for variations in the source function. In this case the source function in the remaining integral may be expanded in a Taylor series

$$S_\omega(t') = S_\omega(\tau'_\omega(z)) + (t'-\tau'_\omega(z))\frac{\mathrm{d}S_\omega(\tau'_\omega(z))}{\mathrm{d}\tau'_\omega} + \cdots .$$

In the deep interior the short mean free collision lengths will maintain LTE conditions. The source function will therefore be well represented with the Planck function, $S_\omega = B_\omega$. The integrals may now be performed with the result

$$\mathcal{I}_\omega(\tau'_\omega(z),\mu) = B_\omega(\tau'_\omega(z)) + \mu\frac{\mathrm{d}B_\omega(\tau'_\omega(z))}{\mathrm{d}\tau'_\omega} + \cdots \tag{8.9}$$

valid for all directions $-1 \leq \mu \leq 1$. The corresponding mean intensity and net radiative energy flux are

$$\mathcal{J}_\omega(\tau'_\omega(z)) = B_\omega(\tau'_\omega(z)) \tag{8.10}$$

$$\mathcal{F}_\omega(\tau'_\omega(z)) = \frac{4\pi}{3}\frac{\mathrm{d}B_\omega(\tau'_\omega(z))}{\mathrm{d}\tau'_\omega}. \tag{8.11}$$

The total net radiative energy flux, found by integrating over all frequencies, is conveniently written in terms of the *Rosseland mean extinction coefficient* α_R. For this we note that

$$\begin{aligned} F(z) &\equiv \int_0^\infty \mathcal{F}_\omega(z)\,\mathrm{d}\omega \\ &\approx -\frac{4\pi}{3}\int_0^\infty \frac{1}{\alpha_\omega}\frac{\mathrm{d}B_\omega}{\mathrm{d}T}\frac{\mathrm{d}T}{\mathrm{d}z}\,\mathrm{d}\omega \\ &= -\frac{16\sigma T^3}{3\alpha_R}\frac{\mathrm{d}T}{\mathrm{d}z}, \end{aligned} \tag{8.12}$$

where we defined

$$\frac{1}{\alpha_R} \equiv \frac{\displaystyle\int_0^\infty \frac{1}{\alpha_\omega}\frac{\mathrm{d}B_\omega}{\mathrm{d}\mathcal{T}}\,\mathrm{d}\omega}{\displaystyle\int_0^\infty \frac{\mathrm{d}B_\omega}{\mathrm{d}\mathcal{T}}\,\mathrm{d}\omega}. \tag{8.13}$$

We see that the inverse Rosseland mean extinction coefficient is the frequency average of the inverse extinction coefficient with the normalized $\mathrm{d}B_\omega/\mathrm{d}\mathcal{T}$ as weighting function. The inverse weighting means that spectral intervals with small extinction play important roles in the total energy transport in the medium, acting as efficient radiation energy drainage channels. The result (8.12) takes the form of a "heat conduction" equation with a temperature dependent "heat conduction coefficient". The result demonstrates the importance of small deviations from isotropy of the radiation field for the energy transport from the star interiors to their surface layers from where radiation may escape into space.

8.2.2 Surface approximation

Near the surface of the plane-parallel medium a similar power series expansion of the source function is convenient (Taylor series around $\tau'_\omega = 0$),

$$S_\omega(\tau'_\omega) = a_0 + a_1\tau'_\omega + a_2\tau'^2_\omega + \cdots . \tag{8.14}$$

For an observer located at $\tau'_\omega = 0$, the emerging specific intensity is then

$$\mathcal{I}^+_\omega(\tau'_\omega = 0, \mu) = a_0 + a_1\mu + 2a_2\mu^2 + \cdots \approx S_\omega(\tau'_\omega = \mu). \tag{8.15}$$

The latter approximation is known as the *Eddington-Barbier approximation* . The approximation is exact if $S_\omega(\tau'_\omega)$ varies linearly with τ'_ω. According to this approximation the observed intensity in the direction μ equals the source function of the medium at a radial optical depth equal to $\tau'_\omega = \mu$.

The result (8.15) has a natural physical interpretation. At a radial optical depth $\tau'_\omega = 1$ the mean free optical length is equal to the remaining optical distance out of the plane-parallel medium for a radially directed photon. Thus a photon traveling radially outward at

this depth has a high probability of escaping the medium. For obliquely traveling photons the corresponding optical depth is $\tau'_{\omega\mu} = \mu^{-1}$.

The corresponding expressions for mean intensity and unidirectional radiative energy flux according to the Eddington-Barbier approximation are

$$\mathcal{J}_\omega(\tau'_\omega = 0) = \frac{1}{2}(a_0 + \frac{1}{2}a_1 + \cdots) \approx \frac{1}{2}S_\omega(\tau'_\omega = \frac{1}{2}) \tag{8.16}$$

$$\mathcal{F}^+_\omega(\tau'_\omega = 0) = \pi(a_0 + \frac{2}{3}a_1 + \cdots) \approx \pi S_\omega(\tau'_\omega = \frac{2}{3}). \tag{8.17}$$

For these expressions to be valid we assumed the plane-parallel medium to be laterally extended, that is, that the μ-integrals are to be taken over the interval (0,1). This is equivalent to an assumption that the observer is located close enough to the local plane-parallel medium. The opposite limit, where the observer is located far from the locally plane-parallel medium is discussed in quiz 8.5

Quiz 8.2 : Verify (8.12), in particular, show that

$$\int_0^\infty \frac{\mathrm{d}B_\omega}{\mathrm{d}T}\,\mathrm{d}\omega = \frac{8\pi^4\kappa^4T^3}{15c^2h^3} = \frac{4}{\pi}\sigma T^3.$$

Quiz 8.3 : To realize the importance of the inverse weighting procedure in the definition of the Rosseland mean extinction coefficient (8.13), we consider the simplified model where $\mathrm{d}B_\omega/\mathrm{d}T$ is constant over a finite frequency range (and vanishes outside). Let $\alpha_\omega = 1$ over 90 % of this frequency range and $\alpha_\omega = .1$ over the remaining range. What is the value of α_R? Discuss.

Quiz 8.4 : Verify (8.16) and (8.17).

Quiz 8.5 : Show that the unidirectional flux from a distant star, radius R at distance $r \gg R$ in the Eddington-Barbier approximation is given as

$$\mathcal{F}^+_\omega = \frac{\pi R^2}{r^2}S_\omega(\tau' = \frac{2}{3}). \tag{8.18}$$

What is the corresponding expression for the mean intensity $\mathcal{J}_\omega$? [*Hint:* Note that $r\sin\psi \approx R\sin\theta$ where ψ is the arrival angle of a ray that departs from the star under an angle θ relative to the local surface normal.]

8.3 Line Formation for Optically Thick Medium

In our discussion so far we have considered a monochromatic case. Let us now discuss the formation of spectral lines as a function of frequency in an optically thick, plane-parallel medium. We assume the extinction coefficient α_ω, plotted in the upper left part of figure 8.3, to be a slowly varying function of frequency as a result of continuous processes in the medium, but superimposed a narrow structure due to one bound-bound transition. The value near the line centre α^ℓ_ω may easily exceed the corresponding continuum value α^c_ω by a factor 10-1000. In addition, α_ω may be a rapidly varying function of altitude. For simplicity we will, however, here assume that the extinction coefficient does not vary with position in the

medium. With this assumption the radial optical depth $\tau'_\omega(z)$ as a function of position z will be a straight line for each given frequency, but different lines for different frequencies. In the upper right part of figure 8.3 the results for two particular frequencies, one at the center of the line ω_0, and one just to the right of the line ω_1 is given. In the lower right part of figure 8.3 the source function $S_\omega(z)$ as a function of position is given. For simplicity we assume that the source function is independent of frequency, at least within the small frequency interval of interest. We have furthermore assumed the source function $S_\omega(\mathcal{T}(z))$ to be increasing linearly with depth $(-z)$ in the medium. This means that the Eddington-Barbier approximation is exact for the case considered. The particular shape of the source function is the result of an increasing temperature profile with depth in the medium as illustrated by the dash-dotted curve in the the lower right panel.

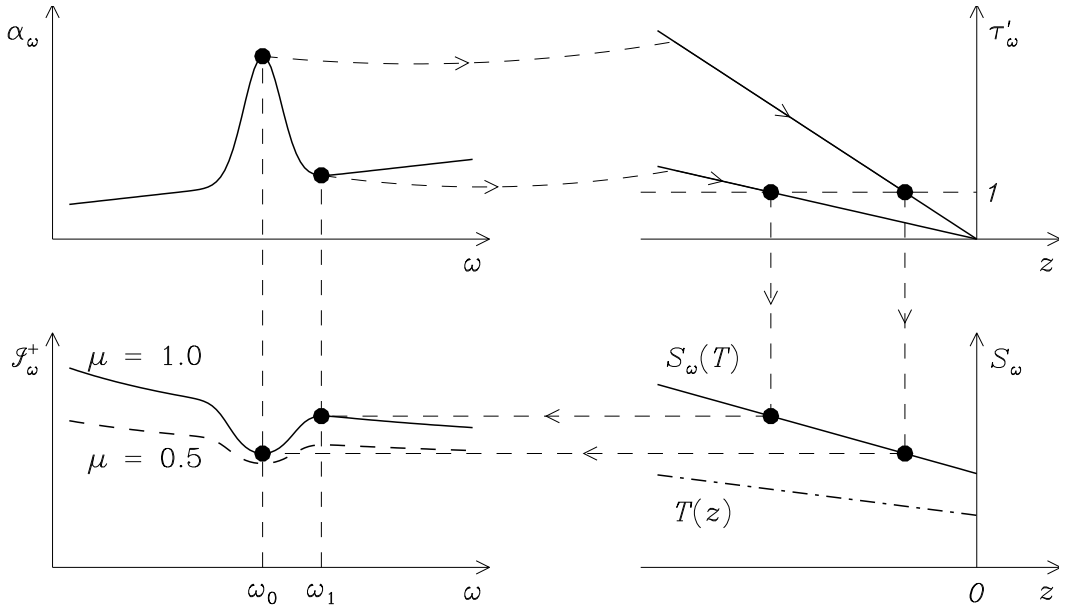

Figure 8.3: Line formation in optically thick, plane-parallel medium

In the lower left part of figure 8.3 the resulting spectral intensities of the radiation from the medium for normal ($\mu = 1$, full line) and oblique propagation ($\mu = .5$, dashed line) are given. The derivation is illustrated graphically in the figure for two particular frequencies, ω_0 and ω_1, and for the case of normal propagation. For a given frequency ω the value of α_ω determines the proper radial optical depth curve $\tau'(z)$. The position z where this curve takes the unity value will determine the relevant value of the source function $S_\omega(z)$ to be substituted in the Eddington-Barbier approximation (8.15) to give the resulting intensity $\mathcal{I}^+_\omega(\omega) = S_\omega(z)$. The same construction applies for the oblique case, the only difference is that the relevant position z for the source function is where the radial optical depth curve takes the value $\mu = \cos\theta$. The result for $\mu = .5$ is dotted in the figure.

The situation depicted in figure 8.3 represents a rather idealized case, yet it illustrates some general properties of the line formation process. We notice for instance that the wings and the core of the line are formed at different depths in the atmosphere, the core at the higher altitude. This property remains valid even if a z-dependence of α_ω is taken into account, and therefore that the straight lines of the upper right panel are replaced by slightly curved lines.

In the example considered in figure 8.3 the result is an absorption line spectrum. The absorption character is produced by the particular form of the source function chosen, the source function increasing with increasing depth in the medium. For a case where this function decreases with depth, an emission line spectrum would result. With a temperature minimum in the line forming altitude range, as is the case near the photosphere-chromosphere interphase, a line profile with depressed shoulders but enhanced central core can be expected – at least as long as the source function S_ω can be approximated by the Planck radiation function B_ω. With increasing altitude in the stellar atmosphere this assumption may not hold.

Quiz 8.6 : The intensity of the Sun in the visible spectral region decreases from the middle of the solar disk to the limb. What does this tell us about the variation in the source function with height in the solar atmosphere?

Chapter 9

Dimensional analysis

The principle of relativity requires the laws of physics to be form invariant (that is, to take the same appearance) whatever the choice of reference (or inertial) system. This requirement must be satisfied whether we are dealing with Galilean relativity or the special or general relativity according to Einstein.

Dimensional analysis represents an analog to the principle of relativity, requiring the laws of physics also to exhibit form invariance with respect to the choice of system of units. The form of the laws of physics cannot depend on the choice of 1 m as the unit of length instead of for instance 2 m. In this chapter we shall outline the contents of this statement and also demonstrate that dimensional analysis can be a powerful tool in science and technology.

Dimensional analysis can be used in two different ways. The *passive* use consists in controlling that any given relation between physical quantities has correct physical dimensions. This means that any two terms that are to be added or set equal to each other, should have identical physical dimensions. This simple statement is known as the *principle of homogeneity*. This principle represents a powerful method for discovering errors at an early stage in any derivation and constitutes a strongly recommended daily routine for any scientist!

The *active* application of dimensional analysis involves seeking *possible* relations that may exist between physical quantities Q_i, $i = 1, \cdots, q$. This means that dimensional analysis can be used to find possible forms that a physical law *may* take, even in situations where certain aspects of the problem are not yet understood. This will be demonstrated in the following examples. First, however, we need to derive the consequences of the principle of form invariance of the laws of physics with respect to the choice of basic units.

9.1 The Π-Theorem

We are familiar with the fact that every physical quantity Q_i has a certain physical dimension. By this we mean that the given quantity can be written in the form

$$Q_i = \{Q_i\}\,[Q_i] \tag{9.1}$$

where $\{Q_i\}$ is a dimensionless constant or variable and $[Q_i]$ represents the physical dimension of the quantity Q_i. $[A_i]$ can always be expressed in terms of a set of basic units $[A_j]$, $j = 1, \cdots, n$,

$$[Q_i] = [A_1]^{b_{i1}} \cdots [A_n]^{b_{in}}. \tag{9.2}$$

In the MKSA (SI) system the basic units are length $[L]$ measured in meter (m), mass $[M]$ in kilogram (kg), time $[T]$ in second (s) and electric current $[I]$ in ampére (A). All other physical dimensions can be expressed in terms of these four basic units. The physical dimension for force (N) is thus $[M][L][T]^{-2}$. In the following we shall refer to the vector

$$\mathbf{b}_i = [b_{i1}, \cdots, b_{in}]$$

formed from the exponents appearing in (9.2) as the *dimensional vector* of the quantity Q_i. A quantity whose dimensional vector is null will be called *dimensionless*.

A physical law is a relation involving a number of physical quantities, Q_i, $i = 1, \cdots, q$. The relation may be in the form of a function or differential equation. We write this relation formally as

$$F(Q_1, \cdots, Q_q) = 0. \tag{9.3}$$

According to the homogeneity principle this relation must also be expressible in the equivalent form in terms of the corresponding dimensionless quantities $\{Q_i\}, i = 1, \cdots, q$, that is,

$$F(\{Q_1\}, \cdots, \{Q_q\}) = 0. \tag{9.4}$$

This is equivalent to the requirement that the dimensional factors in (9.3) must appear in such a way as to form a common factor – or in subgroups such as to cancel each other. This observation in turn means that (9.3) must also be of such a form that it can be rewritten as a relation

$$G(\Pi_1, \Pi_2, \cdots) = 0 \tag{9.5}$$

between a number of dimensionless products Π_j formed from the original quantities Q_i.

To illustrate these points, consider as an example the mathematical pendulum of length $\ell = \{\ell\}[L]$ in a location with gravitational acceleration $g = \{g\}[L][T]^{-2}$. The differential equation for the angular excursion $\theta = \{\theta\}$ is

$$\ell \frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} = -g \sin\theta \quad \text{or} \quad \{\ell\} \frac{\mathrm{d}^2\{\theta\}}{\mathrm{d}\{t\}^2} = -\{g\} \sin\{\theta\}.$$

The right hand equation follows as $[L][T]^{-2}$ is seen to be a common factor to both sides of the equation of motion. In terms of dimensionless products, the equation can be written in the form

$$\frac{\mathrm{d}^2\theta}{\mathrm{d}\Pi^2} = -\sin\theta \quad \text{with} \quad \Pi = \sqrt{\frac{g}{\ell}}\, t.$$

The oscillation period T of the pendulum can most generally be shown to be given by

$$\frac{g}{\ell} T^2 = 8\pi K(\Theta/2) \quad \text{or} \quad \frac{\{g\}}{\{\ell\}} \{T\}^2 = 8\pi K(\{\Theta\}/2). \tag{9.6}$$

Here Θ is the maximum angular excursion of the pendulum and K is the complete elliptical integral of first order. (For small x we have $K(x) \approx \pi/2$.) Thus, we see that both the differential equation and the expression for the oscillation period T satisfy the principle of homogeneity.

To complete the discussion of dimensional analysis it remains to find the number of dimensional products that (9.5) may contain, together with a general method for deriving the actual

form of these products. The problem of identifying independent dimensionless products is equivalent to finding exponent vectors $\mathbf{z} = [z_1, \cdots, z_q]$ such that

$$[Q_1]^{z_1} \cdots [Q_q]^{z_q} = 1 \quad \text{(dimensionless!)}.$$

Expressing the different $[Q_i]$ in terms of the basic units $[A_j]$, $j = 1, \cdots, n$, this requirement takes the form

$$[A_1]^{z_1 b_{11} + \cdots + z_q b_{q1}} \cdots [A_n]^{z_1 b_{1n} + \cdots + z_q b_{qn}} = 1.$$

This means that finding dimensionless products is equivalent to finding solutions $\mathbf{z}$ of the system of equations

$$z_1 b_{1j} + \cdots + z_q b_{qj} = 0 \quad \text{for} \quad j = 1, \cdots, n$$

or in compact notation

$$\mathbf{z} \cdot \mathbf{B} = \mathbf{0}. \tag{9.7}$$

The matrix

$$\mathbf{B} = \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{q1} & \cdots & b_{qn} \end{bmatrix} \tag{9.8}$$

is called the *dimensional matrix* for the set of physical quantities Q_i, $i = 1, \cdots, q$. It is seen that $\mathbf{B}$ is formed from the dimensional vectors $\boldsymbol{b}_i$ of the set, each vector representing one row of the matrix. The number of independent dimensionless products is equal to the number independent non-trivial solutions of (9.7), namely $q - r$ where $r = \text{rang}(\boldsymbol{B})$. $\text{rang}(\boldsymbol{B})$ indicates the number of the equations (9.7) that are linearly independent. It is equal to the dimension of the largest square sub-matrix with non-vanishing determinant that can be formed from $\boldsymbol{B}$.

The main result of our discussion is summarized in the so-called **Π-theorem**:
Any physical law $F(Q_i, \cdots, Q_q) = 0$ relating physical quantities $Q_1, \cdots, Q_q$ with corresponding dimensional matrix $\mathbf{B}$ of rang r can be written as a relation

$$G(\Pi_1, \cdots, \Pi_{q-r}) = 0 \tag{9.9}$$

between a set of $q - r$ independent dimensionless products $\Pi_1, \cdots, \Pi_{q-r}$ formed from the physical quantities involved.

The Π-theorem leads to the following conclusions. If $q - r = 0$, no dimensionless product can be found. This means that *some* quantity, relevant for the problem studied, is missing from the set $Q_1, \cdots, Q_q$. The list of physical quantities should be extended and the analysis repeated. In particular, the number of physical quantities in any physical law must exceed the number of basic units necessary to express the physical dimensions of the quantities involved.

If $q - r > 0$ and *all relevant quantities are included* among the set $Q_1, \cdots, Q_q$, then the physical law will be writable in the general form (9.9). Two important special cases should be mentioned. *First,* in the case that $q - r = 1$, only one dimensionless product can be formed. The relation (9.9) then reduces to the simple form

$$\Pi_1 = C \tag{9.10}$$

where C is an unknown constant. *Second*, in the case that $q - r = 2$, two independent products can be formed. The relation (9.9) this time can be written in the form

$$\Pi_1 = f(\Pi_2) \tag{9.11}$$

where f is an unknown function.

The value of the constant C, the form of the function f, or in the general case, the function G, cannot be found from the dimensional analysis alone. The role of dimensional analysis and the Π-theorem is to tell how many dimensionless products should be sought for and to limit possible functional relationships that may exist between such products. This is not in itself a small achievement on behalf of dimensional analysis. In an experimental situation there is an enormous difference between determining the one function of one variable in (9.11) instead of start looking for a general functional relationship between a larger number of variables.

For the success of dimensional analysis, it is crucial that the initial set of physical quantities $Q_1, \cdots, Q_q$ includes all quantities relevant for the problem at hand. A too large set of physical quantities leads to unnecessary general forms of (9.9). A missing quantity is sometimes indicated as a problem in finding dimensionless products. An irrelevant quantity substituted for a relevant one will lead to wrong conclusions. Dimensional analysis must therefore be treated with proper care.

Before turning to applications of these results let us verify our introductory statement that dimensional analysis can be viewed as an analog to the principle of relativity, namely that the laws of physics should exhibit form invariance with respect to the choice of system of units. To this end let us now perform an arbitrary stretching of the basic units

$$\overline{[A_j]} = \alpha_j [A_j] \quad \text{for} \quad j = 1, \ldots, n$$

where the stretching parameters α_j can take any positive value. This implies that the value of the dimensionless quantities $\{Q_i\}$ also changes

$$\overline{\{Q_i\}} = \alpha_1^{-b_{i1}} \alpha_2^{-b_{i2}} \cdots \alpha_n^{-b_{in}} \{Q_i\}.$$

If we now introduce the notation

$$\mathbf{x} = \begin{bmatrix} \ln\{Q_1\} \\ \vdots \\ \ln\{Q_q\} \end{bmatrix}, \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} \ln \alpha_1 \\ \vdots \\ \ln \alpha_n \end{bmatrix},$$

the transformation equation for the dimensionless quantities can be expressed in the linear form

$$\overline{\mathbf{x}} = \mathbf{x} - \mathbf{B} \cdot \mathbf{y}.$$

If the physical law (9.4) is to take the same form irrespective of the choice of basic units, this must be so also for our modified basic units. That is, we must require

$$F(\overline{\{Q_1\}}, \cdots, \overline{\{Q_q\}}) = 0. \tag{9.12}$$

Stated differently, the function $\Psi(\mathbf{x}) \equiv F(\{Q_1\}, \cdots, \{Q_q\})$ must satisfy the requirement

$$\Psi(\overline{\mathbf{x}}) = \Psi(\mathbf{x}) \quad \text{for any choice of} \quad \mathbf{y}. \tag{9.13}$$

In the language of linear algebra, this means that $\Psi(\mathbf{x})$ can only depend on the projection $\mathbf{x}_{\mathcal{U}}$ of $\mathbf{x}$ into a subspace $\mathcal{U}$ being perpendicular to the subspace spanned by $\mathbf{B} \cdot \mathbf{y}$ as $\mathbf{y}$ is allowed to take any value. The dimension of the subspace $\mathcal{U}$ is given by $\dim(\mathcal{U}) = q - r$ where $r = \text{rang}(\mathbf{B})$. The subspace $\mathcal{U}$ will therefore be spanned by a set of $q - r$ linearly independent

basic vectors. Each of these basic vectors corresponds to a certain linear combination of $\ln\{Q_1\}, \cdots, \ln\{Q_q\}$,

$$z_1 \ln\{Q_1\} + \cdots + z_q \ln\{Q_q\} = \ln\left(\{Q_1\}^{z_1} \cdots \{Q_q\}^{z_q}\right).$$

But a product of dimensionless quantities $\{Q_i\}$ will be independent of any stretching of basic units only when the product of the corresponding physical quantities Q_i is dimensionless. In this way the present line of reasoning again leads to our previous result as summarized in the Π-theorem.

9.2 Simple Applications

Some examples will demonstrate the strength and the shortcomings of the Π-theorem and the method of identifying dimensionless products.

Example 1: The mathematical pendulum.

Let us first return to the mathematical pendulum example, this time pretending that we known neither the differential equation describing the behavior of the pendulum, nor its solution. We want to find the possible forms that an expression for the oscillation period T may take. Let us assume that T can be expressed in terms of the mass of the pendulum m, the length ℓ, the acceleration of gravity g and the maximum amplitude Θ – or a suitable subset of these quantities.

These $q = 5$ physical quantities are all expressible in terms of $n = 3$ basic units: $[L]$, $[M]$ and $[T]$. The dimensional matrix for the set m, ℓ, g, Θ and T takes the form

$$\mathbf{B} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Since $r = \text{rang}(\mathbf{B}) = 3$, we shall need $q - r = 2$ dimensionless products Π_k. It is immediately seen that m cannot form part of any such product. There is no way in which the dimension of m can be compensated in any product with the other variables in the set. Furthermore, one of our quantities, Θ, is already dimensionless and thus represents by itself a valid dimensionless product. A possible choice of products is therefore

$$\Pi_1 = \frac{T^2 g}{\ell} \quad \text{and} \quad \Pi_2 = \Theta.$$

Assuming that all relevant physical quantities were included in our list, any physically allowed relationship between these quantities must be expressible in the form $\Pi_1 = f(\Pi_2)$, or explicitly

$$T^2 = \frac{\ell}{g} f(\Theta).$$

The result is seen to be in agreement with the mathematical solution of the pendulum problem (9.6). The form of the unknown function $f(\Theta)$ can *not* be determined with the help of

dimensional analysis. From an experimental point of view the important point is that even if the mathematical solution is not available, dimensional analysis tells that an empirical determination of the oscillation period T reduces to the experimental determination of one single function f of the single variable Θ.

In this example dimensionless products could be found by simple inspection. According to the more systematic method, we have to seek exponent vectors $\mathbf{z}$ satisfying the set of equations

$$\begin{aligned} z_1 &= 0 \\ z_2 + z_3 &= 0 \\ -2z_3 + z_5 &= 0. \end{aligned}$$

Because Θ is already dimensionless, z_4 will be subject to no requirements. Two possible exponent vectors are therefore $\boldsymbol{z}_1 = [0, -1, 1, 0, 2]$ and $\boldsymbol{z}_2 = [0, 0, 0, 1, 0]$ as remarked above.

Example 2. Planetary motion.

As a second example, consider a planet of mass M_p orbiting a star with mass M_s. We want to find possible relations between the universal constant of gravitation G, M_p, M_s, the distance R between the two bodies and the orbital period T. Basic units are again $[L]$, $[M]$ and $[T]$, while the dimensional matrix is

$$\mathbf{B} = \begin{bmatrix} 3 & -1 & -2 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Since $r = \text{rang}(\mathbf{B}) = 3$, two independent dimensionless products can be formed, for example

$$\Pi_1 = \frac{T^2 G M_s}{R^3} \quad \text{and} \quad \Pi_2 = \frac{M_p}{M_s}.$$

This means that if all relevant quantities were included, the orbital period shall have to be expressible in the form

$$T^2 = \frac{R^3}{GM_s} f\left(\frac{M_p}{M_s}\right) \tag{9.14}$$

where f is an unknown function. This is as far as dimensional analysis goes. By combining the result from dimensional analysis with obvious symmetry requirements, we may proceed a little further. Thus, even if the mass of the star and the mass of the planet may be vastly different, there are reasons to believe that the two masses should appear in a symmetric manner. We would expect the formula for the orbital period to be valid also in the case that the two masses were equal. A symmetric formula can only be achieved if the unknown function has the particular form $f(x) = C(1 + x)/x$, where C is a constant. The formula for the orbital period now reduces to

$$T^2 = C\frac{R^3}{G\mu} \tag{9.15}$$

where $\mu = M_s M_p/(M_s + M_p)$ is the reduced mass.

Quiz 9.1 : Bolometric luminosity L_* for a star with radius R expresses the energy radiated by the star per unit time. If $L_* R^{-2}$ only depends on the surface temperature $\theta = \kappa T$, the speed of light c and Planck's constant h, what is then the possible relation between these quantities?

Quiz 9.2 : According to the Rydberg formula the energy levels of hydrogen-like atoms are given by

$$W_n = -Rhc\, Z^2/n^2$$

where Z is the charge number of the nucleus and n is the principal quantum number. Make use of dimensional analysis to derive possible expressions for the constant Rc in terms of e, h, the masses M and m of the nucleus and the electron and the vacuum permittivity ϵ_0 .

Quiz 9.3 : It is claimed that a light ray passing a star with mass M at minimum distance r will suffer a deviation d (measured in radians). The situation is illustrated in figure 9.1. Discuss the following two statements:
a) d can be expressed as a function of M, r and G.
b) d can be expressed as a function of M, r, G and c.
Here G is the universal gravitational constant and c is the speed of light. What is your expression for the deviation d?

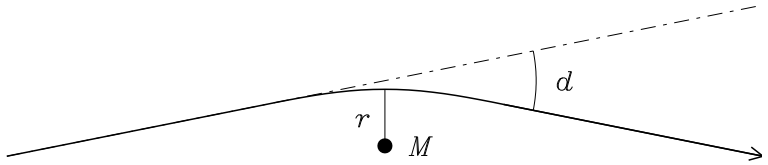

Figure 9.1: Deviation of a light ray by a mass

Appendix A

Vector Calculus and the ∇ operator

One must learn by doing the thing;
for though you think you know it,
you have no certainty until you try.
Sophocles

A large fraction of physical laws involves scalar or vector fields and derivatives of these fields. Spatial derivatives normally appear in the form of the ∇ operator. To understand the contents of these laws or to make use of these laws in solving physical problems, it is necessary to be familiar with basic properties of this operator. In the following review of such properties, we assume the underlying space to be three dimensional and Euclidean, that is, the length of a vector $\boldsymbol{r}$ as expressed in Cartesian components (x, y, z) is given by $|\boldsymbol{r}| = \sqrt{x^2 + y^2 + z^2}$.

A.1 The grad, div and curl Operators

Let $\Phi(\boldsymbol{r})$ be a continuous, differentiable scalar function (scalar field). The equation $\Phi(\boldsymbol{r}) = \Phi_0$, where Φ_0 is a constant, represents a constant Φ surface. The *directional derivative* of Φ at a given point $\boldsymbol{r}$ and along an arbitrary unit vector $\hat{\boldsymbol{\ell}}$ is defined by

$$\frac{\partial \Phi}{\partial \ell} \equiv \lim_{\ell \to 0} \frac{1}{\ell} \left[\Phi(\boldsymbol{r} + \ell\, \hat{\boldsymbol{\ell}}) - \Phi(\boldsymbol{r}) \right]. \tag{A.1}$$

The directional derivative vanishes for directions in the constant Φ surface containing the point $\boldsymbol{r}$. The directional derivative takes its maximum value

$$\frac{\partial \Phi}{\partial n} = \max_{\hat{\boldsymbol{\ell}}} \frac{\partial \Phi}{\partial \ell}$$

for the direction $\hat{\boldsymbol{\ell}} = \hat{\boldsymbol{n}}$ where $\hat{\boldsymbol{n}}$ is the unit normal vector to the constant Φ surface, pointing towards increasing values of $\Phi(\boldsymbol{r})$. The *gradient* of Φ is defined as

$$\text{grad}\, \Phi \equiv \frac{\partial \Phi}{\partial n} \hat{\boldsymbol{n}}. \tag{A.2}$$

The quantities entering the definitions (A.1) and (A.2) are illustrated in figure A.1. The gradient of Φ is itself a vector function, at every point in space where the derivatives of Φ is defined. As a vector the gradient is represented by a magnitude and a direction.

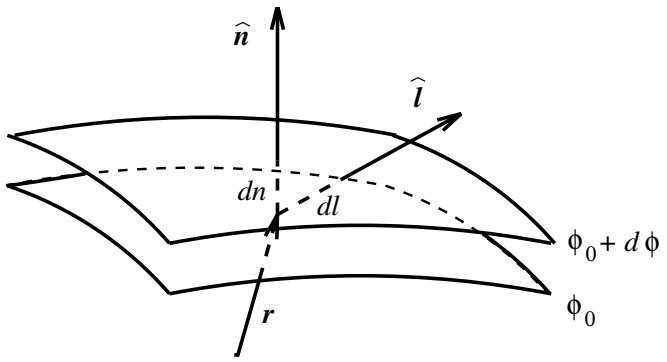

Figure A.1: Directional derivatives and gradients

Next, let $\boldsymbol{F}(\boldsymbol{r})$ be a continuous, differentiable vector function (vector field). The *divergence* of $\boldsymbol{F}$ at a given point $\boldsymbol{r}$ is defined by

$$\operatorname{div} \boldsymbol{F} \equiv \lim_{V \to 0} \frac{1}{V} \oint_{\mathcal{A}} \mathrm{d}^2 \boldsymbol{\mathcal{A}} \cdot \boldsymbol{F}. \tag{A.3}$$

Here $\mathcal{A}$ is the closed surface bounding the volume V containing the point $\boldsymbol{r}$. The infinitesimal surface element $\mathrm{d}^2\boldsymbol{\mathcal{A}}$ is directed out of V. The geometry is illustrated in figure A.2a. By its definition the divergence of $\boldsymbol{F}$ represents the outward directed *flux* $\oint_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}} \cdot \boldsymbol{F}$ of the vector field $\boldsymbol{F}$ per unit volume V. The divergence of $\boldsymbol{F}$ is a scalar function.

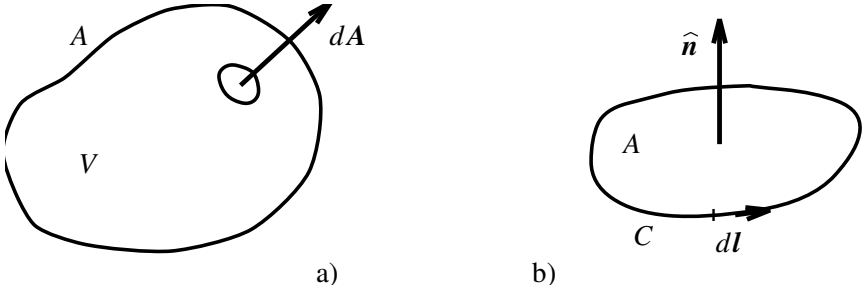

Figure A.2: Geometry of divergence and curl

The *curl* of the continuous, differentiable vector function $\boldsymbol{F}(\boldsymbol{r})$, in contrast, is itself a vector function. The component (projection) of this vector function in an arbitrary direction $\hat{\boldsymbol{n}}$ at a given point $\boldsymbol{r}$ is defined by

$$\hat{\boldsymbol{n}} \cdot \operatorname{curl} \boldsymbol{F} \equiv \lim_{\mathcal{A} \to 0} \frac{1}{\mathcal{A}} \oint_{\mathcal{C}} \mathrm{d}\boldsymbol{\ell} \cdot \boldsymbol{F}. \tag{A.4}$$

Here $\mathcal{C}$ is the directed contour bounding the surface A, containing the point $\boldsymbol{r}$ and with unit normal vector $\hat{\boldsymbol{n}}$. The directions of $\mathcal{C}$ and the surface $\mathcal{A}$ are chosen according to the right-hand rule illustrated in figure A.2b. By its definition $\hat{\boldsymbol{n}} \cdot \operatorname{curl} \boldsymbol{F}$ measures the tendency of the field $\boldsymbol{F}$ to rotate along the contour $\mathcal{C}$. For this reason $\operatorname{curl} \boldsymbol{F}$ is also referred to as the *rotation* of $\boldsymbol{F}$, or as the *circulation* $\oint_{\mathcal{C}} \mathrm{d}\ell \cdot \boldsymbol{F}$ of the vector field $\boldsymbol{F}$ per unit area $\mathcal{A}$.

Quiz A.1 : Verify that

$$\frac{\partial \Phi}{\partial \ell} = \hat{\boldsymbol{\ell}} \cdot \operatorname{grad} \Phi = \frac{\partial \Phi}{\partial n} \hat{\boldsymbol{\ell}} \cdot \hat{\boldsymbol{n}}. \tag{A.5}$$

A.2 Orthogonal Curvilinear Coordinates

A given scalar or vector field may be expressed in any suitable coordinate system. By their definitions the results of applying the grad, divergence or curl operators to the given field must necessarily be independent of the particular coordinate system employed. The explicit expression for these operators, however, *do* depend on the choice of coordinate system.

To derive these expressions it is convenient to introduce a general right-handed orthogonal curvilinear coordinate system with coordinates (q_1, q_2, q_3). These coordinates are associated with corresponding scale factors (Q_1, Q_2, Q_3) and unit vectors $\hat{\boldsymbol{u}}_1$, $\hat{\boldsymbol{u}}_2$ and $\hat{\boldsymbol{u}}_3$. The scale factor Q_i is defined such that $Q_i \mathrm{d}q_i$ represents the length of the infinitesimal translation $q_i \to q_i + \mathrm{d}q_i$, keeping the remaining coordinates constant. The direction of this translation is given by $\hat{\boldsymbol{u}}_i$. The scale factors and the unit vectors are generally functions of position, the unit vectors remaining mutually orthogonal everywhere, $\hat{\boldsymbol{u}}_3 = \hat{\boldsymbol{u}}_1 \times \hat{\boldsymbol{u}}_2$. The geometry of the curvilinear coordinates are illustrated in figure A.3.

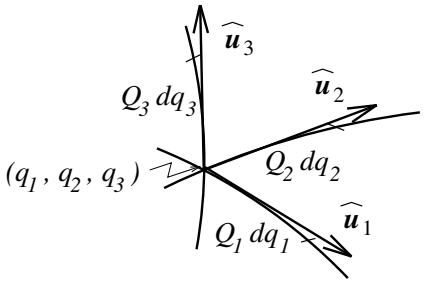

Figure A.3: Curvilinear coordinates

Three simple examples will illustrate the above notation. For a Cartesian coordinate system an arbitrary position vector $\boldsymbol{r}$ is determined by specifying the three coordinates (x, y, z). The unit vectors $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{z}}$ are constant vectors and the scale factors are all equal to unity, $Q_x = Q_y = Q_z = 1$.

In a cylindrical coordinate system the position vector $\boldsymbol{r}$ is specified by giving coordinates (ρ, φ, z) as indicated in figure A.4a. Expressions for the unit vectors $\hat{\boldsymbol{\rho}}$, $\hat{\boldsymbol{\varphi}}$ and $\hat{\boldsymbol{z}}$ in terms of the Cartesian unit vectors $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{z}}$, the scale factors (Q_ρ, Q_φ, Q_z) and the derivatives of the unit vectors with respect to the cylindrical coordinates are summarized in table A.1.

	$\hat{\boldsymbol{x}}$	$\hat{\boldsymbol{y}}$	$\hat{\boldsymbol{z}}$	Q	$\partial/\partial\rho$	$\partial/\partial\varphi$	$\partial/\partial z$
$\hat{\boldsymbol{\rho}}$	$\cos\varphi$	$\sin\varphi$	0	1	0	$\hat{\boldsymbol{\varphi}}$	0
$\hat{\boldsymbol{\varphi}}$	$-\sin\varphi$	$\cos\varphi$	0	ρ	0	$-\hat{\boldsymbol{\rho}}$	0
$\hat{\boldsymbol{z}}$	0	0	1	1	0	0	0

Table A.1: Cylindrical coordinates

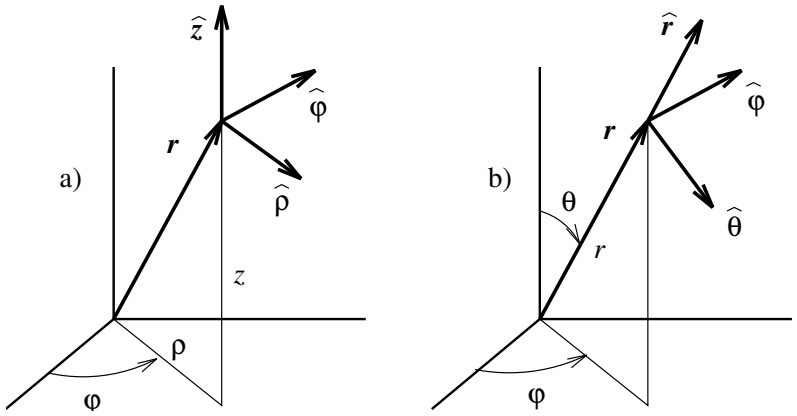

Figure A.4: Cylindrical and spherical coordinates

In a spherical coordinate system the position vector $\boldsymbol{r}$ is specified by giving coordinates (r, θ, φ) as indicated in figure A.4b. Expressions for the unit vectors $\hat{\boldsymbol{\rho}}$, $\hat{\boldsymbol{\varphi}}$ and $\hat{\boldsymbol{z}}$ and the scale factors $(Q_r, Q_\theta, Q_\varphi)$ and the derivatives of the unit vectors are summarized in table A.2.

	$\hat{\boldsymbol{x}}$	$\hat{\boldsymbol{y}}$	$\hat{\boldsymbol{z}}$	Q	$\partial/\partial r$	$\partial/\partial\theta$	$\partial/\partial\varphi$
$\hat{\boldsymbol{r}}$	$\sin\theta\cos\varphi$	$\sin\theta\sin\varphi$	$\cos\theta$	1	0	$\hat{\boldsymbol{\theta}}$	$\hat{\boldsymbol{\varphi}}\sin\theta$
$\hat{\boldsymbol{\theta}}$	$\cos\theta\cos\varphi$	$\cos\theta\sin\varphi$	$-\sin\theta$	r	0	$-\hat{\boldsymbol{r}}$	$\hat{\boldsymbol{\varphi}}\cos\theta$
$\hat{\boldsymbol{\varphi}}$	$-\sin\varphi$	$\cos\varphi$	0	$r\sin\theta$	0	0	$-\hat{\boldsymbol{r}}\sin\theta - \hat{\boldsymbol{\theta}}\cos\theta$

Table A.2: Spherical coordinates

The explicit expressions for the grad, div and curl operators are now easily calculated. By definition we have

$$\lim_{\mathrm{d}q_1\to 0} \frac{1}{Q_1\,\mathrm{d}q_1}\left[\Phi(\boldsymbol{r} + Q_1\,\mathrm{d}q_1\,\hat{\boldsymbol{u}}_1) - \Phi(\boldsymbol{r})\right] = \frac{1}{Q_1}\frac{\partial\Phi}{\partial q_1}$$

and therefore also

$$\hat{\boldsymbol{u}}_1 \cdot \operatorname{grad}\Phi = \frac{1}{Q_1}\frac{\partial\Phi}{\partial q_1}, \tag{A.6}$$

with similar expressions for the other components of the grad operator.

To derive the expression for the div operator it is convenient to choose an infinitesimal volume d^3V limited by constant coordinate surfaces as illustrated in figure A.5. From the figure we find

$$\mathrm{d}^3V = Q_1Q_2Q_3\,\mathrm{d}q_1\,\mathrm{d}q_2\,\mathrm{d}q_3 \quad \text{and} \quad \mathrm{d}^2\mathcal{A}_1 = Q_2Q_3\,\mathrm{d}q_2\,\mathrm{d}q_3$$

and thus

$$\begin{aligned}\oint_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}}\cdot\boldsymbol{F} &= \left[(F_1Q_2Q_3)(q_1+\mathrm{d}q_1) - (F_1Q_2Q_3)(q_1)\right]\,\mathrm{d}q_2\,\mathrm{d}q_3 + \cdots \\ &= \left[\frac{\partial}{\partial q_1}(F_1Q_2Q_3) + \cdots\right]\,\mathrm{d}q_1\,\mathrm{d}q_2\,\mathrm{d}q_3.\end{aligned}$$

In the middle expression the contributions to the surface integral from the two constant q_1 surfaces at q_1 and $q_1 + \mathrm{d}q_1$ are given, the former surface oriented in the $-\hat{\boldsymbol{u}}_1$-direction. The contribution from the remaining sides are found by cyclic permutation of indices. According to definition (A.3) we therefore find

$$\operatorname{div} \boldsymbol{F} = \frac{1}{Q_1 Q_2 Q_3} \left[\frac{\partial}{\partial q_1}(F_1 Q_2 Q_3) + \frac{\partial}{\partial q_2}(F_2 Q_3 Q_1) + \frac{\partial}{\partial q_3}(F_3 Q_1 Q_2) \right]. \tag{A.7}$$

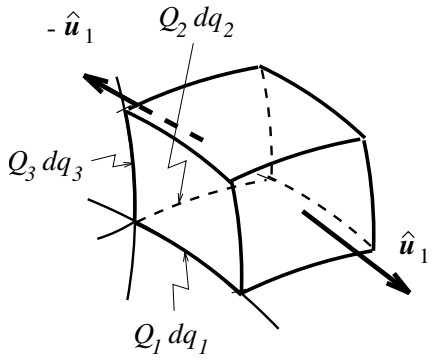

Figure A.5: Geometry for divergence operator

The expression for the curl operator is found in the similar way. With the unit normal vector $\hat{\boldsymbol{u}}_1$ chosen as indicated in figure A.6 we may write the corresponding infinitesimal surface element

$$\mathrm{d}^2\mathcal{A}_1 = Q_2 Q_3 \, \mathrm{d}q_2 \, \mathrm{d}q_3$$

and therefore

$$\begin{aligned} \oint_{\mathcal{C}_1} \mathrm{d}\boldsymbol{\ell} \cdot \boldsymbol{F} &= [(F_3 Q_3)(q_2 + \mathrm{d}q_2) - (F_3 Q_3)(q_2)] \, \mathrm{d}q_3 \\ &\quad - [(F_2 Q_2)(q_3 + \mathrm{d}q_3) - (F_2 Q_2)(q_3)] \, \mathrm{d}q_2 \\ &= \left[\frac{\partial}{\partial q_2}(F_3 Q_3) - \frac{\partial}{\partial q_3}(F_2 Q_2) \right] \mathrm{d}q_2 \, \mathrm{d}q_3. \end{aligned}$$

According to definition (A.4) we now find

$$\hat{\boldsymbol{u}}_1 \cdot \operatorname{curl} \boldsymbol{F} = \frac{1}{Q_2 Q_3} \left[\frac{\partial}{\partial q_2}(F_3 Q_3) - \frac{\partial}{\partial q_3}(F_2 Q_2) \right]. \tag{A.8}$$

The remaining components in the $\hat{\boldsymbol{u}}_2$ and $\hat{\boldsymbol{u}}_3$ directions are found by cyclic permutation of indices.

By combining the above expressions we further derive

$$\operatorname{div} \operatorname{grad} \Phi = \frac{1}{Q_1 Q_2 Q_3} \left[\frac{\partial}{\partial q_1} \left(\frac{Q_2 Q_3}{Q_1} \frac{\partial \Phi}{\partial q_1} \right) + \frac{\partial}{\partial q_2} \left(\frac{Q_3 Q_1}{Q_2} \frac{\partial \Phi}{\partial q_2} \right) + \frac{\partial}{\partial q_3} \left(\frac{Q_1 Q_2}{Q_3} \frac{\partial \Phi}{\partial q_3} \right) \right]. \tag{A.9}$$

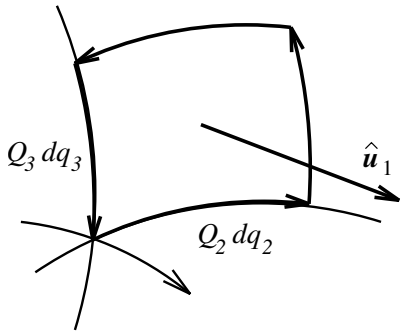

Figure A.6: Geometry for curl operator

Finally, for arbitrary fields Φ and $\boldsymbol{F}$ the identities,

$$\operatorname{curl}\operatorname{grad}\Phi = 0 \tag{A.10}$$

$$\operatorname{div}\operatorname{curl}\boldsymbol{F} = 0, \tag{A.11}$$

may easily be verified.

Quiz A.2 : Verify (A.10) and (A.11).

Quiz A.3 : Verify that the general expression for the curl operator can formally be written as the determinant

$$\operatorname{curl}\boldsymbol{F} = \begin{vmatrix} \dfrac{\hat{\boldsymbol{u}}_1}{Q_2Q_3} & \dfrac{\hat{\boldsymbol{u}}_2}{Q_3Q_1} & \dfrac{\hat{\boldsymbol{u}}_3}{Q_1Q_2} \\ \dfrac{\partial}{\partial q_1} & \dfrac{\partial}{\partial q_2} & \dfrac{\partial}{\partial q_3} \\ Q_1F_1 & Q_2F_2 & Q_3F_3 \end{vmatrix}.$$

A.2.1 Cartesian coordinates

For Cartesian coordinates the above formulas simplify considerably:

$$\operatorname{grad}\Phi = \hat{\boldsymbol{x}}\frac{\partial\Phi}{\partial x} + \hat{\boldsymbol{y}}\frac{\partial\Phi}{\partial y} + \hat{\boldsymbol{z}}\frac{\partial\Phi}{\partial z} \tag{A.12}$$

$$\operatorname{div}\boldsymbol{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z} \tag{A.13}$$

$$\operatorname{curl}\boldsymbol{F} = \hat{\boldsymbol{x}}\left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}\right) + \hat{\boldsymbol{y}}\left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x}\right) + \hat{\boldsymbol{z}}\left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}\right) \tag{A.14}$$

$$\operatorname{div}\operatorname{grad}\Phi = \frac{\partial^2\Phi}{\partial x^2} + \frac{\partial^2\Phi}{\partial y^2} + \frac{\partial^2\Phi}{\partial z^2}. \tag{A.15}$$

A.2.2 Cylinder coordinates

For cylindrical coordinates the corresponding results are:

$$\operatorname{grad}\Phi = \hat{\boldsymbol{\rho}}\frac{\partial\Phi}{\partial\rho} + \frac{\hat{\boldsymbol{\varphi}}}{\rho}\frac{\partial\Phi}{\partial\varphi} + \hat{\boldsymbol{z}}\frac{\partial\Phi}{\partial z} \tag{A.16}$$

$$\operatorname{div}\boldsymbol{F} = \frac{1}{\rho}\frac{\partial}{\partial\rho}(\rho F_\rho) + \frac{1}{\rho}\frac{\partial F_\varphi}{\partial\varphi} + \frac{\partial F_z}{\partial z} \tag{A.17}$$

$$\operatorname{curl}\boldsymbol{F} = \hat{\boldsymbol{\rho}}\left(\frac{1}{\rho}\frac{\partial F_z}{\partial\varphi} - \frac{\partial F_\varphi}{\partial z}\right) + \hat{\boldsymbol{\varphi}}\left(\frac{\partial F_\rho}{\partial z} - \frac{\partial F_z}{\partial\rho}\right) + \frac{\hat{\boldsymbol{z}}}{\rho}\left(\frac{\partial}{\partial\rho}(\rho F_\varphi) - \frac{\partial F_\rho}{\partial\varphi}\right) \tag{A.18}$$

$$\operatorname{div}\operatorname{grad}\Phi = \frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial\Phi}{\partial\rho}\right) + \frac{1}{\rho^2}\frac{\partial^2\Phi}{\partial\varphi^2} + \frac{\partial^2\Phi}{\partial z^2}. \tag{A.19}$$

A.2.3 Spherical coordinates

For spherical coordinates the formulas are:

$$\operatorname{grad}\Phi = \hat{\boldsymbol{r}}\frac{\partial\Phi}{\partial r} + \frac{\hat{\boldsymbol{\theta}}}{r}\frac{\partial\Phi}{\partial\theta} + \frac{\hat{\boldsymbol{\varphi}}}{r\sin\theta}\frac{\partial\Phi}{\partial\varphi} \tag{A.20}$$

$$\operatorname{div}\boldsymbol{F} = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2F_r) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta F_\theta) + \frac{1}{r\sin\theta}\frac{\partial F_\varphi}{\partial\varphi} \tag{A.21}$$

$$\operatorname{curl}\boldsymbol{F} = \frac{\hat{\boldsymbol{r}}}{r\sin\theta}\left[\frac{\partial}{\partial\theta}(\sin\theta F_\varphi) - \frac{\partial F_\theta}{\partial\varphi}\right] + \frac{\hat{\boldsymbol{\theta}}}{r}\left[\frac{1}{\sin\theta}\frac{\partial F_r}{\partial\varphi} - \frac{\partial}{\partial r}(rF_\varphi)\right] \tag{A.22}$$

$$+\frac{\hat{\boldsymbol{\varphi}}}{r}\left[\frac{\partial}{\partial r}(rF_\theta) - \frac{\partial F_r}{\partial\theta}\right] \tag{A.23}$$

$$\operatorname{div}\operatorname{grad}\Phi = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\Phi}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\Phi}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\Phi}{\partial\varphi^2}. \tag{A.24}$$

A.3 Introduction of the ∇-operator

Going back to the case of the Cartesian coordinates it is seen that the grad, div and curl operators may all be expressed in terms of a single vector operator

$$\nabla = \hat{\boldsymbol{x}}\frac{\partial}{\partial x} + \hat{\boldsymbol{y}}\frac{\partial}{\partial y} + \hat{\boldsymbol{z}}\frac{\partial}{\partial z}. \tag{A.25}$$

With the interpretations $\hat{\boldsymbol{x}}\partial/\partial x\cdot\boldsymbol{F} = \hat{\boldsymbol{x}}\cdot\partial\boldsymbol{F}/\partial x$ and $\hat{\boldsymbol{x}}\partial/\partial x\times\boldsymbol{F} = \hat{\boldsymbol{x}}\times\partial\boldsymbol{F}/\partial x$, we find

$$\operatorname{grad}\Phi = \nabla\Phi, \tag{A.26}$$

$$\operatorname{div}\boldsymbol{F} = \nabla\cdot\boldsymbol{F}, \tag{A.27}$$

$$\operatorname{curl}\boldsymbol{F} = \nabla\times\boldsymbol{F}. \tag{A.28}$$

$$\operatorname{div}\operatorname{grad}\Phi = \nabla\cdot\nabla\Phi = \nabla^2\Phi \tag{A.29}$$

This is a convenient and compact notation that is used for any choice of coordinate system. This notation will allow general vector operator manipulations in coordinate-free form to be performed easily. When particular coordinates must eventually be introduced, however, care must be exercised to substitute the proper expressions for the different combinations of ∇-operators and fields.

A.4 General ∇-operator Relations

For general ∇-operator calculus it is essential to observe that the ∇ operator (A.25) is at the same time a vector and a differential operator. The basic rules of vector algebra and differential calculus, exemplified by

$$\boldsymbol{a} \times \boldsymbol{b} = -\boldsymbol{b} \times \boldsymbol{a} \tag{A.30}$$

$$\boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c}) = \boldsymbol{b} \cdot (\boldsymbol{c} \times \boldsymbol{a}) = \boldsymbol{c} \cdot (\boldsymbol{a} \times \boldsymbol{b}) \tag{A.31}$$

$$\boldsymbol{a} \times (\boldsymbol{b} \times \boldsymbol{c}) = (\boldsymbol{a} \cdot \boldsymbol{c})\boldsymbol{b} - (\boldsymbol{a} \cdot \boldsymbol{b})\boldsymbol{c} \tag{A.32}$$

and

$$\frac{\partial}{\partial x}(fg) = \frac{\partial f}{\partial x}g + f\frac{\partial g}{\partial x},$$

must therefore be satisfied concurrently.

One example will demonstrate the basic technique. We want to expand the expression $\nabla \times (\boldsymbol{F} \times \boldsymbol{G})$ where $\boldsymbol{F}$ and $\boldsymbol{G}$ are both vector fields. The differential part of the ∇-operator must once be applied to $\boldsymbol{F}$ and once to $\boldsymbol{G}$. As a reminder we write

$$\nabla \times (\boldsymbol{F} \times \boldsymbol{G}) = \nabla \times (\check{\boldsymbol{F}} \times \boldsymbol{G}) + \nabla \times (\boldsymbol{F} \times \check{\boldsymbol{G}}).$$

The ˇ-sign indicates where the differentiation is to be applied. We next turn to the vector part of the ∇-operator. We therefore apply the rule for expanding the double cross product (A.32) — *middle* vector times the scalar product of the other vectors *minus the other* vector in parenthesis times the scalar product of the other vectors — to each term, making sure to chose an ordering of factors in each expression such that it is evident where the ∇-operator is to be applied. The result is

$$\nabla \times (\boldsymbol{F} \times \boldsymbol{G}) = (\boldsymbol{G} \cdot \nabla)\boldsymbol{F} - \boldsymbol{G}(\nabla \cdot \boldsymbol{F}) + \boldsymbol{F}(\nabla \cdot \boldsymbol{G}) - (\boldsymbol{F} \cdot \nabla)\boldsymbol{G}.$$

Notice that for instance the first and third term on the right-hand side are identical from a vector algebra point of view. By rearranging the ordering of the factors the ∇-operator in the first term is acting on $\boldsymbol{F}$, and in the third term on $\boldsymbol{G}$.

Making use of the techniques exemplified above the following expressions may be verified

$$\nabla(\Phi\Psi) = \Psi\nabla\Phi + \Phi\nabla\Psi \tag{A.33}$$

$$\nabla \cdot (\Phi\boldsymbol{F}) = \Phi\nabla \cdot \boldsymbol{F} + \boldsymbol{F} \cdot \nabla\Phi \tag{A.34}$$

$$\nabla \times (\Phi\boldsymbol{F}) = \Phi\nabla \times \boldsymbol{F} + \nabla\Phi \times \boldsymbol{F} \tag{A.35}$$

$$\nabla(\boldsymbol{F} \cdot \boldsymbol{G}) = \boldsymbol{G} \times (\nabla \times \boldsymbol{F}) + (\boldsymbol{G} \cdot \nabla)\boldsymbol{F} + \boldsymbol{F} \times (\nabla \times \boldsymbol{G}) + (\boldsymbol{F} \cdot \nabla)\boldsymbol{G} \tag{A.36}$$

$$\nabla \cdot (\boldsymbol{F} \times \boldsymbol{G}) = \boldsymbol{G} \cdot \nabla \times \boldsymbol{F} - \boldsymbol{F} \cdot \nabla \times \boldsymbol{G} \tag{A.37}$$

$$\nabla \times (\boldsymbol{F} \times \boldsymbol{G}) = (\boldsymbol{G} \cdot \nabla)\boldsymbol{F} - \boldsymbol{G}\,\nabla \cdot \boldsymbol{F} + \boldsymbol{F}\,\nabla \cdot \boldsymbol{G} - (\boldsymbol{F} \cdot \nabla)\boldsymbol{G}. \tag{A.38}$$

Quiz A.4 : Verify that (A.33) – (A.38) are indeed correct.

Quiz A.5 : Find the unit normal vector $\hat{\boldsymbol{n}}$ to the surface $z = f(x, y)$.

Quiz A.6 : Show that

$$\nabla \cdot \boldsymbol{r} = 3, \quad \nabla \cdot \hat{\boldsymbol{r}} = \frac{2}{r}, \quad \boldsymbol{F} \cdot \nabla \boldsymbol{r} = \boldsymbol{F} \quad \text{and} \quad \nabla \times \boldsymbol{r} = 0$$

using representations for the ∇-operator in cartesian coordinates.

Quiz A.7 : Show that if $\boldsymbol{a}$ is a constant vector then

$$\nabla \times (\boldsymbol{a} \times \boldsymbol{r}) = 2\boldsymbol{a}.$$

Quiz A.8 : Show that

$$\boldsymbol{F} \cdot \nabla \boldsymbol{F} = \frac{1}{2}\nabla \boldsymbol{F}^2 - \boldsymbol{F} \times (\nabla \times \boldsymbol{F})$$

and

$$\nabla^2 \boldsymbol{F} = \nabla \nabla \cdot \boldsymbol{F} - \nabla \times (\nabla \times \boldsymbol{F}).$$

A.5 The ∇-operator in curvilinear, orthogonal coordinates

The representation (A.25) for the ∇-operator explicitly assumed the use of cartesian coordinates. The question then arises whether a similar form will also apply for orthogonal curvilinear coordinates. A natural choice would seem to be

$$\nabla = \sum_\ell \hat{\boldsymbol{u}}_\ell (\hat{\boldsymbol{u}}_\ell \cdot \nabla) = \sum_\ell \frac{\hat{\boldsymbol{u}}_\ell}{Q_\ell} \frac{\partial}{\partial q_\ell}. \tag{A.39}$$

We recognize the general expression for the directional derivative (A.6) in this formula. When this operator is applied to the vector field $\boldsymbol{F} = \sum_\ell F_\ell \hat{\boldsymbol{u}}_\ell$ it must, however, be remembered that the unit vectors $\hat{\boldsymbol{u}}_\ell$ are themselves functions of position. In tables A.1 and A.2 the derivatives of the unit vectors are given for cylindrical and spherical coordinates. In quiz A.9 steps to verify that the representation (A.39) is valid, that is, that relations (A.27) and (A.28) are satisfied are indicated.

Quiz A.9 : With the representation (A.39) in any given orthogonal, right-handed curvilinear coordinate system show that

$$\nabla \cdot \boldsymbol{F} = \sum_\ell \frac{1}{Q_\ell} \frac{\partial F_\ell}{\partial q_\ell} - \boldsymbol{F} \cdot \sum_\ell \frac{1}{Q_\ell} \frac{\partial \hat{\boldsymbol{u}}_\ell}{\partial q_\ell}$$

$$\hat{\boldsymbol{u}}_k \cdot \nabla \times \boldsymbol{F} = \sum_{\ell,m} (\hat{\boldsymbol{u}}_k \times \hat{\boldsymbol{u}}_\ell) \cdot \frac{1}{Q_\ell} \frac{\partial}{\partial q_\ell} (F_m \hat{\boldsymbol{u}}_m) =$$

$$\frac{1}{Q_i}\frac{\partial F_j}{\partial q_i} - \frac{1}{Q_j}\frac{\partial F_i}{\partial q_j} + \frac{F_i}{Q_i}\hat{\boldsymbol{u}}_j \cdot \frac{\partial \hat{\boldsymbol{u}}_i}{\partial q_i} - \frac{F_j}{Q_j}\hat{\boldsymbol{u}}_i \cdot \frac{\partial \hat{\boldsymbol{u}}_j}{\partial q_j} + F_k \left(\frac{1}{Q_i}\hat{\boldsymbol{u}}_j \cdot \frac{\partial u_k}{\partial q_i} - \frac{1}{Q_j}\hat{\boldsymbol{u}}_i \cdot \frac{\partial u_k}{\partial q_j} \right)$$

where $\hat{\boldsymbol{u}}_i \times \hat{\boldsymbol{u}}_j = \hat{\boldsymbol{u}}_k$. For cylindrical and spherical coordinates make use of tables A.1 and A.2 to verify that relations (A.27) and (A.28) are satisfied.

A.6 Integral Theorems

From the definition of the divergence operator (A.3) very useful integral theorems may be derived. Consider an arbitrary finite volume V bounded by the surface $\mathcal{A}$ with outward directed unit normal. Let the derivatives of $\boldsymbol{F}$ exist and be continuous within the volume V. Let V be divided into infinitesimal volumes d^3V_i, $i = 1, \cdots$ with corresponding bounding surfaces S_i. From (A.3) we may now write

$$\sum_i \mathrm{d}^3V_i (\nabla \cdot \boldsymbol{F})_i = \sum_i \oint_{\mathcal{A}_i} \mathrm{d}^2\boldsymbol{\mathcal{A}} \cdot \boldsymbol{F}.$$

The contributions to the surface integrals from the adjoining surfaces of any two neighboring elementary volumes, as illustrated in figure A.7a, are equal and opposite and therefore cancel. The right-hand side sum of surface integrals thus reduces to the corresponding surface integral over $\mathcal{A}$. The left-hand sum is the volume integral over the total volume V. The result

$$\int_V \mathrm{d}^3\boldsymbol{r}\, \nabla \cdot \boldsymbol{F} = \oint_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}} \cdot \boldsymbol{F} \tag{A.40}$$

is known as the *Gauss integral theorem.*

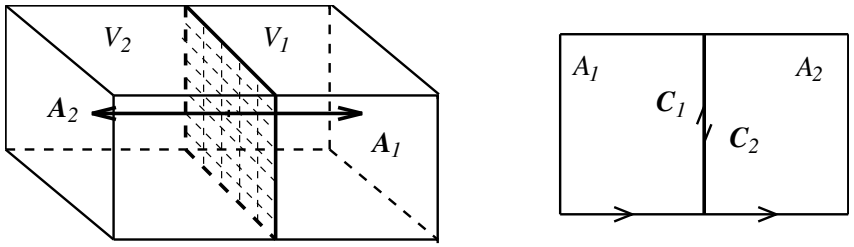

Figure A.7: Illustrating Gauss and Stoke integral theorems

A similar integral theorem applies to the curl operator. This time consider an arbitrary surface $\boldsymbol{\mathcal{A}}$ bounded by the contour $\mathcal{C}$. The directions of $\boldsymbol{\mathcal{A}}$ and $\mathcal{C}$ should be related through the right-hand rule as discussed in relation to (A.4). Let the derivatives of $\boldsymbol{F}$ exist and be continuous over the surface $\boldsymbol{\mathcal{A}}$. Let $\boldsymbol{\mathcal{A}}$ be divided into infinitesimal elementary surfaces $\mathrm{d}^2\boldsymbol{\mathcal{A}}_i$, $i = 1, \cdots$ with corresponding bounding contours C_i. According to the definition of the curl operator (A.4) we may now write

$$\sum_i \mathrm{d}^2\boldsymbol{\mathcal{A}}_i \cdot (\nabla \times \boldsymbol{F})_i = \sum_i \oint_{\mathcal{C}_i} \mathrm{d}\boldsymbol{\ell} \cdot \boldsymbol{F}.$$

The contributions to the contour integrals from the adjoining parts of the contours of any two neighboring elementary surfaces cancel. This fact is illustrated in figure A.7b. The right-hand side sum of contour integrals thus reduces to the corresponding contour integral around $\mathcal{C}$. The left-hand side sum is the surface integral over the total surface A. The result

$$\int_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}} \cdot \nabla \times \boldsymbol{F} = \oint_{\mathcal{C}} \mathrm{d}\boldsymbol{\ell} \cdot \boldsymbol{F} \tag{A.41}$$

is called the *Stoke integral theorem.*

Quiz A.10 : Show that

$$\frac{1}{3}\oint_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}}\cdot\boldsymbol{r} = V.$$

Can you give a geometrical interpretation for this result?

Quiz A.11 : Show that

$$\oint_{\mathcal{C}} \mathrm{d}\boldsymbol{\ell}\cdot\boldsymbol{F} = 0 \quad \text{if} \quad \boldsymbol{F} = \nabla\Phi.$$

Quiz A.12 : Let $\boldsymbol{F} = (y\hat{\boldsymbol{x}} - x\hat{\boldsymbol{y}})/(x^2+y^2)$. Calculate $\nabla\times\boldsymbol{F}$. Calculate $\oint_{\mathcal{C}}\boldsymbol{F}\cdot\mathrm{d}\boldsymbol{\ell}$, where $\mathcal{C}$ is a circle in the (x,y)-plane with center at the origin. Is the Stoke integral theorem satisfied? Explain.

Quiz A.13 : Show that

$$\oint_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}}\cdot(\Phi\nabla\Psi) = \int_V \mathrm{d}^3\boldsymbol{r}\,\left[\Phi\nabla^2\Psi + (\nabla\Phi)\cdot(\nabla\Psi)\right] \tag{A.42}$$

and

$$\oint_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}}\cdot(\Phi\nabla\Psi - \Psi\nabla\Phi) = \int_V \mathrm{d}^3\boldsymbol{r}\,(\Phi\nabla^2\Psi - \Psi\nabla^2\Phi). \tag{A.43}$$

The results (A.42) and (A.43) are known as the *first and second form of the Green theorem.*

A.7 Generalizations

The different combinations of ∇-operators and scalar and vector fields studied so far have always resulted in another scalar or vector field. This is not always so. As an example again consider the first term on the right-hand side of (A.38) as discussed above. This term can be written in two different but equivalent ways

$$(\boldsymbol{G}\cdot\nabla)\boldsymbol{F} \quad \text{or} \quad \boldsymbol{G}\cdot(\nabla\boldsymbol{F}).$$

In the first form the directional derivative is a scalar operator. When applied to the vector field $\boldsymbol{F}$, the result is another vector field. In the second form we encounter the scalar product between the vector field $\boldsymbol{G}$ and the new type of entity $\nabla\boldsymbol{F}$.

The scalar and vector fields are themselves special cases of the more general concept of a *tensor* field. A scalar field is a tensor field of rang zero. A vector field is a tensor field of rang one. The vector field is characterized by a magnitude and a direction at every position $\boldsymbol{r}$. A tensor field of rang two, also called a *dyadic* field, is characterized by a magnitude and two directions at every position. The entity $\nabla\boldsymbol{F}$ is an example of a dyadic field for which we shall introduce the general notation $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$, $\cdots$. Higher order tensor fields are defined in similar fashion. For a general tensor field we introduce the notation $\mathcal{F}$, $\mathcal{G}$, $\cdots$.

For a given orthonormal basis vector set $\{\hat{\boldsymbol{u}}_1, \hat{\boldsymbol{u}}_2, \hat{\boldsymbol{u}}_3\}$ the most general representation of a vector field is $\boldsymbol{F} = \sum_i F_i\hat{\boldsymbol{u}}_i$. The jth component of the vector is found as $F_j = \boldsymbol{F}\cdot\hat{\boldsymbol{u}}_j = \hat{\boldsymbol{u}}_j\cdot\boldsymbol{F}$. The vector $\boldsymbol{F}$ may be uniquely identified as a column or row matrix

$$\boldsymbol{F} = \left\{\begin{array}{c} F_1 \\ F_2 \\ F_3 \end{array}\right\} \quad \text{or} \quad \boldsymbol{F} = \{F_1, F_2, F_3\}.$$

In matrix form the scalar product of two vectors $\boldsymbol{F}$ and $\boldsymbol{G}$ is represented by

$$\boldsymbol{F}\cdot\boldsymbol{G} = \{F_1, F_2, F_3\}\left\{\begin{array}{c} G_1 \\ G_2 \\ G_3 \end{array}\right\}.$$

The most general representation of a dyadic field is the *nonion* form

$$\boldsymbol{\Phi} = \sum_{i,j} \Phi_{ij}\hat{\boldsymbol{u}}_i\hat{\boldsymbol{u}}_j. \tag{A.44}$$

Here $\hat{\boldsymbol{u}}_i\hat{\boldsymbol{u}}_j$ represents the *tensor* or *outer* product of the unit vectors $\hat{\boldsymbol{u}}_i$ and $\hat{\boldsymbol{u}}_j$. In this product the former factor $\hat{\boldsymbol{u}}_i$ is the *antecedent*, the latter factor $\hat{\boldsymbol{u}}_j$ the *consequent*. The *transpose* or the *conjugate* of the product $\hat{\boldsymbol{u}}_i\hat{\boldsymbol{u}}_j$ is defined as

$$\widetilde{\hat{\boldsymbol{u}}_i\hat{\boldsymbol{u}}_j} = \hat{\boldsymbol{u}}_j\hat{\boldsymbol{u}}_i.$$

The ij–th component of the dyadic $\boldsymbol{\Phi}$ is found as

$$\Phi_{ij} = \hat{\boldsymbol{u}}_i\cdot\boldsymbol{\Phi}\cdot\hat{\boldsymbol{u}}_j = \hat{\boldsymbol{u}}_j\hat{\boldsymbol{u}}_i : \boldsymbol{\Phi} = \boldsymbol{\Phi} : \hat{\boldsymbol{u}}_j\hat{\boldsymbol{u}}_i$$

where we applied the tensor algebra rules

$$\boldsymbol{a}\cdot(\boldsymbol{bc}) = (\boldsymbol{a}\cdot\boldsymbol{b})\boldsymbol{c} \tag{A.45}$$

$$(\boldsymbol{bc})\cdot\boldsymbol{d} = (\boldsymbol{c}\cdot\boldsymbol{d})\boldsymbol{b} \tag{A.46}$$

$$(\boldsymbol{ab})\cdot(\boldsymbol{cd}) = (\boldsymbol{b}\cdot\boldsymbol{c})(\boldsymbol{ad}) \tag{A.47}$$

and the definition

$$(\boldsymbol{ab}) : (\boldsymbol{cd}) = (\boldsymbol{b}\cdot\boldsymbol{c})(\boldsymbol{a}\cdot\boldsymbol{d}). \tag{A.48}$$

In matrix form the tensor product of two vectors $\boldsymbol{F}$ and $\boldsymbol{G}$ is defined by

$$\boldsymbol{FG} = \left\{\begin{array}{c} F_1 \\ F_2 \\ F_3 \end{array}\right\}\{G_1, G_2, G_3\}.$$

The general dyadic $\boldsymbol{\Phi}$ is identified as the square matrix

$$\boldsymbol{\Phi} = \left\{\begin{array}{ccc} \Phi_{11} & \Phi_{12} & \Phi_{13} \\ \Phi_{21} & \Phi_{22} & \Phi_{23} \\ \Phi_{31} & \Phi_{32} & \Phi_{33} \end{array}\right\}.$$

The scalar product of the dyadic $\boldsymbol{\Phi}$ with the vector $\boldsymbol{F}$ reduces in matrix form to the matrix product of a square matrix with a column matrix

$$\boldsymbol{\Phi}\cdot\boldsymbol{F} = \left\{\begin{array}{ccc} \Phi_{11} & \Phi_{12} & \Phi_{13} \\ \Phi_{21} & \Phi_{22} & \Phi_{23} \\ \Phi_{31} & \Phi_{32} & \Phi_{33} \end{array}\right\}\left\{\begin{array}{c} F_1 \\ F_2 \\ F_3 \end{array}\right\}.$$

If the order of the dyadic and vector is reversed it is necessary to make use of the transposed form for the vector

$$\boldsymbol{F}\cdot\boldsymbol{\Phi} = \{F_1, F_2, F_3\}\left\{\begin{array}{ccc} \Phi_{11} & \Phi_{12} & \Phi_{13} \\ \Phi_{21} & \Phi_{22} & \Phi_{23} \\ \Phi_{31} & \Phi_{32} & \Phi_{33} \end{array}\right\}.$$

The special dyadic

$$\boldsymbol{I} = \hat{\boldsymbol{u}}_1\hat{\boldsymbol{u}}_1 + \hat{\boldsymbol{u}}_2\hat{\boldsymbol{u}}_2 + \hat{\boldsymbol{u}}_3\hat{\boldsymbol{u}}_3$$

is the *idemfactor* or the *identical dyadic* with the properties

$$\boldsymbol{I}\cdot\boldsymbol{r} = \boldsymbol{r}\cdot\boldsymbol{I} = \boldsymbol{r}$$
$$\boldsymbol{\Phi}\cdot\boldsymbol{I} = \boldsymbol{I}\cdot\boldsymbol{\Phi} = \boldsymbol{\Phi}.$$

The rules for applying the ∇-operator to scalar and vector fields developed above are also valid in the more general case, provided the tensor algebra rules are observed. Thus, an alternative form to (A.36) is now

$$\nabla(\boldsymbol{F}\cdot\boldsymbol{G}) = (\nabla\boldsymbol{F})\cdot\boldsymbol{G} + (\nabla\boldsymbol{G})\cdot\boldsymbol{F} \tag{A.49}$$

The Gauss and Stoke integral theorems (A.40) and (A.41) are easily generalized to general tensor fields $\mathcal{F}$. The generalized Gauss integral theorem reads

$$\int_V \mathrm{d}^3\boldsymbol{r}\,\nabla * \mathcal{F} = \oint_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}} * \mathcal{F} \tag{A.50}$$

while the corresponding generalized Stoke integral theorem takes the form

$$\int_A (\mathrm{d}^2\boldsymbol{\mathcal{A}}\times\nabla) * \mathcal{F} = \oint_{\mathcal{C}} \mathrm{d}\boldsymbol{\ell} * \mathcal{F}. \tag{A.51}$$

In these formulas the binary operator '$*$' may represent a scalar product '$\cdot$', a vector product '$\times$' if the tensor $\mathcal{F}$ has at least rang one, or a tensor product. The formulas (A.50) and (A.51) are easily proved making use of Cartesian coordinates by either direct integration or by reduction to the standard forms (A.40) or (A.41) through scalar multiplications with the Cartesian unit vectors $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{z}}$.

Quiz A.14 : What is the matrix representation of the identical dyadic $\boldsymbol{I}$?

Quiz A.15 : In analogy with (A.46) how would you define $\boldsymbol{I}\times\boldsymbol{r}$? What is the rang and the matrix representation of the new tensor?

Quiz A.16 : Show that

$$\boldsymbol{\Phi}\cdot\boldsymbol{\Psi} = \tilde{\boldsymbol{\Psi}}\cdot\boldsymbol{\Phi} = \boldsymbol{\Psi}\cdot\tilde{\boldsymbol{\Phi}}.$$

Quiz A.17 : Show that

$$\nabla\cdot(\Phi\boldsymbol{I}) = \nabla\Phi.$$

Quiz A.18 : Evaluate the expression $\boldsymbol{F}\cdot\nabla\boldsymbol{F}$ for the case that $\boldsymbol{F} = \boldsymbol{F}(\rho)$ where ρ is the distance from a symmetry axis

- by transforming the given expression into a form containing only standard grad, div or curl expressions.
- by evaluating $(\boldsymbol{F}\cdot\nabla)\boldsymbol{F}$ directly using table A.1.
- by evaluating $\boldsymbol{F}\cdot(\nabla\boldsymbol{F})$ directly again making use of table A.1.

Do you get the same answer?

Quiz A.19 : For an arbitrary closed, plane curve $\mathcal{C}$ with unit normal vector $\hat{\boldsymbol{n}}$, show that

$$\oint_{\mathcal{C}} \boldsymbol{r} \times \mathrm{d}\boldsymbol{\ell} = 2A\,\hat{\boldsymbol{n}}$$

where A is the area of the surface inside $\mathcal{C}$, both by making use of the generalized Stoke integral theorem and from direct geometrical considerations.

Quiz A.20 : Prove by direct computation that the following special cases of (A.50) and (A.51),

$$\int_V \mathrm{d}^3\boldsymbol{r}\,\nabla\Phi = \oint_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}}\,\Phi$$

and

$$\int_{\mathcal{A}} \mathrm{d}^2\boldsymbol{\mathcal{A}} \times \nabla\Phi = \oint_{\mathcal{C}} \mathrm{d}\boldsymbol{\ell}\,\Phi,$$

are indeed valid. [*Hint:* For simplicity choose a right parallelepiped and a rectangle as integration volume and area, respectively.]

Quiz A.21 : Definitions of the directional derivative and gradient operators were given in (A.1) and (A.2). Argue from (A.50) that alternative definitions could be

$$\operatorname{grad}\Phi \equiv \lim_{V\to 0} \oint_{\mathcal{A}} \Phi\,\mathrm{d}^2\boldsymbol{\mathcal{A}} \tag{A.52}$$

and (compare (A.5))

$$\frac{\partial\Phi}{\partial\ell} \equiv \hat{\boldsymbol{\ell}} \cdot \operatorname{grad}\Phi. \tag{A.53}$$

Starting from the new definition (A.52), convince yourself that $\operatorname{grad}\Phi$ in any point $\boldsymbol{r}$ will be perpendicular to the surface $\Phi =$ constant passing through that point.

A.8 The Inverse Problem

So far we have been discussing the rules for applying the ∇-operator to different scalar, vector and tensor fields. Let us now consider the inverse problem, that is, can the original field be re-derived if the expressions for the gradient, the divergence or the curl of the field are known over the volume V together with suitable boundary conditions imposed at the boundary surface A?

The simplest case is represented by the gradient $\nabla\Phi$ of the scalar field Φ. With $\nabla\Phi(\boldsymbol{r})$ given in any point $\boldsymbol{r}$ in the volume V, the definition of the directional derivative (A.1), expressed in terms of the gradient operator (A.5), immediately leads to

$$\Phi(\boldsymbol{r}) = \Phi(\boldsymbol{r}_0) + \int_\Gamma \mathrm{d}\boldsymbol{\ell}' \cdot \nabla\Phi(\boldsymbol{r}'). \tag{A.54}$$

Here Γ is any integration path inside V going from any given point $\boldsymbol{r}_0$ on the boundary A where the scalar field takes the value $\Phi(\boldsymbol{r}_0)$ to the arbitrary point $\boldsymbol{r}$. In particular, it is

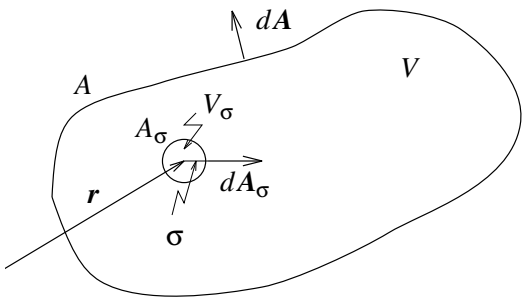

Figure A.8: Geometry for inverse problem

sufficient for a complete determination of $\Phi(\boldsymbol{r})$, in addition to knowing the gradient of Φ in *any* point of the volume V, to know the value of Φ in *one* point at the surface A. Specification of Φ in two points at the surface A generally leads to an inconsistent problem.

Similarly, any vector field $\boldsymbol{F}(\boldsymbol{r})$ can be reconstructed uniquely from the values of the divergence *and* the curl of the field. Thus, let the divergence and the curl of the vector field $\boldsymbol{F}$ be specified inside the volume V together with the value of the field itself at the surface A. Let $\boldsymbol{r}$ be an arbitrary point inside V and let V_σ denote an infinitesimal sphere of radius σ centered on $\boldsymbol{r}$ and with surface A_σ, as illustrated in figure A.8. With $\boldsymbol{r}'$ in $V - V_\sigma$ now consider the identity

$$\nabla' \cdot (\nabla' \psi \boldsymbol{F} - \psi \nabla' \boldsymbol{F}) = \nabla'^2 \psi \, \boldsymbol{F} - \psi \, \nabla'^2 \boldsymbol{F} \tag{A.55}$$

where $\boldsymbol{F} = \boldsymbol{F}(\boldsymbol{r}')$. With $\psi = 1/\,|\boldsymbol{r} - \boldsymbol{r}'|$ and therefore $\nabla'^2 \psi = 0$ and $\nabla' \psi = -\nabla \psi$, the right hand side can be written

$$-\nabla \left(\frac{\nabla' \cdot \boldsymbol{F}}{|\boldsymbol{r} - \boldsymbol{r}'|} \right) + \nabla \times \left(\frac{\nabla' \times \boldsymbol{F}}{|\boldsymbol{r} - \boldsymbol{r}'|} \right) + \nabla' \times \left(\frac{\nabla' \times \boldsymbol{F}}{|\boldsymbol{r} - \boldsymbol{r}'|} \right) - \nabla' \left(\frac{\nabla' \cdot \boldsymbol{F}}{|\boldsymbol{r} - \boldsymbol{r}'|} \right).$$

Integrating (A.55) over the volume $V - V_\sigma$ with this particular choice of ψ, making use of the generalized Gauss integral theorem (A.50) and letting $\sigma \to 0$, the only non-vanishing contribution from the surface A_σ comes from the first term on the left hand side of (A.55) and equals $4\pi \boldsymbol{F}(\boldsymbol{r})$. The result of the integration is therefore

$$\boldsymbol{F}(\boldsymbol{r}) = -\nabla \frac{1}{4\pi} \int_V \mathrm{d}^3 \boldsymbol{r}' \, \frac{\nabla' \cdot \boldsymbol{F}(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} + \nabla \times \frac{1}{4\pi} \int_V \mathrm{d}^3 \boldsymbol{r}' \, \frac{\nabla' \times \boldsymbol{F}(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} + \boldsymbol{B}(\boldsymbol{r}) \tag{A.56}$$

where the contributions from the surface A can be written in the form

$$\boldsymbol{B}(\boldsymbol{r}) = \frac{1}{4\pi} \oint_{\mathcal{A}} \mathrm{d}^2 \boldsymbol{\mathcal{A}}' \times \left(\nabla' \frac{1}{|\boldsymbol{r} - \boldsymbol{r}'|} \times \boldsymbol{F}(\boldsymbol{r}') \right) - \frac{1}{4\pi} \oint_{\mathcal{A}} \mathrm{d}^2 \boldsymbol{\mathcal{A}}' \, \nabla' \frac{1}{|\boldsymbol{r} - \boldsymbol{r}'|} \cdot \boldsymbol{F}(\boldsymbol{r}').$$

The identity (A.56) is sometimes referred to as the *Helmholtz formula*. If the volume V extends to infinity where the vector field is known to vanish, the contributions from the surface term $\boldsymbol{B}(\boldsymbol{r})$ vanish. The vector field $\boldsymbol{F}$ can then be written as a sum of a laminar and a solenoidal part

$$\boldsymbol{F} = -\nabla \Phi + \nabla \times \boldsymbol{H}.$$

The laminar scalar potential

$$\Phi(\boldsymbol{r}) = \frac{1}{4\pi} \int_V \mathrm{d}^3 \boldsymbol{r}' \, \frac{\nabla' \cdot \boldsymbol{F}(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|}$$

and the solenoidal vector potential

$$\boldsymbol{H}(\boldsymbol{r}) = \frac{1}{4\pi}\int_V \mathrm{d}^3\boldsymbol{r}'\ \frac{\nabla'\times\boldsymbol{F}(\boldsymbol{r}')}{|\,\boldsymbol{r}-\boldsymbol{r}'\,|}$$

only depends on the divergence or the curl of the vector field, respectively.

Quiz A.22 : Do you from (A.56) recognize the Coulomb law of electrostatics

$$\boldsymbol{E} = -\frac{1}{4\pi\epsilon_0}\int_V \mathrm{d}^3\boldsymbol{r}'\ \rho(\boldsymbol{r}')\frac{\boldsymbol{r}-\boldsymbol{r}'}{|\,\boldsymbol{r}-\boldsymbol{r}'\,|^3}$$

and the Ampére law of magneto-statics

$$\boldsymbol{B} = \frac{\mu_0}{4\pi}\int_V \mathrm{d}^3\boldsymbol{r}'\ \boldsymbol{j}(\boldsymbol{r}')\times\frac{\boldsymbol{r}-\boldsymbol{r}'}{|\,\boldsymbol{r}-\boldsymbol{r}'\,|^3}\,?$$

Quiz A.23 : For a vector field $\boldsymbol{F}(\boldsymbol{r})$ vanishing at infinity, "prove" (A.56) by calculating the divergence and curl of the right hand side.

A.9 The Dirac δ–function

Consider the complex harmonic segment of length τ,

$$f_\tau(t) = \begin{cases} \exp(-\imath\omega_0 t) & \text{for} \quad t\in[-\frac{\tau}{2},\frac{\tau}{2}] \\ 0 & \text{elsewhere.}\end{cases} \tag{A.57}$$

The Fourier transform (see table 1.1) of $f(t)$ is

$$\tilde{f}_\tau(\omega) = \frac{\tau}{2\pi}\frac{\sin(\omega-\omega_0)\frac{\tau}{2}}{(\omega-\omega_0)\frac{\tau}{2}} \tag{A.58}$$

and has the following properties:

$$\int_{-\infty}^{\infty}\tilde{f}_\tau(\omega)\,\mathrm{d}\omega = 1, \tag{A.59}$$

$$\lim_{\tau\to\infty}\tilde{f}_\tau(\omega) = 0 \quad \text{for} \quad \omega\neq\omega_0. \tag{A.60}$$

It is further evident for any function $g(\omega)$ varying slowly compared to $\tilde{f}_\tau(\omega)$ that

$$\lim_{\tau\to\infty}\int_{-\infty}^{\infty} g(\omega)\tilde{f}_\tau(\omega)\,\mathrm{d}\omega = g(\omega_0). \tag{A.61}$$

The *Dirac δ-function* $\delta(\omega-\omega_0)$ is a (generalized) function having identical properties as the function $\tilde{f}_\tau$ in the limit $\tau\to\infty$, that is:

$$\int_{-\infty}^{\infty}\delta(\omega-\omega_0)\,\mathrm{d}\omega = 1, \tag{A.62}$$

$$\delta(\omega-\omega_0) = 0 \quad \text{for} \quad \omega\neq\omega_0, \tag{A.63}$$

$$\int_{-\infty}^{\infty} g(\omega)\,\delta(\omega-\omega_0)\,\mathrm{d}\omega = g(\omega_0). \tag{A.64}$$

The δ-function may be represented mathematically in many different ways, for instance

$$\delta(\omega-\omega_0)=\lim_{\tau\to\infty}\frac{1}{2\pi}\int_{-\tau/2}^{\tau/2}\exp(\pm\iota(\omega-\omega_0)t)\,\mathrm{d}t=\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(\pm\iota(\omega-\omega_0)t)\,\mathrm{d}t. \tag{A.65}$$

We refer in particular to (1.44).

Some care must be exercised when performing integrals containing δ-functions with more complicated arguments. The standard procedure in these cases is to change the integration variable to the argument of the δ-function, thus

$$\int g(\omega)\,\delta(f(\omega))\,\mathrm{d}\omega=\int\frac{g(\omega)}{J(\omega)}\,\delta(f)\,\mathrm{d}f=\sum_i\frac{g(\omega_i)}{J(\omega_i)}\bigg|_{f(\omega_i)=0} \tag{A.66}$$

where J is the absolute value of the derivative of f with respect to ω, appearing in the variable change from ω to $f(\omega)$

$$\mathrm{d}f=J(\omega)\,\mathrm{d}\omega=\left|\frac{\mathrm{d}f(\omega)}{\mathrm{d}\omega}\right|\mathrm{d}\omega, \tag{A.67}$$

and the summation in (A.66) is taken over all solutions ω_i of $f(\omega)=0$ within the integration range.

The δ-function defined in (A.62)-(A.64) is one-dimensional, it depends on one single variable. Multi-dimensional δ-functions, depending on several independent variables, are similarly useful. An important example is considered below.

A simple calculation will show that

$$\nabla^2\frac{1}{r}=0\quad\text{for}\quad\boldsymbol{r}\neq0. \tag{A.68}$$

For $\boldsymbol{r}=0$ the expression is strictly speaking not defined. At the same time,

$$\oint_{\mathcal{A}}\mathrm{d}^2\boldsymbol{\mathcal{A}}\cdot\nabla\frac{1}{r}=-4\pi \tag{A.69}$$

for any surface $\mathcal{A}$ enclosing the origin. If we apply Gauss integral theorem (A.40) to (A.69), the surface integral is transformed into a volume integral for which the integrand according to (A.68) vanishes identically for $\boldsymbol{r}\neq0$, yet the integral takes a non-vanishing value. This dilemma can be amended by realizing that

$$\delta(\boldsymbol{r}-\boldsymbol{r}_0)=-\frac{1}{4\pi}\nabla^2\frac{1}{|\,\boldsymbol{r}-\boldsymbol{r}_0\,|} \tag{A.70}$$

is a three-dimensional δ-function satisfying the requirements

$$\int_{\mathcal{V}}\delta(\boldsymbol{r}-\boldsymbol{r}_0)\,\mathrm{d}^3\boldsymbol{r}=1\quad\text{if}\quad\boldsymbol{r}_0\in\mathcal{V} \tag{A.71}$$

$$\delta(\boldsymbol{r}-\boldsymbol{r}_0)=0\quad\text{for}\quad\boldsymbol{r}\neq\boldsymbol{r}_0 \tag{A.72}$$

$$\int_{\mathcal{V}}g(\boldsymbol{r})\,\delta(\boldsymbol{r}-\boldsymbol{r}_0)\,\mathrm{d}^3\boldsymbol{r}=g(\boldsymbol{r}_0)\quad\text{if}\quad\boldsymbol{r}_0\in\mathcal{V}. \tag{A.73}$$

For the evaluation of integrals containing multi-dimensional δ-functions of more complex arguments again a change of integration variables is required. This time the Jacobian, the

absolute value of the determinant formed from all partial derivatives between the two sets of integration variables, will appear

$$\int_{\mathcal{V}} g(\boldsymbol{r})\,\delta(\boldsymbol{f}(\boldsymbol{r}))\,\mathrm{d}^3\boldsymbol{r} = \int \frac{g(\boldsymbol{r})}{J(\boldsymbol{r})}\,\delta(\boldsymbol{f})\,\mathrm{d}^3\boldsymbol{f} = \sum_i \left.\frac{g(\boldsymbol{r}_i)}{J(\boldsymbol{r}_i)}\right|_{\boldsymbol{f}(\boldsymbol{r}_i)=0 \text{ and } \boldsymbol{r}_i\in\mathcal{V}} \tag{A.74}$$

with

$$J(\boldsymbol{r}) = \frac{\partial(\boldsymbol{f}(\boldsymbol{r}))}{\partial(\boldsymbol{r})} = \left\| \begin{array}{ccc} \dfrac{\partial f_x}{\partial x} & \dfrac{\partial f_x}{\partial y} & \dfrac{\partial f_x}{\partial z} \\ \dfrac{\partial f_y}{\partial x} & \dfrac{\partial f_y}{\partial y} & \dfrac{\partial f_y}{\partial z} \\ \dfrac{\partial f_z}{\partial x} & \dfrac{\partial f_z}{\partial y} & \dfrac{\partial f_z}{\partial z} \end{array} \right\|. \tag{A.75}$$

In (A.75) the explicit form for the Jacobi determinant was written in terms of cartesian coordinates, $\boldsymbol{r} = (x, y, z)$.

Quiz A.24 : Show that $\int_{-\infty}^{\infty} g(x)\,\delta(x^2-1)\,\mathrm{d}x = \frac{1}{2}(g(-1)+g(1))$. What is the corresponding value of $\int_{-\infty}^{\infty} g(x)\,\delta(x^3-1)\,\mathrm{d}x$?

Appendix B

Solutions to Selected Problems

From chapter 3:

Quiz 3.6: We note that

$$[\nabla^2, x]\,\psi \equiv \nabla^2(x\psi) - x\nabla^2\psi = \nabla\cdot(\psi\nabla x + x\nabla\psi) - x\nabla^2\psi = 2\frac{\partial}{\partial x}\psi.$$

From this result it follows that

$$[H, \boldsymbol{r}] = -\frac{\imath\hbar}{m}\boldsymbol{p}.$$

Similarly, we note that

$$[H, \boldsymbol{p}]\,\psi = [U(r), \boldsymbol{p}]\,\psi = U\boldsymbol{p}\,\psi - \boldsymbol{p}\,(U\psi) = \imath\hbar\nabla U\,\psi.$$

Substituting $Q = \boldsymbol{r}$ and $Q = \boldsymbol{p}$ in (4.12) and noting that the operators $\boldsymbol{r}$ and $\boldsymbol{p}$ are both independent of t (the wave function ψ and therefore $\langle\,\boldsymbol{r}\,\rangle$ and $\langle\,\boldsymbol{p}\,\rangle$ may depend on t!) we find

$$m\frac{\mathrm{d}}{\mathrm{d}t}\langle\,\boldsymbol{r}\,\rangle = \langle\,\boldsymbol{p}\,\rangle \quad \text{and} \quad \frac{\mathrm{d}}{\mathrm{d}t}\langle\,\boldsymbol{p}\,\rangle = -\langle\,\nabla U(r)\,\rangle.$$

That is, the mean values of quantum mechanical operators $\boldsymbol{r}$ and $\boldsymbol{p}$ behaves as the classical quantities $\boldsymbol{r}$ and $\boldsymbol{p}$!

Quiz 3.7:
With $A' \equiv A - \langle\,A\,\rangle$ and similarly for B we may write

$$\begin{aligned}
(\Psi, [A, B]\,\Psi) &= (\Psi, [A', B']\,\Psi) \quad \langle\,A\,\rangle \text{ and } \langle\,B\,\rangle \text{ are numbers} \\
&= (A'\Psi, B'\Psi) - (B'\Psi, A'\Psi) \quad \text{operators } A' \text{ and } B' \text{ are Hermitian} \\
&= (A'\Psi, B'\Psi) - (A'\Psi, B'\Psi)^* \quad \text{from definition of scalar product} \\
&= 2\imath\,\mathrm{Im}\left((A'\Psi, B'\Psi)\right)
\end{aligned}$$

and thus

$$\begin{aligned}
\frac{1}{2}\,|(\Psi, [A, B]\,\Psi)| &\leq |(A'\Psi, B'\Psi)| \quad \text{imaginary part less than absolute value} \\
&\leq \langle\,\Delta A\,\rangle\langle\,\Delta B\,\rangle \quad \text{from Schwartz inequality, see section 2.6.2}
\end{aligned}$$

The minimum value of the product of the uncertainties in mean values $\langle\,A\,\rangle$ and $\langle\,B\,\rangle$ is determined by the commutator of the corresponding operators.

Appendix C

Physical Constants

Fundamental and derived constants (1986–standard)

Speed of light (in vacuum)	$c = 2.997\,924\,58 \cdot 10^{8}$ m/s
Permeability of vacuum	$\mu_0 = 4\pi \cdot 10^{-7}$ H/m
Permittivity of vacuum, $1/\mu_0 c^2$	$\epsilon_0 = 8.854\,187\,817 \cdot 10^{-12}$ F/m
Gravitational constant	$G = 6.672\,59(85) \cdot 10^{-11}$ N m^2/kg^2
Planck constant	$h = 6.626\,075\,5(40) \cdot 10^{-34}$ J s
$h/2\pi$	$\hbar = 1.054\,572\,66(63) \cdot 10^{-34}$ J s
Elementary charge	$e = 1.602\,177\,33(49) \cdot 10^{-19}$ C
Electron mass	$m_e = 9.109\,389\,7(54) \cdot 10^{-31}$ kg
Proton mass	$m_p = 1.672\,623\,1(10) \cdot 10^{-27}$ kg
Neutron mass	$m_n = 1.674\,928\,6(10) \cdot 10^{-27}$ kg
Atomic mass unit, $m(^{12}C)/12$	$m_u = 1.660\,540\,2(10) \cdot 10^{-27}$ kg
Boltzmann constant	$\kappa_B = 1.380\,658(12) \cdot 10^{-23}$ J/K
Fine–structure constant, $\alpha = \dfrac{e^2}{4\pi\epsilon_0\hbar}$	$\alpha = 7.297\,353\,08(33) \cdot 10^{-3}$
inverse fine–structure constant	$\alpha^{-1} = 137.035\,989\,5(61)$
Rydberg constant, $R_\infty = \dfrac{mc^2}{2hc} \cdot \alpha^2$	$R_\infty = 10\,973\,731.534(13)$ m^{-1}
Bohr radius, $a_B = \dfrac{4\pi\epsilon_0\hbar^2}{me^2}$	$a_B = 0.529\,177\,249(24) \cdot 10^{-10}$ m
Stefan–Boltzmann constant, $\sigma = \dfrac{2\pi^5}{15h^3c^2}$	$\sigma = 1.56055 \cdot 10^{84}$ W m^{-2} J^{-4} (or $\sigma = 5.67051 \cdot 10^{-8}$ W m^{-2} K^{-4})

Selected astronomical constants

Solar mass	$M_\odot = 1.99 \cdot 10^{30}$ kg
Solar radius	$R_\odot = 6.96 \cdot 10^{8}$ m
Solar central temperature	$T_{\odot c} = 13.6 \cdot 10^{6}$ K
Solar surface temperature	$T_{\odot s} = 5760$ K
Astronomical unit	AU $= 1.496 \cdot 10^{11}$ m
Parsec	pc $= 3.09 \cdot 10^{16}$ m

Index