

Astrophysical Radiation Processes
— May you grasp the message from light

(This book is meant to be free for everyone)

Wenbin Lu

June 23, 2025

Contents

1	Radiation Fields	1
1.1	Primer	1
1.2	EM waves far from the source	2
1.2.1	Wave equation in vacuum	2
1.2.2	Fourier transformation and plane-wave solutions	3
1.2.3	Gaussian wavepacket	7
1.2.4	Fluence and flux density	8
1.3	Polarization	10
1.3.1	Jones vector for monochromatic wave (elliptically polarized)	10
1.3.2	Stokes parameters and Poincaré sphere	12
1.3.3	*Depolarization (under construction)	16
1.4	Phase-space distribution function for photons	16
1.4.1	Photon occupation number	16
1.4.2	Intensity and specific intensity	18
1.4.3	Moments of intensity: energy density, energy flux & pressure	21
1.5	Homework	25
2	Radiative Transfer	27
2.1	Thermal equilibrium and blackbody radiation	27
2.1.1	Local thermodynamic equilibrium (LTE)	27
2.1.2	Saha-Boltzmann equation	29
2.1.3	*Mixture of hydrogen and helium in LTE	36
2.1.4	Blackbody and Bose-Einstein distribution	38
2.2	Einstein relations	43
2.2.1	Spontaneous emission, stimulated emission, true absorption	43
2.2.2	Einstein coefficients: A , B_{12} , and B_{21}	47
2.2.3	Excitation temperature and brightness temperature	48
2.2.4	Jargons on different temperatures	51
2.3	Radiative transfer equation (without scattering)	53

2.3.1	Emission, absorption, mean free path, and optical depth	54
2.3.2	Formal solution	58
2.3.3	Kirchhoff's law	59
2.3.4	Absorption cross-section for a spectral line	60
2.3.5	*Generalized Kirchhoff's law for anisotropic distribution function .	62
2.4	Emission and absorption lines	64
2.4.1	Emission or absorption?	65
2.4.2	Mechanisms for line broadening	65
2.4.3	Voigt profile	71
2.4.4	Equivalent width and curve of growth	73
2.4.5	Collisional (de-)excitation (under construction)	75
2.4.6	Critical electron density (under construction)	75
2.4.7	P-Cygni profile	75
2.4.8	Hydrogen 21cm line	76
2.5	Photon diffusion under repetitive scatterings	77
2.5.1	Random walk	77
2.5.2	Fick's law for diffusion	79
2.5.3	*Fokker-Planck equation	81
2.5.4	*Infinite uniform medium	83
2.5.5	*Finite spherical uniform medium	85
2.6	Scattering and absorption	91
2.6.1	Eddington approximation and effective optical depth	92
2.6.2	Rosseland-mean opacity and Planck-mean opacity	95
2.6.3	Eddington luminosity	97
2.6.4	*Gravity darkening (under construction)	98
2.6.5	*Limb darkening for grey atmosphere	98
2.6.6	*Uniform slab with two-stream boundary conditions	102
2.6.7	*Moments of radiative transfer equation	109
2.7	Radiative transfer in a spherical dense wind	115
2.8	Homework	119
3	Special Relativity	122
3.1	Events and Minkowski spacetime	122
3.2	Proper time, 4-velocity, and Lorentz factor	128
3.3	Lorentz transformation	129
3.4	Four-vectors	132
3.5	Various relativistic effects	136
3.5.1	Apparent motion of a moving source	136
3.5.2	Doppler factor	138
3.5.3	Time dilation and length contraction	140
3.5.4	*Relative velocity	142

3.5.5	*Threshold energy for inelastic collisions	143
3.5.6	*Extreme elastic and inelastic collisions	147
3.5.7	*A rigid rod moving in both frames	149
3.6	Lorentz transformation of EM fields	151
3.6.1	Monochromatic plane EM waves	151
3.6.2	EM fields	155
3.6.3	Coulomb field of a moving charge	157
3.7	Lorentz invariant quantities of radiation field	160
3.7.1	Phase-space distribution function	160
3.7.2	Optical depth, I_ν/ν^3 , j_ν/ν^2 , emitting power, etc.	162
3.7.3	Lorentz transformation of radiation energy density	164
3.7.4	*Lorentz invariance of polarization state	165
3.8	Energy-momentum tensor (or stress-energy tensor)	166
3.9	*Pair production at relativistic temperatures (LTE)	170
3.10	Homework	172
4	Moving Charges	175
4.1	EM fields of moving charges	175
4.1.1	Causality and retarded time	175
4.1.2	Larmor formula	179
4.1.3	Retarded 4-potential	183
4.1.4	Liénard-Wiechart 4-potential for a point charge	185
4.2	Motion of charged particles in EM fields	190
4.2.1	Equation of motion	190
4.2.2	Uniform B (helical orbit, cyclotron frequency, Larmor radius)	191
4.2.3	Uniform E & B (E-cross-B drift plus helical orbit)	196
4.2.4	Lagrangian and Hamiltonian	198
4.2.5	*Strong EM wave (“figure-8” motion)	200
4.2.6	*Strong EM wave plus uniform B-field (stochastic acceleration)	202
4.2.7	*EM wave and Coulomb potential (under construction)	205
4.2.8	*Energy-momentum tensor for EM fields	207
4.2.9	*Abraham-Lorentz force for radiation reaction	209
4.3	Thomson scattering	210
4.3.1	Linearly polarized incident waves	210
4.3.2	Arbitrarily polarized incident waves	211
4.3.3	Plasma dispersion by a thin sheet	215
4.3.4	*Connection to the Optical Theorem	218
4.3.5	*Scattering in the presence of a strong B-field	220
4.4	Multipolar fields	223
4.4.1	Spatially confined, slowly varying, non-relativistic current system	223
4.4.2	Electric/Magnetic dipole radiation	228

4.4.3	Rayleigh scattering	229
4.4.4	Rotating diatomic molecule as an electric dipole	233
4.4.5	Hydrogen recombination lines as electric dipole radiation	234
4.4.6	Hydrogen 21cm line as magnetic dipole radiation	237
4.4.7	Rotating neutron star as a magnetic dipole	238
4.5	Homework	245
5	Plasma	248
5.1	Introduction	248
5.1.1	Charge neutrality and Debye shielding	248
5.1.2	Plasma oscillation (Langmuir wave)	251
5.1.3	Linear plane waves in quasi-homogeneous, quasi-stationary plasma .	254
5.2	Cold non-magnetized plasma	255
5.2.1	Linearized wave equation	255
5.2.2	Langmuir and EM modes	257
5.2.3	Phase speed, group speed, and wave dispersion	259
5.3	Cold magnetized plasma	263
5.3.1	General dispersion relation — an eigenvalue problem	264
5.3.2	Propagation parallel to B-field	268
5.3.3	Propagation along an arbitrary direction	270
5.3.4	Weakly magnetized plasma: Faraday rotation and conversion	270
5.3.5	Infinite magnetization: Alfvén, fast magnetosonic, O, and X modes .	279
5.4	Weakly inhomogeneous plasma	283
5.4.1	Geometrical optics: Eikonal approximation	284
5.4.2	*Adiabatic walking (under construction)	287
5.4.3	*Wave diffraction by a phase screen (under construction)	287
5.5	*Warm plasma (under construction)	287
5.6	Homework	287
6	Free-free Process	290
6.1	Coulomb collision	290
6.1.1	Impulse approximation and Coulomb relaxation	290
6.1.2	Coulomb logarithm	296
6.1.3	*Deflection angle in the reduced 2-body problem	300
6.1.4	*Rutherford's formula for the differential cross-section	302
6.2	Free-free emission	303
6.2.1	Single-velocity case	303
6.2.2	Thermal velocity distribution and cooling timescale	309
6.2.3	Emission measure and volume-filling factor	314
6.3	Free-free absorption	315
6.3.1	Absorption opacity for a thermal plasma	315

6.3.2	Rosseland- and Planck-mean opacities	317
6.3.3	Free-free spectrum without electron scattering	317
6.3.4	*Free-free spectrum with electron scattering	318
6.3.5	*Absorption of strong EM waves	323
6.4	Radio/Infrared continuum spectrum of an ionized wind	323
6.5	*Relativistic bremsstrahlung and Coulomb losses	327
6.5.1	*Emitting power and spectrum	328
6.5.2	*Coulomb energy-loss rate	333
6.5.3	*Bethe formula and mass stopping power	334
6.6	Homework	338
7	Compton Scattering	342
7.1	Kinematics of the Compton process	342
7.2	Klein-Nishina differential cross-section	344
7.3	Single inverse-Compton (IC) scattering	348
7.3.1	Two Lorentz transformations	348
7.3.2	Emitting power (<i>isotropic</i> photon/electron distribution)	352
7.3.3	Emitting spectrum (<i>isotropic</i> photon distribution)	356
7.3.4	Thermal and power-law electron distributions	363
7.4	Repeated scatterings and Comptonization	367
7.4.1	Compton- y parameter	367
7.4.2	Optically thin thermal electrons	368
7.4.3	Non-relativistic Comptonization and Compton temperature	369
7.4.4	*Stimulated Compton scattering for cold electrons	371
7.4.5	Kompaneets equation and Compton equilibrium	375
7.5	Homework	380
8	Synchrotron and Curvature Emission	382
8.1	Emitting power from an electron in a helical orbit	382
8.1.1	Helical orbit	382
8.1.2	Emitting power and cooling timescale	383
8.2	Synchrotron/curvature emission	384
8.2.1	Characteristic frequency and peak spectral power	384
8.2.2	Emission from instantaneous circular orbit	386
8.2.3	From circular to helical orbits	392
8.3	Synchrotron spectrum	397
8.3.1	Single and isotropic pitch angles	397
8.3.2	Power-law and thermal electron distributions	399
8.3.3	Effects of electron cooling	401
8.3.4	*Synchrotron burn-off limit — maximum photon energy	405
8.4	Synchrotron absorption	406

8.4.1	Applying the generalized Kirchhoff's law	406
8.4.2	Combining emission and absorption	411
8.4.3	*Conditions for synchrotron maser (under construction)	413
8.5	*Polarization	413
8.6	Synchrotron self-Compton	415
8.6.1	Compton- y parameter again	415
8.6.2	Maximum brightness temperature	416
8.7	Homework	418

Chapter 1

Radiation Fields

This chapter provides an overview of how radiation fields are described, including concepts like Fourier decomposition of a wavepacket, polarization, distribution function of photons, photon occupation number, intensity, energy density/flux, and radiation pressure. Throughout this book, we will use CGS/Gaussian units. To go from SI to Gaussian units, simply set the vacuum permittivity $\epsilon_0 \rightarrow (4\pi)^{-1}$ and vacuum permeability $\mu_0 \rightarrow 4\pi$, and then add appropriate factors of speed of light c so that the electric field strength E and magnetic field strength B have the same dimension.

1.1 Primer

We start by briefly reviewing the Maxwell equations and Poynting flux vector in classical electrodynamics. The Maxwell equations are

$$\begin{cases} \nabla \cdot \mathbf{E} = 4\pi\rho, & \text{— Coulomb's law,} \\ c\nabla \times \mathbf{E} = -\partial_t \mathbf{B}, & \text{— Faraday's law of induction,} \\ \nabla \cdot \mathbf{B} = 0, & \text{— Absence of magnetic charge,} \\ c\nabla \times \mathbf{B} = 4\pi\mathbf{J} + \partial_t \mathbf{E}, & \text{— Ampere's law (inc. displacement current),} \end{cases} \quad (1.1)$$

where ρ is the electric charge density and \mathbf{J} is the electric current density. The energy density of the electromagnetic (EM) field is

$$U_{\text{EM}} = (E^2 + B^2)/(8\pi). \quad (1.2)$$

The work needed to assemble a static charge distribution $\rho(\mathbf{x})$ is given by the total energy stored in the electric fields, i.e., $\int_V (E^2/8\pi) d^3x$. Likewise, the work needed to start the

current density $\mathbf{J}(\mathbf{x})$ is given by the total magnetic energy $\int_V (B^2/8\pi) d^3x$. The energy flux carried by the EM fields is given by the *Poynting vector*

$$\mathbf{S} = (\mathbf{E} \times \mathbf{B})c/(4\pi), \quad (1.3)$$

which describes the amount of energy per unit area per unit time flowing across a surface perpendicular to the vector \mathbf{S} .

1.2 EM waves far from the source

In this section, we study the solutions to the Maxwell equations sufficiently far away from any charged particles where the source terms ρ and \mathbf{J} vanish.

1.2.1 Wave equation in vacuum

By taking the curl of Faraday's law of induction, we obtain the *general wave equation*

$$\nabla \times (\nabla \times \mathbf{E}) + \frac{1}{c^2} \partial_t^2 \mathbf{E} = -\frac{4\pi}{c^2} \partial_t \mathbf{J}. \quad (1.4)$$

This is an inhomogeneous wave equation where the right-hand side is the source term (changing current density) and the left-hand describes the wave propagation as we will see below. If the current density \mathbf{J} scales linearly with the electric field \mathbf{E} (as is the case for low-amplitude EM waves propagating through a plasma), then the wave equation is linear. The defining property of a linear equation is the rule of *linear superposition*: if \mathbf{E}_1 and \mathbf{E}_2 are solutions, then any linear combination of the two, $a_1 \mathbf{E}_1 + a_2 \mathbf{E}_2$ (a_1 and a_2 are constants), is also a solution.

We focus on the solution of the electric field $\mathbf{E}(t, \mathbf{x})$ that our telescopes (which is effectively an antenna) are sensitive to, and it is straightforward to recover the magnetic field $\mathbf{B}(t, \mathbf{x})$ if needed (see eq. 1.23 later). In later chapters, we will use the wave equation to calculate the waves emitted from a system of currents $\mathbf{J}(t, \mathbf{x})$. For now, let us consider EM fields in vacuum with $\rho = J = 0$. Since $\nabla \times (\nabla \times \mathbf{E}) = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E}$ and $\nabla \cdot \mathbf{E} = 0$, we obtain the *vacuum wave equation*

$$\nabla^2 \mathbf{E} - \frac{1}{c^2} \partial_t^2 \mathbf{E} = 0. \quad (1.5)$$

This is a homogeneous linear equation as the source term vanishes and both terms on the left-hand side depend linearly on \mathbf{E} .

The rule of linear superposition allows us to use complex notations for mathematical convenience and a physical measurable quantity corresponds to the real part of its complex solution. Note that some (rarely encountered) non-linear effects may cause the rule of linear superposition to break down, typically in a plasma instead of vacuum. For instance,

consider a large-amplitude EM wavepacket interacting with a plasma, the plasma current density driven by the wave electric field has non-linear terms that depend on the wave amplitude squared (e.g., E^2 , see Chapter 4), and if the non-linear terms become important, we must always use the real physical quantities instead of the complex ones. In the following, we assume that linear superposition is always valid and hence each wave component can be analyzed independently.

At sufficiently large distances from a given source¹, all wave quantities can only depend on the coordinate z along the direction from the source to the observer and time t . In other words, we choose our z -axis to be along the direction of wave propagation $\hat{\mathbf{k}}$. For this one-dimensional case, the vacuum wave equation becomes

$$\partial_z^2 \mathbf{E} - \frac{1}{c^2} \partial_t^2 \mathbf{E} = 0. \quad (1.6)$$

From $\nabla \cdot \mathbf{E} = \partial_z(\hat{\mathbf{k}} \cdot \mathbf{E}) = 0$, $\nabla \cdot \mathbf{B} = \partial_z(\hat{\mathbf{k}} \cdot \mathbf{B}) = 0$, we know that the vacuum EM waves far from the source must be *transverse* such that $\mathbf{E} \perp \hat{\mathbf{k}}$ and $\mathbf{B} \perp \hat{\mathbf{k}}$. The solution $\mathbf{E}(t, z)$ only has transverse components $E_j(t) = \hat{\mathbf{e}}_j \cdot \mathbf{E}(t)$, where $j = x, y$ and $\hat{\mathbf{e}}_j$ is the corresponding unit base vector. For this reason, the above vector equation can be written as two independent scalar equations (for $j = x, y$)

$$\partial_z^2 E_j - \frac{1}{c^2} \partial_t^2 E_j = 0. \quad (1.7)$$

This shows that EM waves in vacuum can be decomposed into *linearly polarized eigenmodes*, each of which preserves its polarization along the propagation. For this reason, we often use linear basis. As we show later in §1.3, a superposition of two orthogonal linearly polarized monochromatic waves with different phases produces elliptical polarization, so vacuum EM waves can also be decomposed into circularly polarized eigenmodes and hence circularly polarized basis is also commonly used.

In the following, we will use the method of Fourier decomposition to show that the solution to the 1D scalar vacuum wave equation (1.7) is in the form of $E(t, z) \propto e^{i(kz - \omega t)}$, which describes a plane wave with a fixed linear polarization (e.g., $E = E_x$).

1.2.2 Fourier transformation and plane-wave solutions

The Fourier transformation of a generally complex function $g(t)$ is

$$\tilde{g}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} g(t) e^{i\omega t} dt, \quad (1.8)$$

¹For multiple sources, the final solution is given by the linear superposition of the waves from each individual source.

and the inverse transformation is

$$g(t) = \int_{-\infty}^{\infty} \tilde{g}(\omega) e^{-i\omega t} d\omega. \quad (1.9)$$

These transformations make the time (t) and frequency (ω) domains equivalent to each other. The magnitudes² $|g(t)|$ and $|\tilde{g}(\omega)|$ are related by the *Parseval's theorem*, which states that the total “energy” can be expressed either by a time-integral of the “intensity” (proportional to the amplitude squared) or by a frequency-integral of the spectral power (proportional to the Fourier amplitude squared)

$$\int_{-\infty}^{\infty} |g(t)|^2 dt = 2\pi \int_{-\infty}^{\infty} |\tilde{g}(\omega)|^2 d\omega. \quad (1.10)$$

The real part of $g(t)$ is given by

$$\operatorname{Re} g(t) = \frac{g(t) + g^*(t)}{2}, \quad (1.11)$$

where g^* is the complex conjugate of g . Later on when we discuss the fluence and flux density of EM waves (which depend on the real field squared), we will need the Fourier transformation of the real part of $g(t)$, which is defined as

$$\tilde{g}_R(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \operatorname{Re} g(t) e^{i\omega t} dt = \frac{\tilde{g}(\omega) + \tilde{g}^*(-\omega)}{2}, \quad (1.12)$$

and the inverse transformation is given by

$$\operatorname{Re} g(t) = 2 \operatorname{Re} \int_0^{\infty} \tilde{g}_R(\omega) e^{-i\omega t} d\omega. \quad (1.13)$$

The Parseval's theorem that relates $\operatorname{Re} g(t)$ and $|\tilde{g}_R(\omega)|$ is given by

$$\int_{-\infty}^{\infty} [\operatorname{Re} g(t)]^2 dt = 2\pi \int_{-\infty}^{\infty} |\tilde{g}_R(\omega)|^2 d\omega = 4\pi \int_0^{\infty} |\tilde{g}_R(\omega)|^2 d\omega, \quad (1.14)$$

where we have used $|\tilde{g}_R(-\omega)|^2 = |\tilde{g}_R(\omega)|^2$ because $\tilde{g}_R(-\omega) = \tilde{g}_R^*(\omega)$.

After reviewing Fourier transformation, let us then use the *method of Fourier decomposition* to solve the 1D scalar wave equation $\partial_z^2 E - c^{-2} \partial_t^2 E = 0$ (eq. 1.7). Any solution $E(t, z)$ can be written by a superposition of its Fourier components

$$E(t, z) = \int_{-\infty}^{+\infty} \tilde{E}(z, \omega) e^{-i\omega t} d\omega. \quad (1.15)$$

²Hereafter, $|a| = \sqrt{aa^*}$ is the magnitude of a complex number a and a^* is the complex conjugate of a .

We then plug the above form into the 1D scalar wave equation and obtain

$$\int_{-\infty}^{+\infty} \tilde{g}(z, \omega) e^{-i\omega t} d\omega = g(t, z) = 0, \quad \forall t \in (-\infty, +\infty), \quad (1.16)$$

where we have defined a new function \tilde{g} that depends on $\tilde{E}(z, \omega)$ as follows

$$\tilde{g}(z, \omega) \equiv \partial_z^2 \tilde{E}(z, \omega) + \frac{\omega^2}{c^2} \tilde{E}(z, \omega), \quad (1.17)$$

and $g(t, z)$ is the inverse Fourier transformation of the function $\tilde{g}(z, \omega)$. The fact that $g(t, z) = 0$ at all times t means that $\tilde{g}(z, \omega) = 0$ at all frequencies ω . Thus, we obtain

$$\partial_z^2 \tilde{E}(z, \omega) + \frac{\omega^2}{c^2} \tilde{E}(z, \omega) = 0, \quad \forall \omega \in (-\infty, +\infty). \quad (1.18)$$

For a given frequency ω , the above equation (1.18) describes a simple harmonic oscillator sliding along the z -axis and its solution is in the form of

$$\tilde{E}(z, \omega) = \tilde{E}_0(\omega) e^{ikz}, \quad (1.19)$$

where $\tilde{E}_0(\omega)$ is a complex amplitude and k is the spatial wavenumber. In order for the equation (1.18) to hold for all positions z , we require

$$k^2 = \omega^2/c^2 \quad \text{or} \quad k = \pm\omega/c, \quad (1.20)$$

which is called the *dispersion relation* for vacuum EM waves. This is a special case for a more general dispersion relation where the wavenumber depends non-linearly on the wave frequency. The linear dispersion relation (eq. 1.20) here describes a *non-dispersive medium*, and in this case, the shape of a wavepacket is exactly preserved along the propagation at speed of light (see Fig. 1.1). For this reason, the time-dependent solution of $E(t, z)$ is entirely determined by the initial condition $E(t = 0, z)$.

In the following, we will focus on the case with wavenumber $k = \omega/c$ without losing generality, because the other case with $k = -\omega/c$ simply means that the wave propagates in the $-z$ direction (we may reverse our z -axis to make $k = \omega/c$).

Finally, the general solution of the vacuum wave equation is obtained by an inverse Fourier transformation of $\tilde{E}(z, \omega)$, i.e.,

$$E_j(t, z) = \int_{-\infty}^{+\infty} \tilde{E}_{j,0}(\omega) e^{i(kz - \omega t)} d\omega, \quad (1.21)$$

where $j = x, y$ are the two orthogonal linear polarizations in the transverse directions. We conclude that the EM fields far away from an emitting source can be described by the linear

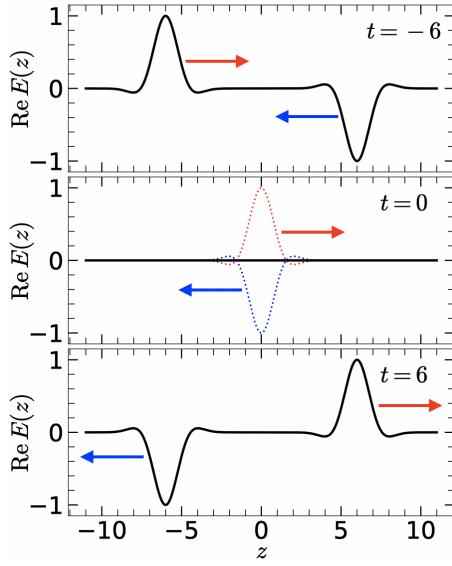


Figure 1.1: Linear superposition of two wavepackets propagating in opposite directions.

superposition of *monochromatic plane-wave components*. For each component, $\tilde{E}_{j,0}(\omega)d\omega$ is the *complex amplitude* and $e^{i(kz-\omega t)}$ is the *phase factor* which contains the spatial and temporal dependences. This is what we would have expected based on the linearity of the vacuum wave equation (1.6).

For our choice of $k = \omega/c$, we see that $kz - \omega t = -\omega(t - z/c)$ and hence the electric field only depends on the spatial and time coordinates through a single variable $\xi \equiv t - z/c$ — this is called the *light-cone coordinate* in Special Relativity. If we use ξ as the single coordinate, the solution takes the exact form of an inverse Fourier transformation

$$E_j(\xi) = \int_{-\infty}^{+\infty} \tilde{E}_{j,0}(\omega) e^{-i\omega\xi} d\omega. \quad (1.22)$$

The function $\tilde{E}_{j,0}(\omega)$ depends on the initial and boundary conditions. In §1.2.3, we will study the Gaussian wavepacket, where $\tilde{E}_{j,0}(\omega)$ is given by a Gaussian function.

Once the wave electric field has been found, the wave magnetic field can be obtained from $c^{-1}\partial_t \mathbf{B} = -\nabla \times \mathbf{E}$, and the result is

$$\mathbf{B}(\xi) = \hat{\mathbf{z}} \times \mathbf{E}(\xi). \quad (1.23)$$

Since \mathbf{E} is perpendicular to $\hat{\mathbf{z}}$ (the direction of propagation), we know that the magnetic field has the same *amplitude and phase* as the electric field's oscillations.

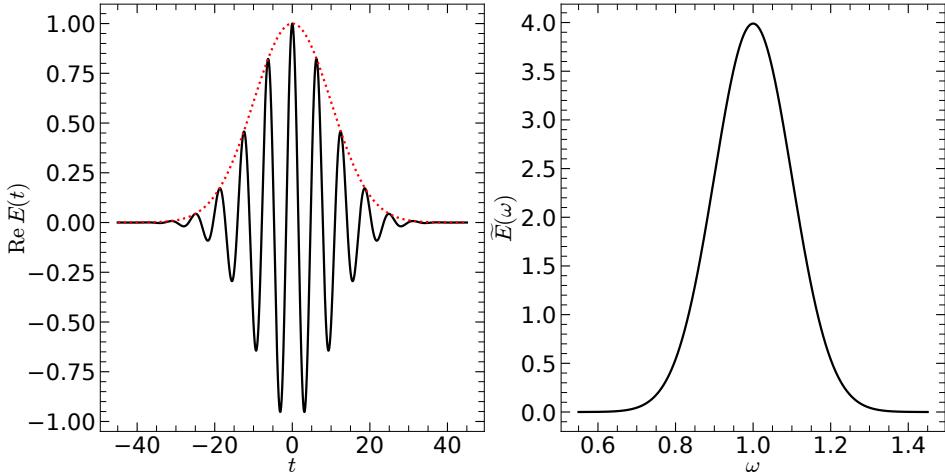


Figure 1.2: A Gaussian wavepacket centered at $t = 0$ and frequency $\omega_0 = 1$, and has frequency dispersion $\sigma_\omega = 0.1$. The left panel shows the real waveform $\text{Re } E(t)$ (black solid line) and the Gaussian amplitude envelope (red dotted line). The right panel shows the Fourier spectrum $\tilde{E}(\omega)$.

A potential concern is on the negative frequencies $\omega < 0$ in the inverse Fourier transform integral (e.g., eq. 1.22). In fact, there are four mathematically distinct plane-wave modes that satisfy the wave equation, and they have phase factors: $e^{-i\omega(t-z/c)}$, $e^{i\omega(t-z/c)}$, $e^{-i\omega(t+z/c)}$, and $e^{i\omega(t+z/c)}$. The first two have the equal phase front propagating along the $+z$ direction as time progresses forward, and the latter two propagate along $-z$. Our final physical solution corresponds to the real part of the complex notation, which is only used as an intermediate step for mathematical convenience (otherwise, the equations will be lengthy when we deal with many sine and cosine functions and their derivatives). The real parts of the first two modes, which are proportional to $e^{-i\omega(t-z/c)}$, $e^{i\omega(t-z/c)}$, are identical, and this means that they are the same *physical mode*. Thus, we do not lose generality by choosing the mode $e^{-i\omega(t-z/c)}$ (or $e^{-i\omega\xi}$ using our light-cone coordinate ξ), because in the end the physical initial and boundary conditions only constrain the physical mode.

1.2.3 Gaussian wavepacket

Let us consider a wavepacket whose Fourier spectrum is given by a Gaussian

$$\tilde{E}_0(\omega) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_\omega} \exp \left[-\frac{(\omega - \omega_0)^2}{2\sigma_\omega^2} \right], \quad (1.24)$$

where ω_0 is the central frequency (also called the *carrier frequency*) and σ_ω is the frequency dispersion that is related to width of the spectrum. The spectrum is normalized such that $\int_{-\infty}^{+\infty} \tilde{E}(\omega) d\omega = 1$.

The complex E-field waveform is then given by an inverse Fourier transformation

$$E(\xi) = \int_{-\infty}^{+\infty} \tilde{E}_0(\omega) e^{-i\omega\xi} d\omega = e^{-\sigma_\omega^2 \xi^2 / 2} e^{-i\omega_0 \xi}, \quad (1.25)$$

where we have used the light-cone coordinate $\xi = t - z/c$. The waveform has a Gaussian amplitude profile $e^{-\sigma_\omega^2 \xi^2 / 2}$ centered at $\xi = 0$ and an oscillating phase factor of $e^{-i\omega_0 \xi}$. When carrying out the above integral, we have used the following theorem for general Gaussian integral ([proof](#))

$$\int_{-\infty}^{\infty} e^{-p(y+c)^2} dy = \sqrt{\frac{\pi}{p}}, \quad \forall p, c \in \mathbb{C}, \operatorname{Re} p > 0. \quad (1.26)$$

An example of Gaussian wavepacket is shown in Fig. 1.2. A Gaussian wavepacket in the frequency domain is also a Gaussian in the time domain multiplied by a phase factor.

If we make the characteristic duration $\Delta t = \sigma_\omega^{-1}$ of a wavepacket very narrow, then σ_ω must be large and the wavepacket has a broad frequency spectrum with character bandwidth $\Delta\omega = \sigma_\omega$. The de Broglie wave nature of a given particle means that a state with energy ϵ is related to the wave frequency ω by $\epsilon = \hbar\omega$, so we have $\Delta\epsilon\Delta t \sim \hbar$ — this is the *Heisenberg Uncertainty Principle*.

1.2.4 Fluence and flux density

For radio observers, recording the full waveform $E(t)$ at a sampling rate of 2ν (*= the Nyquist rate*) using 8 bits = 1 bite per sample would give a data rate of 2 giga-bite/s for frequency $\nu = 1$ GHz. This would give a very large data volume that is difficult to store. A more convenient way is to store the time-averaged spectrum in a number of pre-defined frequency channels. If the total bandwidth is $\Delta\nu$ and the frequency resolution is $\delta\nu$, and we average the spectrum over an integration time T , then the data rate is given by $\Delta\nu/\delta\nu \times 1$ bite/ T = 1 kilo-bite/s for $\Delta\nu = 1$ GHz, $\delta\nu = 1$ MHz, and $T = 1$ s. This is much more manageable than storing the raw voltages, but the disadvantage is that the phase information is lost (the phase information is required for polarization measurement or interferometry).

In the following, we discuss how a simple-minded observer calculates time-averaged spectrum of a source, which is described by the flux density — energy received per unit area per unit time per unit frequency interval.

For a complex waveform $\mathbf{E}(t)$ passing through a given spatial position, the energy flux = transported energy per unit area per unit time is given by the Poynting vector

$$\begin{aligned} \mathbf{S}(t) &= \frac{c}{4\pi} (\operatorname{Re} \mathbf{E}(t)) \times (\operatorname{Re} \mathbf{B}(t)) = \frac{c}{4\pi} (\operatorname{Re} \mathbf{E}(t)) \times [\hat{\mathbf{z}} \times (\operatorname{Re} \mathbf{E}(t))] \\ &= \frac{c}{4\pi} [\operatorname{Re} E(t)]^2 \hat{\mathbf{z}}, \end{aligned} \quad (1.27)$$

where $\text{Re } E = \sqrt{(\text{Re } E_x)^2 + (\text{Re } E_y)^2}$ is the total real E-field strength, and we have made use of the fact that $\mathbf{B} = \hat{\mathbf{z}} \times \mathbf{E}$ and that $\mathbf{E} \perp \hat{\mathbf{z}}$. We see that the Poynting vector $\mathbf{S} = S\hat{\mathbf{z}}$ is along the direction of wave propagation. It is important to note that the energy flux depends on the wave amplitude squared, so we must use the real parts.

In the following, we focus on one component of linear polarization by taking $E = E_x$. In the end, the total fluence or flux density will simply be the sum of the contributions from the x- and y-polarization components.

It is convenient to *define* another Fourier transformation of the *real part* of the E-field as follows

$$\tilde{E}_R(\omega) \equiv \frac{1}{2\pi} \int_{-\infty}^{+\infty} \text{Re } E(t) e^{i\omega t} dt, \quad (1.28)$$

which is related to the Fourier transformation $\tilde{E}(\omega)$ of the complex waveform $E(t)$ by $\tilde{E}_R(\omega) = \tilde{E}(\omega) + \tilde{E}^*(-\omega)$. The *fluence* of the entire wavepacket = total transported energy per unit area is

$$\mathcal{F} = \frac{c}{4\pi} \int_{-\infty}^{\infty} [\text{Re } E(t)]^2 dt = c \int_0^{\infty} |\tilde{E}_R(\omega)|^2 d\omega, \quad (1.29)$$

where we have used the Parseval's theorem for a real waveform (eq. 1.14). Then, the *fluence density* = transported energy per unit area per unit frequency interval is

$$\mathcal{F}_{\nu} = \frac{d\mathcal{F}}{d\nu} = 2\pi\mathcal{F}_{\omega} = 2\pi \frac{d\mathcal{F}}{d\omega} = 2\pi c |\tilde{E}_R(\omega)|^2, \quad (1.30)$$

where the frequency ν is related to the angular frequency ω by $\nu = \omega/2\pi$. For the unit-amplitude Gaussian wavepacket discussed in §1.2.3, in the limit of a narrow frequency spectrum $\sigma_{\omega} \ll \omega_0$, we have $\tilde{E}_R(\omega) \approx \tilde{E}(\omega)/2$ (as $\tilde{E}^*(-\omega) \approx 0$ at negative frequencies), and the fluence and fluence density are given by

$$\mathcal{F} \approx \frac{c/\sigma_{\omega}}{8\sqrt{\pi}}, \quad \mathcal{F}_{\nu} \approx \frac{c}{4\sigma_{\omega}^2} e^{-(\omega-\omega_0)^2/\sigma_{\omega}^2}, \quad \text{for } \sigma_{\omega} \ll \omega_0. \quad (1.31)$$

If the statistical properties of the signal remain stable for a very long time $T \gg \omega^{-1}$ for typical frequencies in the Fourier spectrum (as is the case for a steady source), then the time-averaged *flux density* = transported energy per unit area per unit time per unit frequency is

$$F_{\nu} = \frac{dF}{d\nu} = 2\pi F_{\omega} = 2\pi c \lim_{T \rightarrow \infty} \frac{|\tilde{E}_T(\omega)|^2}{T}, \quad (1.32)$$

where

$$\tilde{E}_T(\omega) = \frac{1}{2\pi} \int_0^T \text{Re } E(t) e^{i\omega t} dt$$

is the Fourier transform of the real E-field over a finite time T .

The above discussion only involves one of the linearly polarized components E_x or E_y , and the total flux is given by the sum of the two components $F = F_x + F_y$ because $[\text{Re } E(t)]^2 = [\text{Re } E_x]^2 + [\text{Re } E_y]^2$. In the next section, we will consider the polarization, which carries rich information about the phase interplay between individual components $E_x(t)$ and $E_y(t)$.

1.3 Polarization

Wave polarization contains interesting information about the emitting source as well as the intervening matter along the line of sight. In this section, we discuss how polarization is measured by astronomers. The Cartesian coordinates we use here are conventionally defined such that \hat{z} (also the wavevector) points from the source towards the observer, \hat{x} points to the celestial north, and \hat{y} points to the celestial east. Regardless of the actual design details, a radio telescope can be effectively considered as two orthogonal antennae — knowing the actual orientation of the antennae with respect to the celestial north and east, one can easily apply a coordinate rotation to achieve this. Practically, the two antenna arms may have different gains, relative phase delays, and coupling — here let us assume that these can be calibrated out (see Problem 2).

1.3.1 Jones vector for monochromatic wave (elliptically polarized)

The whole problem of wave propagation in vacuum comes down to the understanding of individual monochromatic wave components. In the following, we consider one such wavelet at frequency ω . We adopt the convention of a positive wavenumber $k = \omega/c$, and a monochromatic wave describes an infinitely long sinusoidal wave train and can be written in the form of

$$E_j(t, z) = E_{j,0} e^{i\phi_j} e^{i(kz - \omega t)}, \quad (1.33)$$

where $E_{j,0}e^{i\phi_j}$ is the *complex amplitude*, $E_{j,0}$ is the *real amplitude* that equals to the maximum electric field strength along \hat{e}_j , and ϕ_j is the reference phase at $z = t = 0$. We study how the interplay among the two real amplitudes $E_{j,0}$ and the two phases ϕ_j describes the polarization of the wave.

The polarization state of a simple wave is fully described by the pair of complex amplitudes which form the *Jones vector* in linear basis³

$$\bar{E} = \begin{pmatrix} E_{x,0} e^{i\phi_x} \\ E_{y,0} e^{i\phi_y} \end{pmatrix}. \quad (1.34)$$

The physical field is the real part of the *full* complex expression $\text{Re}(\vec{E} e^{i(kz - \omega t)})$. The polarization state described by eq. (1.33) or (1.34) is in general *elliptically polarized* and

³It is also possible to express the Jones vector in circular polarization basis, and the analysis is similar.

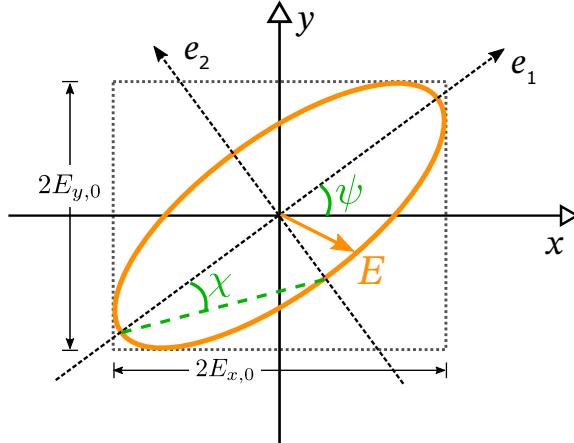


Figure 1.3: The tip of the electric vector of an elliptically polarized wave follows the orange ellipse, whose principle axes are inclined from the x - y frame by an angle ψ .

it depends on three independent variables: $E_{x,0}$, $E_{y,0}$, and $\Delta\phi = \phi_x - \phi_y$ (the absolute phase does not affect the description of a polarization state). The special case of $\Delta\phi = 0$ corresponds to linear polarization.

At position $z = 0$ (without losing generality), the real physical electric field is

$$E_j(t) = E_{j,0} \cos(\omega t - \phi_j). \quad (1.35)$$

The tip of this vector makes an ellipse with inclined principle axes, \hat{e}_1 and \hat{e}_2 , as shown in Fig. 1.3. Let us define the angle between the principle axis \hat{e}_1 and \hat{x} to be ψ (and $-\pi/2 < \psi \leq \pi/2$), and it is given by

$$\tan 2\psi = \tan 2\alpha \cos \Delta\phi, \quad \psi \equiv \arccos(\hat{e}_1 \cdot \hat{x}), \quad (1.36)$$

where the angle $0 \leq \alpha \leq \pi/2$ is defined as

$$\tan \alpha = E_{y,0}/E_{x,0}. \quad (1.37)$$

The angle ψ describes the orientation of the principle axes (\hat{e}_1 and \hat{e}_2) with respect to our adopted coordinate system.

The E-field components along the principle axes are given by

$$E_1 = E_x \cos \psi + E_y \sin \psi, \quad E_2 = -E_x \sin \psi + E_y \cos \psi. \quad (1.38)$$

The semimajor (a) and semiminor (b) axes of the ellipse are given by

$$\begin{aligned} a^2 &= E_{x,0}^2 \cos^2 \psi + E_{y,0}^2 \sin^2 \psi + 2E_{x,0}E_{y,0} \sin \psi \cos \psi \cos \Delta\phi, \\ b^2 &= E_{x,0}^2 \sin^2 \psi + E_{y,0}^2 \cos^2 \psi - 2E_{x,0}E_{y,0} \sin \psi \cos \psi \cos \Delta\phi. \end{aligned} \quad (1.39)$$

If we define another auxiliary angle $\pi/4 \leq \chi \leq \pi/4$ (whose meaning will be clear later in Fig. 1.4) such that

$$\sin 2\chi = \sin 2\alpha \sin \Delta\phi, \quad (1.40)$$

then the ellipticity is given by $b/a = |\tan \chi|$. We see that $\chi = 0$ corresponds to linear polarization and $\chi = \pm\pi/4$ gives right/left circular polarizations.

1.3.2 Stokes parameters and Poincaré sphere

Since the rapid oscillations on wave timescales are simply described by sinusoidal functions, it is more efficient, and practically only possible for faint sources, to measure and record the auto-correlation ($i = j$) or cross-correlation ($i \neq j$) functions averaged over a finite period of time T which contains a large number of wave oscillations. The correlation function between the i th and j th components of the complex waveform (at the same spatial point $z = 0$) is defined using the time-average notation $\langle \dots \rangle \equiv T^{-1} \int_0^T (\dots) dt$,

$$\langle E_i E_j^* \rangle = \frac{1}{T} \int_0^T E_i(t) E_j^*(t) dt, \quad (1.41)$$

where the rapid oscillations on a timescale of ω^{-1} has been removed from the integral (by multiplying the complex conjugate) and the time average is only applied to the potentially slowly varying complex amplitudes. For a monochromatic wave with a single polarization state as described by eq. (1.34), we have $\langle E_j E_j^* \rangle = E_{j,0}^2$ (auto-correlation) and $\langle E_x E_y^* \rangle = E_{x,0} E_{y,0} \exp(i\Delta\phi)$ (cross-correlation). The relationship between the correlation functions for complex waveform and for real fields is given by

$$\langle \text{Re } E_i \cdot \text{Re } E_j \rangle = \langle E_i E_j^* \rangle / 2, \text{ for monochromatic wave,} \quad (1.42)$$

where the factor of $1/2$ comes from $\int_0^T \cos^2 \omega t dt = T/2$ in the limit of $T \gg \omega^{-1}$.

Based on these correlation functions, we define a set of four real quantities — the *time-averaged Stokes parameters* for *quasi-monochromatic* waves,

$$\begin{aligned} I &= \langle E_x E_x^* \rangle + \langle E_y E_y^* \rangle, \\ Q &= \langle E_x E_x^* \rangle - \langle E_y E_y^* \rangle, \\ U &= 2 \text{Re} \langle E_x E_y^* \rangle, \\ V &= -2 \text{Im} \langle E_x E_y^* \rangle, \end{aligned} \quad (1.43)$$

which form the *Stokes vector*

$$\vec{s} = (I, Q, U, V).$$

Here, “quasi-monochromatic” means that we are only considering a narrow frequency bin $\Delta\omega (\ll \omega)$ near frequency ω in the Fourier spectrum of the waves within the duration T ,

and ‘‘time-averaged’’ means that we are averaging the signal over a duration T that is much longer than $(\Delta\omega)^{-1}$ but much shorter than the variability timescale of the source. Thus, the Stokes parameters are functions of time (in the T -averaged sense) and frequency.

For a monochromatic elliptical polarization state described by eq. (1.34), we obtain

$$\begin{aligned} I &= E_{x,0}^2 + E_{y,0}^2, \quad Q = E_{x,0}^2 - E_{y,0}^2, \\ U &= 2E_{x,0}E_{y,0} \cos \Delta\phi, \quad V = -2E_{x,0}E_{y,0} \sin \Delta\phi, \end{aligned} \quad (1.44)$$

where $\Delta\phi = \phi_x - \phi_y$ is the phase difference in the complex amplitudes in the x and y directions. Since $I^2 = Q^2 + U^2 + V^2$ (one can easily verify this), a monochromatic wavelet is 100% polarized. The meanings of the Stokes parameters are explained as follows.

The Stokes- I parameter is related to the total energy flux of the waves. If the width of a given frequency channel is $\delta\nu = \delta\omega/(2\pi)$ and the flux density near frequency ν is F_ν , then the flux inside this channel is $F_\nu\delta\nu = \langle (\text{Re } E)^2 \rangle c/4\pi$, so we obtain

$$I = 8\pi F_\nu \delta\nu / c. \quad (1.45)$$

It is often loosely stated that Stokes- I means the ‘‘intensity’’, but it should not be confused with the (*specific*) *intensity* of the radiation field (see §1.4.2).

The Stokes- Q measures the *difference* in the fluxes carried by the electric field components in the $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ directions. The Stokes- U measures the *difference* in the fluxes carried by the electric field components in the two diagonal directions described by the basis $\hat{\mathbf{e}}_> = (\hat{\mathbf{x}} + \hat{\mathbf{y}})/\sqrt{2}$ and $\hat{\mathbf{e}}_< = (\hat{\mathbf{x}} - \hat{\mathbf{y}})/\sqrt{2}$,

$$E_> = \mathbf{E} \cdot \hat{\mathbf{e}}_> = (E_x + E_y)/\sqrt{2}, \quad E_< = \mathbf{E} \cdot \hat{\mathbf{e}}_< = (E_x - E_y)/\sqrt{2}. \quad (1.46)$$

It can be shown that Stokes- Q can be expressed in the following form

$$\begin{aligned} U &= \langle E_> E_>^* \rangle - \langle E_< E_<^* \rangle \\ &= [\langle (E_x + E_y)(E_x + E_y)^* \rangle - \langle (E_x - E_y)(E_x - E_y)^* \rangle] / 2. \end{aligned} \quad (1.47)$$

We see that Q and U measure linear polarizations in orthogonal directions.

It should be noted that I , Q , and U only depends on the four auto-correlation functions: $\langle E_x E_x^* \rangle$, $\langle E_y E_y^* \rangle$, $\langle (E_x + E_y)(E_x + E_y)^* \rangle$, and $\langle (E_x - E_y)(E_x - E_y)^* \rangle$ — the latter two can be measured if the antenna arms are rotated. In fact, the orientations of the antenna arms do not have to be along $\hat{\mathbf{x}}$ (0°), $\hat{\mathbf{y}}$ (90°), $\hat{\mathbf{x}} + \hat{\mathbf{y}}$ (45°), and $\hat{\mathbf{x}} - \hat{\mathbf{y}}$ (-45°) — any four distinct angles will in principle be sufficient. The fact that I , Q , and U do not rely on any cross-correlation functions means that they do not require the calibration of the relative phase delay between the two antenna arms. Thus, these three Stokes parameters are usually the easiest to obtain, and they give the degree of linear polarization.

Finally, the Stokes- V parameter measures the flux *difference* between right and left circular polarizations. This is because the base vectors for *right circular polarization* (RCP) and *left circular polarization* (LCP) are

$$\hat{\mathbf{R}} = (\hat{\mathbf{x}} - i\hat{\mathbf{y}})/\sqrt{2}, \quad \hat{\mathbf{L}} = (\hat{\mathbf{x}} + i\hat{\mathbf{y}})/\sqrt{2}. \quad (1.48)$$

Therefore, the RCP electric field component is given by $E_R = \mathbf{E} \cdot \hat{\mathbf{R}} = (E_x - iE_y)/\sqrt{2}$, and the LCP component is $E_L = \mathbf{E} \cdot \hat{\mathbf{L}} = (E_x + iE_y)/\sqrt{2}$. It can be shown that Stokes- V can be written in the following way

$$\begin{aligned} V &= \langle E_R E_R^* \rangle - \langle E_L E_L^* \rangle \\ &= [\langle (E_x - iE_y)(E_x - iE_y)^* \rangle - \langle (E_x + iE_y)(E_x + iE_y)^* \rangle] / 2. \end{aligned} \quad (1.49)$$

Our definition follows the Institute of Electrical and Electronics Engineers (IEEE) convention such that $V > 0$ corresponds to RCP, for which the sense of rotation of the electric vector follows the *right-hand rule* with the thumb pointing in the direction of the wavevector $\hat{\mathbf{k}}$. This convention is commonly adopted by radio astronomers and is in agreement with definition of handedness for the quantum mechanical spin. However, this can be very confusing because, if the wavevector is along the $\hat{\mathbf{z}}$ axis, the electric vector of a RCP mode at a given position (e.g., $z = 0$) rotates *counter-clockwise* in the x - y plane (in the sense that the polarization position angle increases with time).

We note that, for data collected by linear antenna arms, a phase shift of $\pm\pi/2$ must be added to $E_y(t)$ to obtain the RCP (E_R) and LCP (E_L) components, this is because $\pm i = e^{\pm i\pi/2}$. Thus, it is crucial to obtain an accurate calibration of the relative instrumental phase delay between the two antenna arms (see Exercise 2).

Sometimes we are only interested in the polarization properties (e.g., polarization fractions and polarization angle) instead of the absolute flux. It is useful to define the *normalized Stokes parameters*: $q = Q/I$, $u = U/I$, and $v = V/I$. In the following, we discuss a few important polarization properties that are defined based on q, u, v .

Based on their definitions, one can easily show that the Stokes parameters satisfy

$$Q^2 + U^2 + V^2 \leq I^2 \Leftrightarrow q^2 + u^2 + v^2 \leq 1. \quad (1.50)$$

The *total polarization fraction* (also called the polarization fraction) is defined as

$$\Pi = \frac{\sqrt{Q^2 + U^2 + V^2}}{I} = \sqrt{q^2 + u^2 + v^2} \leq 1. \quad (1.51)$$

Partially polarized waves have $0 < \Pi < 1$ and *unpolarized* waves correspond to $\Pi = 0$. The *linear polarization fraction* is defined as

$$\Pi_{\text{lin}} = \frac{\sqrt{Q^2 + U^2}}{I} = \sqrt{q^2 + u^2} \leq 1, \quad (1.52)$$

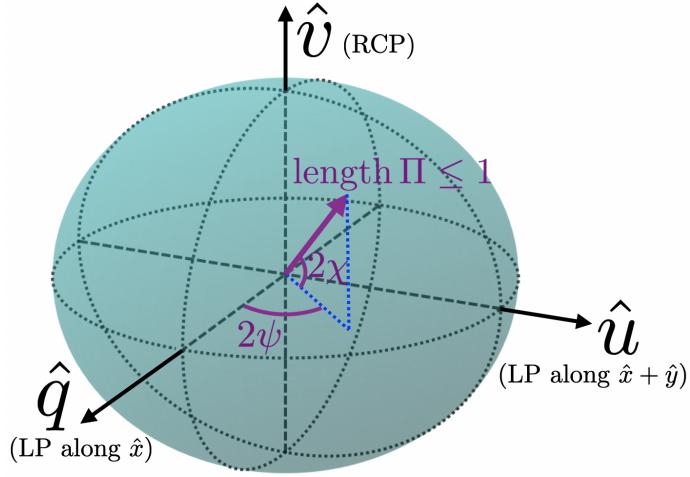


Figure 1.4: Poincaré sphere representation of the 3D polarization vector $\mathbf{p} = (q, u, v)$. The length of the vector is $|\mathbf{p}| = \Pi \leq 1$, where Π is the total polarization fraction. The “longitude” angle is 2ψ and “latitude” angle is 2χ , which are defined in eq. (1.54).

and the *circular polarization fraction* is given by

$$\Pi_{\text{cir}} = \frac{|V|}{I} = |v| \leq 1. \quad (1.53)$$

Based on the Stokes parameters $\vec{s} = (I, Q, U, V)$, we can define two angles ψ and χ by

$$\begin{aligned} \tan 2\psi &\equiv \frac{U}{Q} = \frac{u}{q}, \\ \sin 2\chi &\equiv \frac{V}{\sqrt{Q^2 + U^2 + V^2}} = \frac{v}{\sqrt{q^2 + u^2 + v^2}} = \frac{v}{\Pi}, \end{aligned} \quad (1.54)$$

which reduces to eqs. (1.36, 1.40) for the case of a single elliptical polarization state.

It is convenient to define the *polarization vector* as a 3D vector $\mathbf{p} = (q, u, v)$, whose physical meaning can be intuitively illustrated using the Poincaré sphere (Fig. 1.4), where the three Cartesian basis are $(\hat{\mathbf{q}}, \hat{\mathbf{u}}, \hat{\mathbf{v}})$, the polarization vector \mathbf{p} is contained within the sphere of radius 1, the length of the vector stands for the total polarization fraction $|\mathbf{p}| = \Pi \leq 1$, the linear polarization angle is $\psi \in (-\pi/2, \pi/2]$, and the ratio between circular and total polarization fractions is described by the angle $\chi \in [-\pi/4, \pi/4]$.

1.3.3 *Depolarization (under construction)

1.4 Phase-space distribution function for photons

In the previous section, we described the radiation field as the linear superposition of many monochromatic plane waves. Such a description is often adopted by radio observers because (i) the number density of photons is extremely large such that the classical wave picture holds and (ii) one can (at least try to) resolve the wave oscillations by rapidly sampling the electric fields. However, at higher frequencies (e.g., in the visible or X-ray bands), the number density of photons is much lower and also resolving the wave oscillations becomes impractical. A monochromatic wave have infinite spatial extent ($e^{i(kz-\omega t)}$ applies to all z 's), which becomes inconvenient when we want to describe the spatial variations of the radiation field. For these reasons, we use a more fundamental way of describing the radiation field — the *phase-space distribution function* for photons

$$f(\mathbf{x}, \mathbf{p}, t) = \frac{dN}{d^3x d^3p}, \quad (1.55)$$

which means the number of photons dN within the phase volume element $d^3x d^3p$ at time t . The energy of each photon is $h\nu$ (ν =frequency), and the photon momentum is

$$\mathbf{p} = \hbar\mathbf{k} = (h\nu/c)\hat{\mathbf{n}},$$

where $h = 2\pi\hbar$ is the Planck constant, $\hat{\mathbf{n}}$ is the unit vector along the instantaneous direction of motion, and $\mathbf{k} = (2\pi\nu/c)\hat{\mathbf{n}}$ is the wavevector. The only drawback of using the distribution function (or its descendant functions like η_γ and I_ν below) is that the phase information is lost — this might be quite serious depending on the reader's purpose. For instance, polarization can not be fully represented by the the distribution function.

1.4.1 Photon occupation number

From quantum theory, the wave nature of a particle with momentum p is characterized by its de Broglie wavelength $\lambda = h/p$. A particle within a cubic volume of length L (physical volume = L^3) can only have *discrete momentum states*

$$\mathbf{p} = (h/L)\mathbf{n}$$

so as to be considered as standing waves or modes (with L/λ = integer to satisfy the periodic boundary condition), where $\mathbf{n} \in \mathbb{Z}^3$ is a vector on the 3D integer lattice with all three components being non-zero integers. The volume of each discrete lattice cube in the momentum space is h^3/L^3 . Therefore, a 6D phase-space volume of $L^3 d^3p$ contains $L^3 d^3p/h^3$ distinct free-motion quantum states. Effectively, the entire 6D phase-space is not continuous but consists of quantized volume elements of size h^3 (like unit cells in a

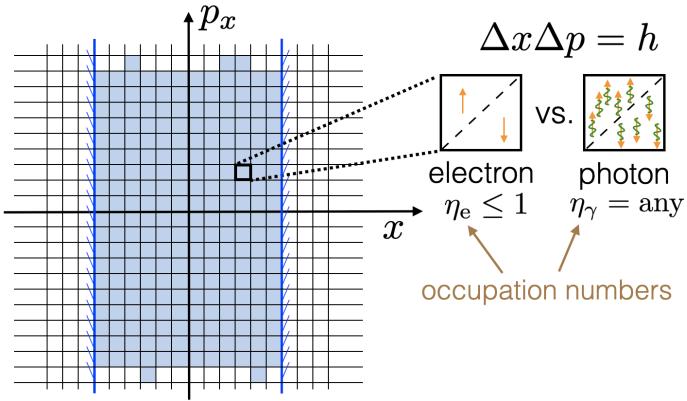


Figure 1.5: The phase-space consists of quantized volume elements of size h^3 (here showing only two dimensions $x-p_x$). The occupation numbers η_e (for electrons) and η_γ (for photons) are defined as the average number of particles per quantum state. For electrons ($s = 1/2$) or photons ($s = 1$), there are $g_s = 2$ spin states per h^3 volume. The blue shaded region shows a distribution of (degenerate) electrons confined in a rigid box whose boundaries are shown by blue solid lines. For photons, there can be any number of particles in each h^3 volume. A high photon occupation number $\eta_\gamma \gg 1$ allows for strong stimulated emission (see §2.2).

lattice), regardless of whether the particles are relativistic or not. Each quantized volume element can have g_s internal states due to the particle's internal degree of freedom — here the subscript s is for “spin” (the only internal degree of freedom for photons). An intuitive understanding from the *Uncertainty Principle* is that one cannot constrain the position of a photon in phase space to better than within a *phase-space volume* of $\sim h^3$.

Thus, another way of describing radiation fields that is equivalent to the phase-space distribution function $f(\mathbf{x}, \mathbf{p}, t)$ is the dimensionless *photon occupation number*,

$$\eta_\gamma(\mathbf{x}, \mathbf{p}, t) \equiv \frac{h^3}{g_s} f, \quad g_s = 2, \quad (1.56)$$

which means the *time-averaged* number of photons in a given phase-space volume of $h^3/2$ near position \mathbf{r} and momentum \mathbf{p} , and near time t . At a given time t , a given quantum state can only have an integer number of particles (for bosons, it is 0, 1, 2, ...). The time-averaging is carried out on a timescale T that is much longer than the inverse of the photon frequency ν^{-1} but much shorter than the variability timescale of the distribution function. Under such a time averaging, we see that η_γ may be any positive real number (not necessarily an integer). The factor of $g_s = 2$ is the statistical weight due to spin, since photons have spin $s = 1$ (in units of \hbar). Two photons of spins +1 and -1 (corresponding to right and left circular polarizations) within the same volume of h^3 are distinguishable.

However, if two photons with the same spin (e.g., +1) are placed inside the same phase-space volume of h^3 , they become *indistinguishable* by any possible means.

The momentum-space volume element near momentum \mathbf{p} is given by

$$d^3\mathbf{p} = p^2 dp d\Omega = (h/c)^3 \nu^2 d\nu d\Omega, \quad (1.57)$$

where we have used $p = h\nu/c$ and $d\Omega$ is the solid angle within which the photons are moving. The photon number density near position \mathbf{x} at time t is given by

$$n_{\text{ph}}(\mathbf{x}, t) = \int f d^3\mathbf{p} = \int d\Omega \int d\nu \frac{2\nu^2}{c^3} \eta_\gamma. \quad (1.58)$$

Then, the total radiation energy density of photons near position \mathbf{x} at time t is

$$U(\mathbf{x}, t) = \int h\nu f d^3\mathbf{p} = \int d\Omega \int d\nu \frac{2h\nu^3}{c^3} \eta_\gamma. \quad (1.59)$$

Hereafter, integral without limits (for brevity) simply means the entire allowed range.

An intuitive way of understanding the photon occupation number in eq. (1.56) is as follows. From eq. (1.58), we see that the number density of photons with wavelengths near $\lambda = c/\nu$ is roughly given by

$$n_{\text{ph}}(\lambda) \sim \eta_\gamma / \lambda^3, \quad (1.60)$$

which means that the occupation number η_γ near wavelength λ is of the order the number of photons in a volume of λ^3 . If $\eta_\gamma \gg 1$, there are a large number of photons with wavelengths near λ in a λ^3 volume and they blend together into classical electromagnetic waves. When $\eta_\gamma \ll 1$, photons are discrete as their wave packets do not overlap significantly, so we may treat them as classical particles. Note that $\eta \gg 1$ is only possible for *bosons* (particles with integer spins of 0, 1, 2, etc). For *fermions* (with half integer spins, e.g., electrons, neutrons, protons), the *Pauli Exclusion Principle* tells that no two particles of the same kind⁴ can occupy the same quantum state, so $\eta \leq 1$.

1.4.2 Intensity and specific intensity

Physicists likely prefer the distribution function f or occupation number η_γ , but astronomers commonly use a third and equivalent way to describe the radiation field — the *specific intensity function* $I_\nu(\mathbf{x}, \hat{\mathbf{n}}, t)$, where the dependence on photon momentum \mathbf{p} is split into two parts: magnitude $p(\nu) = h\nu/c$ and direction ($\hat{\mathbf{n}} = \mathbf{p}/p$). It is one of the most confusing concepts in astrophysics, so we need to gain a clear understanding of it.

⁴Apart from the intrinsic property of spin, any two electrons (or other fundamental particles) in the Universe are fundamentally indistinguishable, and the only way to tell them apart is by looking at their different positions in the phase-space.

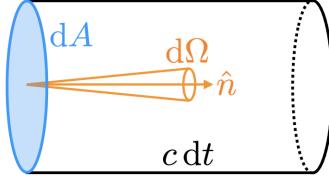


Figure 1.6: Geometry for the definition of intensity function (eq. 1.63), which describes the energy dE of photons propagating within a solid angle $d\Omega$ near direction \hat{n} into the volume $c dA dt$.

Let us only consider photons moving within a small solid angle $d\Omega$ around the unit vector \hat{n} , and the energy density per solid angle is given by (cf. eq. 1.59)

$$\frac{dU}{d\Omega} = \int d\nu \frac{2h\nu^3}{c^3} \eta_\gamma. \quad (1.61)$$

Consider a cylinder with area dA and length $c dt$, and it is oriented such that the circular faces are perpendicular to the unit vector \hat{n} , as shown in Fig. 1.6. The radiation energy of the photons moving in the solid angle $d\Omega$ inside the cylinder is

$$dE = \frac{dU}{d\Omega} d\Omega dA c dt. \quad (1.62)$$

Then, the frequency-integrated *intensity function* is defined as

$$I(\mathbf{x}, \hat{n}, t) \equiv \frac{dE}{d\Omega dA dt} = \frac{dU}{d\Omega} c = \int d\nu \frac{2h\nu^3}{c^2} \eta_\gamma, \quad (1.63)$$

which describes the amount of energy per solid angle ($d\Omega$) through a unit area (dA) per unit time (dt). The *specific intensity function* is then defined as

$$I_\nu(\mathbf{x}, \hat{n}, t) \equiv \frac{dI}{d\nu} = \frac{dE}{d\Omega dA dt d\nu} = \frac{2h\nu^3}{c^2} \eta_\gamma. \quad (1.64)$$

Here, we follow the convention by referring to “... per unit frequency” as “specific ...”. Also, “(specific) intensity function” is often simply called “(specific) intensity”. However, it is very important for the reader to understand the functional dependence of I_ν on frequency ν , position \mathbf{x} , direction \hat{n} , and time t .

Going back to the energy density in eq. (1.61), we see that the specific intensity I_ν is related to the specific energy density U_ν by

$$U_\nu = c^{-1} \int d\Omega I_\nu. \quad (1.65)$$

and the total energy density is

$$U = \int d\nu U_\nu = c^{-1} \int d\Omega I. \quad (1.66)$$

In Euclidean or flat spacetime and in the absence of interactions, the distribution function of photons are described by the *collisionless Boltzmann equation* (or the *Liouville theorem*)

$$\frac{df}{dt} = \partial_t f + \dot{\mathbf{x}} \cdot \nabla f + \dot{\mathbf{p}} \cdot \nabla_{\mathbf{p}} f = 0, \quad (1.67)$$

where $\dot{\mathbf{x}} = \mathbf{v} = c\hat{\mathbf{n}}$ is the velocity of photons (along direction $\hat{\mathbf{n}}$), ∇ describes the spatial gradient, $\dot{\mathbf{p}}$ describes the acceleration, and $\nabla_{\mathbf{p}}$ describes the gradient in momentum space. Since photon frequency is conserved along the propagation, we have $\dot{\mathbf{p}} = 0$, and since $I_\nu \propto \nu^3 f$, we find that the specific intensity stays constant along a ray or photons' local direction of motion in the absence of interactions (e.g., emission, absorption, or scattering)

$$\frac{dI_\nu}{dt} = \partial_t I_\nu + (c\hat{\mathbf{n}} \cdot \nabla) I_\nu = 0. \quad (1.68)$$

Here, d/dt is the Lagrangian derivative along the ray, whereas the temporal variation of the specific intensity at a fixed position is given by $\partial_t I_\nu$ which is generally non-zero for a time-dependent system.

It should be noted that, in non-Euclidean spacetime (e.g., near a black hole or over cosmological distances), the frequency of a given photon changes along the propagation (due to e.g., gravitational redshift or cosmic time dilation), and we need to use the more general expression based on the conservation of photon number. Since $df/dt = d\eta_\gamma/dt = 0$ holds *locally* in any spacetime, the photon occupation number, $\eta_\gamma \propto f \propto I_\nu/\nu^3$, is conserved along a given ray or null geodesic. For instance, if a group of photons propagate from the initial position \mathbf{x}_1 to the final position \mathbf{x}_2 and the corresponding gravitational redshift factor z_g is given by $\lambda_2 = (1 + z_g)\lambda_1$ or $\nu_2 = \nu_1/(1 + z_g)$, then the corresponding specific intensities at these two positions are related by $I_{\nu_2}(\mathbf{x}_2) = I_{\nu_1}(\mathbf{x}_1)(\nu_2/\nu_1)^3 = I_{\nu_1}(\mathbf{x}_1)(1 + z_g)^{-3}$.

Throughout this book, we will ignore spacetime curvature for simplicity, so the intensity is conserved along any given ray in the absence of interactions.

The constancy of intensity in flat spacetime can also be shown geometrically. In Fig. 1.7, we define the following quantities: dA_{em} is the area of the emitting source (or a patch on its surface), $dA_{\text{em},\perp}$ is the projected emitting area perpendicular to the line of sight, d is the distance from the source to the observer, dA_{obs} is area of the observer's detector (which is oriented perpendicular to the line of sight), $d\Omega_{\text{em}} = dA_{\text{obs}}/d^2$ is the solid angle spanned by the detector as viewed from the source, and finally $d\Omega_{\text{obs}} = dA_{\text{em},\perp}/d^2$ is the solid angle spanned by the source as viewed by the observer. From the above definitions, we find

$$dA_{\text{em},\perp} d\Omega_{\text{em}} = dA_{\text{obs}} d\Omega_{\text{obs}}. \quad (1.69)$$

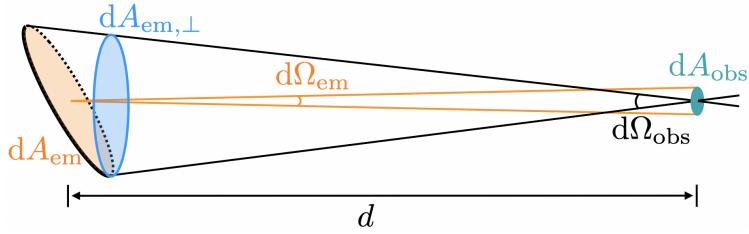


Figure 1.7: Geometry for the constancy of intensity, which describes that the source-observer distance squared d^2 is equal to the ratio between the observer's detector area dA_{obs} over the solid angle $d\Omega_{\text{em}}$ that the observed photons are emitted into and also the ratio between the projected area of the emitting surface $dA_{\text{em},\perp}$ and the solid angle $d\Omega_{\text{obs}}$ spanned by the source on the observer's sky, i.e., $d^2 = dA_{\text{obs}}/d\Omega_{\text{em}} = dA_{\text{em},\perp}/d\Omega_{\text{obs}}$.

The *emitted intensity* along the line of sight at the surface of the source is given by

$$I_{\text{em}} = \frac{dE}{dA_{\text{em},\perp} d\Omega_{\text{em}} dt}, \quad (1.70)$$

where dE is the amount of radiation energy that flows past the area $dA_{\text{em},\perp}$ within the solid angle $d\Omega_{\text{em}}$ and in a time interval dt . Similarly, the *observed intensity* along the line of sight at the surface of the detector is given by

$$I_{\text{obs}} = \frac{dE}{dA_{\text{obs}} d\Omega_{\text{obs}} dt}. \quad (1.71)$$

From eq. (1.69), we see that

$$I_{\text{em}} = I_{\text{obs}}, \quad (1.72)$$

and therefore the intensity stays constant along a given ray (= the line of sight).

1.4.3 Moments of intensity: energy density, energy flux & pressure

Here, we consider the angular dependence of the specific intensity function $I_\nu(\theta, \phi)$ in a fixed spherical coordinate with polar axis along the normal direction \hat{z} of a given infinitesimal surface area dA . Hereafter, θ and ϕ are the polar and azimuthal angles. This is shown in Fig. 1.8.

The 0th moment or angle-average of $I_\nu(\theta, \phi)$, which may have arbitrary angular dependences on θ and ϕ , is called the *specific mean intensity*

$$J_\nu = \frac{1}{4\pi} \int d\Omega I_\nu(\theta, \phi), \quad (1.73)$$

which is related to the specific energy density U_ν by

$$U_\nu = c^{-1} \int d\Omega I_\nu = \frac{4\pi}{c} J_\nu, \quad (1.74)$$

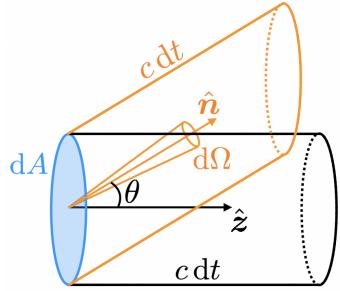


Figure 1.8: The specific intensity function $I_\nu(\theta, \phi)$ described in a polar coordinate whose polar axis \hat{z} is along the normal direction of a small surface area dA . In this fixed coordinate system, we consider a particular beam along a solid angle $d\Omega$ near direction \hat{n} . In a time interval dt , the photons passing through dA along the $d\Omega$ beam are contained in a volume of $c dA dt \cos \theta$, as the height of the oblique cylinder is $c dt \cos \theta$.

as can be obtained from eq. (1.61).

The *net* energy per unit frequency interval passing through a unit area per unit time is called the *specific energy flux* or the *flux density*, and it is the 1st moment of $I_\nu(\theta, \phi)$ as given by the following integral

$$F_\nu = \int d\Omega I_\nu(\theta, \phi) \cos \theta, \quad (1.75)$$

where $\theta = \text{acos}(\hat{n} \cdot \hat{z})$ is the projection angle between the radiation beam \hat{n} and the unit vector \hat{z} perpendicular to the area in consideration. It is important to understand the projection factor $\cos \theta$. The photons passing through the area dA within a solid angle $d\Omega$ near \hat{n} in a time interval dt is contained in a volume of $c dA dt \cos \theta$, as shown by the orange oblique cylinder in Fig. 1.8. When our telescope is pointed directly at a given source (such that $\hat{n} \parallel \hat{z}$ and $\cos \theta \approx 1$) and the source's angular size is sufficiently small ($\ll 1$ rad), then the flux density is related to the specific intensity by $F_\nu \approx \int I_\nu d\Omega$. If the solid angle spanned by the source Ω_* , then the average specific intensity on the emitting surface is $\langle I_\nu \rangle \approx F_\nu / \Omega_*$.

The net *specific momentum flux* or *specific radiation pressure* perpendicular to a given surface along direction \hat{z} is described by the 2nd moment of $I_\nu(\theta, \phi)$ as follows

$$P_\nu = \frac{1}{c} \int d\Omega I_\nu(\theta, \phi) \cos^2 \theta. \quad (1.76)$$

which has the physical meaning of the \hat{z} -component of momentum passing through a unit area per unit time. Here, we have used two factors of $\cos \theta$, one due to the fact the number of photons passing through the surface per unit time is reduced by a factor of $\cos \theta$,

and the other one is because we are only considering the \hat{z} -component of the momentum $p_z/p = \cos \theta$.

The flux density F_ν in the above discussion describes the rate of energy flow (per unit frequency bin) along a given \hat{z} direction. To obtain a more complete description of the energy flow in a given radiation field, we would like to know the flux density along the \hat{x} and \hat{y} directions as well. This motivates us to define the *flux density vector*

$$\mathbf{F}_\nu = \int d\Omega I_\nu \hat{\mathbf{n}}, \quad (1.77)$$

which is independent of our choice of coordinate system. The energy flux flowing along the \hat{z} direction is then given by the following projection

$$F_\nu^z = \mathbf{F}_\nu \cdot \hat{\mathbf{z}} = \int d\Omega I_\nu \cos \theta, \quad (1.78)$$

where $\hat{\mathbf{n}} \cdot \hat{\mathbf{z}} = n_z = \cos \theta$. The radiation pressure P_ν in the above discussion describes the flux of the z -component of photon momentum along the \hat{z} direction. More generally, we can define a 3×3 *radiation pressure tensor* $\overset{\leftrightarrow}{P}_\nu$, such that P_ν^{ij} describes the flux of the i -component of photon momentum along the j direction, where $i, j = x, y, z$. The radiation pressure tensor is then given by

$$\overset{\leftrightarrow}{P}_\nu = \frac{1}{c} \int d\Omega I_\nu \hat{\mathbf{n}} \hat{\mathbf{n}}, \quad (1.79)$$

where $\hat{\mathbf{n}} \hat{\mathbf{n}}$ (or $\hat{\mathbf{n}} \otimes \hat{\mathbf{n}}$) should be understood as an outer product. The radiation pressure tensor is symmetric such that $P_\nu^{ij} = P_\nu^{ji}$. The flux of z -momentum along the \hat{z} direction is then given by the following double projection

$$P_\nu^{zz} = (\overset{\leftrightarrow}{P}_\nu \cdot \hat{\mathbf{z}}) \cdot \hat{\mathbf{z}} = \frac{1}{c} \int d\Omega I_\nu \cos^2 \theta. \quad (1.80)$$

where $\hat{\mathbf{n}} \cdot \hat{\mathbf{z}} = n_z = \cos \theta$. An important property of the radiation pressure tensor is that its diagonal components P_ν^{ii} (for $i = x, y, z$) are related to the radiation energy density U_ν as follows

$$\text{Tr} \left(\overset{\leftrightarrow}{P}_\nu \right) = \sum_{i=x,y,z} P_\nu^{ii} = \frac{1}{c} \int d\Omega I_\nu = U_\nu, \quad (1.81)$$

where Tr means the matrix trace and we have used $\sum_{i=x,y,z} n_i n_i = \hat{\mathbf{n}} \cdot \hat{\mathbf{n}} = 1$. For the above reason, the frequency-integrated radiation pressure tensor $\overset{\leftrightarrow}{P}$ can be factored into the energy density U times a dimensionless tensor $\overset{\leftrightarrow}{D}$ that describes the anisotropy,

$$\overset{\leftrightarrow}{D} \equiv \overset{\leftrightarrow}{P}/U, \quad (1.82)$$

which is called the (dimensionless) *Eddington tensor*. One may also define the frequency-dependent version of the Eddington tensor $\overset{\leftrightarrow}{D}(\nu) = \overset{\leftrightarrow}{P}_\nu/U_\nu$.

Astronomers often use Jansky⁵ (Jy) as the unit for the specific energy flux, although at high frequencies, e.g., in the gamma-ray band where photons are sparse, one should be aware that the F_ν only describes the average specific flux and the actual number of detected photons in a given integration time follow the Poisson distribution. Optical astronomers use the AB magnitude to describe the flux density of sources

$$m_{\text{AB}} = -2.5 \log(F_\nu/\text{Jy}) + 8.90 \approx -2.5 \log(F_\nu/3631\text{Jy}). \quad (1.83)$$

A special case is an isotropic radiation field, e.g. in the interior of an optically thick gas. The net flux density along the \hat{z} direction vanishes as $F_\nu = I_\nu \int d\Omega \cos \theta = 0$, but the outwards flux density in one hemisphere (along $+\hat{z}$) is

$$F_\nu^{\text{out}} = 2\pi I_\nu \int_0^1 \mu d\mu = \pi I_\nu, \quad (1.84)$$

where $\mu = \cos \theta$. The flux density in the inwards direction (along $-\hat{z}$) is $-\pi I_\nu$ such that there is no net flux. The radiation pressure along the \hat{z} direction is given by

$$P_\nu = 2\pi \frac{I_\nu}{c} \int_{-1}^1 \mu^2 d\mu = \frac{4\pi}{3} \frac{I_\nu}{c} = U_\nu/3. \quad (1.85)$$

The total pressure P is related to the total energy density U by

$$P = U/3.$$

Finally, we discuss the three moments of intensity in the context of an infinite expansion of $I_\nu(\theta, \phi)$ by the [spherical harmonics](#), $Y_{\ell m}(\theta, \phi)$. For an *axis-symmetric intensity function* $I_\nu(\theta)$, we can in general describe it by the Legendre series expansion

$$I_\nu(\theta) = \sum_0^\infty c_\ell P_\ell(\cos \theta),$$

where $P_\ell(\cos \theta)$ are [Legendre polynomials](#) of degree ℓ and c_ℓ are the coefficients given by orthogonality condition

$$c_\ell = \frac{2\ell+1}{2} \int_0^\pi I_\nu(\theta) P_\ell(\cos \theta) \sin \theta d\theta = \frac{2\ell+1}{4\pi} \int d\Omega I_\nu(\theta) P_\ell(\cos \theta), \quad (1.86)$$

where the second expression describes an integral of $d\Omega = 2\pi \sin \theta d\theta$ over the entire 4π solid angle. The first three Legendre polynomials are $P_0 = 1$ (“monopole”), $P_1 = \cos \theta$

⁵1 Jy = 10^{-23} erg cm $^{-2}$ s $^{-1}$ Hz $^{-1}$.

(“dipole”), $P_2 = (3 \cos^2 \theta - 1)/2$ (“quadrupole”), and the corresponding coefficients are given by

$$c_0 = \frac{1}{4\pi} \int d\Omega I_\nu(\theta), \quad c_1 = \frac{3}{4\pi} \int d\Omega I_\nu(\theta) \cos \theta, \quad c_2 = \frac{5}{8\pi} \int d\Omega I_\nu(\theta)(3 \cos^2 \theta - 1). \quad (1.87)$$

We see that c_0 , c_1 , and c_2 are related to the three *physical* moments of the intensity function: mean intensity = 0th moment, energy flux = 1st moment, and pressure (or momentum flux) = 2nd moment.

1.5 Homework

Prob. 1. An observer measured the following polarization properties of a spatially unresolved optical transient in a nearby host galaxy: $Q/I = 0.7\%$, $U/I = 0.7\%$, $V = 0$. From this the observer inferred that the source is partially linearly polarized and the polarization fraction is $\Pi = \sqrt{Q^2 + U^2 + V^2}/I = 1\%$. Later on, however, it was realized that the stars in the host galaxy contributed a fraction $I_{\text{host}}/I = 30\%$ of the total intensity (as obtained based on the brightness profile of the host galaxy after the transient has faded away). Under the *assumption* that the star light in the host galaxy is unpolarized, calculate the intrinsic polarization fraction of the transient. Do a quick literature search to see if “host-galaxy star light is unpolarized” is a good assumption.

Prob. 2. Suppose that a radio telescope has two independent, orthogonal antenna arms in the \hat{x} and \hat{y} directions, and the EM waves from a source propagate in the $+\hat{z}$ direction. An observer recorded so-called “raw voltage” data in a given narrow frequency channel of width $\Delta\omega$ centered at frequency ω . For simplicity, we assume that the source has a single polarization state and that the data is written in the complex form $E_j(t) = E_{j,0}e^{i\phi_j}e^{-i\omega t}$ for $j = x, y$, where $E_{j,0}$ are the amplitudes and ϕ_j are the phases. Let us assume that $E_{j,0}$ and ϕ_j stay constant over a duration T that is much longer than ω^{-1} but much shorter than the variability timescale of the source (so it is possible to average the data over the timescale of T). The two antenna arms have different, unknown gain factors A_j and time delays δt_j , because of the differences in the electronic systems (e.g., amplifiers, wire lengths) in the two signal pathways. The true amplitudes of the source are $E_{j,0}/A_j$ and the true phases are $\phi_j + \omega\delta t_j$. The goal here is to calculate A_j and δt_j by observing a source with known brightness and polarization properties. In fact, we are only interested in the relative time delays between the two arms $\Delta t = \delta t_x - \delta t_y$ (the absolute phases are only important if we want to jointly analyze the data from multiple telescopes at different sites), so there are only three unknowns: A_x , A_y and Δt .

- (i) Calculate the true Stokes- I parameter of the source at the given frequency channel using eq. (1.43). Suppose the source has a known flux density F_ν near frequency $\nu = \omega/(2\pi)$. Using $I = 8\pi c^{-1} F_\nu \Delta\nu$, you should obtain an equation that contains the gain factors A_j .
- (ii) Calculate the other three (Q, U, V) parameters of the source at the given frequency channel using eq. (1.43). Suppose the source is known to be 100% linearly polarized, which one(s) of the

three unknowns can be determined? Calculate the polarization angle (PA, ψ defined in eq. 1.36 and shown in Fig. 1.3) and find its dependence on the remaining unknowns.

(iii) Next we repeat the above exercise for to obtain the Stokes parameters in two frequency channels centered at ω_n for $n = 1, 2$ (and one can consider more frequency channels if available). Radio waves propagating through a magnetized plasma (e.g., the interstellar medium) are modified by a number of effects, e.g., multi-path propagation, dispersion delay, and Faraday rotation. Here, we only consider the effect of Faraday rotation, which causes the PA ψ_n in each channel to change in a known, frequency-dependent way: $\Delta\psi_n = \text{RM} \lambda_n^2$, where $\Delta\psi_n$ is the difference between the PA arriving at the observer and the intrinsic PA at the source, RM is called *rotation measure* often expressed in units of rad m^{-2} , and $\lambda_n = 2\pi c/\omega_n$ is the wavelength. Suppose the unmodified waves from the source has a constant but unknown PA across the entire frequency band and that RM is known (from other, independent measurements). Now, you can solve for all three unknowns for our calibration purpose. For our purpose here, you do not need to dive into the math to explicitly solve the unknowns. It is sufficient to simply demonstrate the underlying logic.

(iv) If the RM is unknown, what would you do to solve for the calibration unknowns?

Prob. 3. The Sun has luminosity $L_\odot \approx 4 \times 10^{33} \text{ erg s}^{-1}$ and radius $R_\odot \approx 7 \times 10^{10} \text{ cm}$. Let us assume that, for each surface element of Sun, the intensity in the outward-facing hemisphere is isotropic (we are ignoring an effect called *limb darkening*).

- (i) Calculate the (frequency-integrated) intensity on a surface element of the Sun in the outward-facing hemisphere in terms of L_\odot and R_\odot , and then express it in CGS units.
- (ii) Let F be the flux of Sun light along the radial direction at the distance of the Earth, $d \approx 1 \text{ AU} = 1.5 \times 10^{13} \text{ cm}$. At this distance, the Sun spans a solid angle of Ω on the sky. Express the intensity of Sun light viewed from the Earth in terms of F and Ω , and then express it in CGS units.
- (iii) What would be the intensity of a Sun-like star at a distance of $d = 1 \text{ kpc} \approx 3.1 \times 10^{21} \text{ cm}$? For this problem, please ignore the effects of atmospheric distortion and telescope's finite resolution on the angular size of the star on the sky.
- (iv) For ground-based observations in the optical band, turbulence (i.e., random fluctuations of air density and hence the refractive index) in the Earth's atmosphere causes the star's image on the sky to span a solid angle $\Omega \sim 1 \text{ arcsec}^2$. Note that $1 \text{ arcsec} = \text{AU}/\text{pc} \approx 4.8 \times 10^{-6} \text{ rad}$. Calculate the intensity of a Sun-like star at $d = 1 \text{ kpc}$ as seen by a ground-based telescope. Assume that twinkling can be time-averaged out. P.S., the finite angular resolution of the telescope has a similar effect on the angular size of the source.

References

- Rybicki & Lightman, *Radiative processes in astrophysics* (A Wiley-Interscience Publication, New York, 1979)
- M. Born and E. Wolf, *Principles of Optics* (Cambridge, U.P., Cambridge, 1999).

Chapter 2

Radiative Transfer

This chapter discusses how radiation interacts with and propagates through matter.

2.1 Thermal equilibrium and blackbody radiation

The general distribution functions for matter or radiation (or a mixture of the two) statistically describe the particle distributions along all the degrees of freedom in a given system. To describe the translational degrees of freedom for a given species, we need a 7-dimensional function $f(\mathbf{x}, \mathbf{p}, t)$ for position \mathbf{x} , momentum \mathbf{p} , and time t . The internal degrees of freedom of atoms further include ionization states and bound electron states. In practice, a full description of all these degrees of freedom would be extremely demanding, so we always look for ways to simplify the distribution function. In this section, we will discuss the simplest case of *local thermodynamic equilibrium* (LTE) — this is also the most important case as a system always has the tendency of evolving towards LTE.

2.1.1 Local thermodynamic equilibrium (LTE)

“Local” means that the system is in *quasi-equilibrium*: (i) within a *volume* that is sufficiently large to contain a large number of particles but small such that the average properties (density, pressure, temperature, etc) can be considered to be uniform, and (ii) in a *time interval* that is sufficiently long compared to the time it takes to reach thermodynamical equilibrium (this timescale is known as the *relaxation time*) but short compared to the timescale in which the system’s average properties change substantially. The relaxation processes that cause the particles’ distribution function to approach the thermal distribution are generally due to matter-matter and matter-radiation interactions.

“Thermodynamic equilibrium” means that the rates of all possible state transitions (e.g., $1 \rightarrow 2$) are balanced by the rate of their corresponding inverse transitions ($2 \rightarrow 1$)

— this is called *detailed balance*. Microscopic transitions are constantly occurring but the system has relaxed to a statistically steady state where the macroscopic properties are no longer evolving within the timescales of interest. Thermodynamic equilibrium has a few important properties: (i) the *total entropy* of the system is maximized; (ii) *all* particles follow *thermal distribution* in energy and momentum; (iii) *all* chemical reactions, e.g. $A + B \leftrightarrow C + D$, are in *chemical equilibrium* with the chemical potentials satisfying $\mu_A + \mu_B = \mu_C + \mu_D$. We will make use of these properties later.

Astrophysical gases are generally very dilute such that we can ignore long-range interactions between particles, e.g., Coulomb forces between charged particles and Van der Waals forces between induced electric dipole moments. Therefore, in the following, we will restrict our discussion of the matter component to the case of *ideal gas*. As photons generally do not directly interact with each other, the radiation field (in the form of photon gas) will also be considered as a component of the ideal gas mixture.

For particles in an ideal gas in LTE, the occupation number of a given species follows *thermal distribution*,

$$\eta(\epsilon) = \frac{1}{e^{(\epsilon-\mu)/(k_B T)} \pm 1}, \quad (2.1)$$

where ϵ is the particle's total energy (including excitation energy of internal degrees of freedom as well as kinetic and potential energies), k_B is the Boltzmann constant, T is the *kinetic temperature*, μ is the chemical potential of the given species, and the “+” (or “−”) sign in the denominator is for fermion (or boson) particles¹. For the matter components under LTE, the chemical potential of different species may be positive, negative, or zero. However, for the radiation component under LTE, their chemical potential must be zero — this is the consequence of the fact that photon number is not conserved due to emission and absorption (see §2.1.4).

Compared to the general distribution function $f(\mathbf{x}, \mathbf{p}, t) = (g/h^3) \eta(\mathbf{x}, \mathbf{p}, t)$ (where g = statistical weight and η = occupation number), the thermal distribution in eq. (2.1) is spatially homogeneous (no dependence on \mathbf{x}), temporally stationary (no dependence on t), and isotropic (no dependence on the direction of momentum $\hat{\mathbf{p}}$). More importantly, it has a particular functional shape that is controlled by only two parameters μ and T . We see that the LTE condition provides an enormous simplification of the distribution function that reduces the number of free dimensions from seven to effectively zero (only 2 numbers μ and T)!

A given system may be a mixture of multiple species of atoms, ions, electrons, and photons, and each species will have their own chemical potential. Note that the same atoms at different energy levels (corresponding to different internal excitations) have different

¹A way to remember this rule is to notice that $(e^x + 1)^{-1} < 1$ for any $x \in (-\infty, +\infty)$, so the “+” sign must correspond to the Fermi-Dirac distribution. On the other hand, since $(e^x - 1)^{-1}$ can be anywhere between 0 and $+\infty$, so the “−” sign must correspond to the Bose-Einstein distribution.

chemical potentials, so we should consider the atoms in different energy levels as separate species. If the number density n of a given species is known, then the chemical potential μ is determined by $n = \int 4\pi p^2 f(p) dp$, and the integral can be carried out as long as we know the kinetic temperature T and the energy-momentum relation $\epsilon(p)$ for that species. Thus, the chemical potential generally carries the information about the number density. We will carry out this integral in §2.1.2 where we discuss the Saha-Boltzmann equation.

It is important to understand the difference between LTE and thermal distribution. It is possible that only some species have a thermal distribution function whereas other species do not. For instance, in the Earth's atmosphere, gas particles undergo frequent collisions among themselves to reach the thermal Maxwellian velocity distribution, whereas photons from the Sun light have much less frequent interactions with gas particles due to longer mean-free path, so the photons do not follow a thermal distribution — the blue sky does not have a blackbody spectrum (as required if the radiation-matter mixture is in LTE, see §2.1.4). In fact, astrophysicists also encounter interesting cases of a “two-temperature plasma” where electrons and protons each follow thermal distributions but at different temperatures. These situations are not referred to as LTE in this book as we require that *all* particles, including all species of matter and photons, follow thermal distribution at the same temperature. In fact, for photons, a thermal (Bose-Einstein) distribution is insufficient for LTE, we further require zero chemical potential (see §2.1.4).

If matter follows thermal distribution, we call the radiation generated by matter *thermal emission*, but thermal emission does not mean that the resulting photons will follow the thermal distribution. An example of thermal emission is the free-free emission by ionized gas where the free electrons are in thermal distribution, and unless the gas is optically thick to free-free absorption at all wavelengths, the emerging radiation spectrum is non-thermal (see Ch. 6). The radiation spectrum only becomes thermalized if photons of all wavelengths can interact with matter sufficiently frequently. A good example is the radiation field in the deep interior of stars.

In the following, we discuss the matter and radiation components separately.

2.1.2 Saha-Boltzmann equation

In this section, we consider the matter component. We use the condition of chemical equilibrium to derive the Saha-Boltzmann equation, which links the abundances of atoms in different excitation and ionization states.

We will restrict us to the case of *non-degenerate gas*, where the occupation numbers for all matter species are much smaller than unity at all energies ($\eta \ll 1$). This is equivalent to the condition that the average separation between particles is much less than the de Broglie wavelength, i.e., the particles' wave functions do not “overlap” significantly and we can consider them as distinguishable, classical particles. This corresponds to the limit of

$(\epsilon - \mu)/(k_B T) \gg 1$ and hence $e^{(\epsilon - \mu)/(k_B T)} \gg 1$, so the occupation number is given by the *Maxwell-Boltzmann distribution*

$$\eta \approx e^{-\epsilon/(k_B T)} e^{\mu/(k_B T)}. \quad (2.2)$$

The corresponding phase-space distribution function takes the form of the *Maxwellian distribution*

$$f(p) = \frac{g}{h^3} \eta \approx \frac{g}{h^3} e^{\mu/(k_B T)} e^{-\epsilon/(k_B T)}, \quad (2.3)$$

where g is the statistic weight due to spin or other intrinsic degrees of freedom. For non-relativistic gas particles, the energy ϵ is related to the momentum p by

$$\epsilon(p) = mc^2 + p^2/(2m), \quad (2.4)$$

where mc^2 is the particle's rest-mass energy (including internal excitation energy) and $p^2/(2m)$ is the non-relativistic kinetic energy. The number density of a given species of rest-mass m is given by

$$\begin{aligned} n &= \int_0^\infty f(p) 4\pi p^2 dp = \frac{4\pi g}{h^3} e^{(\mu - mc^2)/(k_B T)} \int_0^\infty e^{-p^2/(2mk_B T)} p^2 dp \\ &= g \left(\frac{\sqrt{2\pi m k_B T}}{h} \right)^3 e^{(\mu - mc^2)/(k_B T)}. \end{aligned} \quad (2.5)$$

The integral in the above expression can be carried out by changing variable $x = p^2/(2mk_B T)$ and then making use of the Gamma function $\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$ and $\Gamma(3/2) = \sqrt{\pi}/2$. We define the *thermal de Broglie wavelength* for a given species of rest-mass m as

$$\lambda_T \equiv \frac{h}{\sqrt{2\pi m k_B T}}, \quad (2.6)$$

so the chemical potential is given by

$$\mu = mc^2 + k_B T \ln \left(\frac{n \lambda_T^3}{g} \right). \quad (2.7)$$

In the limit of $n \lambda_T^3 / g \ll 1$, $(\mu - mc^2)/(k_B T)$ is a large negative number.

Let us consider the simplest case of a mixture of hydrogen gas and radiation in LTE. The excitation energy of hydrogen atoms is determined by the principle quantum number² $n (= 1, 2, 3, \dots)$, so the corresponding rest-mass energy is given by

$$m_{H_n} c^2 = m_p c^2 + m_e c^2 - \chi_n, \quad (2.8)$$

²This should not be confused with number density.

where χ_n is the ionization energy given by

$$\chi_n = \chi_H/n^2, \quad (2.9)$$

and $\chi_H = \chi_{n=1} = 13.6 \text{ eV}$ for the ground state.

The chemical potentials for atomic hydrogen (H_n with principle quantum number n), protons (p), and electrons (e) are given by

$$\mu_i = m_i c^2 + k_B T \ln \left(\frac{n_i \lambda_{T,i}^3}{g_i} \right), \quad i = H_n, p, e. \quad (2.10)$$

where the thermal de Broglie wavelengths are given by

$$\lambda_{T,i} = \frac{\hbar}{\sqrt{2\pi m_i k_B T}}, \quad i = H_n, p, e. \quad (2.11)$$

Since free electrons have the lowest rest-mass $m_e \ll m_p \approx m_{H_n}$, they have the longest thermal de Broglie wavelengths $\lambda_{T,e} \gg \lambda_{T,p} \approx \lambda_{T,H_n}$. Thus, if we consider gas at increasing number densities, free electrons usually tend to become degenerate before protons and atoms (as long as the ionization fraction is higher than $(m_e/m_p)^{3/2} \approx 1.3 \times 10^{-5}$). For this reason, it is usually convenient to define an *electron degeneracy factor* as follows

$$\psi_e \equiv \frac{\mu_e - m_e c^2}{k_B T} = \ln \left(\frac{n_e \lambda_{T,e}^3}{g_e} \right). \quad (2.12)$$

The non-degenerate gas limit corresponds to ψ_e being a large negative number.

The statistical weights g_i describe the internal degrees of freedom of a given species. Free electrons may have two possible spin quantum numbers $m_s = \pm 1/2$, so $g_e = 2$. For protons, we ignore nuclear spins and take $g_p = 1$ here. For hydrogen atoms, each principle quantum number n corresponds to $g_{H_n} = 2n^2$ distinct quantum states, which are described by the orbital quantum number $\ell = 0, 1, \dots, n-1$ (corresponding to different total electron orbital angular momenta of $\ell\hbar$), magnetic quantum number $m = -\ell, -\ell+1, \dots, \ell-1, \ell$ (corresponding to different z -projected electron orbital angular momenta $m\hbar$), and electron spin quantum number $m_s = \pm 1/2$. If we were to take nuclear spins into account (with two possible nuclear spin quantum numbers $m_I = \pm 1/2$), then $g_p = 2$ and $g_{H_n} = 4n^2$, and our final Saha-Boltzmann equation (which involves the ratio g_{H_n}/g_p) will be the same.

Next we consider the photo-ionization and radiative recombination of hydrogen



When the system is in LTE such that photons have zero chemical potential ($\mu_\gamma = 0$), we have the following chemical equilibrium

$$\mu_{H_n} + \mu_\gamma = \mu_{H_n} = \mu_p + \mu_e. \quad (2.14)$$

Thus, we obtain a very important and general result: at LTE, the chemical potential of atoms in a given ionization state does not depend on the excitation energy (e.g., $\mu_{\text{H}_1} = \mu_{\text{H}_2}$)! Going back to the occupation number for non-degenerate ideal gas (eq. 2.2), we find that the occupation numbers at different energy levels follow the *Boltzmann distribution*

$$\eta(\epsilon) \propto e^{-\epsilon/(k_B T)} \Leftrightarrow \frac{\eta_2}{\eta_1} = e^{-(\epsilon_2 - \epsilon_1)/(k_B T)} = e^{-\Delta\epsilon/(k_B T)}, \quad (2.15)$$

where “1” and “2” stands for any two bound energy levels, and $\Delta\epsilon = \epsilon_2 - \epsilon_1$ is the energy difference. If $\epsilon_2 > \epsilon_1$, then there are less particles per quantum state (=occupation number) at the higher energy level by a factor of $e^{-\Delta\epsilon/(k_B T)}$ — this is often called the *Boltzmann factor*. If the statistical weights of the two energy levels are g_1 and g_2 , then the number density ratio is given by

$$\frac{n_2}{n_1} = \frac{g_2}{g_1} \frac{\eta_2}{\eta_1} = \frac{g_2}{g_1} e^{-\Delta\epsilon/(k_B T)}. \quad (2.16)$$

The reader shall not confuse herself between the number density ratio n_2/n_1 and the occupation number ratio η_2/η_1 .

Going back to the chemical equilibrium of hydrogen (eq. 2.14), we obtain a relation among the number densities n_{H_n} , n_{p} , and n_{e} as follows

$$m_{\text{H}_n} c^2 + k_B T \ln \left(\frac{n_{\text{H}_n} \lambda_{\text{H}_n}^3}{g_{\text{H}_n}} \right) = m_{\text{p}} c^2 + k_B T \ln \left(\frac{n_{\text{p}} \lambda_{\text{T,p}}^3}{g_{\text{p}}} \right) + m_{\text{e}} c^2 + k_B T \ln \left(\frac{n_{\text{e}} \lambda_{\text{T,e}}^3}{g_{\text{e}}} \right). \quad (2.17)$$

We then make use of eq. (2.8) that relates the rest-mass energies of H_n, p, and e and the ionization potential energy χ_n (n = principle quantum number), and obtain

$$\frac{n_{\text{H}_n}}{n_{\text{p}}} = \frac{g_{\text{H}_n}}{g_{\text{p}}} \frac{\lambda_{\text{T,p}}^3}{\lambda_{\text{T,H}_n}^3} \frac{n_{\text{e}} \lambda_{\text{T,e}}^3}{g_{\text{e}}} e^{\chi_n/(k_B T)} \approx \frac{g_{\text{H}_n}}{g_{\text{p}}} \frac{n_{\text{e}} \lambda_{\text{T,e}}^3}{g_{\text{e}}} e^{\chi_n/(k_B T)}, \quad (2.18)$$

where we have used $\lambda_{\text{T,H}_n} \approx \lambda_{\text{T,p}}$ as $m_{\text{H}_n} \approx m_{\text{p}}$ to the lowest order (affording a fractional error of the order $m_{\text{e}}/m_{\text{p}} \sim 10^{-3}$). For a pure hydrogen gas composition, we know $n_{\text{p}} = n_{\text{e}}$ from charge neutrality, so eq. (2.18) is all we need to calculate the fraction of hydrogen in each of the ionization and excitation states.

In the following, we re-write eq. (2.18) in a more general form that can be used for other atoms (e.g., He, C, O) with more ionization and excitation states. The excitation energy of a hydrogen atom above the ground ($n = 1$) state is given by

$$\epsilon_n \equiv \chi_{\text{H}} - \chi_n = \chi_{\text{H}} (1 - n^{-2}), \quad \chi_{\text{H}} = 13.6 \text{ eV}. \quad (2.19)$$

The *partition function* for energy level n (including g_{H_n} distinct quantum states) is

$$Z_{\text{H}_n} = g_{\text{H}_n} e^{-\epsilon_n/(k_B T)} = g_{\text{H}_n} e^{-(\chi_{\text{H}} - \chi_n)/(k_B T)}. \quad (2.20)$$

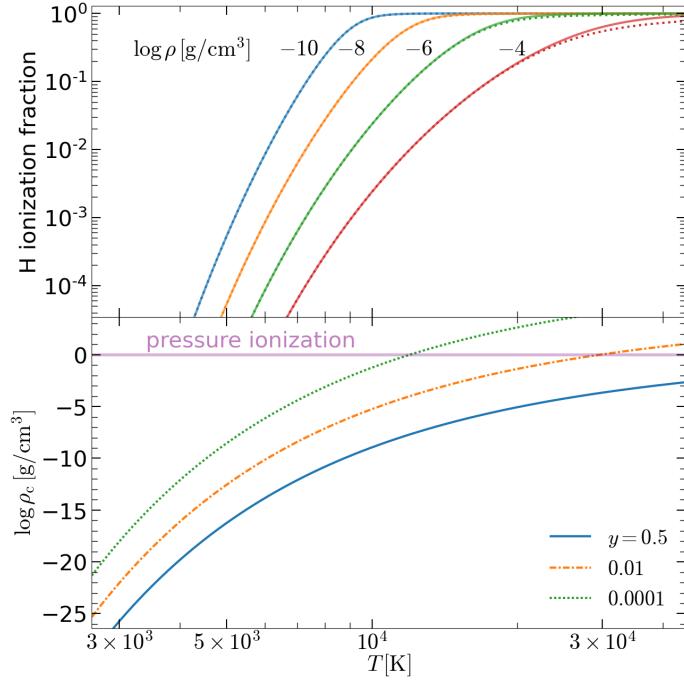


Figure 2.1: Upper panel: ionization fractions of hydrogen in LTE. The solid lines are for simplified partition function $Z_H \approx g_1 = 2$ and the dotted lines are for more realistic partition function $Z_H(\rho, T)$ with maximum principle quantum number given by eq. (2.29). Lower panel: The critical densities ρ_c corresponding to three different ionization fractions $y = 10^{-4}, 10^{-2}, 0.5$. The purple horizontal line marks the density above which pressure ionization (not included here) becomes important.

Thus, the partition function for hydrogen atoms in all bound states up to a maximum principle quantum number n_{\max} (see below) is given by

$$Z_H \equiv \sum_{n=1}^{n_{\max}} Z_{H_n} = e^{-\chi_H/(k_B T)} \sum_{n=1}^{n_{\max}} 2n^2 \exp\left(\frac{1}{n^2} \frac{\chi_H}{k_B T}\right). \quad (2.21)$$

The partition function Z_H depend mainly on temperature and very weakly on the gas density (through n_{\max} , see eq. 2.29) in the limit of non-degenerate ideal gas. For complicated atoms with a large number of bound electrons, the partition functions are often experimentally measured.

Using the partition function Z_{H_n} , we can then write the number density of atoms in energy level n as

$$n_{H_n} \approx Z_{H_n} \frac{n_p}{g_p} \frac{n_e \lambda_{T,e}^3}{g_e} e^{\chi_H/(k_B T)}, \quad (2.22)$$

The above is called the *Saha-Boltzmann equation for hydrogen* — it tells us that the number density of atoms in any given energy level n is linearly proportional to the partition function of that level Z_n , i.e.,

$$n_{\text{H}_n} \propto Z_{\text{H}_n} = g_{\text{H}_n} e^{-\epsilon_n/(k_B T)}. \quad (2.23)$$

This is a direct consequence of the Boltzmann distribution (eq. 2.15).

Summing up the contributions from all bound states, we further obtain the *total* number density of all hydrogen atoms

$$n_{\text{H}} = \sum_{n=1}^{n_{\text{max}}} n_{\text{H}_n} \approx Z_{\text{H}} \frac{n_{\text{p}}}{g_{\text{p}}} \frac{n_{\text{e}} \lambda_{\text{T,e}}^3}{g_{\text{e}}} e^{\chi_{\text{H}}/(k_B T)}. \quad (2.24)$$

And the ratio between the number densities of ionized and atomic hydrogen is

$$\frac{n_{\text{p}}}{n_{\text{H}}} \approx \frac{g_{\text{p}}}{Z_{\text{H}}} \frac{g_{\text{e}}}{n_{\text{e}} \lambda_{\text{T,e}}^3} e^{-\chi_{\text{H}}/(k_B T)}. \quad (2.25)$$

This is the *Saha equation for hydrogen* — it can be used to calculate the ionization fraction. For pure hydrogen composition, we have $n_{\text{p}} = n_{\text{e}}$ from charge neutrality, and then the hydrogen ionization fraction $y \equiv n_{\text{p}}/(n_{\text{p}} + n_{\text{H}})$ is given by the following quadratic equation

$$\frac{y^2}{1-y} = C(\rho, T) \approx \frac{2}{Z_{\text{H}}(\rho, T)} \frac{m_{\text{p}}}{\rho \lambda_{\text{T,e}}^3} e^{-\chi_{\text{H}}/(k_B T)}. \quad (2.26)$$

where we have taken $g_{\text{p}} = 1$, $g_{\text{e}} = 2$, and $n_{\text{p}} + n_{\text{H}} \approx \rho/m_{\text{p}}$ (ignoring electron mass). The positive solution to the quadratic equation is

$$y(\rho, T) = \left[(C^2 + 4C)^{1/2} - C \right] / 2, \quad (2.27)$$

which is shown in Fig. 2.1.

Typically for $T \lesssim 10^4$ K, one can show that Z_{H} is dominated by the ground-state ($n = 1$) term, so it is often sufficiently accurate to simply adopt $Z_{\text{H}} \approx g_1 = 2$. At much higher temperatures $T \gg 10^4$ K, Z_{H} may be much greater than g_1 , due to significant population of high- n levels³ near n_{max} (these are marginally bound atoms), but hydrogen is fully ionized at such high temperatures anyway. The comparison between the results from these two treatments is shown in the upper panel of Fig. 2.1.

In the limit of $n \rightarrow \infty$, we have $\chi_n \rightarrow 0$ and $Z_{\text{H}_n} \approx 2n^2 e^{-\chi_{\text{H}}/(k_B T)} \rightarrow \infty$. This apparent divergence is removed by the following physical effect which imposes a maximum principle quantum number n_{max} for hydrogen atoms.

³These high- n energy levels, due to their large statistical weights $g_n \propto n^2$, can have significant populations in LTE. Even when the system is in non-LTE, 3-body recombination process $p^+ + 2e^- \leftrightarrow \text{H}_n + e^-$ can efficiently populate sufficiently high- n levels. This can sometimes lead to population inversion when the low- n levels are under-populated as compared to LTE.

The high- n states of hydrogen atoms have electron “orbital radii” r_n such that the classical orbital binding energy $-e^2/(2r_n)$ equals to the quantum mechanical binding energy $-\chi_H/n^2$, so we obtain

$$r_n \simeq n^2 e^2 / (2\chi_H) = n^2 a_0 = 5.3 \times 10^{-5} \text{ cm} (n/100)^2, \text{ for } n \gg 1, \quad (2.28)$$

where $a_0 = e^2 / (2\chi_H) = \hbar^2 / (m_e e^2) = 0.53 \text{ \AA}$ is the Bohr radius. When the “orbital radius” r_n is comparable to the average separation between two adjacent nuclei $[4\pi(n_H + n_p)/3]^{-1/3}$, where $n_H + n_p$ is the total number density of hydrogen nuclei, the electron in consideration can no longer be considered to be bound to an atom, so we obtain a conservative limit for the maximum principle quantum number

$$n_{\max} \simeq a_0^{1/2} \left(\frac{3/(4\pi)}{n_H + n_p} \right)^{1/6} \approx 110 \left(\frac{n_H + n_p}{10^{12} \text{ cm}^{-3}} \right)^{-1/6}. \quad (2.29)$$

This gets rid of the diverging behavior of the sum $\sum_n 2n^2 e^{-\epsilon_n/(k_B T)}$ in the total partition function of hydrogen (eq. 2.21).

Another consequence of the finite gas density is that the ionization energy is reduced from $\chi_n = \chi_H/n^2$ (eq. 2.9) for an isolated atom to $\chi_n = \chi_H(1/n^2 - 1/n_{\max}^2)$. This finite-density effect would increase the ionization fraction as compared to the prediction from our Saha equation — the errors are more significant at higher gas densities. In the extreme limit of very high gas densities $n_H + n_p \gtrsim 10^{24} \text{ cm}^{-3}$ (or $\rho \gtrsim 1 \text{ g cm}^{-3}$), the maximum principle quantum number n_{\max} would become order unity, which means that even ground-state hydrogen atoms are destroyed by collisions with nearby particles and that hydrogen atoms would be ionized by such collisions even at relatively low temperatures $T \lesssim 10^4 \text{ K}$. This effect is called *pressure ionization* and is in the regime of strongly coupled plasmas that is beyond the our scope as we focus on very dilute gas in this book. [Hummer & Mihalas \(1988\)](#) provided an in-depth discussion on other effects affecting the maximum principle number n_{\max} , the dominating one being the perturbations to the marginally bound states by the electric fields of nearby ions.

Finally, we generalize the Saha-Boltzmann equation (2.22) to a general ionization and recombination reaction



where X_n^{q+} is an atom with energy excitation level n in charge state q , and $\text{X}^{(q+1)+}$ stands for *all possible energy excitation levels* in the charge state $q + 1$. For this reaction, the *general Saha-Boltzmann equation* is given by

$$n_{q,n} \approx Z_{q,n} \frac{n_{q+1}}{Z_{q+1}} \frac{n_e \lambda_{T,e}^3}{g_e} e^{\chi_q/(k_B T)}, \quad (2.31)$$

where $Z_{q,n}$ is the partition function for the energy level n in charge state q , Z_{q+1} is the total partition function for charge state $q + 1$, n_{q+1} is the total number density of all atoms

in the $q+1$ charge state, $g_e = 2$ accounts for the two spin states of free electrons, and χ_q is the ionization energy of the q charge state — it equals to the minimum energy to remove an electron from the original atom X_n^{q+} in excited energy level n and leave the remaining atom in the *ground state* of $X^{(q+1)+}$.

Summing up all energy excitation levels of X_n^{q+} , we obtain the general *Saha equation*

$$\frac{n_{q+1}}{n_q} \approx \frac{Z_{q+1}}{Z_q} \frac{g_e}{n_e \lambda_{T,e}^3} e^{-\chi_q/(k_B T)}, \quad (2.32)$$

where $Z_q = \sum_n Z_{q,n}$ is the total partition function for the q charge state. Note that, in eqs. (2.31) and (2.32), we have ignored the small difference between the thermal de Broglie wavelengths of X_n^{q+} and $X^{(q+1)+}$ (affording a fractional error $\mathcal{O}(m_e/m_p) \sim \mathcal{O}(10^{-3})$). If we make use of the electron degeneracy factor $\psi_e(n_e, T) = \ln(n_e \lambda_{T,e}^3 / g_e)$ as defined in eq. (2.12), the Saha equation can be as follows

$$\frac{n_{q+1}}{n_q} \approx \frac{Z_{q+1}}{Z_q} e^{-\psi_e - \chi_q/(k_B T)}, \quad (2.33)$$

which is slightly more concise.

2.1.3 *Mixture of hydrogen and helium in LTE

Let us consider a realistic case of a hydrogen-helium mixture in LTE. Here we focus on the temperatures near 10^4 K and ignore the extremely small number density of He^{2+} ions. The total mass fraction of helium is denoted as Y and $Y \simeq 0.3$ for cosmic abundance, so the total number densities of hydrogen and helium at gas density ρ are given by

$$n_H + n_p \approx \rho \frac{1 - Y}{m_p}, \quad n_{\text{He}} + n_{\text{He}^+} \approx \rho \frac{Y}{4m_p}. \quad (2.34)$$

We denote the hydrogen and helium ionization fractions as

$$y_H = \frac{n_p}{n_H + n_p}, \quad y_{\text{He}} = \frac{n_{\text{He}^+}}{n_{\text{He}} + n_{\text{He}^+}}. \quad (2.35)$$

Charge conservation gives

$$n_e = n_p + n_{\text{He}^+} = \left[y_H (1 - Y) + \frac{y_{\text{He}}}{4} Y \right] \frac{\rho}{m_p}. \quad (2.36)$$

From the Saha equation (2.33), we write

$$\frac{n_p}{n_H} = \frac{y_H}{1 - y_H} = \frac{Z_p}{Z_H} e^{-\psi_e - \chi_H/(k_B T)}, \quad \frac{n_{\text{He}^+}}{n_{\text{He}}} = \frac{y_{\text{He}}}{1 - y_{\text{He}}} = \frac{Z_{\text{He}^+}}{Z_{\text{He}}} e^{-\psi_e - \chi_{\text{He}}/(k_B T)}, \quad (2.37)$$

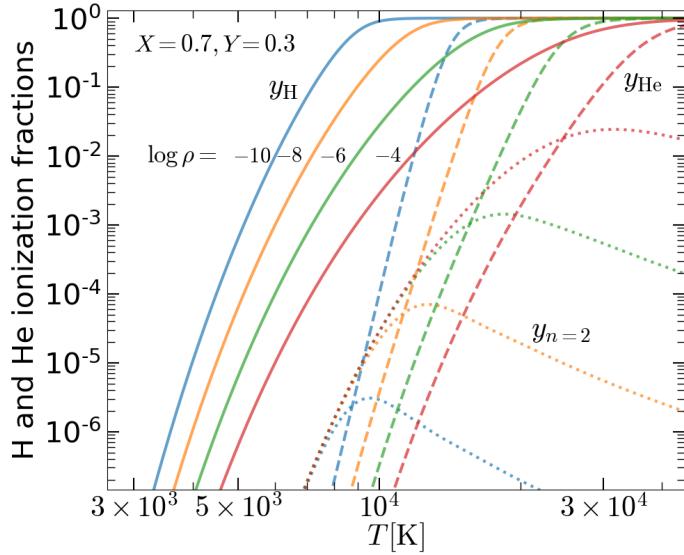


Figure 2.2: LTE ionization fractions of H (solid lines) and He (dashed lines) at different temperatures and densities. The dotted lines show the fraction of hydrogen that are in the $n = 2$ energy level (eq. 2.41). The H and He mass fractions are taken to be $X = 0.7$ and $Y = 0.3$.

where the electron degeneracy factor $\psi_e(n_e, T)$ is defined in eq. (2.12), and the relevant ionization energies are $\chi_H = 13.6 \text{ eV}$ and $\chi_{He} = 24.6 \text{ eV}$. In the partition functions, we ignore nuclear spins and only consider the population of the ground state of each ion (e.g., the two electrons in ground-state neutral He have opposite spins), and hence

$$Z_p = 1, \quad Z_H \approx 2, \quad Z_{He^+} \approx 2, \quad Z_{He} \approx 1. \quad (2.38)$$

We plug these into the expressions in (2.37) and obtain

$$\frac{1 - y_H}{y_H} \approx n_e \lambda_{T,e}^3 e^{\chi_H / (k_B T)}, \quad \frac{1 - y_{He}}{y_{He}} \approx \frac{1}{4} n_e \lambda_{T,e}^3 e^{\chi_{He} / (k_B T)}, \quad (2.39)$$

where the free electron density n_e is given by eq. (2.36). For each given set of (ρ, T) , we then numerically solve these two equations for two variables y_H and y_{He} . The solutions for H mass fraction $X = 0.7$ and He mass fraction of $Y = 0.3$ are shown in Fig. 2.2.

Compared with the case of a pure hydrogen composition, the y_H in a H-He mixture is moderately affected by the existence of He. In particular, at a given density ρ and temperature T , y_H increases as we add more He. This is because a higher He mass fraction decreases the number density of free electrons (which are locked inside He), and this moves the equilibrium of the reaction $H + \gamma \leftrightarrow p^+ + e^-$ towards the right-hand side of the equation according to the *Le Chatelier's principle* (which is familiar to chemists).

Finally, the fraction of hydrogen in the energy level with principle quantum number n is given by

$$y_n \equiv \frac{n_{\text{H}_n}}{n_{\text{p}} + n_{\text{H}}} = (1 - y_{\text{H}}) \frac{Z_{\text{H}_n}}{Z_{\text{H}}} \approx (1 - y_{\text{H}}) \exp \left[-\frac{\chi_{\text{H}}}{k_{\text{B}}T} \left(1 - \frac{1}{n^2} \right) \right]. \quad (2.40)$$

For instance, for $n = 2$, we obtain

$$y_{n=2} \approx (1 - y_{\text{H}}) e^{-10.2 \text{ eV}/(k_{\text{B}}T)}, \quad (2.41)$$

which is shown in Fig. 2.2. At very low temperatures $T \ll 10^4 \text{ K}$, the neutral fraction is near unity ($1 - y_{\text{H}} \approx 1$) but the Boltzmann factor of $e^{-10.2 \text{ eV}/(k_{\text{B}}T)}$ becomes very small, so we see that $y_{n=2}$ decreases towards lower temperatures. At very high temperatures $T \gg 10^4 \text{ K}$, the Boltzmann factor of $e^{-10.2 \text{ eV}/(k_{\text{B}}T)}$ becomes order unity, but the neutral fraction becomes very small ($1 - y_{\text{H}} \ll 1$), so $y_{n=2}$ again decreases towards higher temperatures. Roughly speaking, $y_{n=2}$ peaks near $T \simeq 10^4 \text{ K}$. This is why stars of spectral type A (with photospheric temperature near 10^4 K) have the strongest Balmer absorption lines.

2.1.4 Blackbody and Bose-Einstein distribution

Photons are spin=1 bosons, meaning that they have a spin angular momentum of \hbar . In this section, we show that, in LTE, the photon occupation number in *each spin state* is given by the Bose-Einstein distribution with zero chemical potential ($\mu = 0$),

$$\eta_{\gamma} = \frac{1}{e^{h\nu/(k_{\text{B}}T)} - 1}, \quad (2.42)$$

where $h\nu$ is the photon energy. The zero chemical potential is essentially because photons are constantly created and destroyed at an equal rate when the system is in LTE. For readers who would like to know more details, here we provide a proof of $\mu = 0$ when the photon number is not conserved. Whoever already accepted $\mu = 0$ may skip all the way to eq. (2.51) and onward where we discuss the properties of a blackbody.

The first law of thermodynamics tells us $dE = TdS - PdV + \mu dN$, where E =internal energy, S =entropy, P =pressure, V =volume, and N =number of particles. Let us consider a thermally insulated box of volume V . The inner walls only spatially confine the photons without absorption/emission. Inside the box, there are two ingredients: (1) a very small amount of matter that can emit, absorb, and scatter photons at all frequencies; and (2) a large number of photons with total energy E that is much larger than the energy content of matter. In the astrophysical context, we are simply considering a *radiation-dominated* gas. For fixed E and V , we see that the creation and destruction of photons are associated with the entropy change by

$$dS = -\frac{\mu}{T} dN, \text{ for } dE = 0, dV = 0. \quad (2.43)$$

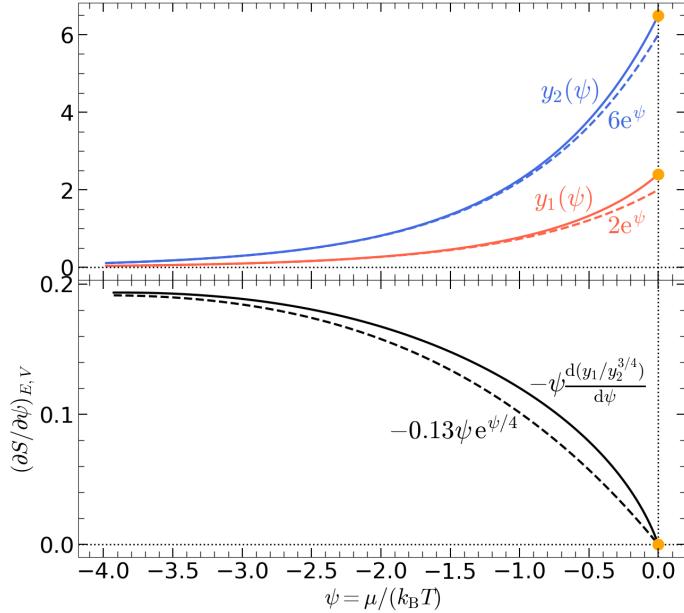


Figure 2.3: Upper panel: the two functions $y_1(\psi)$ and $y_2(\psi)$ related to the photon number density n_{ph} and radiation energy density U (respectively) in a Bose-Einstein distribution. The dashed lines show the approximations. Lower panel: the function $-\psi d(y_1/y_2^{3/4})/d\psi \simeq -0.13\psi e^{\psi/4}$ related to the entropy derivative $(\partial S/\partial \psi)_{E,V}$ is positive for all $\psi < 0$ and the maximum entropy is achieved when $\psi = 0$.

After a sufficiently long time when the radiation-matter system reaches LTE, the entropy would be maximized such that $dS = 0$. However, since photons can be created and destroyed by matter, we have $dN \neq 0$, and this means $\mu = 0$. If you are not still satisfied, here is a longer version of the argument.

Let us consider the general Bose-Einstein distribution (eq. 2.1) with a non-zero chemical potential μ , and the photon number density is

$$n_{\text{ph}} = \int_0^\infty \frac{2\eta_\gamma}{h^3} 4\pi p^2 dp = \frac{8\pi(k_B T)^3}{h^3 c^3} y_1, \quad y_1(\psi) = \int_0^\infty \frac{x^2 dx}{e^{x-\psi} - 1}, \quad \psi \leq 0, \quad (2.44)$$

and the radiation energy density is

$$U = \int_0^\infty \frac{2\eta_\gamma}{h^3} 4\pi p^3 c dp = \frac{8\pi(k_B T)^4}{h^3 c^3} y_2, \quad y_2(\psi) = \int_0^\infty \frac{x^3 dx}{e^{x-\psi} - 1}, \quad \psi \leq 0, \quad (2.45)$$

where we have defined $\psi \equiv \mu/(k_B T)$ and $x \equiv pc/(k_B T)$. It is clear that only $\psi \leq 0$ (or $\mu \leq 0$) is allowed, because otherwise the integrals y_1 and y_2 would diverge. This means

that Bose-Einstein distribution for a radiation field in thermal equilibrium must have non-positive chemical potential $\mu \leq 0$.

As shown in Fig. 2.3, both $y_1(\psi)$ and $y_2(\psi)$ are positive and monotonically increasing functions and their maximum values are given by

$$y_1(0) = \int_0^\infty \frac{x^2 dx}{e^x - 1} = 2\zeta(3) \approx 2.404, \quad y_2(0) = \int_0^\infty \frac{x^3 dx}{e^x - 1} = \frac{\pi^4}{15} \approx 6.494, \quad (2.46)$$

where $\zeta(3)$ is called the Apéry's constant and ζ is the Riemann zeta function. In fact, the two functions y_1 and y_2 can be reasonably approximated by

$$y_1(\psi) \simeq e^\psi \int_0^\infty x^2 e^{-x} dx = 2e^\psi, \quad y_2(\psi) \simeq e^\psi \int_0^\infty x^3 e^{-x} dx = 6e^\psi. \quad (2.47)$$

These approximations are exact in the limit of $|\psi| \gg 1$ (i.e., ψ is a large negative number⁴) and the errors are small even in the worst case of $\psi = 0$ (as one can see by comparing the numerical and approximate results). The fact that $y_2/y_1 \simeq 3$ means that the mean energy per photon $U/n_{\text{ph}} = (y_2/y_1)k_B T \simeq 3k_B T$ for any Bose-Einstein distribution.

From the radiation energy density U , we obtain the equilibrium temperature

$$T(\psi) = \frac{(hc)^{3/4}}{(8\pi)^{1/4} k_B} U^{1/4} [y_2(\psi)]^{-1/4}. \quad (2.48)$$

The photon number density is given by

$$n_{\text{ph}}(\psi) = \frac{(8\pi)^{1/4}}{(hc)^{3/4}} U^{3/4} \frac{y_1(\psi)}{[y_2(\psi)]^{3/4}}. \quad (2.49)$$

For a fixed energy density $U = E/V$ in our consideration, we see that the equilibrium temperature $T(\psi) \propto y_2^{-1/4}$ and the photon number density $n_{\text{ph}}(\psi) \propto y_1/y_2^{3/4}$ are both entirely determined by ψ and hence the chemical potential μ .

To show that the entropy is maximized when $\psi = 0$ (or $\mu = 0$), we study how the entropy S depends on ψ using eq. (2.43)

$$\begin{aligned} \left(\frac{\partial S}{\partial \psi} \right)_{E,V} &= -k_B \psi \left(\frac{\partial N}{\partial \psi} \right)_{E,V} = -k_B V \psi \left(\frac{\partial n_{\text{ph}}}{\partial \psi} \right)_{E,V} \\ &= -\frac{(8\pi)^{1/4} k_B}{(hc)^{3/4}} V U^{3/4} \psi \frac{d}{d\psi} \left(y_1/y_2^{3/4} \right). \end{aligned} \quad (2.50)$$

⁴The limit of $|\psi| \gg 1$ is called the *Wien distribution* for which the radiation energy density is given by $U = (6/6.494)aT^4 e^\psi = 0.924 aT^4 e^\psi \ll aT^4$, where $a = 8\pi^5 k_B^4 / (15h^3 c^3)$ is the radiation density constant (see eq. 2.58).

Using the roughly approximations $y_1 \simeq 2e^\psi$ and $y_2 \simeq 6e^\psi$, we see that $y_1/y_2^{3/4} \propto e^{\psi/4}$ on the RHS of eq. (2.50) is a monotonically increasing function of ψ , so its derivative $d(y_1/y_2^{3/4})/d\psi \propto e^{\psi/4}$ is always positive for any $\psi \leq 0$. Thus, we conclude $dS/d\psi \propto -\psi d(y_1/y_2^{3/4})/d\psi > 0$ for $\psi < 0$ and $dS/d\psi = 0$ at $\psi = 0$.

The above discussion shows that the entropy inside the box monotonically increases with ψ for $\psi < 0$ and it reaches a maximum at $S_{\max} = S(\psi = 0)$. Therefore, we conclude that, when photons can be created or destroyed (and hence N is not conserved), the entropy of the system is maximized at $\psi = 0$ or zero chemical potential $\mu = 0$. The solution with maximum entropy is indeed achieved in nature, as can be seen from the beautiful spectrum of the [cosmic microwave background](#).

We also note that an equilibrium radiation field with a negative chemical potential $\mu < 0$ can be produced if the photons inside the box are *only scattered* by matter such that the total photon number N is conserved. Such a situation will be discussed in Chapter 7. There, we will show that the Bose-Einstein distribution with $\mu < 0$ is the equilibrium solution of Comptonization, and in that case, the chemical potential μ can be obtained from the (given and fixed) photon number density n_{ph} . However, the state of Compton equilibrium is not LTE.

The photon occupation number η_γ at LTE (eq. 2.42) gives the specific intensity for the *blackbody radiation field*, also known as the *Planck function* (see Fig. 2.4)

$$I_\nu(\text{LTE}) = B_\nu(T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/(k_B T)} - 1}. \quad (2.51)$$

In LTE, the radiation field is isotropic and the frequency spectrum only depends on a single variable — the radiation temperature T which equals to the kinetic temperature of gas. The mean photon energy is given by

$$\langle h\nu \rangle = \frac{\int B_\nu d\nu}{k_B T \int B_\nu/(h\nu) d\nu} = \frac{\int_0^\infty x^3 (e^x - 1)^{-1} dx}{\int_0^\infty x^2 (e^x - 1)^{-1} dx} = \frac{\pi^4/15}{2\zeta(3)} \approx 2.70. \quad (2.52)$$

The peak frequency of the Planck function, where B_ν is maximized or $\partial_\nu B_\nu = 0$, is at

$$h\nu_{\text{pk}} \approx 2.82 k_B T. \quad (2.53)$$

To find the peak of $y(x) = x^q/(e^x - 1)$, we solve $(dy/dx)|_{x_{\text{pk}}} = 0$, which gives

$$q = \frac{x_{\text{pk}}}{1 - e^{-x_{\text{pk}}}} \approx \begin{cases} 1, & \text{for } x_{\text{pk}} \ll 1, \\ x_{\text{pk}}, & \text{for } e^{x_{\text{pk}}} \gg 1 \text{ or } x_{\text{pk}} \gtrsim 2. \end{cases} \quad (2.54)$$

For a relatively large $q \gtrsim 2$, the lowest-order solution is $x_{\text{pk}}^{(1)} \simeq q$. In fact, a more accurate solution can be obtained for $q \gtrsim 2$. Let us write the solution in the form of $x_{\text{pk}} = q + \Delta x$,

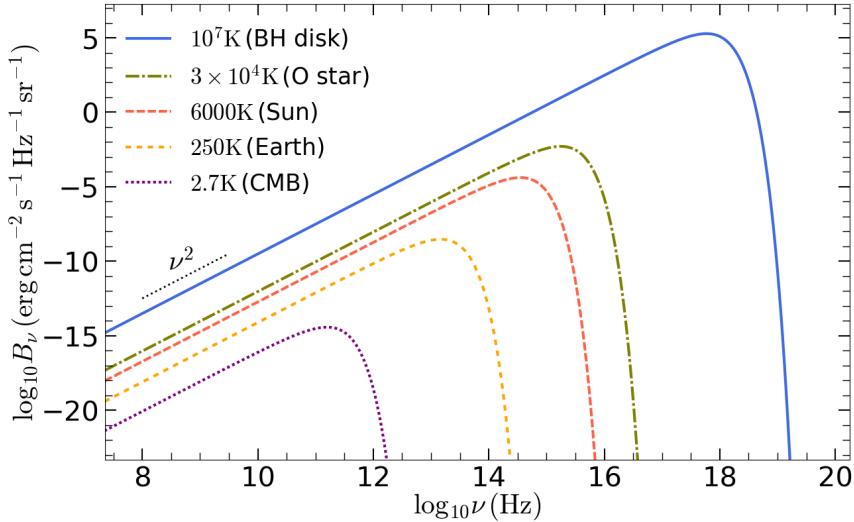


Figure 2.4: Planck function for different temperatures: cosmic microwave background (CMB), the Earth's atmospheric thermal emission, the Sun's photosphere, a typical O-type star's photosphere, and the typical temperature at the innermost regions of a black hole (BH) accretion disk. In reality, CO₂, H₂O and other molecules in the Earth's atmosphere cause absorption lines on top of the continuum emission.

where $|\Delta x| \ll q$. We plug this form into the above equality and obtain

$$q = \frac{q + \Delta x}{1 - e^{-q - \Delta x}} \approx \frac{q + \Delta x}{1 - e^{-q}} \approx (q + \Delta x)(1 + e^{-q}) \approx q + q e^{-q} + \Delta x, \quad (2.55)$$

where we have adopted a few approximations based on $e^{-q} \ll 1$ and $|\Delta x| \ll q$. Then, we obtain $\Delta x \approx -q e^{-q}$ and hence a higher-order solution

$$x_{\text{pk}}^{(2)} \approx q - q e^{-q}. \quad (2.56)$$

The blackbody spectrum has $q = 3$, so the above approximation gives $x_{\text{pk}}^{(2)} \approx 2.85$, close to the numerical result of 2.82. It is not worthwhile to go to an even higher order.

The total energy density in a blackbody radiation field is

$$U = \int U_\nu d\nu = \int \frac{4\pi}{c} B_\nu d\nu = aT^4, \quad (2.57)$$

where a is the radiation density constant, related to the Stefan-Boltzmann constant σ_{SB} by

$$a = \frac{4\sigma_{\text{SB}}}{c} \approx 7.56 \times 10^{-15} \frac{\text{erg}}{\text{cm}^3 \text{K}^4}, \quad \sigma_{\text{SB}} = \frac{2\pi^5 k_{\text{B}}^4}{15c^2 h^3} \approx 5.67 \times 10^{-5} \frac{\text{erg}}{\text{cm}^2 \text{s K}^4}. \quad (2.58)$$

The radiation pressure is

$$P = U/3 = aT^4/3. \quad (2.59)$$

The specific flux flowing through a small surface in one direction is $F_\nu = \pi B_\nu$ and the frequency-integrated flux is

$$F = \sigma_{\text{SB}} T^4, \quad (2.60)$$

The dimensionless ratio between νF_ν (which is sometimes called the *spectral flux*) and the total flux is

$$\frac{\nu F_\nu}{F} = \frac{\pi \nu B_\nu}{\sigma_{\text{SB}} T^4} = \frac{15}{\pi^4} \frac{x^4}{e^x - 1}, \quad x = h\nu/(k_B T). \quad (2.61)$$

The above function peaks at $x_{\text{pk}} \approx 3.92$ and $(\nu F_\nu)_{\text{max}} \approx 0.736F$. Note that the numerical value of $x_{\text{pk}} \approx 3.92$ is very close to the analytic result of 3.93 given by eq. (2.56), as the analytic approximation becomes exact in the limit $e^{-x_{\text{pk}}} \rightarrow 0$.

The entropy can be obtained by considering that an isolated box containing blackbody radiation is gradually heated from zero temperature to T . The entropy change in this process is given by (the initial entropy is zero)

$$S = V \int \frac{dU}{T} = V \int_0^T \frac{4aT^3 dT}{T} = \frac{4aT^3 V}{3}, \quad (2.62)$$

where V is the volume of the box. Since $S \propto P^{3/4}V$, we see that isentropic evolution radiation field can be described by a polytropic equation of state $PV^{4/3} = \text{const}$ with an adiabatic index $\gamma_{\text{ad}} = 4/3$.

2.2 Einstein relations

In this section, we discuss the fundamental connections between emission and absorption processes driven by the system's tendency of evolving towards LTE.

2.2.1 Spontaneous emission, stimulated emission, true absorption

Let us denote the total number of matter particles (e.g., atoms, free electrons) in a given system in two energy levels 1 and 2 as N_1 and N_2 , respectively. The energy levels are denoted as ϵ_1 and ϵ_2 . We will ignore the finite widths of these two energy levels for now (this will be discussed later in 2.2.2). The energy levels each have g_1 and g_2 distinct quantum states. For instance, the first excited energy level of a hydrogen atom with principle quantum number $n = 2$ has $g = 2n^2 = 8$ distinct spin-orbit states. Each $2 \rightarrow 1$ transition emits a photon of energy

$$h\nu_0 = \epsilon_2 - \epsilon_1. \quad (2.63)$$

For simplicity, we restrict our discussion to case where both matter and radiation have isotropic distribution functions, but our conclusions can be generalized to the anisotropic case as well.

Each particle in the higher energy level 2 tends to spontaneously transition down to energy level 1 at a rate that is denoted as A . This is a process called *spontaneous emission*, which the total system would evolve towards a higher entropy, as the emitted photon takes the excessive energy in the *localized* matter particle to other places.

In the *absence of radiation field*, the total downward transition rate⁵ in the system is

$$R_{2 \rightarrow 1} = N_2 A, \text{ if no radiation field.} \quad (2.64)$$

Here, A (in units of sec^{-1}) is called the *Einstein-A coefficient*, which is either theoretically calculated or experimentally measured. For instance, the [NIST Atomic Spectra Database](#) contains extensive tables of Einstein- A 's for various atomic transitions.

We then consider that the matter particles are embedded in an isotropic radiation field which is described by the photon occupation number $\eta_\gamma(\nu)$ or equivalently the intensity $I_\nu = 2h\nu^3\eta_\gamma/c^2$ as a function of frequency ν . Here, we are only interested the occupation number $\eta(\nu_0)$ at the given frequency ν_0 for the transition in consideration.

In the following, we show that the total rate of $2 \rightarrow 1$ transitions in the matter-radiation system is given by

$$R_{2 \rightarrow 1} = N_2 A [1 + \eta_\gamma(\nu_0)], \quad (2.65)$$

where $N_2 A$ is the rate of *spontaneous emission* and $N_2 A \eta_\gamma(\nu_0)$ is the rate of *stimulated emission*. The $1 + \eta_\gamma$ factor is nicely explained by Feynman in his *Lectures on Physics*.

Consider that an *integer number* of $n \geq 1$ photons in the same quantum state, where n is the photon occupation number of that quantum state. Let us label these photons by $1, 2, \dots, n$, despite the fact that they are physically indistinguishable. There can be $n!$ possible ways⁶ of listing the n photons if we *pretend* that they are distinguishable. For *each one* of the $n!$ configurations, when we try to add a new, $(n + 1)$ -th, photon to the list, there are $n + 1$ possible ways of doing it. For instance, if $n = 3$ and the original list is $(1, 2, 3)$, then we can add a 4th photon in the following ways: $(4, 1, 2, 3)$, $(1, 4, 2, 3)$, $(1, 2, 4, 3)$, $(1, 2, 3, 4)$. In statistical mechanics, the probability of achieving a certain final systemic state (e.g., 4 photons in a given quantum state) is linearly proportional to the number of allowed configurations. Therefore, the probability that a matter particle will

⁵Note that we have assumed that the matter particles are non-degenerate in the sense that the occupation numbers η_1 and η_2 in each of the relevant quantum states are much less than unity. If the matter particles are fermions (e.g., free electrons) and end state of a given transition is already partially occupied, the spontaneous transition rate would be given by $N_2 A(1 - \eta_1)$. The factor of $1 - \eta_1$ is due to an effect called *Pauli blocking* — in the extreme case of $\eta_1 = 1$, there can be no $2 \rightarrow 1$ transition at all.

⁶For $n = 3$, these possibilities are: $(3, 1, 2)$, $(1, 3, 2)$, $(1, 2, 3)$, $(3, 2, 1)$, $(2, 3, 1)$, $(2, 1, 3)$.

emit a new photon into a given photon quantum state is enhanced by a factor of $(1 + n)$ if there are already n photons in that state. Note that all possible ways of adding the $(n+1)$ -th photon to the existing list of n photons give the same observational consequence: the photon occupation number in this state is $n+1$ and we can say no more than that because, after all, photons in the same quantum state are indistinguishable.

A puzzled reader might wonder why we only consider an integer number of n photons in a given quantum state, whereas the photon occupation $\eta_\gamma > 0$ can in general be any positive real number. The difference is that η_γ by definition is the *time-averaged photon occupation number* in a given quantum state — we have taken an average on a timescale $T \gg \nu^{-1}$ (but T must be short compared to the timescale in which η_γ varies). At a given moment within the time window T , the number of photons in a given quantum state can only be an integer n (there is no such thing as “half a photon”), but the time-averaged $\eta_\gamma = \langle n \rangle_T$ can be any real positive number.

Knowing the downward transition rate $R_{1 \rightarrow 2}$, we then move on to *calculate* the upward transition rate of $1 \rightarrow 2$. To calculate $R_{1 \rightarrow 2}$, the only thing we will need is that any system has a *tendency* of evolving towards LTE (to maximize the total entropy), as we know from the second law of thermodynamics.

We start by writing the upward transition rate in the following form

$$R_{1 \rightarrow 2} = CN_1\eta_\gamma(\nu_0), \quad (2.66)$$

where C is a constant related to the nature of the given quantum transition. The linear proportionality of $R_{1 \rightarrow 2} \propto N_1$ can be easily understood by considering the transition rate within e.g., half of the system (containing half of the particles). The linear proportionality of $R_{1 \rightarrow 2} \propto \eta_\gamma(\eta_0) \propto I_\nu(\nu_0)$ can be demonstrated by considering the fact each photon at frequency ν_0 has a certain rate of being absorbed.

Then we consider the following thought experiment: suppose the entire radiation-matter system is already in LTE at a common temperature T such that matter particles follow the Boltzmann distribution $N_2/N_1 = (g_2/g_1)\exp(-h\nu_0/k_B T)$ and that photons follow the Planck distribution $\eta_\gamma(\nu_0) = [\exp(h\nu_0/k_B T) - 1]^{-1}$, then it is easy to show

$$\frac{N_2}{g_2} [1 + \eta_\gamma(\nu_0)] = \frac{N_1}{g_1} \eta_\gamma(\nu_0), \quad \text{if in LTE.} \quad (2.67)$$

Multiplying both sides by A and making use of eq. (2.65), we obtain

$$R_{2 \rightarrow 1} = \frac{g_2}{g_1} N_1 A \eta_\gamma(\nu_0), \quad \text{if in LTE.} \quad (2.68)$$

By construction in our thought experiment, we must have $R_{1 \rightarrow 2} = R_{2 \rightarrow 1}$ as a result of

detailed balance in LTE. Therefore, we find the upward transition rate⁷

$$R_{1 \rightarrow 2} = \frac{g_2}{g_1} N_1 A \eta_\gamma(\nu_0), \quad (2.69)$$

which is equal to the rate at which photons are absorbed. Let us compare the above result with eq. (2.66) which shows that the upward transition rate is linearly proportional to $N_1 \eta_\gamma(\nu_0)$, and we find the constant of proportionality to be $C = (g_2/g_1)A$.

Now we step back and *remove* the LTE assumption. The N_1 particles in energy level 1, without knowing the existence of the N_2 particles in level 2, would simply interact with the radiation field and undergo upward transitions at a rate given by eq. (2.69), as a result of intrinsic tendency of doing so. Another way of thinking is that the constant $C = (g_2/g_1)A$ is independent of N_1 , N_2 , $\eta_\gamma(\nu_0)$ and hence the assumption of LTE. The reasoning above, provided by Einstein, is very powerful in that, to calculate upward and downward radiative transition rates, we only need to know the spontaneous emission rate A and the photon occupation number $\eta_\gamma(\nu_0)$ in the radiation field.

The number of photons at frequency ν_0 changes at a rate given by

$$\frac{dN_{\text{ph}}(\nu_0)}{dt} = N_2 A - \left(\frac{g_2}{g_1} N_1 - N_2 \right) A \eta_\gamma(\nu_0), \quad (\text{for monochromatic photons}), \quad (2.70)$$

where the first term ($\propto N_2$) is from *spontaneous emission*, the second term ($\propto N_1 \eta_\gamma$) is due to *true absorption*, and the third term ($\propto N_2 \eta_\gamma$) is due to *stimulated emission* (which is commonly treated as *negative absorption*). The ratio between the rates of absorption and stimulated emission in the whole system is

$$\frac{\text{stimulated emission}}{\text{true absorption}} = \frac{N_2/g_2}{N_1/g_1} = \frac{\eta_2}{\eta_1}, \quad (2.71)$$

where the RHS is simply the ratio between the occupation number of matter particles = the number of particles per quantum state in the two energy levels. The ratio between the rates of stimulated emission and spontaneous emission is given by the photon occupation number

$$\frac{\text{stimulated emission}}{\text{spontaneous emission}} = \eta_\gamma(\nu_0). \quad (2.72)$$

The *net absorption* rate is defined as the rate of true absorption minus that of stimulated emission, and we can write

$$\frac{\text{net absorption}}{\text{spontaneous emission}} = \left(\frac{\eta_1}{\eta_2} - 1 \right) \eta_\gamma(\nu_0). \quad (2.73)$$

⁷Note the absence of a final-state factor of $1 + \eta_\gamma(\nu_0)$ in the upward transition rate $R_{1 \rightarrow 2}$. This is because there is no photon in the final state.

The relations in Eqs. (2.71), (2.72), and (2.73) (only two of three are independent) are called the *Einstein relations*, which hold for all radiative transitions.

It is intriguing to note that the rates of all three processes (spontaneous emission, stimulated emission, and true absorption) related to a given radiative transition are proportional to the Einstein-A coefficient of that transition. This can be summarized as:

$$A \text{ good emitter is also a good absorber.}$$

In the following, we use the concept of *excitation temperature* to demonstrate that the Einstein relations describe the system's tendency of approaching LTE.

2.2.2 Einstein coefficients: A , B_{12} , and B_{21}

In the considerations in the earlier subsection, we have ignored the energy width of the radiative transition — we are effectively taking the line profile to be in the limit of a δ -function centered at frequency ν_0 . Realistically, there will be a line profile function ϕ_ν , which is *nearly* a δ -function near the central frequency ν_0 and is normalized as $\int \phi_\nu d\nu = 1$ by definition. In this case, we would need to replace $\eta_\gamma(\nu_0)$ with the profile-averaged photon occupation number

$$\bar{\eta}_\gamma \equiv \int \eta_\gamma(\nu) \phi_\nu d\nu. \quad (2.74)$$

Then, eq. (2.70) on the evolution of the total photon number needs to be slightly modified as follows

$$\frac{dN_{\text{ph}}}{dt} = N_2 A - \left(\frac{g_2}{g_1} N_1 - N_2 \right) A \bar{\eta}_\gamma. \quad (2.75)$$

We then introduce the definition of the two *Einstein-B coefficients*, which are based on the profile-averaged intensity $\bar{J}_\nu = \int J_\nu \phi_\nu d\nu = (2h\nu_0^3/c^2)\bar{\eta}_\gamma$ and J_ν is the mean intensity.

The first Einstein-B coefficient B_{12} is defined based on the rate of true absorption per matter particle in energy level ϵ_1 as⁸

$$B_{12} \bar{J}_\nu = \frac{g_2}{g_1} A \bar{\eta}_\gamma, \quad (2.76)$$

so we obtain

$$B_{12} = \frac{g_2}{g_1} \frac{c^2}{2h\nu_0^3} A. \quad (2.77)$$

The second Einstein-B coefficient B_{21} is defined based on the stimulated emission rate per matter particle in energy level ϵ_2 as

$$B_{21} \bar{J}_\nu = A \bar{\eta}_\gamma, \quad (2.78)$$

⁸It should be noted that sometimes the Einstein-B coefficients are defined differently based on the profile-averaged specific energy density $\bar{U}_\nu = (4\pi/c)\bar{J}_\nu$ — one should always check.

so we obtain

$$B_{21}/B_{12} = g_1/g_2. \quad (2.79)$$

Eqs. (2.77) and (2.79) are another way of expressing the *Einstein relations* and they are equivalent to eqs. (2.71) and (2.72).

2.2.3 Excitation temperature and brightness temperature

Often, we would like to compare the rate of net absorption and the rate of spontaneous emission, as the ratio of the two determines whether the energy flows from matter to radiation or the other way around. This ratio is given by $(\eta_1/\eta_2 - 1)\eta_\gamma(\nu_0)$ (eq. 2.73), where the first term $(\eta_1/\eta_2 - 1)$ is related to the distribution function of matter particles and the second term $\eta_\gamma(\nu_0)$ is related to the distribution function of photons. In this subsection, we introduce the concepts of *excitation temperature* (for matter) and *brightness temperature* (for radiation) to describe these two terms in a more intuitive way.

There are two different regimes depending on the matter distribution function: *normal population* ($\eta_1/\eta_2 > 1$, for positive net absorption) and *inverted population* ($\eta_1/\eta_2 < 1$ for negative net absorption).

Inverted population with $\eta_1/\eta_2 < 1$ occurs in rare cases where the higher energy level ϵ_2 is a mesa-stable one that is efficiently “pumped” by other radiative/collisional transitions. The consequence is that the net absorption rate becomes negative, so the radiation intensity would get exponentially amplified. This is the case for astrophysical *masers*, where the observed intensities are orders of magnitude higher than expected from normal population at the typical temperatures in those environments. As we will see below, under normal population, the maximum intensity is given by the blackbody at the matter excitation temperature. The first known astrophysical maser comes from the rotational transitions of OH (hydroxyl radical) — which is also the first known interstellar molecule. [This catalog](#) contains known OH masers around cool stars with strong mass loss.

In the following, we focus on the case of normal population where the rate of net absorption is *positive*. This is the most common case as the lower energy level usually has a larger occupation number.

As long as $\eta_1/\eta_2 > 1$, we can define an *excitation temperature* T_{exc} (for matter) in the form the Boltzmann distribution

$$\frac{\eta_2}{\eta_1} \equiv \exp[-h\nu_0/(k_B T_{\text{exc}})] \Leftrightarrow T_{\text{exc}} \equiv \frac{h\nu_0/k_B}{\ln(\eta_1/\eta_2)}. \quad (2.80)$$

One might notice from the above expression that the excitation temperature would become negative ($T_{\text{exc}} < 0$) for inverted population. Mathematically, such a negative excitation temperature is not problematic but it is not used in this book so as to avoid confusion.

On the other hand, we define an *brightness temperature* T_b for radiation in the form of the Planck distribution

$$\eta_\gamma(\nu_0) \equiv \frac{1}{e^{h\nu_0/(k_B T_b)} - 1} \Leftrightarrow T_b(\nu_0) \equiv \frac{h\nu_0/k_B}{\ln [1 + \eta_\gamma^{-1}(\nu_0)]} \quad (2.81)$$

More generally, the brightness temperature is defined as the temperature that a blackbody should have to produce a given specific intensity I_ν at a given frequency ν , i.e.

$$I_\nu = B_\nu(T_b). \quad (2.82)$$

We see that the brightness temperature $T_b(\mathbf{r}, \mathbf{p}, t)$ provides another general way of describing an arbitrary radiation field, equivalent to the specific intensity function I_ν , photon occupation number η_γ , and the phase-space distribution function $f(\mathbf{r}, \mathbf{p}, t)$.

Using these two temperatures T_{exc} and T_b , we can write the ratio between the rates of net absorption and spontaneous emission as (cf. 2.73)

$$\frac{\text{net absorption}}{\text{spontaneous emission}} = \left(\frac{\eta_1}{\eta_2} - 1 \right) \eta_\gamma = \frac{e^{h\nu_0/(k_B T_{\text{exc}})} - 1}{e^{h\nu_0/(k_B T_b)} - 1}. \quad (2.83)$$

In the Rayleigh-Jeans limit $h\nu_0 \ll k_B \min(T_{\text{exc}}, T_b)$, we use Taylor expansion to express the exponential terms $e^x \approx 1 + x$ (for $x \ll 1$) and obtain

$$\frac{\text{net absorption}}{\text{spontaneous emission}} \approx \frac{T_b}{T_{\text{exc}}}, \quad \text{in the R-J limit.} \quad (2.84)$$

If $T_{\text{exc}} > T_b$ (we would say that “matter is hotter than radiation”), the above ratio is less than 1, so the rate of net absorption is slower than the rate of spontaneous emission, and hence the radiation field will become stronger (T_b increases) over time and the population in the higher energy level will go down (T_{exc} decreases) over time. On the other hand, if $T_{\text{exc}} < T_b$ (we would say that “radiation is hotter than matter”), then the net absorption rate exceeds the spontaneous emission rate, and we expect T_{exc} to increase and T_b to decrease over time.

We conclude that the interplay among the rates of spontaneous emission, stimulated emission, and true absorption always tries to push the distribution functions of radiation and matter towards the *stable equilibrium* at which $T_{\text{exc}} = T_b$. In the above discussion, we are only considering a single radiative transition near frequency ν_0 , so the equilibrium solution of $T_{\text{exc}} = T_b$ does not mean LTE for the whole system (e.g., the radiation field does not necessarily have a blackbody spectrum). The system would be in LTE only if such an equilibrium is achieved at all frequencies. Nevertheless, we see that any radiation-matter system has the tendency of evolving towards LTE (where the total entropy is maximized).

Below, we use a simplified example to demonstrate the way equilibrium is achieved. A more in-depth discussion will be provided when we study the radiative transfer equation.

Suppose the matter excitation temperature T_{exc} is fixed (by some physical mechanism), and we consider the Rayleigh-Jeans limit $h\nu_0 \ll k_B \min(T_{\text{exc}}, T_b)$. Our goal is to solve for the time evolution of the brightness temperature $T_b(t)$. The photon number evolves with time as described by (cf. eq. 2.70)

$$\frac{dN_{\text{ph}}(\nu_0)}{dt} = N_2 A \left[1 - \left(\frac{\eta_1}{\eta_2} - 1 \right) \eta_\gamma(\nu_0) \right] \approx N_2 A \left(1 - \frac{T_b(\nu_0)}{T_{\text{exc}}} \right). \quad (2.85)$$

Since $N_{\text{ph}}(\nu_0) \propto I_{\nu_0} \propto \eta_\gamma(\nu_0) = [e^{h\nu_0/(k_B T_b)} - 1]^{-1} \approx k_B T_b / (h\nu_0) \propto T_b$, $N_2 A = \text{const}$, and $T_{\text{exc}} = \text{const}$, we can write the time evolution of the brightness temperature as

$$\frac{dT_b}{dt} = C (T_{\text{exc}} - T_b(t)), \quad (2.86)$$

where C is a constant that has the dimension of sec^{-1} . We then define another dimensionless time variable $\tau \equiv C dt$ (which will become the optical depth in our later discussion) and write

$$\frac{dT_b}{d\tau} = T_{\text{exc}} - T_b(\tau). \quad (2.87)$$

Integrating the above equation with the initial condition $T_b(\tau = 0) = T_{b,0}$, we obtain

$$T_b(\tau) - T_{\text{exc}} = (T_{b,0} - T_{\text{exc}}) e^{-\tau} \Leftrightarrow T_b(\tau) = T_{\text{exc}} + (T_{b,0} - T_{\text{exc}}) e^{-\tau}. \quad (2.88)$$

The solutions are shown in Fig. 2.5.

Let us look at two different regimes, depending on the initial conditions $T_{b,0}$.

- (i) For a very low initial brightness temperature $T_{b,0} \ll T_{\text{exc}}$ (a weak initial radiation field), we obtain

$$T_b(\tau) \approx T_{\text{exc}} (1 - e^{-\tau}), \text{ for } T_{b,0} \ll T_{\text{exc}}. \quad (2.89)$$

The brightness temperature (or intensity) of the radiation field initially grows linearly (as $1 - e^{-\tau} \propto \tau$ for $\tau \ll 1$) and then saturates at $T_b \approx T_{\text{exc}}$ when $\tau \gtrsim 1$.

- (ii) If the initial brightness temperature is very high $T_{b,0} \gg T_{\text{exc}}$ (a very intense initial radiation field), we obtain

$$T_b(\tau) \approx T_{\text{exc}} + T_{b,0} e^{-\tau}, \text{ for } T_{b,0} \gg T_{\text{exc}}. \quad (2.90)$$

We see that the brightness temperature drops *exponentially* with time and levels off near T_{exc} after $\tau \sim$ a few to 10.

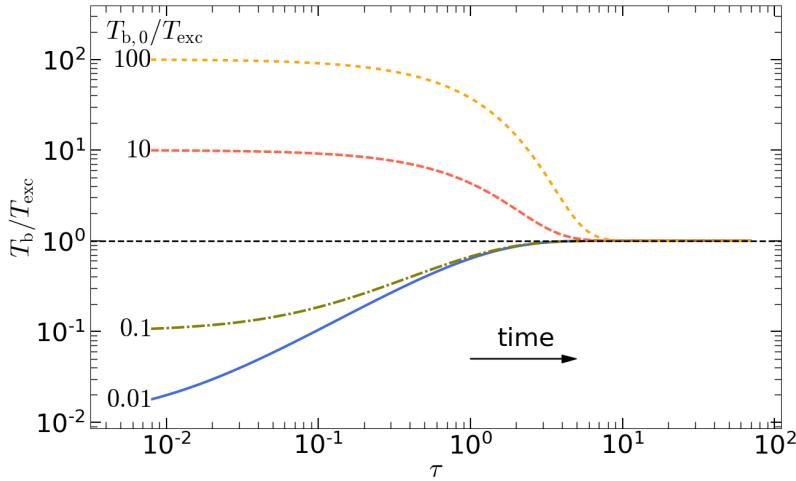


Figure 2.5: Time evolution of the brightness temperature under a fixed matter excitation temperature $T_{\text{exc}} = \text{const}$. Different lines correspond to initial brightness temperatures $T_{\text{b},0}/T_{\text{exc}} \in [0.01, 100]$. The time axis has been scaled to be dimensionless as in eq. (2.87).

2.2.4 Jargons on different temperatures

Below is a list of temperatures commonly used in astronomy.

- *Kinetic temperature* T is defined if the energies of a given matter species follow the Boltzmann distribution $\eta(\epsilon) \propto e^{-\epsilon/(k_B T)}$ in the limit of non-degenerate, ideal gas, where ϵ includes internal excitation energy, potential energy, and kinetic energy associated with microscopic random motions.
- *Radiation temperature* T_r is defined as the temperature in the Planck function when radiation field is thermalized (otherwise, strictly speaking, the radiation temperature is not well defined). Sometimes, T_r is loosely defined based on the total energy density of the radiation field $T_r = (U_{\text{rad}}/a)^{1/4}$, where a is the radiation density constant. When reading the literature, one should be cautious about the definition of T_r .
- *Brightness temperature* T_b is a measurement of specific intensity I_ν at a given frequency $\nu = c/\lambda$, as defined by the temperature of a blackbody that produces a given intensity,

$$B_\nu(T_b) \equiv I_\nu. \quad (2.91)$$

Using the photon occupation number $\eta_\gamma = [e^{h\nu/(k_B T_b)} - 1]^{-1} = I_\nu c^2 / (2h\nu^3)$, we obtain

$$T_b \equiv \frac{h\nu/k_B}{\ln [1 + 2h\nu^3/(c^2 I_\nu)]}. \quad (2.92)$$

In the Rayleigh-Jeans limit $h\nu \ll k_B T_b$, we write

$$T_b \approx \frac{c^2}{2k_B\nu^2} I_\nu = \frac{h\nu}{k_B} \eta_\gamma, \quad \text{in the R-J limit.} \quad (2.93)$$

For a spatially resolved source, one can directly measure the intensity and hence T_b . When the source is unresolved (with an upper limit on its size), one can put a lower limit on T_b . If the source emits *incoherently* (i.e., the waves from individual particles have random phases), T_b cannot exceed⁹ the excitation temperature of the emitting particles (see Fig. 2.5). Radio observers like using the brightness temperature because radio telescopes often have a thermal calibration scale — the total specific power P_ν (in units of Watt per Hz) recorded by the antenna system is referred to using an *antenna temperature* $T_a \equiv P_\nu/k_B$.

- *Excitation temperature* T_{exc} refers to relative populations of two quantized energy levels. If the occupation numbers of matter particles in the two energy levels $\epsilon_1 < \epsilon_2$ are η_1 and η_2 (number of particles per distinct quantum state), then the T_{exc} is defined as

$$\exp(-\Delta\epsilon/k_B T_{\text{exc}}) \equiv \frac{\eta_2}{\eta_1} = \frac{N_2/g_2}{N_1/g_1}, \quad (2.94)$$

where $\Delta\epsilon = \epsilon_2 - \epsilon_1$, g_1 , g_2 are the statistic weights of the two energy levels, and N_1 , N_2 are the total number of particles in the two energy levels.

- *Effective temperature* T_{eff} is a measurement of the total flux emerging from a given surface of the source (usually the photosphere), defined as

$$\sigma_{\text{SB}} T_{\text{eff}}^4 \equiv F. \quad (2.95)$$

Only when matter and radiation near the surface is in LTE with temperature T , we have $T_{\text{eff}} = T$. Realistically, due to non-LTE effects near the photosphere, the effective temperature T_{eff} generally differs from the kinetic temperature of matter T and the color temperature T_c (see below) of radiation.

- *Color temperature* T_c describes the shape of the spectrum or *spectral energy distribution* (the latter is usually based on photometric measurements in a few narrow/broad-band filters). If the spectral energy distribution is close to the shape of the Planck function or if there is a physical reason to believe so, then one can fit the data with the Planck function and then the best-fit temperature is usually called the color temperature, which provides a rough estimate of the average photon energy $\sim 2.7k_B T_c$.

⁹However, three classes of sources have abnormally high brightness temperatures: (1) sources moving towards the observer at relativistic speeds and the observed intensity is strongly enhanced as compared to that in the source's comoving frame due to relativistic beaming effect (see Ch 3); (2) population inversion of the emitting particles in astrophysical masers causes the brightness temperature to grow exponentially along a given ray; (3) if the matter particles emit *coherently* like the electrons inside a [half-wave transmitting antenna](#), the resulting intensity can have extremely high brightness temperatures.

2.3 Radiative transfer equation (without scattering)

In the rest of this chapter, we discuss how radiation propagates in matter. Since the radiation field is described by the specific intensity function $I_\nu(\mathbf{x}, \hat{\mathbf{n}}, t)$, the whole problem of radiative transfer comes down to solving the spatial (\mathbf{x}), angular ($\hat{\mathbf{n}}$), and temporal (t) dependences of I_ν . Ignoring refractive/diffractive effects¹⁰ and spacetime curvatures, the general radiative transfer equation is

$$\frac{dI_\nu}{dt} = \partial_t I_\nu + c\hat{\mathbf{n}} \cdot \nabla I_\nu = \text{source} - \text{sink}, \quad (2.96)$$

where the $d/dt = \partial_t + c\hat{\mathbf{n}} \cdot \nabla$ denotes the Lagrangian derivative along the propagation of a given photon packet at the speed of light c , and the source/sink terms describes the amount of intensity added/removed from the photon packet *per unit time*.

In the absence of interactions (i.e., no source/sink terms), we know that the intensity stays constant along a given ray, which propagates at the speed of light in a straight line along $\hat{\mathbf{n}}$. Let s be the linear coordinate on the axis $\hat{\mathbf{n}}$ along a given ray. For this ray, we will solve for $I_\nu(s, t)$ using the *method of ray tracing*. We follow the position of *a given photon packet* using $ds = c dt$, then the radiative transfer equation becomes one-dimensional

$$\frac{dI_\nu}{ds} = c^{-1} \partial_t I_\nu + \partial_s I_\nu = \text{source}(s) - \text{sink}(s), \quad (2.97)$$

where the source/sink terms depend on the coordinate s and describe the amount of intensity added/removed from the photon packet *per unit length*.

Note that the time-dependence is implicitly contained in eq. (2.97). Suppose we are interested in $I_\nu(s_{\text{obs}}, t_{\text{obs}})$ at the observer's position s_{obs} and time t_{obs} . We first trace the observed photon packet backwards to find the retarded time t_{ret} when the photons passed by position s ,

$$t_{\text{ret}}(s) = t_{\text{obs}} - (s_{\text{obs}} - s)/c. \quad (2.98)$$

We then calculate the source/sink terms at position s and at the corresponding retarded time $t_{\text{ret}}(s)$ in the form of intensity change per unit length $(dI/ds)_{\text{src}} - (dI/ds)_{\text{sink}}$. Finally, we integrate over the entire path with boundary condition $I_\nu(s = 0, t_{\text{ret}}(0))$ and obtain

$$I_\nu(s_{\text{obs}}, t_{\text{obs}}) = I_\nu(s = 0, t_{\text{ret}}(0)) + \int_0^{s_{\text{obs}}} ds \left[\left(\frac{dI}{ds} \right)_{\text{src}} - \left(\frac{dI}{ds} \right)_{\text{sink}} \right] (s, t_{\text{ret}}(s)). \quad (2.99)$$

Our task is to calculate the source/sink terms $(dI/ds)_{\text{src}} - (dI/ds)_{\text{sink}}$.

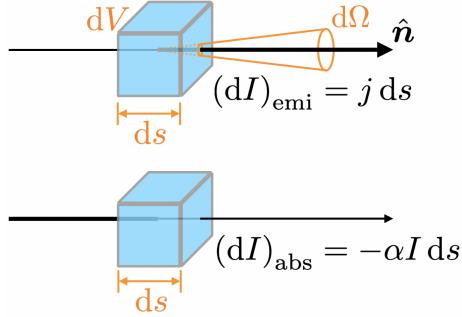


Figure 2.6: Upper panel: emission from matter adding photons to an existing beam. Bottom panel: absorptive matter removing photons from an existing beam.

2.3.1 Emission, absorption, mean free path, and optical depth

Here, we study how emission by the matter along a given direction \hat{n} adds to the existing intensity and how the intensity along the ray is attenuated due to absorption (see Fig. 2.6). We postpone the discussion of photon scattering to later sections (§2.5 and 2.6).

The *emissivity* (also called *volume emissivity*¹¹) is defined as the energy emitted by matter per unit volume per unit time per unit solid angle,

$$j = \frac{dE}{dV dt d\Omega}. \quad (2.100)$$

Note that the emissivity j refers to the *spontaneous emission* of matter — the stimulated emission will included as part of the *net absorption*. Scattered photons are not included in the emissivity either (see §2.6 later). Let us consider a volume $dV = dA ds$, where dA is an area perpendicular to the ray and ds is along the ray. Over a time interval $T \gg ds/c$ (but $T \ll$ variability timescale), the amount of energy emitted by matter in the dV volume into the solid angle $d\Omega$ is given by $E = jT dV d\Omega$. Since $T \gg ds/c$, almost all the energy E will have passed through the area on the right boundary of the volume over the time interval of T . Thus, the extra intensity added to the beam is given by

$$(dI)_{\text{emi}} = \frac{E}{T dA d\Omega} = \frac{jT dV d\Omega}{T dA d\Omega} = j ds. \quad (2.101)$$

The reader should check that the emissivity indeed has the same dimension as dI/ds . If the effect of photon absorption is negligible inside the source (this case is called *optically thin*, then the intensity escaping from the source is given by an integral along the line of

¹⁰For dilute astrophysical gases, the refractive/diffractive effects are negligible at high frequencies. The propagation of low-frequency (radio) waves will be discussed later in Chapter 5.

¹¹The *mass emissivity* refers to the energy emitted *per unit mass* per unit time per solid angle.



Figure 2.7: Flame of a lighter. In the heated air-fuel mixture, molecules are collisionally excited and then emit visible photons when they make downward transitions. Here, the lighter is under external illumination by a source of white light, and we see that the flame is *optically thin* to visible light, because the flare does not form a shadow (whereas the gas tube has a shadow). The yellow light produced from the flame shines on a piece of paper and produces diffuse reflection.

sight

$$I = \int j \, ds, \text{ for optically thin source.} \quad (2.102)$$

An example of an optically thin source is the flame of a lighter (as shown in Fig. 2.7).

Next, we discuss the attenuation by matter. Attenuation may be due to absorption or scattering of photons, but here we only consider absorption effect for simplicity.

Let us denote $P(s)$ as the *cumulative probability* of absorption *within* a distance of s from where the photon was emitted at $s = 0$. Then, the probability that a photon is absorbed between s and $s + ds$ is given by the product of two probabilities: (i) the photon *does not* get absorbed when traveling from 0 (the emission position) to s , and (ii) it *does* get absorbed from s to $s + ds$. The probability (i) is given by $1 - P(s)$. We will show later that the probability (ii) is given by $n\sigma_{\text{abs}}ds$, where n is the number density of absorbers and σ_{abs} is the absorption cross-section of each particle. Putting these together, we write

$$dP = \underbrace{[1 - P(s)]}_{(i)} \times \underbrace{n\sigma_{\text{abs}}ds}_{(ii)}. \quad (2.103)$$

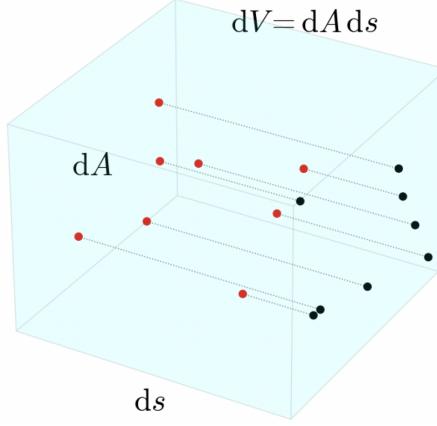


Figure 2.8: The absorption probability within a distance ds is given by the ratio of the area of the shadows cast by the absorbers over the total area dA .

Integrating eq. (2.103), we obtain an *exponential distribution*¹²

$$P(s) = 1 - e^{-s/\ell_{\text{mfp}}}, \quad \text{or} \quad \frac{dP}{ds} = \frac{1}{\ell_{\text{mfp}}} e^{-s/\ell_{\text{mfp}}}, \quad (2.105)$$

where ℓ_{mfp} is the *mean free path* (mfp) given by

$$\ell_{\text{mfp}} = (n\sigma_{\text{abs}})^{-1}. \quad (2.106)$$

One can verify that the mean path length is indeed given by $\langle s \rangle = \int_0^\infty s(dP/ds)ds = \ell_{\text{mfp}}$. The *mean squared free path* is $\langle s^2 \rangle = \int_0^\infty s^2(dP/ds)ds = 2\ell_{\text{mfp}}^2$, which will be useful for the understanding of random walk of photons under repetitive scatterings (see §2.5).

Our remaining task is to demonstrate the probability (ii) in the discussion above is given by $n\sigma_{\text{abs}}ds$ — this is the probability that a freshly released photon will be absorbed within a short distance of ds . As shown in Fig. 2.8, let us consider an area dA perpendicular to the ray, and the total number of absorbers in the volume $dV = dA ds$ is given by $dN = ndV$. These absorbers (shown in red circles), each with a cross-section σ_{abs} , will cast shadows

¹²Another way of obtaining the exponential distribution is to directly calculate the probability $1 - P(s)$ that the photon does not get absorbed on its way from 0 to s . We divide the path length s into equal segments of length Δs , and the number of segments is $N = s/\Delta s$. We require that the photon does not get absorbed on each of these segments, which means that

$$\begin{aligned} 1 - P(s) &= \lim_{\Delta s \rightarrow 0} (1 - \Delta s/\ell_{\text{mfp}})^N = 1 - N\Delta s/\ell_{\text{mfp}} + \frac{N(N-1)}{2} (\Delta s/\ell_{\text{mfp}})^2 + \dots \\ &= 1 - s/\ell_{\text{mfp}} + (s/\ell_{\text{mfp}})^2/2 + \dots = e^{-s/\ell_{\text{mfp}}}. \end{aligned} \quad (2.104)$$

Here we have used Taylor expansion and then calculated each term in the limit of $\Delta s \rightarrow 0$.

(shown in black circles) on the right boundary of the volume dV . Under the condition that the shadows do not overlap each other (which is the case as $ds \rightarrow 0$), the total area of their shadows is $dN\sigma_{\text{abs}}$. Thus, the absorption probability is given by the fraction of the area dA covered by the shadows, which is

$$\frac{dN\sigma_{\text{abs}}}{dA} = \frac{n\sigma_{\text{abs}} dV}{dA} = n\sigma_{\text{abs}} ds = \frac{ds}{\ell_{\text{mfp}}}. \quad (2.107)$$

This causes the intensity entering the left boundary of the volume to be reduced by a factor of ds/ℓ_{mfp} by the time the ray exits from the right boundary.

It is also convenient to define the *absorption coefficient* $\alpha \equiv \ell_{\text{mfp}}^{-1}$ (in units of cm^{-1}) to make the radiative transfer equation linear-looking. The loss of intensity in a given beam as it propagates a distance ds is given by

$$(dI)_{\text{abs}} = -\frac{Ids}{\ell_{\text{mfp}}} = -\alpha Ids, \quad (2.108)$$

Note that the RHS of the above is proportional to the current intensity I , because $\alpha ds = ds/\ell_{\text{mfp}}$ describes the *fractional loss*.

Putting the absorption and emission terms together, we obtain the total change in intensity in the ds interval

$$dI = (dI_{\nu})_{\text{emi}} + (dI)_{\text{abs}} = jds - \alpha Ids. \quad (2.109)$$

Often, we are interested in the change in specific intensity $I_{\nu} \equiv dI/d\nu$. For this purpose, we define the *specific emissivity*

$$j_{\nu} \equiv \frac{dj}{d\nu} = \frac{dE}{dVdt\Omega d\nu}. \quad (2.110)$$

We also define a *frequency-dependent* absorption coefficient α_{ν} , which is still in units of cm^{-1} . Then the change in specific intensity in the ds interval is given by

$$dI_{\nu} = (dI_{\nu})_{\text{abs}} + (dI_{\nu})_{\text{emi}} = -\alpha_{\nu} I_{\nu} ds + j_{\nu} ds. \quad (2.111)$$

Finally, we arrive at the *radiative transfer equation*

$$\frac{dI_{\nu}}{ds} = j_{\nu} - \alpha_{\nu} I_{\nu}. \quad (2.112)$$

It is often convenient to use *opacity* κ_{ν} , which is defined as the absorption cross-section per unit mass at frequency ν , and it is related to the intrinsic property of the constituents of the matter. According to this definition, the absorption coefficient can be written as

$$\alpha_{\nu} = \rho\kappa_{\nu} = \frac{1}{\ell_{\text{mfp}}(\nu)} = n\sigma_{\text{abs}}(\nu), \quad (2.113)$$

where $\sigma_{\text{abs}}(\nu)$ is the frequency-dependent absorption cross-section of matter particles.

The *optical depth* along a given ray is defined as

$$\tau_\nu = \int \alpha_\nu ds = \int \rho \kappa_\nu ds. \quad (2.114)$$

The matter along a given ray is called *optically thick* at frequency ν if $\tau_\nu \gg 1$ and *optically thin* if $\tau_\nu \ll 1$. We often use the optical depth τ_ν as our *dimensionless coordinate* along a given ray, and the radiative transfer equation can be rewritten as

$$\frac{dI_\nu}{d\tau_\nu} = \frac{j_\nu}{\alpha_\nu} - I_\nu. \quad (2.115)$$

Since j_ν (spontaneous emission rate) and α_ν (coefficient for net absorption) are only related to the distribution function of matter, it is convenient to denote their ratio as

$$S_\nu \equiv j_\nu / \alpha_\nu, \quad (2.116)$$

which is called the *source function* given by the source/matter properties.

Later on, when we consider the effects of scattering in §2.6, the scattered photons can be considered as re-emission by matter, and the meanings of j_ν and S_ν will need to be generalized in that context.

2.3.2 Formal solution

If for some reason we know the source function $S_\nu(\tau_\nu)$ at all optical depths τ_ν inside a given medium, then the solution to eq. (2.115) can be written analytically. We multiply both sides of the equation by e^{τ_ν} and then carry out integral-by-parts. This gives the *formal solution* to the radiative transfer equation

$$I_\nu(\tau_\nu) = I_\nu(0)e^{-\tau_\nu} + e^{-\tau_\nu} \int_0^{\tau_\nu} S_\nu(\tau'_\nu) e^{\tau'_\nu} d\tau'_\nu, \quad (2.117)$$

where $I_\nu(0)$ is the incident/background intensity at the beginning of the ray ($\tau_\nu = 0$), and $I_\nu(\tau_\nu)$ is the intensity at coordinate τ_ν .

The above formal solution can be intuitively understood as follows. The first term $I_\nu(0)e^{-\tau_\nu}$ describes the exponential attenuation of the incident intensity by along the ray. The effect of the exponential attenuation can be clearly seen in Fig. 2.9. The second term is the superposition of many small segments of the ray each contributing an intensity $S_\nu(\tau'_\nu)d\tau'_\nu$ but is then attenuated by a factor of $e^{-(\tau_\nu - \tau'_\nu)}$ where $\tau_\nu - \tau'_\nu$ is the *remaining optical depth* from the position of emission at τ'_ν to the end of the ray at τ_ν .

Unfortunately, in most cases, we do not have prior knowledge on the source function. In the most general case where the medium can absorb, emit, and scatter the radiation, the



Figure 2.9: Fog consists of small water droplets which scatter light in all directions. This image shows that the intensities of the distant trees and road are exponentially suppressed when the scattering optical depth $\tau \gtrsim 1$.

scattered emission depends on the local intensity (which is yet to be solved), and the local intensity can affect the distribution function (e.g., temperature) of the medium and hence the emissivity and opacity depend on the local intensity as well. Solving the whole problem of radiative transfer usually requires an iterative method: one first solves the distribution function of the medium for an unattenuated radiation field, then calculates the emissivity and opacity using the distribution function, afterwards an updated version of the radiation field is obtained by integrating the radiative transfer equation through the medium, and this process is repeated until convergence. The reader is referred to the [Cloudy code](#) for an example of such an iterative method for a one-dimensional medium.

An important special case is for a constant source function (i.e., for a uniform medium without scattering), and we obtain the following simple solution

$$I_\nu(\tau_\nu) = I_\nu(0) e^{-\tau_\nu} + S_\nu(1 - e^{-\tau_\nu}), \text{ for } S_\nu = \text{const}, \quad (2.118)$$

where $I_\nu(0)$ is the incident intensity at $\tau_\nu = 0$ and τ_ν is the total optical depth of the medium. Later on, we will show a few examples of the observed spectrum from a medium with a constant source function.

2.3.3 Kirchhoff's law

Since j_ν describes the spontaneous emission and α_ν is related to the net absorption rate (true absorption – stimulated emission), their ratio is given by the Einstein relations. From

the radiative transfer equation (2.112), the ratio between the net absorption rate and the spontaneous emission rate is given by

$$\frac{\alpha_\nu I_\nu}{j_\nu} = \frac{\text{net absorption rate}}{\text{spontaneous emission rate}} = \left(\frac{\eta_1}{\eta_2} - 1 \right) \eta_\gamma = \left(e^{h\nu/(k_B T_{\text{exc}})} - 1 \right) \eta_\gamma, \quad (2.119)$$

where we have used one of the Einstein relations (eq. 2.73) and the definition of the excitation temperature T_{exc} (eq. 2.80). Since the specific intensity is related to the photon occupation number by $I_\nu = (2h\nu^3/c^2)\eta_\gamma$, we see that the photon occupation number η_γ cancels out on both sides of the above equation, and we obtain the *Kirchhoff's law*¹³

$$\frac{j_\nu}{\alpha_\nu} = \frac{2h\nu^3/c^2}{e^{h\nu/(k_B T_{\text{exc}})} - 1} = B_\nu(T_{\text{exc}}), \quad (2.120)$$

where $B_\nu(T_{\text{exc}})$ is the Planck function (eq. 2.51) at the matter excitation temperature. As a direct consequence of the Einstein relations, the Kirchhoff's law also describes a matter-radiation system's tendency of reaching LTE. This can be demonstrated as follows.

Consider that matter is in thermal distribution at temperature T , and we have $T_{\text{exc}} = T$ for all frequencies. For a uniform medium, the source function $S_\nu = B_\nu(T)$ is a constant (no dependence on the optical depth). Then, the solution to the radiative transfer equation is given by (cf. eq. 2.118)

$$I_\nu = I_\nu(0) e^{-\tau_\nu} + B_\nu(T)(1 - e^{-\tau_\nu}), \text{ for } T = \text{const.} \quad (2.121)$$

At small optical depths $\tau_\nu \ll 1$, we have $I_\nu = I_\nu(0) + B_\nu\tau_\nu$, which means that the initial intensity $I_\nu(0)$ is not yet substantially attenuated and that the thermal intensity $B_\nu\tau_\nu$ is growing linearly with optical depth. However, if $\tau_\nu \gg 1$, we the initial intensity will be completely attenuated and the thermal intensity will grow to its maximum, and we obtain a thermal radiation spectrum $I_\nu \approx B_\nu$. This shows that the system reaches LTE at sufficiently large optical depth $\tau_\nu \gg 1$.

2.3.4 Absorption cross-section for a spectral line

Let us use the Kirchhoff's law to calculate the absorption cross-section for a narrow spectral line from a given atomic or molecular transition. When particles at a higher energy level ϵ_2 make transitions to a lower energy level ϵ_1 , the emissivity is related to the Einstein-A coefficient of this transition by

$$j_\nu = \frac{h\nu_0}{4\pi} n_2 A \phi_\nu, \quad (2.122)$$

¹³In some other textbooks (e.g., Rybicki & Lightman), the Kirchhoff's law is only restricted to the case where matter has a thermal distribution function, but eq. (2.120) is defined more generally and applies to non-thermal matter distribution functions as well as long as we properly account for the frequency dependence of the excitation temperature $T_{\text{exc}}(\nu)$.

where the factor of 4π is due to isotropic emission, n_2 is the number density of emitting particles at energy level ϵ_2 , ϕ_ν is a narrow line-profile function that is non-zero only near frequency ν_0 (line center), and ϕ_ν is normalized by $\int \phi_\nu d\nu = 1$.

The absorption coefficient due to this transition is given by the Kirchhoff's law

$$\alpha_\nu = \frac{j_\nu}{B_\nu(T_{\text{exc}})} = \frac{h\nu_0 n_2 A}{4\pi B_{\nu_0}(T_{\text{exc}})} \phi_\nu, \quad (2.123)$$

where we have used $B_\nu(T_{\text{exc}}) \approx B_{\nu_0}(T_{\text{exc}})$, so the only frequency dependence is from the line profile function ϕ_ν .

The *net* absorption cross-section is given by the absorption coefficient divided by the number density of electrons n_1 in the lower energy level ϵ_1 ,

$$\sigma_{\text{abs}}(\nu) = \frac{\alpha_\nu}{n_1} = \frac{\lambda_0^2 g_2}{8\pi g_1} A \left(1 - e^{-h\nu_0/(k_B T_{\text{exc}})} \right) \phi_\nu, \quad (2.124)$$

where we have used $n_2/n_1 = (g_2/g_1)e^{h\nu_0/(k_B T_{\text{exc}})}$ and the wavelength at the line center $\lambda_0 = c/\nu_0$. The absorption optical depth is then given by

$$\tau_\nu = N_{\text{col}} \sigma_{\text{abs}}(\nu), \quad (2.125)$$

where $N_{\text{col}} = \int n_1 ds$ is the column density of particles in energy level ϵ_1 along the line of sight. We see that the net absorption cross-section depends on three things: (i) the Einstein-A coefficient — a stronger transition with a larger A corresponds to a larger absorption cross-section; (ii) the line profile function ϕ_ν — a broader line profile reduces the absorption cross-section near the line center frequency by a factor of the line width $\Delta\nu \simeq \phi_{\nu_0}^{-1}$; (iii) the excitation temperature T_{exc} . Regarding the two terms in the factor of $1 - e^{-h\nu_0/(k_B T_{\text{exc}})}$: the “1” term accounts for true absorption and the “ $e^{-h\nu_0/(k_B T_{\text{exc}})}$ ” term is due to stimulated emission (which is treated here as “negative absorption”). At very high excitation temperatures ($k_B T_{\text{exc}} \gg h\nu_0$), stimulated emission is very strong such that it almost cancels with the true absorption, leaving behind a small residual $1 - e^{-h\nu_0/(k_B T_{\text{exc}})} \approx h\nu_0/(k_B T_{\text{exc}})$. At very low excitation temperature $k_B T_{\text{exc}} \gg h\nu_0$, the effect of stimulated emission can be ignored. Thus, the cross-section for *true absorption* (only the “1” term) is given by

$$\sigma_{\text{abs}}^{\text{true}} = \frac{\lambda_0^2 g_2}{8\pi g_1} A \phi_\nu. \quad (2.126)$$

If we do an experiment like what is shown in Fig. 2.8 and measure the fractional absorption of an incident light beam with a low photon occupation number $\eta_\gamma \ll 1$, then we would measure the true absorption cross-section $\sigma_{\text{abs}}^{\text{true}}$ (eq. 2.126), which generally differs from the net absorption cross-section in astrophysical environments σ_{abs} (eq. 2.124).

2.3.5 *Generalized Kirchhoff's law for anisotropic distribution function

Let us use the Einstein relations (eqs. 2.71, 2.72, 2.73) to further derive the relation between emissivity and the *net absorption coefficient* (true absorption minus stimulated emission) for the case of an anisotropic and non-thermal distribution function.

Consider a small volume filled with matter which has emissivity $j_\nu(\mathbf{k})$ near the direction and photon frequency specified by the wavevector \mathbf{k} . Let $\xi_\nu(\mathbf{p}, \mathbf{k})$ (unit: $\text{erg s}^{-1} \text{Hz}^{-1} \text{sr}^{-1}$) be the *single-particle specific emissivity* of an electron with momentum \mathbf{p} (and for bound electrons, one should understand \mathbf{p} as specifying a quantum state near a given energy level ϵ), and then we have

$$j_\nu(\mathbf{k}) = \int d^3\mathbf{p} f(\mathbf{p}) \xi_\nu(\mathbf{p}, \mathbf{k}), \quad (2.127)$$

where $f(\mathbf{p})$ is the Boltzmann distribution function. Let us also denote the *net absorption cross-section* of an electron with momentum \mathbf{p} near photon wavevector \mathbf{k} as $\sigma_{\text{abs}}(\mathbf{p}, \mathbf{k})$ (unit: cm^2), then the *net absorption coefficient* is given by

$$\alpha_\nu(\mathbf{k}) = \int d^3\mathbf{p} f(\mathbf{p}) \sigma_{\text{abs}}(\mathbf{p}, \mathbf{k}). \quad (2.128)$$

Note that the α_ν is always defined based on net absorption (= true absorption – stimulated emission).

We write the radiative transfer equation for a beam along direction $\hat{\mathbf{k}}$

$$\frac{dI_\nu}{ds} = -\alpha_\nu I_\nu + j_\nu = \int d^3\mathbf{p} f(\mathbf{p}) [\sigma_{\text{abs}}(\mathbf{p}, \mathbf{k}) I_\nu + \xi_\nu(\mathbf{p}, \mathbf{k})]. \quad (2.129)$$

It is easy to identify the first term $\sigma_{\text{abs}} I_\nu$ as the rate at which energy is taken away from the beam (along $\hat{\mathbf{k}}$) per particle and the second term ξ_ν as the rate at which energy is added to the beam per unit time per particle. From the Einstein relations, the ratio between these two terms is given by

$$\frac{\alpha_\nu I_\nu}{j_\nu} = \frac{\sigma_{\text{abs}} I_\nu}{\xi_\nu} = \left[\frac{\eta(\mathbf{p}')}{\eta(\mathbf{p})} - 1 \right] \eta_\gamma, \quad (2.130)$$

where $\eta_\gamma(\mathbf{k}) = I_\nu(\mathbf{k})(c^2/2h\nu^3)$ is the photon occupation number near wavevector \mathbf{k} , and $\eta(\mathbf{p})$ and $\eta(\mathbf{p}')$ are the electron occupation numbers in the two quantum states associated with the radiative transition $\mathbf{p} \rightarrow \mathbf{p}'$ that generates a photon with wavevector \mathbf{k} (note that \mathbf{p}' is determined by \mathbf{p} and \mathbf{k} through momentum and energy conservation laws for the transition). This allows us to obtain the net absorption cross-section

$$\sigma_{\text{abs}}(\mathbf{p}, \mathbf{k}) = \frac{c^2}{2h\nu^3} \left[\frac{\eta(\mathbf{p}')}{\eta(\mathbf{p})} - 1 \right] \xi_\nu(\mathbf{p}, \mathbf{k}), \quad (2.131)$$

The key step in this section is to define the *excitation temperature* for the $\mathbf{p} \rightarrow \mathbf{p}'$ transition $T_{\text{exc}}(\mathbf{p})$ as follows

$$\frac{\eta(\mathbf{p}')}{\eta(\mathbf{p})} \equiv \exp\left(\frac{h\nu}{k_{\text{B}}T_{\text{exc}}(\mathbf{p})}\right) \Rightarrow \frac{c^2}{2h\nu^3} \left[\frac{\eta(\mathbf{p}')}{\eta(\mathbf{p})} - 1 \right] = \frac{1}{B_{\nu}(T_{\text{exc}}(\mathbf{p}))}, \quad (2.132)$$

where $B_{\nu}(T_{\text{exc}})$ is our familiar Planck function evaluated at the excitation temperature. Using our newly defined excitation temperature, we see that the single-particle specific emissivity ξ_{ν} and its net absorption cross-section σ_{abs} are related by

$$\frac{\xi_{\nu}(\mathbf{p}, \mathbf{k})}{\sigma_{\text{abs}}(\mathbf{p}, \mathbf{k})} = B_{\nu}(T_{\text{exc}}(\mathbf{p})). \quad (2.133)$$

This is the generalized Kirchhoff's law for *single particles*, and it looks similar to the usual Kirchhoff's law (eq. 2.120) for the entire population with an isotropic distribution function.

The next step is to calculate the absorption coefficient for the entire population (instead of a single particle). We plug $\sigma_{\text{abs}} = \xi_{\nu}/B_{\nu}(T_{\text{exc}})$ into eq. (2.128) and then the absorption coefficient is given by

$$\alpha_{\nu}(\mathbf{k}) = \frac{c^2}{2h\nu^3} \int d^3\mathbf{p} \left[\frac{g(\mathbf{p})}{g(\mathbf{p}')} f(\mathbf{p}') - f(\mathbf{p}) \right] \xi_{\nu}(\mathbf{p}, \mathbf{k}) = \int d^3\mathbf{p} f(\mathbf{p}) \frac{\xi_{\nu}(\mathbf{p}, \mathbf{k})}{B_{\nu}(T_{\text{exc}}(\mathbf{p}))}, \quad (2.134)$$

This can then be combined with the emissivity in eq. (2.127) to obtain the following neat-looking expression

$$\frac{j_{\nu}(\mathbf{k})}{\alpha_{\nu}(\mathbf{k})} = \langle B_{\nu}(T_{\text{exc}}) \rangle_{\mathbf{p}}, \quad (2.135)$$

where the RHS is defined as

$$\langle B_{\nu}(T_{\text{exc}}) \rangle_{\mathbf{p}} \equiv \frac{\int d^3\mathbf{p} f(\mathbf{p}) \xi_{\nu}(\mathbf{p}, \mathbf{k})}{\int d^3\mathbf{p} f(\mathbf{p}) \xi_{\nu} / B_{\nu}(T_{\text{exc}}(\mathbf{p}))}. \quad (2.136)$$

Eq. (2.135) is called the *generalized Kirchhoff's law* for the entire population in nLTE, and the problem essentially comes down to computing the momentum-averaged Planck function at the excitation temperature $\langle B_{\nu}(T_{\text{exc}}) \rangle_{\mathbf{p}}$ (similar to the procedure of computing the Rosseland-mean opacity). We emphasize that the generalized Kirchhoff's law in eqs. (2.133) and (2.135) are derived from the Einstein relations, which relies on the *tendency* of a given system to achieve LTE (so as to maximize the entropy). Therefore, they are applicable to radiative transitions in arbitrary non-LTE systems. In the special case when matter is in LTE, we have $T_{\text{exc}}(\mathbf{p}) = T = \text{const}$, so the Planck function does not depend on the particle momentum \mathbf{p} and the LTE version of the Kirchhoff's law is recovered: $j_{\nu}/\alpha_{\nu} = B_{\nu}(T)$.

In later chapters, we will use either quantum mechanics or classical electrodynamics to calculate the single-particle emissivity $\xi_\nu(\mathbf{p}, \mathbf{k})$ or volume emissivity $j_\nu(\mathbf{k})$, and then, one can use either eq. (2.134) or eq. (2.135) to calculate the absorption coefficient α_ν near a given wavevector \mathbf{k} . It should be noted that the absorption coefficient depends on three things: (i) the ratio between electron occupation numbers at momentum states \mathbf{p} and \mathbf{p}' (or energy states ϵ and ϵ') or equivalently the excitation temperature T_{exc} , (ii) the emissivity ξ_ν or j_ν , and (iii) the momentum distribution function $f(\mathbf{p})$. The readers should convince themselves that these dependences are expected from the Einstein relations.

In the following, we discuss the case of free electrons, for which there are infinite number of combinations of initial states \mathbf{p} that can produce photons near a given momentum $\hbar\mathbf{k}$ (as can be seen in eq. 2.127).

For free electrons, we adopt $g(\mathbf{p}) = g(\mathbf{p}') = 2$ due to spin and then eq. (2.134) becomes

$$\alpha_\nu(\mathbf{k}) = \frac{c^2}{2h\nu^3} \int d^3\mathbf{p} [f(\mathbf{p}') - f(\mathbf{p})] \xi_\nu(\mathbf{p}, \mathbf{k}). \quad (2.137)$$

Since absorption is only significant at low frequencies, the photon energy is usually small compared to the electron energy $\hbar k \ll p$. This leads to the following simplification

$$f(\mathbf{p}') - f(\mathbf{p}) \approx (\partial f / \partial \mathbf{p}) \cdot (\mathbf{p}' - \mathbf{p}), \quad (2.138)$$

where $\partial f / \partial \mathbf{p} = \nabla_{\mathbf{p}} f$ is the gradient of the distribution function near momentum \mathbf{p} . Then, the absorption coefficient is given by

$$\alpha_\nu^{\text{free}}(\mathbf{k}) = \frac{c^2}{2h\nu^3} \int d^3\mathbf{p} \frac{\partial f(\mathbf{p})}{\partial \mathbf{p}} (\mathbf{p}' - \mathbf{p}) \xi_\nu(\mathbf{p}, \mathbf{k}). \quad (2.139)$$

Since the final state momentum \mathbf{p}' is determined by the initial state momentum \mathbf{p} and photon momentum $\hbar\mathbf{k}$ (from energy and momentum conservation laws), the above integral can be carried out as long as we know the distribution function $f(\mathbf{p})$ and the single-particle emissivity $\xi_\nu(\mathbf{p}, \mathbf{k})$. In this book, we will devote a few chapters to calculating $\xi_\nu(\mathbf{p}, \mathbf{k})$ for different emission mechanisms.

2.4 Emission and absorption lines

In this section, we consider the emission/absorption caused by the radiative transitions between two energy levels 1 and 2 of a given species of atoms or molecules, which give rise to a *spectral line* near the central frequency $h\nu_0 = \hbar\omega_0 = \epsilon_2 - \epsilon_1$, where $\omega_0 = 2\pi\nu_0$ is the angular frequency.

2.4.1 Emission or absorption?

Consider *line photons* near frequency ν_0 propagating through a spatially uniform medium where the relative population between the two energy levels ϵ_1 and ϵ_2 is described by the excitation temperature T_{exc} (eq. 2.80). For a constant source function, the solution to the radiative transfer equation gives the change in intensity (cf. eq. 2.118)

$$\Delta I_\nu = I_\nu(\tau_\nu) - I_\nu(0) = (S_\nu - I_\nu(0))(1 - e^{-\tau_\nu}), \quad (2.140)$$

where $I_\nu(0)$ is the background continuum intensity, τ_ν is the optical depth of the cloud, and $S_\nu = j_\nu/\alpha_\nu$ is the source function that is equal to $B_\nu(T_{\text{exc}})$ by definition of the excitation temperature. If we further define a brightness temperature $T_{\text{b},0}$ based on the background continuum intensity $B_\nu(T_{\text{b},0}) = I_\nu(0)$, then the change in intensity is

$$\Delta I_\nu = (B_\nu(T_{\text{exc}}) - B_\nu(T_{\text{b},0}))(1 - e^{-\tau_\nu}). \quad (2.141)$$

In the Rayleigh-Jeans limit $h\nu \ll k_B \min(T_{\text{exc}}, T_{\text{b},0})$ (e.g., for radio frequencies), we obtain the change in brightness temperature

$$\Delta T_b = T_b - T_{\text{b},0} = (T_{\text{exc}} - T_{\text{b},0})(1 - e^{-\tau_\nu}). \quad (2.142)$$

This shows that we will see an *emission line* when $T_{\text{exc}} > T_{\text{b},0}$ and *absorption line* otherwise. The full solution of $T_b(\tau_\nu)$ is shown in Fig. 2.5, where we see that the brightness temperature of the radiation field T_b always tries to approach the matter excitation temperature T_{exc} and that an equilibrium of $T_b = T_{\text{exc}}$ is achieved at $\tau_\nu \gg 1$. The line profile is given by the $(1 - e^{-\tau_\nu})$ factor, as τ_ν is maximized at the line center frequency ν_0 and $\tau_\nu \rightarrow 0$ at frequencies far away ν_0 . This will be discussed in the next subsection.

2.4.2 Mechanisms for line broadening

We discuss a few mechanisms that determine the shape of a spectral line.

(1) Natural broadening

Due to radiative or collisional transitions to other states, any quantum state has a finite lifetime. According to the Uncertainty Principle, the energies of the two states, ϵ_1 and ϵ_2 , we are considering are uncertain. Suppose the lifetimes of these two states are

$$\Delta t_1 \equiv \gamma_1^{-1}, \quad \Delta t_2 \equiv \gamma_2^{-1}, \quad (2.143)$$

where $\gamma_{1/2}$ are called the *damping rates* of the two states based on the classical picture of a harmonic oscillator. Then, the energy uncertainties are $\Delta\epsilon_n = \hbar/\Delta t_n = \hbar\gamma_n$ (for $n = 1, 2$), and this leads to a total uncertainty for the line frequency to be $\Delta\omega = (\Delta\epsilon_1 + \Delta\epsilon_2)/\hbar = \gamma$, where we have used the *total damping rate*

$$\gamma = \gamma_1 + \gamma_2. \quad (2.144)$$

We will later show that the resulting line profile has a *full width at half maximum* of

$$(\Delta\nu)_{\text{fwhm}}^{\text{nat}} = \frac{\Delta\omega}{2\pi} = \frac{\gamma}{2\pi}, \quad (2.145)$$

where ‘‘nat’’ stands for *natural broadening*.

Let us consider the spectral shape of a naturally broadened line, whose central angular frequency is ω_0 and total damping rate is γ . The real part of the scalar electric field of the EM waves produced by a given emitting particle has an exponentially decaying amplitude

$$\text{Re } E(t > 0) = E_0 e^{-\gamma t/2} e^{-i\omega_0 t} + \text{complex conjugate}, \quad (2.146)$$

where E_0 is a complex amplitude that is unimportant for our purpose here. It is important to notice that the amplitude decays as $e^{-\gamma t/2}$, so the energy flux decays¹⁴ as $e^{-\gamma t}$ (analogous to radioactive decay). The Fourier transformation of $\text{Re } E(t)$ is given by

$$\tilde{E}_{\text{R}}(\omega) = \tilde{G}(\omega) + \tilde{G}^*(-\omega), \quad \tilde{G}(\omega) = \frac{E_0}{2\pi} \int_0^\infty e^{i(\omega-\omega_0)t} e^{-\gamma t/2} dt. \quad (2.147)$$

Using the Laplace transforms of sine and cosine functions

$$\int_0^\infty \cos(at) e^{-bt} dt = \frac{b}{a^2 + b^2}, \quad \int_0^\infty \sin(at) e^{-bt} dt = \frac{a}{a^2 + b^2}, \quad (2.148)$$

we obtain

$$\tilde{G}(\omega) = \frac{E_0}{2\pi} \frac{\gamma/2 + i(\omega - \omega_0)}{(\omega - \omega_0)^2 + (\gamma/2)^2}, \quad \tilde{G}^*(-\omega) = \frac{E_0^*}{2\pi} \frac{\gamma/2 - i(\omega + \omega_0)}{(\omega + \omega_0)^2 + (\gamma/2)^2}. \quad (2.149)$$

In the limit of slow damping $\gamma \ll \omega_0$ which gives rise to a narrow line near the central frequency $|\omega - \omega_0| \ll \omega_0$, we see that $|\tilde{G}^*(-\omega)| \ll |\tilde{G}(\omega)|$, and hence

$$\tilde{E}_{\text{R}}(\omega) \approx \tilde{G}(\omega). \quad (2.150)$$

The spectral power, which is the energy emitted per unit angular frequency interval (cf. eq. 1.30), is proportional to $|\tilde{E}_{\text{R}}(\omega)|^2$, so we obtain

$$\frac{dj}{d\omega} \propto |\tilde{E}_{\text{R}}(\omega)|^2 \approx |\tilde{G}(\omega)|^2 = \tilde{G}(\omega) \tilde{G}^*(\omega) = \frac{|E_0|^2}{4\pi^2} \frac{1}{(\omega - \omega_0)^2 + (\gamma/2)^2}. \quad (2.151)$$

The shape of the line profile function is given by $\phi_\nu \propto j_\nu = 2\pi dj/d\omega$ and it is normalized by $\int_0^\infty \phi_\nu d\nu = 1$.

¹⁴For this reason, it might be useful to define an amplitude damping rate as $\gamma' = \gamma/2$, whereas we stick to the flux damping rate of γ .

Therefore, we obtain the following normalized line profile under natural broadening

$$\phi_\nu^{\text{nat}} = \frac{\gamma/(4\pi^2)}{(\nu - \nu_0)^2 + (\gamma/4\pi)^2}, \quad (2.152)$$

and it is easy to show at the FWHM is given by $(\Delta\nu)_{\text{fwhm}}^{\text{nat}} = \gamma/(2\pi)$. The above functional form (eq. 2.152) is called a *Lorentzian profile*.

For instance, the Einstein-A coefficient for the hydrogen Ly α transition ($\nu_0 = 2.4661 \times 10^{15}$ Hz or central wavelength $\lambda_0 = 1215.7 \text{ \AA}$) is $A = 6.26 \times 10^8 \text{ s}^{-1}$, and this corresponds to $\gamma_2 = A$ due to the lifetime of the upper energy state and $\gamma_1 \approx 0$ due to the nearly infinite lifetime of the lower energy state (provided that the timescale for $1 \rightarrow 2$ excitations is much longer than A^{-1}). The total damping rate for this case is $\gamma = \gamma_1 + \gamma_2 \approx A$ and hence the natural width is $(\Delta\nu)_{\text{fwhm}}^{\text{nat}} = A/(2\pi) = 1.0 \times 10^8 \text{ Hz}$ in frequency, or $(\Delta\lambda)_{\text{fwhm}}^{\text{nat}} = \lambda^2 \Delta\nu/c = 4.9 \times 10^{-5} \text{ \AA}$ in wavelength, or $(\Delta v)_{\text{fwhm}}^{\text{nat}} = (\Delta\lambda/\lambda)c = 0.012 \text{ km s}^{-1}$ in Doppler velocity. This is a very small line width that is difficult to observe directly. However, it produces a broad wing at $|\nu - \nu_0| \gg (\Delta\nu)_{\text{fwhm}}^{\text{nat}}$ that is observable when the hydrogen column density is sufficiently large.

(2) Collisional broadening

While an atom is emitting line photons, it may collide with other particles at random times. Suppose the *mean collision time interval* is given by $t_c = \gamma_c^{-1}$, where γ_c is the *collision rate*, then the probability distribution of the time interval T between two adjacent collisions is given by the exponential distribution (cf. eq. 2.105)

$$\frac{dP}{dT} = \gamma_c e^{-\gamma_c T}. \quad (2.153)$$

We focus on the collisions which occur on timescales much shorter than the wave oscillation time ω_0^{-1} , because otherwise the external EM fields provided by the incoming particle only change slowly and the atom can smoothly adjust to the changing background fields (such cases are better analyzed in the limit of static external fields, which lead to e.g., Stark/Zeeman effect). The line emission in between two consecutive collisions is sinusoidal and continuous, as described by

$$\text{Re } E(t) = E_0 e^{-i\omega_0 t} + \text{complex conjugate,} \quad (\text{between adjacent collisions}) \quad (2.154)$$

where E_0 is a complex amplitude. However, each collision causes an abrupt change in the phase of the emission, which can be thought of as a sudden kick on a pendulum causing an abrupt change in the phase of the swinging motion.

The Fourier transformation of the continuous waves between two consecutive collisions separated by T is given by

$$\tilde{E}_R(\omega, T) = \frac{1}{2\pi} \int_0^T \text{Re } E(t) e^{i\omega t} dt = \tilde{G}(\omega, T) + \tilde{G}^*(-\omega, T), \quad (2.155)$$

where

$$\begin{aligned}\tilde{G}(\omega, T) &= \frac{E_0}{2\pi} \int_0^T e^{i(\omega-\omega_0)t} dt = \frac{E_0 (e^{i(\omega-\omega_0)T} - 1)}{2\pi i(\omega - \omega_0)} \\ &= \frac{E_0 T e^{i(\omega-\omega_0)T/2}}{2\pi} \text{sinc} \left[\frac{(\omega - \omega_0)T}{2} \right],\end{aligned}\quad (2.156)$$

and $\text{sinc}(x) = \sin(x)/x$ is the sinc function. For a sufficiently narrow line, we can ignore the small term $\tilde{G}^*(-\omega, T)$ and hence $\tilde{E}_R(\omega, T) \approx \tilde{G}(\omega, T)$. Thus, the spectral power of a wave segment with duration T is given by

$$|\tilde{E}_R(\omega)|^2 \propto \text{sinc}^2 \left[\frac{(\omega - \omega_0)T}{2} \right], \quad (2.157)$$

which is narrowly peaked within a frequency range given by $|\omega - \omega_0| \lesssim 2\pi/T$ — this is identical to the Fourier transform of a box function.

The time-averaged total power spectrum is given by the incoherent sum¹⁵ of all the wave segments in between consecutive collisions,

$$\begin{aligned}\langle |\tilde{E}_R(\omega)|^2 \rangle &= \int_0^\infty |\tilde{E}_R(\omega, T)|^2 \frac{dP}{dT} dT \propto \int_0^\infty \frac{\sin^2[(\omega - \omega_0)T/2]}{(\omega - \omega_0)^2} e^{-\gamma_c T} \gamma_c dT \\ &\propto \frac{1}{(\omega - \omega_0)^2 + \gamma_c^2},\end{aligned}\quad (2.158)$$

where we have used $\sin^2[(\omega - \omega_0)T/2] = [1 - \cos((\omega - \omega_0)T)]/2$ and then the Laplace transform of the cosine function as given in eq. (2.148). Thus, we see that collisions also give rise to a Lorentzian line profile in a similar form as in eq. (2.162),

$$\phi_\nu^{\text{col}} = \frac{\gamma_c/(2\pi^2)}{(\nu - \nu_0)^2 + (\gamma/2\pi)^2}, \quad (2.159)$$

but the effective damping rate is *twice* the collision rate $2\gamma_c$. The FWHM of a collisionally broadened line profile is

$$\Delta\nu_{\text{fwhm}}^{\text{col}} = \gamma_c/\pi, \quad (2.160)$$

where γ_c is the collision rate (=inverse of the mean interval between collisions).

In fact, the finite lifetimes of the two states responsible for the line transition mean that the line emission process may be terminated after an emission episode lasting for T , and T follows the exponential distribution with a mean value of the $(\gamma/2)^{-1}$. This is why natural

¹⁵The reason for the incoherent sum is that the phases of different wave segments are not mutually correlated, so their cross-correlations cancel out when we carry out a long-term time average.

broadening and collisions produce the same Lorentzian line profile (but with different damping rates). Therefore, we can combine the two effects using a total damping rate

$$\Gamma = \gamma + 2\gamma_c, \quad (2.161)$$

and the combined line profile is given by

$$\phi_\nu = \frac{\Gamma/4\pi^2}{(\nu - \nu_0)^2 + (\Gamma/4\pi)^2}. \quad (2.162)$$

The corresponding FWHM of the combined line profile is

$$\Delta\nu_{\text{fwhm}} = \frac{\Gamma}{2\pi}, \quad (2.163)$$

We need to make a technical remark here. For a given atomic system in the upper energy state, how can it emit a continuous EM wave train lasting for a duration of T that is of the order Γ^{-1} ? This is physically impossible considering that a classical wave train consists of a large number of photons, whereas a given atomic system in an excited state can at most emit one photon! Quantum mechanically, the “EM wave” here should be understood as the wave function of the emitted photon, which has a temporal span of T or spatial span of cT . The wave function tells us that it is equally likely to find the photon anywhere within a spatial segment of length cT along the direction of propagation. When the observer detects a large number of photons, and the wave functions of them have different temporal durations T which follow an exponential distribution with a mean value of $(\Gamma/2)^{-1}$. Thus, the spectrum will be given by the Lorentzian profile in eq. (2.162).

(3) Recoil effect

The above Lorentzian profile is computed in the comoving frame of the emitting atom *before* the line photon is emitted. The momentum carried away by the photon must give rise to a recoil kick to the atom. For this reason, the energy release from the transition between the two states is distributed between the photon and atom, so the line center is in fact always slightly shifted to a lower frequency. We know from energy conservation $\Delta\epsilon = \epsilon_2 - \epsilon_1 = h\nu_0 + (1/2)mv^2$ and momentum conservation $h\nu_0/c = mv$, where m is the atomic mass in the low-energy state and v is the speed of the recoil kick. In the limit $v \ll c$, we obtain $v \approx \Delta\epsilon/(mc)$ and the fractional line energy shift

$$\frac{h\nu_0 - \Delta\epsilon}{\Delta\epsilon} = \frac{h\nu_0}{\Delta\epsilon} - 1 = -\frac{mv^2}{2\Delta\epsilon} \approx -\frac{\Delta\epsilon}{2mc^2}, \quad (2.164)$$

which is always a *redshift* as compared to the case without considering a kick. We see that the recoil redshift is extremely small as it is roughly given by the ratio between the photon energy and the atomic rest-mass energy. For typical atomic transitions in the optical band,

$\Delta\epsilon \lesssim 10$ eV and $mc^2 \gtrsim$ GeV, the recoil effect leads to a redshift $z(\text{recoil}) \lesssim 10^{-8}$. This redshift is inherent and uniform to all photons emitted in all directions (independent of the observer's line of sight), so the observable result is a systematic redshift of the line without broadening the frequency distribution of photons.

As an example, let us consider the Ly α transition for a H-like atom with nuclear charge Z . The total transition energy is $\Delta\epsilon = 3Z^2\chi_H/4$, where $\chi_H = 13.6$ eV is the ionization energy for hydrogen. Taking the atomic mass to be $m = Am_p$ (A shall not be confused with the Einstein-A coefficient), we obtain the fractional frequency shift due to recoiling (retaining only the lowest order term)

$$\frac{\Delta\nu_0(\text{recoil})}{\nu_0} \approx -\frac{\Delta\epsilon}{2mc^2} = 5.4 \times 10^{-9} \frac{Z^2}{A}. \quad (2.165)$$

This is smaller than the fractional natural width of the line $(\Delta\nu)_{\text{fwhm}}^{\text{nat}}/\nu_0 = 4.0 \times 10^{-8}Z^2$ by an order of magnitude or more. Note that the Einstein-A coefficient for Ly α transition scales as Z^4 , whereas the line energy scales as Z^2 . Although the recoil effect is often negligible for atomic transitions, it becomes important for nuclear transitions where the typical photon energy is of the order MeV.

(4) Doppler broadening

The line-of-sight component of velocity of a given emitting atom produces a Doppler shift. Let us denote the component of the velocity along the line of sight as v_z (and conventionally $v_z > 0$ for redshift), and the observed frequency is given by

$$\nu = \nu_0(1 - v_z/c). \quad (2.166)$$

For non-relativistic thermal motions, the line-of-sight velocity distribution is described by the Maxwellian (=Gaussian) form with 1D velocity dispersion σ_v ,

$$\frac{dP}{dv_z} = \frac{1}{\sqrt{2\pi}\sigma_v} \exp\left(-\frac{v_z^2}{2\sigma_v^2}\right), \quad \sigma_v = \sqrt{k_B T/m}, \quad (2.167)$$

where m is the atomic mass and T is the kinetic temperature.

In the absence of damping $\Gamma \rightarrow 0$, we obtain a Gaussian line profile due to *thermal Doppler broadening*

$$\phi_\nu = \frac{1}{\sqrt{\pi}\Delta\nu_D} \exp\left[-\frac{(\nu - \nu_0)^2}{\Delta\nu_D^2}\right], \quad (2.168)$$

where $\Delta\nu_D$ is called the *Doppler width* as given by

$$\Delta\nu_D = \sqrt{2} \nu_0 \frac{\sigma_v}{c}. \quad (2.169)$$

The Gaussian profile due to thermal Doppler broadening has a FWHM of

$$(\Delta\nu)_{\text{fwhm}} = 2\sqrt{\ln 2} \Delta\nu_D = 2\sqrt{2\ln 2} \nu_0 \frac{\sigma_v}{c} = 2\sqrt{2\ln 2} \nu_0 \sqrt{\frac{k_B T}{mc^2}}. \quad (2.170)$$

In reality, the motions of atoms consist of two parts: thermal velocity due to finite temperature and bulk motion of the fluid. If the bulk motions of all contributing fluid elements are coherent, then the bulk velocity simply causes a global shift in the entire line profile but not broadening. However, if the bulk motions are random and unresolved (for e.g., small-scale turbulent motions), then the line width will be broader than that given by thermal Doppler broadening alone. Thus, the *total Doppler width* is given by

$$\Delta\nu_D = \sqrt{\Delta\nu_T^2 + \Delta\nu_{\text{turb}}^2}, \quad (2.171)$$

where $\nu_T = \nu_0 \sqrt{2k_B T/m}$ is the thermal contribution and $\Delta\nu_{\text{turb}}$ is the contribution due to turbulence.

How can we tell apart the contributions from these two terms if they are mixed together? The answer is that, if we have lines from different species (with different atomic masses) located in the same physical emitting region, then the thermal Doppler width is mass dependent $\Delta\nu_T \propto m^{-1/2}$ whereas the contribution due to micro-turbulence $\Delta\nu_{\text{turb}}$ does not depend on the atomic mass. By simultaneously fitting multiple lines, it is in principle possible to disentangle these two components.

2.4.3 Voigt profile

Here, we combine the effects of natural/collision broadening (as described by the total damping rate Γ) and the Doppler effect due to the atomic motion, and then the line shape for an atom of given line-of-sight velocity v_z is given by

$$\phi_\nu(v_z) \propto \frac{1}{[\nu(1 - v_z/c) - \nu_0]^2 + (\Gamma/4\pi)^2} \approx \frac{1}{[\nu - \nu_0 - \nu_0 v_z/c]^2 + (\Gamma/4\pi)^2}, \quad (2.172)$$

where we have ignored 2nd-order terms $\mathcal{O}(v_z^2/c^2)$ and relativistic effects.

For a general non-zero damping rate Γ , the averaged line profile is given by

$$\phi_\nu = \int_{-\infty}^{\infty} dv_z \frac{dP}{dv_z} \phi_\nu(v_z) = \frac{1}{\sqrt{\pi}\Delta\nu_D} V(a, x), \quad a \equiv \frac{\Gamma}{4\pi\Delta\nu_D}, \quad x \equiv \frac{\nu - \nu_0}{\Delta\nu_D}, \quad (2.173)$$

where $V(a, x)$ is the *Voigt function* given by

$$V(a, x) = \frac{a}{\pi} \int_{-\infty}^{\infty} \frac{e^{-y^2} dy}{a^2 + (y - x)^2}. \quad (2.174)$$

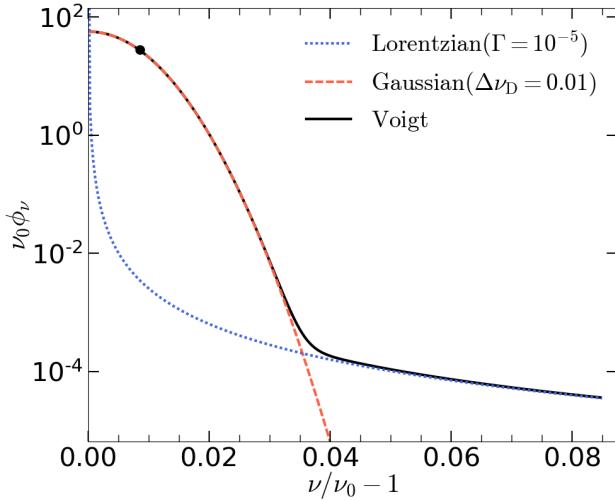


Figure 2.10: Voigt profile (black) for a weakly damped line and the corresponding Gaussian (red) and Lorentzian (blue) profile. The black circle marks the half maximum point.

The Voigt function is the convolution of the Lorentzian profile with a Gaussian and is symmetric between x and $-x$. The Voigt function also satisfies the normalization $\int_0^\infty V(a, x)dx = \sqrt{\pi}/2$ for any $a > 0$. The Lorentzian, Gaussian, and Voigt profiles are shown in Fig. 2.10. A typical line has a Gaussian core due to thermal/turbulent motions of the emitting atoms and a Lorentzian wing due to collisional/natural broadenings.

There are many proposed methods to numerically evaluate the Voigt function (no simple analytic form). The one adopted in making Fig. 2.10 is as follows ([Rotson & Obaid 2005](#))

$$V(a, x) = (1 - \text{erf}(a)) \exp(-x^2 + a^2) \cos(2ax) - \frac{2e^{x^2}}{\sqrt{\pi}} \left(\cos(2ax) \int_0^x e^{u^2} \sin(2au) du - \sin(2ax) \int_0^x e^{u^2} \cos(2au) du \right), \quad (2.175)$$

where $\text{erf}(a) = (2/\sqrt{\pi}) \int_0^a e^{-t^2} dt$ is the Gaussian error function. In the limit of weak damping $a \ll 1$, we have $V(a, x = 0) \approx 1 - 2a/\sqrt{\pi}$ at the line center. It is straightforward to numerically evaluate the FWHM for the Voigt function Δx_{fwhm} , which is defined by $V(a, \Delta x_{\text{fwhm}}/2) = V(a, x = 0)/2$. The result is given by the following polynomial fit

$$\Delta x_{\text{fwhm}} = \frac{(\Delta\nu)_{\text{fwhm}}}{\Delta\nu_D} = 2\sqrt{\ln 2} + 1.078a + 0.236a^2, \quad a = \Gamma/(4\pi\nu_D), \quad (2.176)$$

which applies for $0 < a < 1$ with a maximum fractional error of $< 0.1\%$.

2.4.4 Equivalent width and curve of growth

The equivalent width of an emission or absorption line is defined by the frequency integrated photon energy above/below the the continuum. In the following, we consider an absorption line as an example.

An absorption line profile is determined by the optical depth τ_ν , which is non-zero only near the line frequency ν_0 . The observed spectrum can be written in this form (cf. eq. 2.141)

$$F_\nu = F_\nu(0)(1 - e^{-\tau_\nu}), \quad (2.177)$$

where $F_\nu(0)$ means the continuum (for which $\tau_\nu = 0$). The continuum-subtracted line flux profile is given by

$$\Delta F_\nu = F_\nu - F_\nu(0) = -F_\nu(0)e^{-\tau_\nu}, \quad (2.178)$$

where the negative sign indicates an absorption line. The *frequency equivalent width* of such an absorption line is defined as

$$W = \int_0^\infty d\nu \frac{|\Delta F_\nu|}{F_\nu(0)} = \int_0^\infty (1 - e^{-\tau_\nu})d\nu \Rightarrow WF_{\nu_0}(0) = \int_0^\infty |\Delta F_\nu|d\nu. \quad (2.179)$$

It is more common to define the *wavelength equivalent width* (in units of Å) as

$$W_\lambda = \int_0^\infty (1 - e^{-\tau_\lambda})d\lambda = \frac{\lambda_0^2}{c} W \Rightarrow W_\lambda F_{\lambda_0}(0) = \int_0^\infty |\Delta F_\lambda|d\lambda. \quad (2.180)$$

The second expression shows that W is equal to the spectral width of the underlying continuum (near the line wavelength λ_0) that is needed to make up for the missing energy. This also applies to emission lines, where ΔF_λ and ΔF_ν are positive. Since W_λ is linearly proportional to W , they are both referred to as the equivalent width.

The equivalent width can be measured even when the line profile is not well resolved by a low-resolution spectrograph. In the following, we discuss how the equivalent width of an absorption line W depends on and may be used to measure the column density of absorbers N_{col} along the line of sight.

In the following, we consider a Voigt profile ϕ_ν with a narrow Gaussian core as specified by the Doppler width $\Delta\nu_D$ and a broad Lorentzian wing as specified by the damping rate Γ . We take the limit of weak damping $\Gamma \ll \nu_D$ or $a = \Gamma/(4\pi\nu_D) \ll 1$, which is usually the case in various astrophysical environments.

The line optical depth profile has the following scaling (cf. eq. 2.125)

$$\tau_\nu \propto N_{\text{col}}\phi_\nu. \quad (2.181)$$

We denote the optical depth at the line center is as $\tau_0 \equiv \tau_{\nu_0}$, which is proportional to the column density of absorbers N_{col} . Our goal is to study how the equivalent width W depends on the line-center optical depth τ_0 .

At the line center and in the weak damping limit, (cf. 2.175)

$$V(a, x = 0) = (1 - \text{erf}(a))e^{a^2} \approx 1, \quad (2.182)$$

so we obtain

$$\phi_{\nu_0} \approx \frac{1}{\sqrt{\pi}\nu_D}. \quad (2.183)$$

Thus, we can write the optical depth profile in terms of the line-center optical depth τ_0 ,

$$\tau_\nu \approx \sqrt{\pi}\Delta\nu_D\tau_0\phi_\nu. \quad (2.184)$$

The dependence of the equivalent width $W = \int(1 - e^{-\tau_\nu})d\nu$ on τ_0 can be divided into the following three regimes, depending on whether the Gaussian core region and the Lorentzian wings are optically thick.

(i) In the limit $\tau_0 \ll 1$, the entire line is optically thin, so we obtain $W \approx \tau_\nu d\nu = \sqrt{\pi}\Delta\nu_D\tau_0$, which is linearly proportional to τ_0 or the absorbers' column density N_{col} .

(ii) In the limit $1 \ll \tau_0 \ll \tau_{\text{damp}}$, the system is optically thick near the line center but optically thin at frequencies in the so-called *damped wings* where the Lorentzian wings dominate over the Gaussian core. We will later on provide an estimate of the characteristic optical depth τ_{damp} , above which the damped wings become optical thick. In the current case, the line is optical thick at frequencies that satisfy $\phi_\nu \gtrsim (\sqrt{\pi}\Delta\nu_D\tau_0)^{-1}$. Since we are still in the Gaussian core of the line, the profile is roughly given by eq. (2.168), so we can solve $\phi_\nu \simeq (2\sqrt{\pi}\Delta\nu_D\tau_0)^{-1}$ for the critical frequency at which $\tau_\nu = 1/2$. The result is

$$|\nu(\tau_\nu = 1/2) - \nu_0| \simeq \Delta\nu_D\sqrt{\ln 2\tau_0}, \quad \text{for } 1 \ll \tau_0 \ll \tau_{\text{damp}}. \quad (2.185)$$

Thus, the equivalent width is roughly given by $W \simeq 2\Delta\nu_D\sqrt{\ln 2\tau_0} \propto \sqrt{\ln \tau_0}$ in the regime of $1 \ll \tau_0 \ll \tau_{\text{damp}}$. In this regime, the equivalent width depends strongly on the Doppler width but very weakly on the line-center optical depth, so W cannot be used to accurately measure the column density.

(iii) In the limit $\tau_0 \gg \tau_{\text{damp}}$, the line is optically thick all the way into the damped wings. To estimate the frequency at which $\tau_\nu = 1/2$, we adopt the Lorentzian profile in (2.162) to solve $\phi_\nu \simeq (\sqrt{\pi}\Delta\nu_D\tau_0)^{-1}/2$. In the limit $|\nu - \nu_0| \gg \Gamma/4\pi$, we obtain

$$|\nu(\tau_\nu = 1/2) - \nu_0| \simeq \sqrt{\frac{\Delta\nu_0\tau_0\Gamma}{2\pi^{3/2}}}, \quad \text{for } \tau_0 \gg \tau_{\text{damp}}. \quad (2.186)$$

Then, the equivalent width is roughly given by $W \simeq 2\sqrt{\Delta\nu_0\tau_0\Gamma/(2\pi^{3/2})} \propto \tau_0^{1/2}$. Again, W can be used to accurately measure the column density.

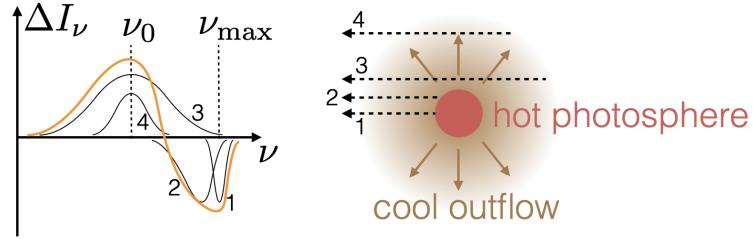


Figure 2.11: The P-Cygni line profile as the superposition of the intensity profiles at different lines of sight (4 examples shown). The relative strength of emission and absorption depends on the solid angle spanned by the outflow and the continuum photosphere, as well as the temperature profile of the outflow. In the very late so-called “nebular” phase of a supernova, the continuum photosphere is relatively small and the contributions by sightlines like “1” and “2” are negligible, so the spectrum is dominated by emission lines.

To summarize, we have the following three regimes

$$W = \frac{W_{\lambda}c}{\lambda_0^2} \simeq \begin{cases} \sqrt{\pi}\Delta\nu_D\tau_0 \propto \tau_0, & \text{for } \tau_0 \ll 1, \\ 2\Delta\nu_D\sqrt{2\ln 2\tau_0} \propto \sqrt{\ln \tau_0}, & \text{for } 1 \ll \tau_0 \ll \tau_{\text{damp}} \\ 2\sqrt{\Delta\nu_0\tau_0\Gamma/(2\pi^{3/2})} \propto \tau_0^{1/2}, & \text{for } \tau_0 \gg \tau_{\text{damp}} \end{cases} \quad (2.187)$$

The dependence of W on τ_0 is called the *curve of growth*. Finally, the critical line-center optical depth τ_{damp} is roughly given by equating the equivalent widths in regimes (ii) and (iii), and we obtain

$$\frac{\tau_{\text{damp}}}{\ln 2\tau_{\text{damp}}} \simeq \frac{2\pi^{3/2}}{\Gamma/\Delta\nu_D} \rightarrow \tau_{\text{damp}} \simeq \frac{2\pi^{3/2}}{\Gamma/\Delta\nu_D} \ln \left(\frac{4\pi^{3/2}}{\Gamma/\Delta\nu_D} \right). \quad (2.188)$$

[Todo: Show curve of growth for Lyα.]

2.4.5 Collisional (de-)excitation (under construction)

[TODO:] collision rate coefficient, law of detailed balance

2.4.6 Critical electron density (under construction)

[TODO:] Coulomb focusing in electron-ion collisions, critical electron density

2.4.7 P-Cygni profile

An interesting application of our theory of emission/absorption lines is the P-Cygni line profile, which is commonly observed in stars with strong winds and supernovae. The physical picture is a central source surrounded by an outflow with the brightness temperature of

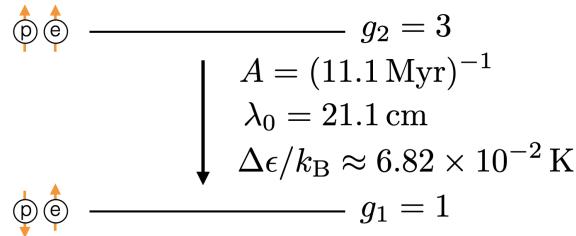


Figure 2.12: The 21cm line from the electron spin-flip transition of a hydrogen atom.

the source's photosphere $T_{b,0}$ (defined at the line frequency) higher than the line excitation temperature T_{exc} (defined by the ratio of the electron occupation numbers in the energy levels involved) in the outflow. The outflow is extended in that it covers a volume much larger than the central source, as shown in Fig. 2.11. For sightlines towards the continuum photosphere, we obtain an absorption line at blue-shifted frequencies — up to a maximum frequency of $\nu_{\text{max}} = \nu_0(1 + v/c)$, where v is the outflow velocity and ν_0 is the line center frequency. For sightlines that only intersect the outflow, we get a symmetric emission line centered at ν_0 in the optically thin case¹⁶. The superposition of all sightlines yield a complicated line profile with blue absorption and red emission. Note that the zero point, where $\Delta I_\nu = 0$, is not necessarily located at the line center. The diversity of P-Cygni line profiles has been extensively studied observationally and theoretically (see e.g., Beals 1953).

2.4.8 Hydrogen 21cm line

The 21cm line comes from the electron spin-flip transition in the ground configuration ($1s^1$) of neutral hydrogen (HI), as shown in Fig. 2.12. It is convenient to describe the ratio between electron occupation numbers by the spin temperature T_{spin} (which is just another name for the excitation temperature in this context)

$$\frac{\eta_2}{\eta_1} \equiv \exp\left(-\frac{\Delta\epsilon}{k_B T_{\text{spin}}}\right) \approx 1 - \frac{6.82 \times 10^{-2} \text{ K}}{T_{\text{spin}}}, \quad (2.189)$$

where the second expression is a Taylor expansion that is always valid since we are in the limit $\Delta\epsilon/(k_B T_{\text{spin}}) \ll 1$. Thus, the number ratio between atoms in the upper (spin aligned) and lower (spin anti-aligned) levels is roughly $N_2/N_1 \approx g_2/g_1 = 3$, to very good accuracy. If the *total* number density of HI is n_{HI} , then 3/4 of them are in the upper level and hence the emissivity is given by

$$j_\nu = \frac{1}{4\pi} \frac{3}{4} n_{\text{HI}} A h \nu_0 \phi_\nu, \quad (2.190)$$

where ϕ_ν is the line profile given by broadening effects.

¹⁶If outflow is optically thick, the red part of the line emission for sightlines 3 and 4 will be missing.

For the 21cm line, broadening is dominated by Doppler shifts associated with the motion of the atoms, so we take

$$\phi_\nu = \frac{1}{\sqrt{2\pi}\sigma_v} \frac{c}{\nu_0} e^{-v_z^2/2\sigma_v^2}, \quad (2.191)$$

where v_z is the line-of-sight (LOS) velocity component and σ_v is the LOS velocity dispersion. Since $B_\nu(T_{\text{spin}}) \approx (2\nu_0^2/c^2)k_B T_{\text{spin}}$, we obtain the absorption coefficient

$$\alpha_\nu = \frac{j_\nu}{B_\nu(T_{\text{spin}})} \approx 2.19 \times 10^{-19} \text{ cm}^{-1} \frac{n_{\text{HI}}}{\text{cm}^{-3}} \left(\frac{T_{\text{spin}}}{\text{K}} \right)^{-1} \left(\frac{\sigma_v}{\text{km s}^{-1}} \right)^{-1} e^{-v_z^2/2\sigma_v^2}. \quad (2.192)$$

Typical HI gas in the ISM has $\sigma_v \sim \text{few km s}^{-1}$ and $T_{\text{spin}} \sim 10^2 \text{ K}$ (set by kinetic temperature due to collisions), so we know that the 21cm line may be optically thick if the column density is higher than about 10^{21} cm^{-2} .

2.5 Photon diffusion under repetitive scatterings

In this section, we consider the dynamics of photons moving in a homogeneous scattering medium. Each scattering is assumed to be *isotropic* in that the outgoing direction is randomly distributed independent of the incident direction. Our goal is to show that, for a medium with a large scattering optical depth, photon propagation can be described by diffusion where the number flux is proportional to the spatial gradient of the number density and the diffusion coefficient is related to the mean free path (mfp) ℓ_{mfp} .

2.5.1 Random walk

Let us consider a photon initially released at the coordinate origin. The scattering mean free path is denoted as ℓ_{mfp} . The photon's position after the N th scattering is

$$\mathbf{r}_N = \sum_{i=1}^N \Delta \mathbf{r}_n, \quad (2.193)$$

where $\Delta \mathbf{r}_n$ is a 3D vector pointing from the position of the $(n-1)$ th scattering to the n th one. The length of each $\Delta \mathbf{r}_n$ is drawn from the exponential distribution in eq. (2.105), which means: the mean value of the magnitude is $\langle |\Delta \mathbf{r}_n| \rangle = \ell_{\text{mfp}}$, the vectorial mean is $\langle \Delta \mathbf{r}_n \rangle = 0$, and the mean-squared distance is $\langle |\Delta \mathbf{r}_n|^2 \rangle = 2\ell_{\text{mfp}}^2$. Let us then consider that initially there are a large number of photons at the coordinate origin. After each of them undergoing N scatterings, they would be at different displacements \mathbf{r}_N . The vectorial mean of all photons vanishes, $\langle \mathbf{r}_N \rangle = 0$, because each scattering is isotropic. However, the mean squared displacement is non-zero,

$$\langle \mathbf{r}_N^2 \rangle = \sum_n \langle |\Delta \mathbf{r}_n|^2 \rangle + 2 \sum_{n < m} \langle \Delta \mathbf{r}_n \cdot \Delta \mathbf{r}_m \rangle. \quad (2.194)$$

Since any two scatterings (for $n \neq m$) are not correlated, the cross terms must be zero if we average over sufficiently large number of photons. This leads to an important result for isotropic scattering

$$\langle \mathbf{r}_N^2 \rangle = 2N\ell_{\text{mfp}}^2, \quad (2.195)$$

and the root-mean-squared radial displacement from the origin after N scattering is $\langle \mathbf{r}_N^2 \rangle^{1/2} = \sqrt{2N}\ell_{\text{mfp}}$. In fact, the above result applies to any scattering with forward-backward symmetry (e.g. Thomson scattering). In the limit of $N \gg 1$, we can approximate

$$N \approx ct/\ell_{\text{mfp}}, \quad (2.196)$$

where t is the time since the photons are released. The mean-squared displacement from the origin grows with time as

$$\langle \mathbf{r}^2 \rangle = 2c\ell_{\text{mfp}}t, \quad (2.197)$$

which is the characteristic property of a *diffusion process*.

For a uniform sphere of scattering-dominate gas with radius $R = \tau\ell_{\text{mfp}}$ and optical depth $\tau \gg 1$, photons released from the center will escape after an average number of scatterings (by traveling a root-mean-squared displacement of R)

$$\langle N \rangle_{\text{sphere}} = R^2/(2\ell_{\text{mfp}}^2) = \tau^2/2. \quad (2.198)$$

This corresponds to a diffusion timescale

$$t_{\text{dif}} = \frac{R^2}{2c\ell_{\text{mfp}}} = \frac{\tau}{2} \frac{R}{c}, \quad (2.199)$$

which is a factor of $\tau/2$ longer than the light-crossing time R/c .

On the other hand, let us consider photons released at the center of a uniform scattering slab of *half* thickness $L = \tau\ell_{\text{mfp}}$ (again $\tau \gg 1$), the mean-squared displacement along the normal ($\hat{\mathbf{z}}$) direction of the slab after N scatterings is given by

$$\langle z_N^2 \rangle = \langle (\mathbf{r}_N \cdot \hat{\mathbf{z}})^2 \rangle = \sum_n \langle (\Delta \mathbf{r}_n \cdot \hat{\mathbf{z}})^2 \rangle = \sum_n \langle \Delta r_n^2 \rangle \langle \mu_n^2 \rangle = 2N\ell_{\text{mfp}}^2/3, \quad (2.200)$$

where we have dropped the ensemble average of the cross-correlation terms and expressed $\Delta \mathbf{r}_n \cdot \hat{\mathbf{z}} = \Delta r_n \mu_n$ with a projection factor μ_n drawn *independently* from the Poisson-distributed length Δr_n for each scattering. The second moment of the projection factor is $\langle \mu^2 \rangle = \int_{-1}^1 \mu^2 (dP/d\mu) d\mu = 1/3$ for a uniform distribution $dP/d\mu = 1/2$ in the range $\mu \in [-1, 1]$. After traveling a root-mean-squared displacement of L in the $\hat{\mathbf{z}}$ -direction, an average photon will be able to escape, so we obtain the average number of scatterings for the slab geometry

$$\langle N \rangle_{\text{slab}} = (3/2)L^2/\ell_{\text{mfp}}^2 = 3\tau^2/2, \quad (2.201)$$

which is slightly different from the result for a spherical gas distribution (eq. 2.198).

2.5.2 Fick's law for diffusion

From the earlier subsection, we conclude that, for a typical photon to escape from a region with scattering optical depth $\tau \gg 1$, it must be scattered $N \sim \tau^2$ times (for any geometry). The time it takes to cross a distance $L \gg \ell_{\text{mfp}}$ is called the *diffusion time*

$$t_{\text{dif}} \sim \tau^2 \ell_{\text{mfp}} / c = \tau L / c, \quad (2.202)$$

and hence the effective *diffusion speed* is

$$v_{\text{dif}} \sim L / t_{\text{dif}} \sim c / \tau, \quad (2.203)$$

which is much slower than c since the photon must travel for a *total path length* of $\sim \tau L$ in order to escape. The number flux = the number of passing photons per unit area per unit time is of the order

$$F \sim n v_{\text{dif}} \sim \ell_{\text{mfp}} c n / L. \quad (2.204)$$

Below, we quantify the number flux a bit more.

Let us consider a linear gradient of photon number density along the \hat{z} direction. The net flux passing through the x-y plane located at position z is given by

$$F(z) = \frac{c}{6} [n(z - \ell_{\text{mfp}}) - n(z + \ell_{\text{mfp}})] = -\frac{1}{3} c \ell_{\text{mfp}} \frac{dn}{dz}(z), \quad (2.205)$$

where $n(z - \ell_{\text{mfp}})c/6$ describes the flux of photons moving along the $+\hat{z}$ direction (because photons on average stream freely for a distance of ℓ_{mfp}), $n(z + \ell_{\text{mfp}})c/6$ is the flux in the $-\hat{z}$ direction, and, in 3D, 1/6 of the particles at any given position are moving along the $+\hat{z}$ (or $-\hat{z}$) direction.

Therefore, the number flux can be written in the following form

$$\mathbf{F} = -D \nabla n, \quad (2.206)$$

which is called the *Fick's law for diffusion*. From eq. (2.205), we find the *diffusion coefficient* to be

$$D = c \ell_{\text{mfp}} / 3. \quad (2.207)$$

Our eq. (2.205) seems a bit hand-waving. For readers who are interested in the details, we derive the Fick's law and the diffusion coefficient using the Fokker-Planck equation in the next subsection. In fact, the diffusion coefficient in eq. (2.207) applies to any isotropic scattering in 3D, as long as we replace “ c ” with the particles' average speed and “ ℓ_{mfp} ” with the mean free path.

An interesting example of photon diffusion is as follows. Clouds in the Earth's atmosphere are made of tiny water droplets which primarily scatter visible light from the Sun (and



Figure 2.13: Dark clouds (left panel) viewed from the ground and white clouds (right panel, credit: Emiko Gardiner) viewed from an airplane's window.

absorption can be ignored here). Let us consider a vast horizontal layer of cloud with scattering optical depth τ in the vertical (z) direction. If $\tau \ll 1$, an observer underneath the cloud can of course directly see the Sun. When $\tau \gg 1$ (which is typically the case for the cloud to be easily noticeable by the observer), one might naïvely think that the observer will be left in nearly complete darkness because $e^{-\tau} \approx 0$.

However, in reality, we can still see things well even in cloudy days (but cannot tell the shape or precise position of the Sun). Armed with Fick's law (eq. 2.206) and the diffusion coefficient (eq. 2.207), we find that the net flux penetrating through an optically thick scattering layer is given by

$$F \sim c\ell_{\text{mfp}}|\text{d}U/\text{d}z| \sim c\ell_{\text{mfp}}U/z \sim U_0c/\tau,$$

where U_0 is the energy density of photons near the *upper boundary* of the cloud that is directly illuminated by the Sun. Another way of understanding is that photons deeply inside an optically thick cloud move at the diffusion speed $v_{\text{dif}} \sim c/\tau$ (eq. 2.203), so the flux is given by $F \sim Uv_{\text{dif}} \sim U_0c/\tau$, which is a factor of τ^{-1} less than the *incident flux* $F_0 = U_0c$. We see that the *transmitted flux* is a factor of τ^{-1} (instead of $e^{-\tau}$) smaller than the incident flux because photons undergo many scatterings before reaching the observer. Still, the suppression factor of $\tau^{-1} \ll 1$ will make the bottom of the cloud look dark because the intensity is much reduced (see the left panel of Fig. 2.13). From the window of an airplane overseeing the cloud from above, the observer sees that the cloud is white and shiny, and this is because most of the Sun light is reflected back into the space by the optically thick cloud, as shown in the right panel of Fig. 2.13.

If you are not satisfied with the arguments above, the τ^{-1} suppression factor can also be understood in a more rigorous way as follows. Let us denote the incident, reflected, and transmitted fluxes to be F_{in} , F_{refl} , and F_{tran} , respectively. In a steady state, the *net flux* F normal to the slab is constant throughout the entire (1D) domain and it is

possible to convince yourself that $F = F_{\text{in}} - F_{\text{refl}} = F_{\text{tran}}$. Our intuition tells us that $F_{\text{in}} \approx F_{\text{refl}} \gg F_{\text{tran}}$, since nearly all the incident light is reflected, but the question is what fraction gets transmitted. From Fick's law (eq. 2.206) and the diffusion coefficient (eq. 2.207), we use the optical depth $d\tau \equiv dz/\ell_{\text{mfp}}$ as our vertical coordinate and write

$$dU/d\tau = -3F/c.$$

Since F is a constant, this can be integrated into

$$U(\tau) = -3F\tau/c + C,$$

where C is a constant of integration. On the incident boundary at $\tau = 0$, we know that the radiation field is nearly isotropic (since $F_{\text{in}} \approx F_{\text{refl}}$), and we can estimate $U(\tau = 0) = C \simeq 4F_{\text{in}}/c$ — the factor of “4” is analogous to the relationship between the energy density aT^4 and one-sided flux $\sigma_{\text{SB}}T^4$ for the case of a blackbody (since $\sigma_{\text{SB}} = ac/4$, but here the radiation field does not have a blackbody spectrum). Thus, we obtain

$$U(\tau) = -3F\tau/c + 4F_{\text{in}}/c. \quad (2.208)$$

On the other boundary at $\tau \gg 1$, we know that the radiation energy density is of the order $F_{\text{tran}}/c = F/c$, which is much smaller than the absolute values of the two terms on the right-hand side of the above equation. We then find $3F\tau/c \simeq 4F_{\text{in}}/c$ and hence

$$F_{\text{tran}} = F \simeq (4/3)F_{\text{in}}/\tau. \quad (2.209)$$

This shows that the transmitted flux is a factor of τ^{-1} smaller than the incident flux. We refer to §2.6.6 for a more detailed discussion of radiative transfer in a slab.

2.5.3 *Fokker-Planck equation

We would like to explicitly solve the time evolution of the photon number density $n(\mathbf{r}, t)$ inside a scattering medium. Let us consider a time interval δt that is much shorter than the diffusion time t_{dif} across the lengthscale L in consideration but much longer than the interval ℓ_{mfp}/c between consecutive scatterings for a given photon. Let $p(\mathbf{r}, \Delta\mathbf{r}) = dP/d^3\Delta\mathbf{r}$ be the probability density for a given photon to get scattered from a volume $d^3\Delta\mathbf{r}$ near position \mathbf{r} to the same volume near another position $\mathbf{r} + \Delta\mathbf{r}$ *within a time interval of δt* . The probability density function $p(\mathbf{r}, \Delta\mathbf{r})$ depends on the time interval δt and we leave this dependence implicit because the final results do not depend on δt . The probability density function is normalized such that

$$\int p(\mathbf{r}, \Delta\mathbf{r}) d^3\Delta\mathbf{r} = 1, \quad (2.210)$$

which applies to all initial positions \mathbf{r} . Based on the discussion in the previous subsection, we know that, for isotropic scattering, the mean and mean-squared displacements after a time interval δt are given by, respectively,

$$\begin{aligned}\langle \Delta\mathbf{r} \rangle &= \int p(\mathbf{r}, \Delta\mathbf{r}) \Delta\mathbf{r} d^3\Delta\mathbf{r} = 0, \\ \langle \Delta\mathbf{r}^2 \rangle &= \langle \Delta x^2 \rangle + \langle \Delta y^2 \rangle + \langle \Delta z^2 \rangle = 3 \langle \Delta x^2 \rangle = \int p(\mathbf{r}, \Delta\mathbf{r}) \Delta\mathbf{r}^2 d^3\Delta\mathbf{r} = 2c\ell_{\text{mfp}}\delta t.\end{aligned}\quad (2.211)$$

The following discussion generally applies to any diffusion process, but the special case of isotropic scattering is a good example.

Since no photons are created or destroyed in the system, the change in the photon number density near position \mathbf{r} in the time interval δt is given by the difference between those scattered from other positions $\mathbf{r} - \Delta\mathbf{r}$ (where $\Delta\mathbf{r} \neq 0$) to \mathbf{r} and the ones scattered from position \mathbf{r} to other positions. This can be written as

$$n(\mathbf{r}, t + \delta t) - n(\mathbf{r}, t) = \int n(\mathbf{r} - \Delta\mathbf{r}, t) p(\mathbf{r} - \Delta\mathbf{r}, \Delta\mathbf{r}) d^3\Delta\mathbf{r} - \int n(\mathbf{r}, t) p(\mathbf{r}, \Delta\mathbf{r}) d^3\Delta\mathbf{r}, \quad (2.212)$$

which is called the *Master equation*. For a small displacement $|\Delta\mathbf{r}| \ll |\mathbf{r}|$, one can write the following Taylor expansions near position \mathbf{r} ,

$$\begin{aligned}n(\mathbf{r} - \Delta\mathbf{r}, t) &\approx n(\mathbf{r}, t) - \sum_{i=x,y,z} \Delta r_i \partial_i n + (1/2) \sum_{i,j=x,y,z} \Delta r_i \Delta r_j \partial_i \partial_j n, \\ p(\mathbf{r} - \Delta\mathbf{r}, \Delta\mathbf{r}) &\approx p(\mathbf{r}, \Delta\mathbf{r}) - \sum_{i=x,y,z} \Delta r_i \partial_i p + (1/2) \sum_{i,j=x,y,z} \Delta r_i \Delta r_j \partial_i \partial_j p,\end{aligned}\quad (2.213)$$

where $\partial_i \equiv \partial/\partial x_i$, and $i, j = x, y, z$ are for the three Cartesian axes. Plugging the truncated expansion series above into the Master equation, we re-arrange the terms and obtain

$$n(\mathbf{r}, t + \delta t) - n(\mathbf{r}, t) = - \sum_{i=x,y,z} \partial_i [\langle \Delta r_i \rangle n(\mathbf{r}, t)] + (1/2) \sum_{i,j=x,y,z} \partial_i \partial_j [\langle \Delta r_i \Delta r_j \rangle n(\mathbf{r}, t)], \quad (2.214)$$

where we have defined the mean displacement, auto-correlation (for $i = j$), and cross-correlation (for $i \neq j$) of the displacement,

$$\langle \Delta r_i \rangle = \int \Delta r_i p(\mathbf{r}, \Delta\mathbf{r}) d^3\Delta\mathbf{r}, \quad \langle \Delta r_i \Delta r_j \rangle = \int \Delta r_i \Delta r_j p(\mathbf{r}, \Delta\mathbf{r}) d^3\Delta\mathbf{r}. \quad (2.215)$$

Non-zero mean displacement gives rise to a drift, and the auto-/cross- correlations leads to (generally anisotropic) diffusion. This motivates us to define the *drift velocity*

$$V_i = \langle \Delta r_i \rangle / \delta t, \quad (2.216)$$

and the *diffusion coefficient tensor*

$$D_{ij} = \langle \Delta r_i \Delta r_j \rangle / (2\delta t). \quad (2.217)$$

For a small time interval $\delta t \ll t_{\text{dif}}$ (t_{dif} being the diffusion timescale), we write $n(\mathbf{r}, t + \delta t) - n(\mathbf{r}, t) \approx \delta t \partial_t n(\mathbf{r}, t)$ and then arrive at

$$\partial_t n(\mathbf{r}, t) = - \sum_i \partial_{i=x,y,z} [V_i n(\mathbf{r}, t)] + \sum_{ij=x,y,z} \partial_i \partial_j [D_{ij} n(\mathbf{r}, t)], \quad (2.218)$$

which is called the *Fokker-Planck equation* or more intuitively the *drift-diffusion equation*. This is widely used in many fields of astrophysics (e.g., particle transport in plasma, cosmic ray acceleration, dynamics of stellar cluster, orbital migration of planets) where a stochastic process is involved. In other contexts, the drift and diffusion might be occurring in energy, momentum or angular-momentum space, so the vector \mathbf{r} would need to be replaced accordingly.

Going back to the simple case of isotropic diffusion, the drift velocity vanishes $V_i = 0$, and the diffusion coefficient tensor is given by $\overset{\leftrightarrow}{D} = D \overset{\leftrightarrow}{I}$, where $\overset{\leftrightarrow}{I}$ is the identity matrix and the *scalar diffusion coefficient* D is given by (for $i = x, y, z$)

$$D = D_{ii} = \langle \Delta r_i^2 \rangle / (2\delta t) = c \ell_{\text{mfp}} / 3. \quad (2.219)$$

Then, the Fokker-Planck equation simplifies into the following *diffusion equation*

$$\partial_t n(\mathbf{r}, t) = \nabla^2 [D n(\mathbf{r}, t)]. \quad (2.220)$$

Comparing the diffusion equation with the continuity equation $\partial_t n + \nabla \cdot \mathbf{F} = 0$, we find that the number flux is proportional to the gradient of the local gradient of the particle number density as given by the Fick's law (eq. 2.206).

2.5.4 *Infinite uniform medium

For a uniform, time-independent scattering medium, the diffusion coefficient D is a constant that pulled out of the diffusion equation, and we write the diffusion equation as

$$\partial_t n(\mathbf{r}, t) = D \nabla^2 [n(\mathbf{r}, t)]. \quad (2.221)$$

In the following, we solve the diffusion equation (2.221) using the method of Fourier transformation. Let us write the spatial Fourier transformation of the solution $n(\mathbf{r}, t)$ as

$$\tilde{n}(\mathbf{k}, t) = \frac{1}{(2\pi)^3} \int n(\mathbf{r}, t) e^{i\mathbf{k} \cdot \mathbf{r}} d^3 \mathbf{r}, \quad (2.222)$$

and the inverse Fourier transformation is

$$n(\mathbf{r}, t) = \int \tilde{n}(\mathbf{k}, t) e^{-i\mathbf{k}\cdot\mathbf{r}} d^3\mathbf{k}. \quad (2.223)$$

Plugging the inverse Fourier form of $n(\mathbf{r}, t)$ into the diffusion equation (and for a constant D), we obtain

$$\int [\partial_t \tilde{n}(\mathbf{k}, t) + Dk^2 \tilde{n}(\mathbf{k}, t)] e^{-i\mathbf{k}\cdot\mathbf{r}} d^3\mathbf{k} = 0. \quad (2.224)$$

For the above equation to hold at all \mathbf{k} and t , we require

$$\partial_t \tilde{n}(\mathbf{k}, t) = -Dk^2 \tilde{n}(\mathbf{k}, t), \quad (2.225)$$

which can be easily integrated to obtain the following solution

$$\tilde{n}(\mathbf{k}, t) = \tilde{n}(\mathbf{k}, t=0) e^{-Dk^2 t}, \quad (2.226)$$

where $\tilde{n}(\mathbf{k}, t=0)$ is given by the initial condition $n(\mathbf{r}, t=0)$ through Fourier transformation above.

In the special case of an initial condition where all photons are concentrated at the coordinate origin, the initial photon distribution $n(\mathbf{r}, t=0)$ is given by a delta-function

$$\tilde{n}(\mathbf{k}, t=0) = N_{\text{tot}} \delta(\mathbf{r}), \quad (2.227)$$

where the delta function is normalized by $\int \delta(\mathbf{r}) d^3\mathbf{r} = \int_0^\infty \delta(r) 4\pi r^2 dr = 1$ such that there are N_{tot} photons in the entire domain. In this case, the Fourier transformation of the initial density distribution is given by

$$\tilde{n}(\mathbf{k}, t=0) = \frac{N_{\text{tot}}}{(2\pi)^3} \int \delta(\mathbf{r}) e^{i\mathbf{k}\cdot\mathbf{r}} d^3\mathbf{r} = \frac{N_{\text{tot}}}{(2\pi)^3}. \quad (2.228)$$

Thus, we obtain the Fourier transformation of the time-dependent photon distribution

$$\tilde{n}(\mathbf{k}, t) = \frac{N_{\text{tot}}}{(2\pi)^3} e^{-Dk^2 t}. \quad (2.229)$$

Finally, we use an inverse Fourier transformation to obtain the photon distribution

$$\begin{aligned} n(\mathbf{r}, t) &= \frac{N_{\text{tot}}}{(2\pi)^3} \int e^{-Dk^2 t} e^{-i\mathbf{k}\cdot\mathbf{r}} d^3\mathbf{k} \\ &= \frac{N_{\text{tot}}}{(2\pi)^3} e^{-r^2/(4Dt)} \int \exp \left[-Dt \left(\mathbf{k} + \frac{i}{2Dt} \mathbf{r} \right)^2 \right] d^3\mathbf{k} \\ &= \frac{N_{\text{tot}}}{8\pi^{3/2} (Dt)^{3/2}} e^{-r^2/(4Dt)}, \end{aligned} \quad (2.230)$$

where we have used a trick by ‘‘completing the square’’ to rewrite the expression into the form of a Gaussian integral and then used the theorem for Gaussian integral in eq. (1.26) along each of the Cartesian axis (k_x , k_y , and k_z).

We find that the photons follow a spherically symmetric Gaussian distribution around their initial position. It is convenient to define the spatial dispersion of the Gaussian as

$$\sigma(t) = \sqrt{2Dt}, \quad (2.231)$$

and then the time-dependent photon distribution is given by

$$n(\mathbf{r}, t) = \frac{N_{\text{tot}}}{(2\pi\sigma^2)^{3/2}} e^{-r^2/(2\sigma^2)}, \quad (2.232)$$

where N_{tot} is the total number of photons and the normalization is given by $\int n d^3\mathbf{r} = N_{\text{tot}}$. We see that an average photon moves away from the origin at a rate given by $\sigma(t) \propto t^{1/2}$.

2.5.5 *Finite spherical uniform medium

In this section, we consider photons released from the center of a uniform spherical static medium of radius R , and our discussion also applies to the case of an infinite medium discussed earlier as long as we take the limit of $R \rightarrow \infty$. The photons undergo a random walk due to scattering with a constant mean free path of $\ell_{\text{mfp}} \ll R$. We assume isotropic scattering, so the diffusion coefficient is given by $D = \ell_{\text{mfp}} c / 3 = \text{const}$. In the interior of the medium, the photon number flux is given by the Fick’s law $\mathbf{F} = D \nabla n(\mathbf{r}, t)$, and the time evolution of the photon number density profile $n(\mathbf{r}, t)$ is governed by the diffusion equation $\partial_t n = D \nabla^2 n$.

The initial condition of our problem is given by

$$n(\mathbf{r}, t=0) = N_{\text{tot}} \delta(\mathbf{r}), \quad (2.233)$$

where the delta-function is normalized such that $\int \delta(\mathbf{r}) d^3\mathbf{r} = 1$ and total number of photons is $N_{\text{tot}} = 1$. When photons reach the boundary at $r = R$, they are assumed to rapidly escape, so we take the boundary condition as

$$n(r = R, t) = 0. \quad (2.234)$$

Note that our boundary condition is an approximation. In reality, photons propagate at a finite speed even at $r > R$, so we have $n(r = R) \neq 0$ in general. However, as we demonstrate below, our simplified boundary condition is an excellent approximation for the case with a very large optical depth $\tau = R/\ell_{\text{mfp}} \gg 1$, so their diffusion speed is much less than the speed of light in the interior of the medium.

In spherical symmetry, the diffusion equation can be written as

$$\partial_t n(r, t) = \frac{D}{r^2} \partial_r [r^2 \partial_r n(r, t)]. \quad (2.235)$$

In the following, we solve the above equation using separation of variables.

Let us take the ansatz of

$$n(r, t) = X(r)T(t), \quad (2.236)$$

and plug it into the diffusion equation (2.235). This leads to

$$\frac{1}{T} \frac{dT}{dt} = \frac{D}{X} \frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{dX}{dr} \right) = -Dk^2, \quad (2.237)$$

where k is a constant to be determined by the initial and boundary conditions.

The solution to the temporal part is exponential

$$T(t) = e^{-Dk^2 t}, \quad (2.238)$$

where the constant of integral has been taken to be unity as it can be absorbed into the spatial part $X(r)$. To avoid exponentially growing solutions which violate photon number conservation, we require k to be real and hence $k^2 > 0$.

To obtain the solution to the spatial part, we define another function $Y(x) = rX$, and it can be shown that $dX/dr = (rY' - Y)/r^2$, where $Y' \equiv dY/dr$. Then, we find

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{dX}{dr} \right) = \frac{Y''}{r} = -k^2 X = -k^2 \frac{Y}{r}, \quad (2.239)$$

which gives the harmonic oscillator equation $Y'' = -k^2 Y$. The general solution is then

$$Y = rX = A \sin kr + B \cos kr \Rightarrow X = \frac{Y}{r} = \frac{A \sin kr + B \cos kr}{r}. \quad (2.240)$$

where A and B are constants. We also require that $X(r)$ does not diverge as $r \rightarrow 0$ (except for the initial condition at $t = 0$), so the physical solution must have $B = 0$. We therefore obtain $X(r) = Ar^{-1} \sin kr$ and hence

$$n(\mathbf{r}, t) = A \frac{\sin kr}{r} e^{-Dk^2 t}. \quad (2.241)$$

One can plug the above form back into the diffusion equation (2.235) and see that it is indeed a solution for *arbitrary* constants A and k . Moreover, the linearity of the diffusion equation (2.235) means that, if $n_1(\mathbf{r}, t)$ and $n_2(\mathbf{r}, t)$ are solutions, then their sum $n_1 + n_2$ is also a solution. In the following, we use the initial and boundary conditions to determine the allowed values of constants A and k .

At the outer boundary $r = R$, since $n(r = R, t) = 0$ at all times, we must have $\sin kR = 0$, and this means that the constant k takes quantized values

$$k_m = \frac{m\pi}{R}, \quad \text{for } m = 1, 2, 3, \dots \quad (2.242)$$

Note that $m = 0$ has no physical significance as it corresponds to the trivial solution of $n(\mathbf{r}, t) = 0$, which does not satisfy the initial condition. If we sum over all terms with different k_m , the full solution is

$$n(\mathbf{r}, t) = \sum_{m=1}^{\infty} A_m \frac{\sin k_m r}{r} e^{-Dk_m^2 t}, \quad (2.243)$$

where the constants A_m can be considered as the (arbitrary) scale factors for each of the k_m terms. Importantly, the radial functions $X_m(r) = r^{-1} \sin k_m r$ satisfy the following orthogonality condition

$$\begin{aligned} \int_0^R \frac{\sin k_n r}{r} \frac{\sin k_m r}{r} 4\pi r^2 dr &= 4\pi R \int_0^1 \sin(n\pi x) \sin(m\pi x) dx \\ &= 2\pi R \int_0^1 \{\cos[(m-n)\pi x] - \cos[(m+n)\pi x]\} dx \\ &= 2\pi R \delta_{mn}, \end{aligned} \quad (2.244)$$

where we have changed the variable $x = r/R$ and used the Dirac delta

$$\delta_{mn} = \begin{cases} 1, & \text{if } m = n, \\ 0, & \text{if } m \neq n. \end{cases} \quad (2.245)$$

The orthogonality condition is useful in determining the coefficients A_m when we apply the initial condition.

For an arbitrary initial condition $n(\mathbf{r}, t = 0) = n_0(r)$, we require

$$\sum_{m=1}^{\infty} A_m \frac{\sin k_m r}{r} = n_0(r). \quad (2.246)$$

Then, the coefficients A_m (for any $m = 1, 2, 3, \dots$) can be solved from the orthogonality condition

$$A_m = \frac{1}{2\pi R} \int_0^R n_0(r) \frac{\sin k_m r}{r} 4\pi r^2 dr. \quad (2.247)$$

Let us then consider the particular initial condition of a delta-function distribution of photons $n(\mathbf{r}, t) = \delta(\mathbf{r})$. We plug $n_0(r) = N_{\text{tot}}\delta(r)$ into eq. (2.247) and obtain

$$A_m = \frac{N_{\text{tot}}}{2\pi R} \int_0^R \delta(r) \frac{\sin k_m r}{r} 4\pi r^2 dr = \frac{k_m}{2\pi R} = \frac{mN_{\text{tot}}}{2R^2}, \quad \text{for } m = 1, 2, 3, \dots, \quad (2.248)$$

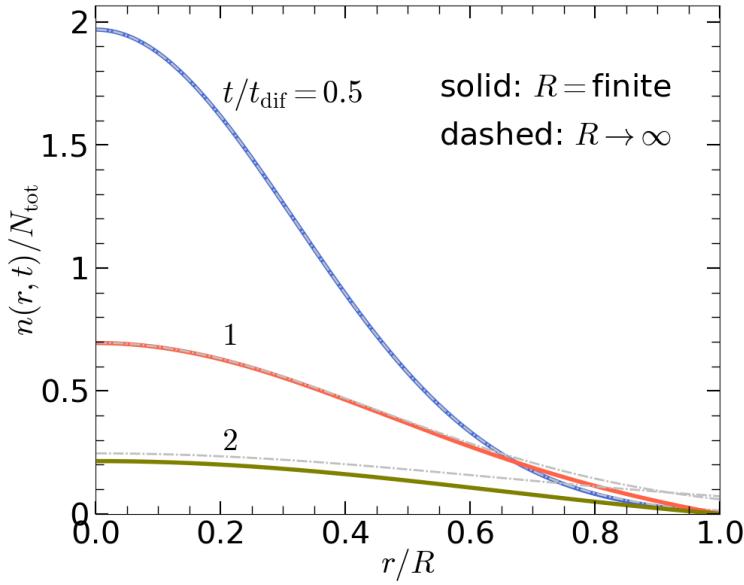


Figure 2.14: The solid lines show the distribution of photons (given by eq. 2.249) at different times in units of the diffusion time t_{dif} (eq. 2.250), for a finite spherical uniform medium. The silver dashed lines show the cases for an infinite medium ($R \rightarrow \infty$) at the same times and the photon positions are given by a Gaussian distribution (eq. 2.232).

where we have used the normalization $\int_0^R \delta(\mathbf{r}) 4\pi r^2 dr = 1$.

Finally, the time-dependent photon number density is given by

$$\begin{aligned} n(\mathbf{r}, t) &= \frac{N_{\text{tot}}}{2R^2} \sum_{m=1}^{\infty} \frac{m \sin(m\pi r/R)}{r} e^{-m^2 \pi^2 D t / R^2} \\ &= \frac{N_{\text{tot}}}{2R^3} \sum_{m=1}^{\infty} \frac{m \sin(m\pi x)}{x} e^{-m^2 t / t_{\text{dif}}}, \end{aligned} \quad (2.249)$$

where we have used the fractional radius $x = r/R$ and defined the diffusion time

$$t_{\text{dif}} \equiv R^2 / \pi^2 D = \frac{3R^2}{\pi^2 \ell_{\text{mfp}} c} = \frac{3\tau R}{\pi^2 c}, \quad (2.250)$$

and $\tau = R/\ell_{\text{mfp}}$ is the radial optical depth. The photon distribution at three different times $t/t_{\text{dif}} = 0.5, 1, 2$ are shown in Fig. 2.14.

It is possible to numerically check that the solution in eq. (2.249) is identical to the Gaussian distribution in eq. (2.232) in the limit $t \ll t_{\text{dif}}$, as can be seen in the $t/t_{\text{dif}} = 0.5$

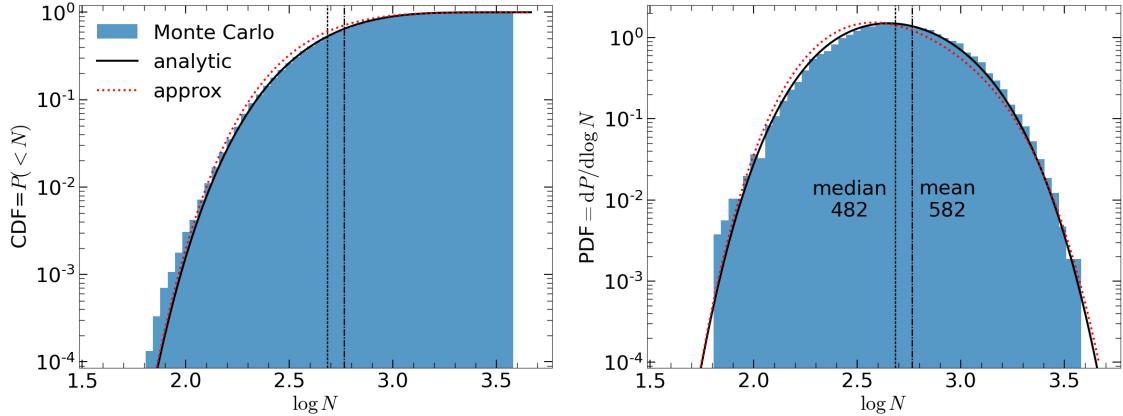


Figure 2.15: The histograms are from a Monte Carlo simulation of $N_{\text{tot}} = 3 \times 10^4$ photons scattering their ways through a spherical, uniform medium with radius $R = 1$ and scattering mean-free path $\ell_{\text{mfp}} = 0.03$. The black solid line shows the analytic solution of the diffusion equation (eq. 2.254) and the red dotted line shows the approximation in eq. (2.258).

case in Fig. 2.14. This is when nearly all photons are still confined in a region with radius much less than R , so photons are not aware of the edge of the spherical medium. However, at $t \gtrsim t_{\text{dif}}$, the solution in eq. (2.249) differs significantly from the Gaussian distribution because of the boundary condition $n(r = R, t) = 0$.

Next, we discuss the probability that an average photon escapes from the finite medium. The survival probability at a given time t is given by

$$\begin{aligned} P_{\text{surv}}(t) &= N_{\text{tot}}^{-1} \int_0^R n(\mathbf{r}, t) 4\pi r^2 dr \\ &= 2\pi \sum_{m=1}^{\infty} m e^{-m^2 t / t_{\text{dif}}} \int_0^1 \sin(m\pi x) x dx \\ &= 2 \sum_{m=1}^{\infty} (-1)^{m+1} e^{-m^2 t / t_{\text{dif}}}. \end{aligned} \quad (2.251)$$

At $t \ll t_{\text{dif}}$, the survival probability approaches unity¹⁷ $P(t \ll t_{\text{dif}}) \approx 1$; whereas at much later times $t \gg t_{\text{dif}}$, the $m = 1$ term dominates and we obtain $P(t \gg t_{\text{dif}}) \approx 2e^{-t/t_{\text{dif}}}$, which means that the fraction of trapped photons within the spherical medium decreases exponentially with time after one diffusion timescale.

Sometimes, we are interested in the number of scatterings N a photon undergo before escaping. This can be calculated with a straight-forward Monte Carlo simulation and the

¹⁷One can numerically verify that $\sum_{m=1}^{\infty} (-1)^{m+1} e^{-am^2} \approx 0.5$ when $0 < a \ll 1$.

results are shown by the histograms in Fig. 2.15 (where log means \log_{10}). In the following, we study the probability distribution for N in more detail.

Let us denote the probability distribution function (PDF) for the number of scatterings before escaping as dP/dN , and the corresponding cumulative distribution function (CDF) is denoted as $P(< N) = \int_0^N (dP/dN) dN$. There is a stochastic relation between the current time t since the release of the photons and the number of scatterings N up to the current time t . In the limit of $N \gg 1$, we expect $t \approx N\ell_{\text{mfp}}c$, because the mean-free time between adjacent scatterings is ℓ_{mfp}/c .

In the following, we take $t = N\ell_{\text{mfp}}/c$ for simplicity. In the limit of $\tau \gg 1$, the average number of scatterings before escaping is given by

$$\langle N \rangle \approx \tau^2/2 + \tau, \quad (2.252)$$

where the linear term τ is included to better account for the cases with the relatively small optical depths ($\tau \lesssim 10$). In our analytical derivation below, we take $\langle N \rangle \simeq \tau^2/2$, but we find our final result (eq. 2.258) to be in better agreement with that from the Monte Carlo simulation if we include the linear term in the end.

The ratio between the current time t and the diffusion timescale t_{dif} is given by

$$\frac{t}{t_{\text{dif}}} \approx \frac{\pi^2 N}{3\tau^2} \approx \frac{\pi^2}{6} \frac{N}{\langle N \rangle}. \quad (2.253)$$

The cumulative distribution for the number of scatterings is related to the $P_{\text{surv}}(t)$ by

$$\begin{aligned} P(< N) &\approx 1 - P_{\text{surv}}(t = N\ell_{\text{mfp}}/c) \\ &= 1 - 2 \sum_{m=1}^{\infty} (-1)^{m+1} \exp\left(-m^2 \frac{\pi^2 N}{3\tau^2}\right). \end{aligned} \quad (2.254)$$

We hope to find an approximate expression for $P(< N)$. If $t \gg t_{\text{dif}}$ or $N \gg \langle N \rangle$, as the $m = 1$ term dominates, we obtain $P_{\text{surv}}(t) \approx 2e^{-t/t_{\text{dif}}} \approx 2e^{-\pi^2 N/(6\langle N \rangle)}$ and this means

$$P(< N) \approx 1 - 2 \exp\left(-\frac{\pi^2 N}{6\langle N \rangle}\right) \Rightarrow \frac{dP}{dN} = \frac{\pi^2}{3\langle N \rangle} \exp\left(-\frac{\pi^2 N}{6\langle N \rangle}\right), \text{ for } N \gg \langle N \rangle. \quad (2.255)$$

The other limit of $t \ll t_{\text{dif}}$ or $N \ll \langle N \rangle$ is slightly more complicated because a large number of m terms contribute to the sum in $P_{\text{surv}}(t)$. However, it turns out that we can use the Gaussian distribution (eq. 2.232) for an infinite medium, and the escaping probability is approximately given by

$$\begin{aligned} P(< N) &\approx 1 - P_{\text{surv}}(t) \approx \frac{1}{(2\pi\sigma^2)^{3/2}} \int_R^\infty e^{-r^2/(2\sigma^2)} 4\pi r^2 dr \\ &\approx \sqrt{\frac{2}{\pi}} \frac{R}{\sigma} e^{-R^2/(2\sigma^2)} \approx \sqrt{\frac{6}{\pi}} \left(\frac{\langle N \rangle}{N}\right)^{1/2} \exp\left(-\frac{3\langle N \rangle}{2N}\right), \text{ for } N \ll \langle N \rangle, \end{aligned} \quad (2.256)$$

where $\sigma = \sqrt{2Dt} \approx R\sqrt{N/3\langle N \rangle}$ and we have approximated the integral in the limit of $R^2/\sigma^2 = 3\langle N \rangle/N \gg 1$. We then obtain the probability distribution by differentiating the cumulative distribution (and ignoring small terms)

$$\frac{dP}{dN} \approx \frac{3}{\langle N \rangle} \sqrt{\frac{3}{2\pi}} \left(\frac{\langle N \rangle}{N} \right)^{5/2} \exp \left(-\frac{3\langle N \rangle}{2N} \right), \text{ for } N \ll \langle N \rangle. \quad (2.257)$$

Finally, we combine the two limits of small and large N and write an approximate probability distribution function for the number of scatterings

$$\frac{dP}{dN} \approx \frac{1}{\langle N \rangle} \left[\frac{\pi^2}{3} + a x^{-5/2} \right] \exp \left(-\frac{\pi^2 x}{6} - \frac{3}{2x} \right), \quad x \equiv N/\langle N \rangle, \quad (2.258)$$

and $a = 9.114$ comes from the normalization $\int_0^\infty (dP/dN)dN = 1$ (error $< 10^{-4}$). As shown in Fig. 2.15, the approximation above (shown as a red dotted line) is in good agreement with the analytical solution (eq. 2.254, shown as a black solid line) of the diffusion equation and that from the Monte Carlo simulations (shown by the histogram). The agreement is better if we use the average number of scatterings given by eq. (2.252) including the linear term (instead of $\langle N \rangle \simeq \tau^2/2$). Note that the condition for eqs. (2.254) and (2.258) to apply is $\tau \gg 1$ or $\langle N \rangle \gg 1$.

2.6 Scattering and absorption

In this section, we consider the effects of scattering and absorption on radiative transfer. For simplicity, we assume that matter has an isotropic, thermal distribution function at temperature T (at least for the energy levels we are interested in), and the temperature may be spatially non-uniform. The medium has absorption coefficient α_ν defined as the inverse of absorption mfp and scattering coefficient σ_ν defined as the inverse of the scattering mfp, and these two coefficients α_ν and σ_ν may be spatially non-uniform. We further make simplifying assumptions that the scattering is coherent (photon frequency stays unchanged before and after scattering) and isotropic (the momentum direction of scattered photons are isotropically distributed). Under the above assumptions, the radiative transfer equation along a given direction \hat{n} is given by

$$\frac{dI_\nu}{ds} = \underbrace{-(\alpha_\nu + \sigma_\nu)I_\nu}_{\text{total attenuation}} + \underbrace{j_\nu + \sigma_\nu J_\nu}_{\text{effective emissivity}} = -(\alpha_\nu + \sigma_\nu)I_\nu + (\alpha_\nu B_\nu + \sigma_\nu J_\nu), \quad (2.259)$$

where $J_\nu = (4\pi)^{-1} \int d\Omega I_\nu$ is the mean intensity, in the second expression we have used the Kirchhoff's law $j_\nu = \alpha_\nu B_\nu$ for thermal matter, and the *effective emissivity* \tilde{j}_ν includes the spontaneous emission from matter j_ν and the scattered emission $\sigma_\nu J_\nu$, i.e.,

$$\tilde{j}_\nu = j_\nu + \sigma_\nu J_\nu = \alpha_\nu B_\nu + \sigma_\nu J_\nu. \quad (2.260)$$

In the following, we focus on the simplest case of a 1D slab geometry with matter density and temperature variations in the \hat{z} direction. Let z be the depth into the slab and

$$\mu \equiv \cos \theta$$

describes the angle between the z -axis and a given ray direction \hat{n} . Since s is the coordinate along a given ray (along \hat{n}), we write $ds = dz/\mu$, and hence

$$\frac{dI_\nu}{ds} = \mu \frac{dI_\nu}{dz}. \quad (2.261)$$

The two attenuation coefficients α_ν and σ_ν only depends on z but not on μ because matter has isotropic distribution function¹⁸. The geometry is shown in Fig. 2.16 where the surface of the medium is at $z = 0$ and the other boundary is at z_0 .

Let us change the integration variable from z to the total optical depth defined as

$$\tau = \int_0^z (\alpha_\nu + \sigma_\nu) dz', \quad (2.262)$$

with $\tau = 0$ on one end and $\tau = \tau_0$ on the other end of the slab. Here we do *not* follow the convention of defining the direction of increasing optical depth as along $-\hat{z}$ (later on, we will consider the slab to be externally illuminated). Note that $\mu > 0$ for rays entering the slab in the \hat{z} direction, and $\mu < 0$ for those emerging from the $z = 0$ surface in the $-\hat{z}$ direction. Then, eq. (2.259) can be written in the following form

$$\mu \partial_\tau I_\nu = -I_\nu + \tilde{S}_\nu, \quad (2.263)$$

where we have define the *effective source function* \tilde{S}_ν as

$$\tilde{S}_\nu = \frac{\tilde{j}_\nu}{\alpha_\nu + \sigma_\nu} = \frac{\alpha_\nu B_\nu + \sigma_\nu J_\nu}{\alpha_\nu + \sigma_\nu}, \quad (2.264)$$

which is isotropic (independent of μ).

2.6.1 Eddington approximation and effective optical depth

In the regions that are at least mildly optically thick (with total optical depth $\tau \gtrsim 1$), the radiation field is nearly isotropic, so we expect the angular dependence to be well described by the *Eddington approximation* (only retaining the monopole and dipole terms)

$$I_\nu(z, \mu) = a_\nu(z) + b_\nu(z)\mu. \quad (2.265)$$

¹⁸If matter has non-negligible motion in the lab frame, then our consideration of an isotropic distribution function is only possible in the comoving frame of matter's bulk motion. For this reason, our equation of radiative transfer (2.259) only applies in the matter's comoving frame.

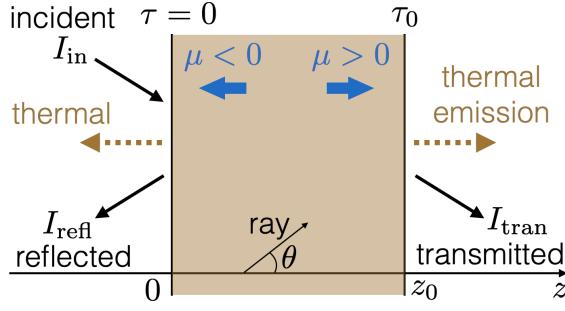


Figure 2.16: Geometry of the slab problem. Note that the direction of increasing optical depth is along $+\hat{z}$ (whereas in many textbooks it is along $-\hat{z}$, just a matter of definition).

The linear angular dependence is convenient in that it provides a closure to the moment equations in the following way

$$\begin{aligned} J_\nu &= \frac{1}{4\pi} \int I_\nu d\Omega = a_\nu, \\ F_\nu &= \int I_\nu \mu d\Omega = \frac{4\pi b_\nu}{3}, \\ K_\nu &= \int I_\nu \mu^2 d\Omega = \frac{4\pi a_\nu}{3} = \frac{4\pi J_\nu}{3}, \end{aligned} \quad (2.266)$$

where the 0th moment J_ν is the mean intensity, the 1st moment F_ν is the specific energy flux along the z -axis, and the 2nd moment K_ν is related to the specific radiation pressure along the z -axis $P_\nu = K_\nu/c$ contributed by the photons in the frequency interval of $(\nu, \nu + d\nu)$. Since the radiation energy density is given by $U_\nu = 4\pi J_\nu/c$, an important consequence of the Eddington approximation is

$$P_\nu = U_\nu/3. \quad (2.267)$$

Note that the above relation does *not* necessarily require an isotropic radiation field — a linear angular dependence as in eq. (2.265) also gives $P_\nu = U_\nu/3$.

The 0th and 1st moments of eq. (2.263) are given by

$$\begin{aligned} \int \mu \partial_\tau I_\nu d\Omega &= \partial_\tau F_\nu = 4\pi(-J_\nu + \tilde{S}_\nu), \\ \int \mu^2 \partial_\tau I_\nu d\Omega &= \partial_\tau K_\nu = -F_\nu = \frac{4\pi}{3} \partial_\tau J_\nu, \end{aligned} \quad (2.268)$$

where we have used $K_\nu = 4\pi J_\nu/3$. The second equation above shows that the radiative flux is proportional to the gradient of the radiation energy density $U_\nu = 4\pi J_\nu/c$,

$$F_\nu = -\frac{c}{3} \partial_\tau U_\nu = -\frac{\ell_{\text{mfp},\nu} c}{3} \nabla U_\nu, \quad (\text{Eddington approximation}) \quad (2.269)$$

which is in agreement with Fick's law (eq. 2.206) as long as we define the mean free path $\ell_{\text{mfp},\nu} = (\alpha_\nu + \sigma_\nu)^{-1}$ based on the total attenuation coefficient.

Combining the two equations in (2.268), we arrive at the following 2nd-order ordinary differential equation (ODE)

$$\frac{1}{3}\partial_\tau^2 J_\nu = J_\nu - \tilde{S}_\nu = \epsilon_\nu(J_\nu - B_\nu), \quad \epsilon_\nu \equiv \frac{\alpha_\nu}{\alpha_\nu + \sigma_\nu}, \quad (2.270)$$

where $1 - \epsilon_\nu$ is the *single-scattering albedo*. The goal is to solve the above equation for the mean intensity $J_\nu(\tau, \mu)$ under reasonable boundary conditions. Before doing that, it is useful to define an *effective optical depth*

$$\tau_{\text{eff}} = \int_0^z \sqrt{3\alpha_\nu(\alpha_\nu + \sigma_\nu)} dz'. \quad (2.271)$$

Then the 2nd-order ODE above can be re-written as

$$\partial_{\tau_{\text{eff}}}^2 J_\nu = J_\nu - B_\nu. \quad (2.272)$$

The effective optical depth is related to the absorption optical depth $\tau_a = \int_0^z \alpha_\nu dz'$ and scattering optical depth $\tau_s = \int_0^z \sigma_\nu dz'$. Often, the following approximations work reasonably well

$$\tau_{\text{eff}} \approx \sqrt{3\epsilon_\nu} \tau \approx \sqrt{3\tau_a(\tau_a + \tau_s)}, \quad (2.273)$$

and the approximations become equal for a uniform medium where α_ν and σ_ν are both constant. Hereafter, we omit the frequency dependences of τ_{eff} , τ_a , and τ_s for brevity.

Eq. (2.272) has exponentially growing or decaying solutions — note the sign difference from the driven harmonic oscillator equation of the form $d^2y/dx^2 + y = f(x)$. At very large optical depths, only the exponentially decaying solution of $J_\nu - B_\nu \propto e^{-\tau_{\text{eff}}}$ is physically possible. Thus, the meaning of the effective optical depth is that, if $\tau_{\text{eff}} \gtrsim 1$ near a given frequency ν , the radiation field *near that frequency* would be “thermalized” such that $J_\nu \approx B_\nu$, because photons near frequency ν will be absorbed before escaping and the energy is re-emitted according to the matter temperature T (and hence the shape of the initial spectrum near frequency ν is lost). This is why τ_{eff} is also called the *thermalization optical depth*.

The optical depth for absorption and re-emission can be easily understood in the absorption-dominated limit $\tau_a \gg \tau_s$. In the opposite, scattering-dominated limit $\tau_s \gg \tau_a$, the total path length is enhanced by a factor of τ_s due to the large number of scatterings (provided that $\tau_s \gg 1$), so the probability of a photon being absorbed while passing through a region of effective optical depth τ_{eff} is roughly given by $\simeq \min(1, \tau_s \tau_a)$ in the limit of $\tau_s \gg \tau_a$. If we put these two regimes together, the absorption probability is given by

$$P_{\text{abs}} \simeq \min(1, \tau_a + \tau_s \tau_a). \quad (2.274)$$

It should be noted that $P_{\text{abs}} \neq \tau_{\text{eff}}$ in general.

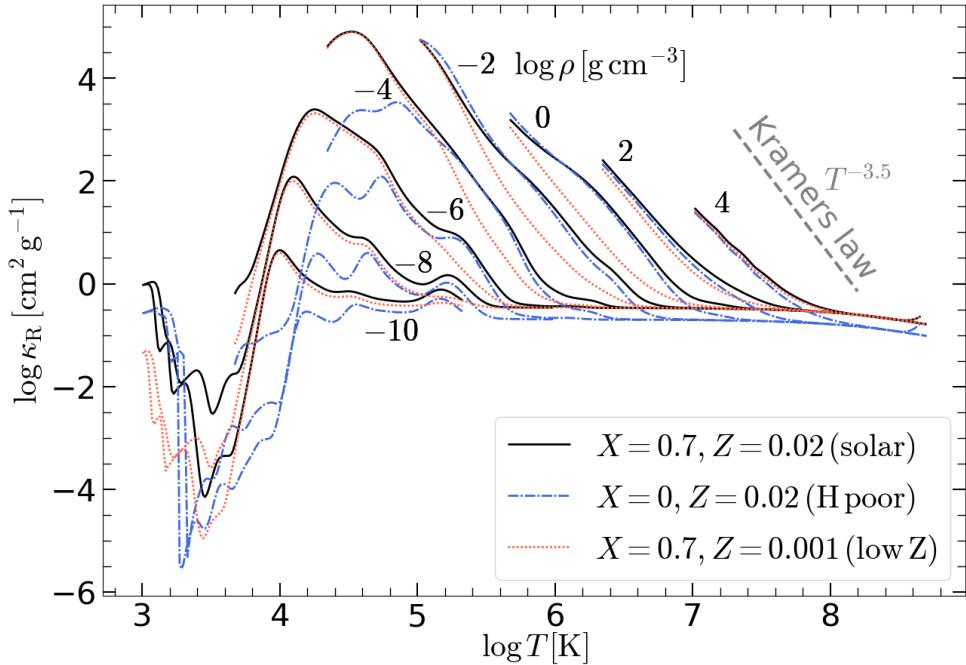


Figure 2.17: Rosseland-mean opacity for three different gas compositions in LTE.

2.6.2 Rosseland-mean opacity and Planck-mean opacity

Based on the above arguments, even without solving the diffusion equation, we know that at depth $\tau_{\text{eff}} \gg 1$, the radiation field is well thermalized due to frequent absorption and re-emission. If the effective optical depth is much greater than unity *at all frequencies*, the mean intensity should approach the blackbody function, i.e.,

$$J_\nu \approx B_\nu, \text{ if } \tau_{\text{eff}} \gg 1 \text{ at all frequencies,} \quad (2.275)$$

which is called the *Rosseland approximation* and is equivalent to LTE. The LTE condition is satisfied in the interior stars, so we can calculate the rate radiative energy transport under the Rosseland approximation and the Rosseland-mean opacity is all we need.

From the second equation in (2.268), we can calculate the specific flux

$$F_\nu(z) \approx -\frac{4\pi}{3}\partial_\tau B_\nu(T(z)) = -\frac{4\pi}{3(\alpha_\nu + \sigma_\nu)}\frac{\partial B_\nu(T)}{\partial T}\frac{\partial T}{\partial z}, \quad (2.276)$$

where we have used $d\tau = (\alpha_\nu + \sigma_\nu)dz$. The total radiative flux is given by

$$F(z) \approx -\frac{4\pi}{3}\frac{\partial T}{\partial z} \int_0^\infty (\alpha_\nu + \sigma_\nu)^{-1} \frac{\partial B_\nu(T)}{\partial T} d\nu. \quad (2.277)$$

For matter in LTE at given density ρ and temperature T , one can numerically calculate the frequency dependent opacities $\kappa_{a,\nu} = \alpha_\nu/\rho$ (absorption) and $\kappa_{s,\nu} = \sigma_\nu/\rho$ (scattering), and then obtain the *Rosseland-mean opacity*

$$\kappa_R^{-1} \equiv \frac{\int d\nu (\kappa_{a,\nu} + \kappa_{s,\nu})^{-1} \partial_T B_\nu}{\int d\nu \partial_T B_\nu}, \quad (2.278)$$

where the denominator is given by $\int d\nu \partial_T B_\nu = acT^3/\pi$ (a = radiation density constant). Then we obtain

$$F(z) \approx -\frac{4acT^3}{3\rho\kappa_R} \partial_z T. \text{ (Rosseland approximation)} \quad (2.279)$$

The above flux can be written in the following form

$$\mathbf{F} = -\frac{c\ell_{\text{mfp},R}}{3} \nabla U, \text{ (Rosseland approximation)} \quad (2.280)$$

where $U = aT^4$ is the radiation energy density, and $\ell_{\text{mfp},R} \equiv (\rho\kappa_R)^{-1}$ is the effective mean free path. We see that, under the Rosseland approximation, the radiative flux is given by the Fick's law of diffusion — the same as the case of isotropic scattering but with a different mfp.

Table 2.1: Conditions for two different forms of Fick's law.

condition	$\tau_a + \tau_s \gtrsim 1$	$\tau_{\text{eff}} \simeq \sqrt{\tau_a(\tau_a + \tau_s)} \gtrsim 1$
approximation	Eddington ($I_\nu \approx a_\nu + b_\nu \mu$)	Rosseland ($J_\nu \approx B_\nu$)
outcome	$\mathbf{F}_\nu = -(c\ell_{\text{mfp},\nu}/3) \nabla U_\nu$	$\mathbf{F} = -(c\ell_{\text{mfp},R}/3) \nabla U$

In highly optically thick regions where the radiation-matter mixture is in LTE, the Rosseland-mean opacity table can be pre-computed for a wide range of gas densities and temperatures. This makes Eq. (2.280) very powerful when one wants to compute the frequency-integrated radiative flux flowing from one region to its neighboring regions. Fig. 2.17 shows the Rosseland-mean opacity table for the LTE mixture of gas and radiation. For the case of solar metallicity (solid black lines), the mass fractions of hydrogen, helium and metals are taken to be $X = 0.7$, $Y = 0.28$, and $Z = 0.02$, respectively. For the hydrogen-poor case (dash-dotted blue lines), the mass fractions of helium and metals are $Y = 0.98$ and $Z = 0.02$. For the low-metallicity case (dotted red lines), the mass fractions are $X = 0.7$, $Y = 0.299$, and $Z = 0.001$. The data is taken from the high-temperature $T \gtrsim 10^4$ K table from OPAL ([Iglesias & Rogers 1996](#)) and low-temperature ($10^3 \lesssim T \lesssim 10^4$ K) table by [Ferguson et al. \(2005\)](#), which includes molecules and dust grains.

Another useful frequency-integrated opacity is the *Planck-mean* defined as follows

$$\kappa_P \equiv \frac{\int B_\nu(T) \kappa_\nu d\nu}{\int B_\nu d\nu} = \frac{\int B_\nu \kappa_\nu d\nu}{\sigma_{\text{SB}} T^4 / \pi}, \quad (2.281)$$

where the Planck function is evaluated at the *gas temperature* T . Like the Rosseland-mean, when the Planck-mean opacity is used, usually the underlying assumption is that the gas and radiation are in LTE such that gas temperature = radiation temperature and that the radiation spectrum $U_\nu \propto B_\nu(T)$. However, if LTE condition does not hold (e.g., in optically thin regions), one can still define the Planck-mean opacity if the radiation spectrum has a blackbody shape $U_\nu \propto B_\nu(T_r)$ at the radiation temperature T_r . In this case, since the opacity function $\kappa_\nu(\rho, T_g)$ depends on the gas density ρ and gas temperature T_g , the resulting Planck-mean opacity $\kappa_P(\rho, T_g, T_r)$ would depend on all three quantities — this becomes very complicated.

We note that, as compared to the Rosseland-mean, the Planck-mean is more heavily weighted at frequencies where the opacity κ_ν is higher, so we expect κ_P to be (usually substantially) larger than the κ_R for the same gas and radiation conditions.

2.6.3 Eddington luminosity

The Eddington luminosity is defined such that the radiation force balances the gravitational force on a given fluid element. The concept of the Eddington luminosity is important in both optically thin and optically thick regions.

Let us first consider an optically thin gas with frequency-dependent total (absorption + scattering) opacity κ_ν . Each absorbed/scattered¹⁹ photon of energy ϵ would deposit a momentum ϵ/c into the gas. If the gas is at a distance r from an isotropically emitting source of specific luminosity L_ν , the radiation force per unit mass on the gas is

$$g_{\text{rad}} = \frac{\int L_\nu \kappa_\nu d\nu}{4\pi r^2 c}. \quad (2.282)$$

Suppose the gas is near a massive body of mass M , then the gravitational acceleration is $g = GM/r^2$. These two forces equal when the source has the *Eddington luminosity*

$$L_{\text{Edd}} = \frac{4\pi G M c}{\kappa_E}, \quad (2.283)$$

where the *energy-averaged opacity* is defined as

$$\kappa_E = \frac{\int L_\nu \kappa_\nu d\nu}{\int L_\nu d\nu}. \quad (2.284)$$

¹⁹Here we assume that the gas scatters the incoming radiation in an isotropic (or at least forward-backward symmetric) manner.

We do not discuss the detailed calculation of κ_E , which is non-trivial as it depends on the gas properties and the radiation spectrum in the problem in hand.

On the other hand, in highly optically thick regions where the radiation field and gas are in LTE, the radiative flux can be obtained from the Rosseland approximation (eq. 2.280) $\mathbf{F} = -c/(3\rho\kappa_R)\nabla U$, where κ_R is the Rosseland-mean opacity (eq. 2.278). Associated with the radiative flux, the gas experiences a *radiation force density* \mathbf{G} that can be obtained by taking the 1st moment of the radiative transfer equation (see eq. 2.340)

$$\mathbf{G} = \rho\kappa_R\mathbf{F}/c, \text{ (under LTE)} \quad (2.285)$$

and the corresponding acceleration is given by

$$\mathbf{g}_{\text{rad}} = \mathbf{G}/\rho = \kappa_R\mathbf{F}/c. \quad (2.286)$$

If \mathbf{g}_{rad} is in the outward radial direction, when it is in balance with the gravitational acceleration $g = GM/r^2$, one obtains a critical flux $F_c = GMc/(\kappa_R r^2)$, and it also corresponds to the Eddington luminosity (for a spherically symmetric system)

$$L_{\text{Edd}} = 4\pi r^2 F_c = \frac{4\pi GMc}{\kappa_R}. \quad (2.287)$$

We see that the Eddington luminosities in the optically thin and optically thick cases are generally different due to the difference in the frequency-averaged opacities (κ_E and κ_R respectively).

If opacity is mainly due to Thomson scattering by free electrons in a fully ionized gas, then we have a gray opacity

$$\kappa_T = \left(X + \frac{1-X}{2} \right) \frac{\sigma_T}{m_p} \approx 0.2(1+X) \text{ cm}^2 \text{ g}^{-1}, \quad (2.288)$$

where X is the mass fraction of hydrogen, $1-X$ is the mass fraction of helium, and σ_T is the Thomson cross-section (see Chapter 4). In this case, the Eddington luminosity corresponding to typical cosmic hydrogen abundance of $X = 0.7$ is given by

$$L_{\text{Edd,T}} \approx \frac{8\pi GMc}{(1+X)\sigma_T/m_p} = 1.48 \times 10^{38} \text{ erg s}^{-1} \frac{1.7}{1+X} \frac{M}{M_\odot}. \quad (2.289)$$

2.6.4 *Gravity darkening (under construction)

2.6.5 *Limb darkening for grey atmosphere

Let us consider an atmosphere extending from $\tau = 0$ to very large optical depth $\tau_0 \gg 1$, throughout which the radiative flux $-F$ (outward going) is a constant. The total opacity is assumed to be independent of frequency (“grey”) $\kappa_\nu = \kappa$.

Plugging the constant flux into the second equation of (2.268), we obtain the frequency-integrated mean intensity as a function of optical depth

$$J = \frac{3}{4\pi} F\tau + J_0, \quad (2.290)$$

where $J_0 = J(\tau = 0)$ is an integration constant to be determined using the boundary conditions.

Since $\partial_\tau F = 0$, from the first equation of (2.268), we know that the frequency-integrated source function $\tilde{S} = J$.

If the source function $\tilde{S}_\nu(\tau)$ is known, then we can use the following formal solution²⁰ (similar to eq. 2.117)

$$I_\nu(\tau, \mu) e^{\tau/\mu} = I_\nu(0, \mu) + \int_0^\tau \tilde{S}_\nu(\tau') e^{\tau'/\mu} d\tau' / \mu = I_\nu(\tau_0, \mu) e^{\tau_0/\mu} - \int_\tau^{\tau_0} \tilde{S}_\nu(\tau') e^{\tau'/\mu} d\tau' / \mu, \quad (2.291)$$

where the two expressions use different boundary conditions at $\tau = 0$ and τ_0 respectively. Unfortunately, in most cases, it is non-trivial to solve for $S_\nu(\tau)$ self-consistently given the boundary conditions.

The intensity emerging from the surface at $\tau = 0$ in the directions $\mu < 0$ is given by

$$\begin{aligned} I(\tau = 0, \mu < 0) &= I(\tau_0, \mu) e^{\tau_0/\mu} - \int_0^{\tau_0} \tilde{S}(\tau') e^{\tau'/\mu} d\tau' / \mu \\ &= \frac{1}{|\mu|} \int_0^\infty \tilde{S}(\tau') e^{-\tau'/|\mu|} d\tau', \\ &= \frac{3}{4\pi} F |\mu| + J_0, \end{aligned} \quad (2.292)$$

where we have taken the limit $\tau_0 \rightarrow \infty$. Then, since the total emerging flux is given by $2\pi \int_{-1}^0 \mu I(\tau = 0) d\mu = F$, we obtain $J_0 = F/(2\pi)$ and hence

$$J = \frac{3}{4\pi} F \left(\tau + \frac{2}{3} \right). \quad (2.293)$$

The effective temperature is defined based on the energy flux $F = \sigma_{\text{SB}} T_{\text{eff}}^4$. Under LTE, the source function $\tilde{S} = J$ equals to $B = \int B_\nu d\nu = \sigma_{\text{SB}} T^4 / \pi$, so we obtain the temperature profile

$$T^4 = \frac{3}{4} T_{\text{eff}}^4 \left(\tau + \frac{2}{3} \right). \quad (2.294)$$

In the above solution, since $T(\tau = 2/3) = T_{\text{eff}}$, it is often said that the *photosphere* of a star is located at $\tau = 2/3$.

²⁰The procedure is to multiply both sides of eq. (2.263) by $e^{\tau/\mu}$ and then integrate by parts.

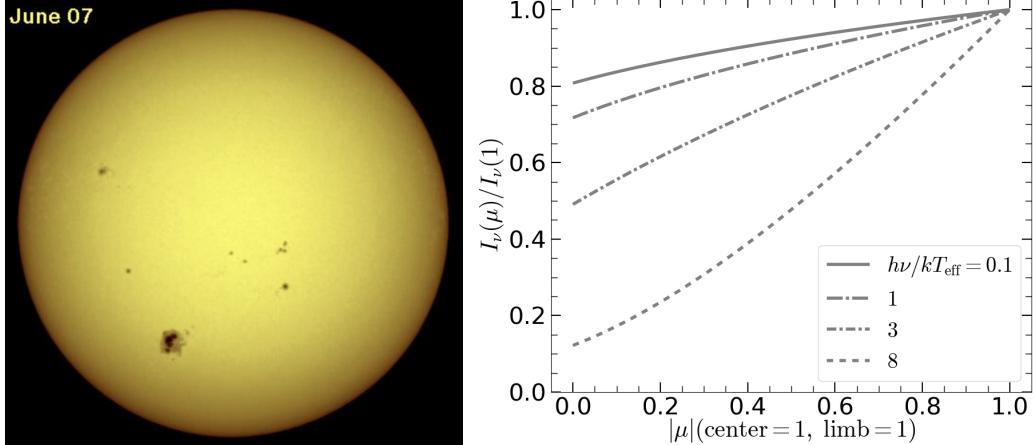


Figure 2.18: Left panel: visible image of the Sun taken by NASA. We see deeper (hotter) layers near the center and shallower (cooler) layers near the limb. Right panel: Limb darkening at different frequencies.

Then, the angle-dependent intensity at the surface is given by

$$I(\tau = 0, \mu < 0) = \frac{3F}{4\pi} \left(|\mu| + \frac{2}{3} \right). \quad (2.295)$$

The intensity ratio between surface normal ($\mu = -1$, center of a spatially resolved star's disk) and tangential ($\mu = 0$, limb of the disk) directions is

$$\frac{I(\tau = 0, \mu = -1)}{I(\tau = 0, \mu = 0)} = 2.5. \quad (2.296)$$

This is the effect of *limb darkening*. This agrees reasonably well with observations of the Sun in the optical continuum (see Fig. 2.18), where the opacity is roughly grey.

The essential reason for limb darkening is that, closer to the limb, the $\tau \sim 1$ region along a given line of sight is located at a shallower depth (in z coordinate) from the stellar surface where the temperature is lower than the deeper interior. The opacity is a strong function of frequency near strong spectral lines (e.g., Balmer lines for the Sun): the opacity is higher near the line center and hence at those wavelengths the observer sees a shallower depth in the atmosphere where the temperature is lower than the position of $\tau \sim 1$ for the continuum. This leads to absorption lines in stellar spectra (Fig. 2.19).

With the temperature profile (eq. 2.294) in hand, we assume LTE so the source function is given by $B_\nu(T)$ at all depths. Thus, the frequency-dependent outward going intensity at the surface is

$$I_\nu(\tau = 0, \mu < 0) = \frac{1}{|\mu|} \int_0^\infty B_\nu(T(\tau)) e^{-\tau/|\mu|} d\tau, \quad (2.297)$$

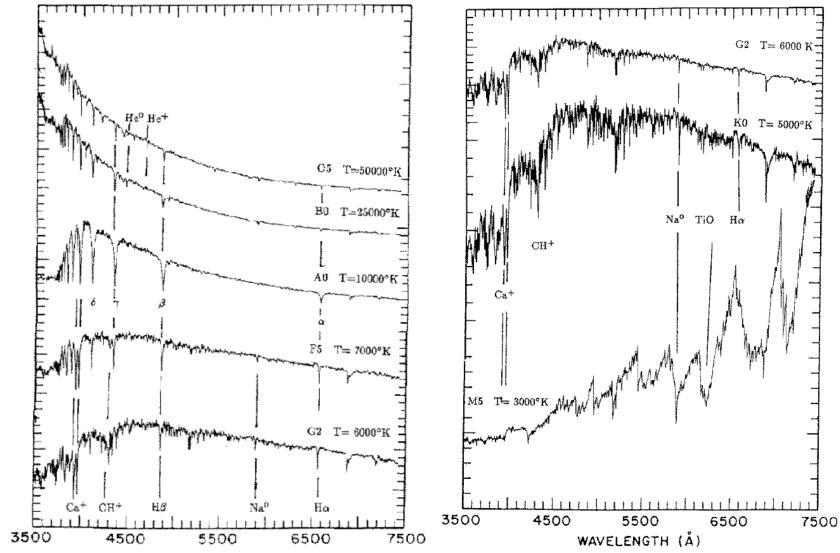


Figure 2.19: Spectra of main-sequence stars of different types from O (hottest) to M (coolest). A common feature is that these spectra all show absorption lines. Note that the spectra of M-type stars show broad absorption bands (due to molecular transitions) that render the unabsorbed parts to like emission features.

which can be written in the following dimensionless way

$$\frac{I_\nu(\tau = 0, \mu < 0)}{B_\nu(T_{\text{eff}})} = \int_0^\infty \frac{e^x - 1}{e^{x[(3t|\mu|+2)/4]^{-1/4}} - 1} e^{-t} dt, \quad (2.298)$$

where $t = \tau/|\mu|$ and $x = h\nu/(k_B T_{\text{eff}})$. The result is that limb darkening is generally stronger at higher frequencies, as shown by the right panel of Fig. 2.18.

However, the Eddington approximation is not suppose to be very accurate at $\tau \lesssim 1$. Another solution can be obtained based on the two-stream approximation at the $\tau = 0$ boundary (see eq. 2.307). Using $I^+(\tau = 0) = 0$ (no incident intensity), we obtain

$$\partial_\tau J(\tau = 0) = \frac{3}{4\pi} F = \sqrt{3} J(\tau = 0) = \sqrt{3} J_0, \quad (2.299)$$

where we have used $J(\tau)$ obtained in eq. (2.290). This gives $J_0 = \sqrt{3}F/(4\pi)$ and hence

$$J = \frac{3}{4\pi} F \left(\tau + \frac{1}{\sqrt{3}} \right). \quad (2.300)$$

Then, the full intensity is given by

$$I(\tau, \mu) = \frac{3F}{4\pi} \left(\tau + \frac{1}{\sqrt{3}} - \mu \right). \quad (2.301)$$

At the surface ($\tau = 0$), the emerging intensity ratio between the center and limb of the star's disk is

$$\frac{I(\tau = 0, \mu = -1)}{I(\tau = 0, \mu = 0)} = 1 + \sqrt{3} \simeq 2.7. \quad (2.302)$$

2.6.6 *Uniform slab with two-stream boundary conditions

In this subsection, we discuss radiative transfer through a uniform slab with constant absorption and scattering coefficients (see Fig. 2.16). Such an idealized case turns out to be highly useful in understanding radiative transfer in optically thick medium in general. The questions we are interested are: What is the emerging flux from the slab's own thermal emission? What fraction of the incident flux is reflected/transmitted? The answers to these questions depend on two parameters: the ratio between the absorption coefficient and the total attenuation (absorption plus scattering) coefficient $\epsilon_\nu = \alpha_\nu / (\alpha + \sigma_\nu)$ and the total (absorption plus scattering) optical depth τ_0 . In terms of these two parameters, the effective optical depth of the entire slab is given by $\tau_{\text{eff},0} = \sqrt{3\epsilon_\nu}\tau_0$. Before going into the details, we would like to have an intuitive, order-of-magnitude understanding of the solutions in different cases.

- If the entire slab has effective optical depth $\tau_{\text{eff},0} \gg 1$, any incident flux from the left boundary must be absorbed before reaching the right boundary. The emerging flux from the right boundary can only come from the slab's thermal emission within an emitting region of thickness $\tau_{\text{eff}} \sim 1$ from the surface (the emission from deeper regions will be absorbed before escaping). There are two limits.
 - (i) If the opacity is dominated by scattering $\epsilon_\nu \ll 1$, then the emitting region with $\tau_{\text{eff}} \sim 1$ is still highly optically thick to scattering, because $\tau_s \sim \epsilon_\nu^{-1/2}\tau_{\text{eff}} \sim \epsilon_\nu^{-1/2} \gg 1$ whereas the absorption optical depth $\tau_a \sim \epsilon_\nu^{1/2} \ll 1$. From Fick's law, we estimate the emerging flux from the right boundary to be $F_\nu \sim c\ell_{\text{mfp},s}|dU_\nu/dz| \sim \pi B_\nu/\tau_s \sim \pi\epsilon_\nu^{1/2}B_\nu$. At the left boundary, the outward-going flux is roughly given by $F_\nu(\text{in}) + \pi\epsilon_\nu^{1/2}B_\nu$ because nearly all the incident flux is reflected by the scattering dominated region of thickness $\tau_s \sim \epsilon_\nu^{-1/2}$.
 - (ii) If the opacity is dominated by absorption $\epsilon_\nu \approx 1$, then the emitting region with $\tau_{\text{eff}} \sim 1$ is also the region with $\tau_a \sim 1$ whereas the scattering optical depth is negligible $\tau_s \ll 1$. In this case, the emerging fluxes from both boundaries are roughly given by $F_\nu \sim \pi B_\nu$. The entire slab is effectively a blackbody which absorbs nearly all the incident flux on the left boundary without significant reflection.
- If the entire slab has effective optical depth $\tau_{\text{eff},0} \ll 1$, then absorption is unimportant everywhere and the only relevant processes are scattering and thermal emission. There are again two limits.

- (i) If the slab is optically thin to scattering $\tau_{s,0} \ll 1$, then the emerging flux on the right boundary is given by $F_\nu \sim F_\nu(\text{in}) + \pi\tau_{a,0}B_\nu$, i.e., the incident flux plus thermal emission along the entire slab (note that $\tau_{a,0} \ll 1$). The outward-going flux at the left boundary has a thermal component of order $\pi\tau_{a,0}B_\nu$ plus a small reflected component of the order $\tau_{s,0}F_\nu(\text{in})$.
- (ii) If the slab is optically thick to scattering $\tau_{s,0} \gg 1$ (which is possible if the slab is scattering dominated $\epsilon_\nu \ll 1$), then the reduced diffusion speed $v_{\text{dif}} \sim c/\tau_{s,0}$ means that the emerging flux on the right boundary is $F_\nu \sim F_\nu(\text{in})/\tau_{s,0} + \pi\tau_{a,0}B_\nu$. In this case, the transmitted fraction of the incident light is $\tau_{s,0}^{-1}$ whereas the vast majority is reflected near the left boundary. The thermal flux contribution of $\pi\tau_{a,0}B_\nu$ has the interesting behavior that it only depends on the total absorption optical depth $\tau_{a,0}$ (independent of the scattering optical depth $\tau_{s,0}$). When the slab is optical thick to scattering $\tau_{s,0} \gg 1$, the total path length enhances emitted flux by a factor of $\tau_{s,0}$ as compared to $\pi\tau_{a,0}B_\nu$. However, due to photon diffusion, the transmitted flux will be suppressed by a factor of $\tau_{s,0}^{-1}$. These two factors $\tau_{s,0}$ and $\tau_{s,0}^{-1}$ cancel with each other.

To summarize the above results, the emerging flux on the right boundary ($\tau = \tau_0$) is

$$F_\nu(\tau_0) \sim \begin{cases} \pi B_\nu / (1 + \epsilon_\nu^{-1/2}), & \text{for } \tau_{\text{eff},0} \gg 1, \\ F_\nu(\text{in}) / (1 + \tau_{s,0}) + \pi\tau_{a,0}B_\nu, & \text{for } \tau_{\text{eff},0} \ll 1. \end{cases} \quad (2.303)$$

The outcomes in these different regimes are summarized in Table 2.2.

Table 2.2: The outgoing flux at the right boundary of a uniform thermal slab. The reflected and transmitted fractions of the incident light (on the left boundary) are denoted as f_{ref} and f_{tran} , respectively. The flux of the thermal emission from the slab are expressed as a fraction of the blackbody flux $\pi B_\nu(T)$ at the matter temperature T .

α_ν/σ_ν	$\tau_{\text{eff},0}$	τ_0	f_{ref}	f_{tran}	thermal flux (πB_ν)
$\gg 1$	$\gg 1$	$\gg 1$	0	0	1
$\gg 1$	$\ll 1$	$\ll 1$	$\tau_{s,0} (\ll 1)$	1	$\tau_0 \approx \tau_{a,0} (\ll 1)$
$\ll 1$	$\gg 1$	$\gg 1$	1	0	$(\alpha_\nu/\sigma_\nu)^{1/2} (\ll 1)$
$\ll 1$	$\ll 1$	$\gg 1$	1	$\tau_0^{-1} \approx \tau_{s,0}^{-1} (\ll 1)$	$\tau_{a,0} (\ll 1)$
$\ll 1$	$\ll 1$	$\ll 1$	$\tau_0 \approx \tau_{s,0} (\ll 1)$	1	$\tau_{a,0} (\ll 1)$

In the following, we present an analytical solution to the diffusion equation (2.270), which is the general form of an inhomogeneous second-order ordinary differential equation

$$y'' + p(\tau)y' + q(\tau)y = f(\tau), \quad (2.304)$$

where $y = J_\nu$, the derivatives are taken wrt. τ , $p(\tau) = 0$, $q(\tau) = -3\epsilon_\nu$, and $f(\tau) = -3\epsilon_\nu B_\nu$. There are many methods designed to solve such an equation, including the cases where $f(\tau)$ is a linear combination of polynomial, exponential and/or sine functions.

The general solution can be written as

$$y(\tau) = c_1 y_1(\tau) + c_2 y_2(\tau) + Y_p(\tau), \quad (2.305)$$

where y_1 and y_2 are solutions to the homogeneous equation (for $f = 0$) and $Y_p(\tau)$ is one particular solution of the full inhomogeneous solution. The two constants c_1 and c_2 are determined by the boundary conditions $y(\tau = 0)$ and $y'(\tau = 0)$. Obtaining analytic forms of y_1 , y_2 and Y_p are only possible for some simple cases.

We will consider a uniform slab where ϵ_ν and B_ν are both constants. Such a case is reasonable for an isothermal atmosphere. In this case, one particular solution to eq. (2.270) or eq. (2.304) is simply $y = J_\nu = B_\nu$. Then, we only need to find solutions to the homogeneous equation $\partial_\tau^2 J_\nu - 3\epsilon_\nu J_\nu = 0$. It is easy to see that the solutions must be exponential $J_\nu^\pm = \exp(\pm\sqrt{3\epsilon_\nu}\tau)$. Thus, the general solution has the following form

$$J_\nu = c_1 \exp(\sqrt{3\epsilon_\nu}\tau) + c_2 \exp(-\sqrt{3\epsilon_\nu}\tau) + B_\nu. \quad (2.306)$$

In the following, we consider one particular set of boundary conditions obtained under the *two-stream approximation*.

For the moment, let us consider that, instead of the full linear angular dependence as in eq. (2.265), there are only two beams at angles $\mu = \pm 1/\sqrt{3}$: the beam going in the $+z$ direction has intensity $I_\nu^+ = I_\nu(\tau, \mu = 1/\sqrt{3})$ and the other beam has $I_\nu^- = I_\nu(\tau, \mu = -1/\sqrt{3})$. In this case, the mean intensity and next flux are given by (respectively)

$$J_\nu = (I_\nu^+ + I_\nu^-)/2, \quad F_\nu = 2\pi(I_\nu^+ - I_\nu^-)/\sqrt{3}.$$

These two particular angles are chosen because they reproduce the key result of the Eddington approximation $P_\nu = U_\nu/3$, since $P_\nu = 2\pi(I_\nu^+ + I_\nu^-)/(3c) = 4\pi J_\nu/(3c) = U_\nu/3$.

Since $F_\nu = -4\pi\partial_\tau J_\nu/3$, we can solve for the two intensities in terms of J_ν and $\partial_\tau J_\nu$,

$$I_\nu^+ = J_\nu - \frac{1}{\sqrt{3}}\partial_\tau J_\nu, \quad I_\nu^- = J_\nu + \frac{1}{\sqrt{3}}\partial_\tau J_\nu. \quad (2.307)$$

Case I — semi-infinite uniform slab without incident radiation

Consider a slab that extends from $\tau = 0$ (left boundary) and $\tau \rightarrow \infty$. In this case, we must have $c_1 = 0$ so that J_ν remains finite. We consider no incident radiation on the left boundary, i.e. $I^+(\tau = 0) = 0$, which means that

$$\partial_\tau J_\nu(\tau = 0) = \sqrt{3}J_\nu(\tau = 0). \quad (2.308)$$

This gives

$$c_2 = -\frac{B_\nu}{1 + \sqrt{\epsilon_\nu}}. \quad (2.309)$$

The final solution is

$$J_\nu = B_\nu \left[1 - \frac{\exp(-\sqrt{3\epsilon_\nu}\tau)}{1 + \sqrt{\epsilon_\nu}} \right]. \quad (2.310)$$

This gives the flux

$$F_\nu = -\frac{4\pi}{3} \partial_\tau J_\nu = -\frac{4\pi B_\nu}{\sqrt{3}} \frac{\sqrt{\epsilon_\nu}}{1 + \sqrt{\epsilon_\nu}} \exp(-\sqrt{3\epsilon_\nu}\tau), \quad (2.311)$$

where the negative sign means the net flux is propagating in the $-z$ direction. Thus, the full intensity under the Eddington approximation (note that the two-stream approximation was only used to set the boundary conditions) is given by

$$I_\nu(\tau, \mu) = J_\nu + \frac{3F_\nu\mu}{4\pi} = B_\nu \left[1 - \frac{\exp(-\sqrt{3\epsilon_\nu}\tau)}{1 + \sqrt{\epsilon_\nu}} (1 + \sqrt{3\epsilon_\nu}\mu) \right]. \quad (2.312)$$

It can be seen that the intensity is nearly isotropic and approaches the blackbody value at high effective optical depths $\tau_{\text{eff}} \gg 1$. The emerging flux at the left boundary of the slab is given by

$$|F_\nu(\tau = 0)| = \frac{4\pi}{\sqrt{3}} \frac{\sqrt{\epsilon_\nu}}{1 + \sqrt{\epsilon_\nu}} B_\nu, \text{ for semi-infinite slab.} \quad (2.313)$$

In the scattering-dominated limit $\epsilon_\nu \ll 1$, the emerging flux is much less than that given by a blackbody photosphere by a factor of the order $\epsilon_\nu^{1/2}$.

In the absorption-dominated limit $\epsilon_\nu \approx 1$, we have $|F_\nu(\tau = 0)| = 2\pi B_\nu/\sqrt{3}$ and the frequency-integrated flux $|F(\tau = 0)| = (2/\sqrt{3})\sigma_{\text{SB}}T^4$, which are slightly different from the case of an isotropic radiation field where the one hemisphere fluxes are $F_\nu^{\text{out}} = \pi B_\nu$ and $F^{\text{out}} = \sigma_{\text{SB}}T^4$. The difference is caused by the effect of limb darkening.

Case II — finite uniform slab with incident radiation

Consider a finite slab that extends from $\tau = 0$ to $\tau = \tau_0$. Under the two-stream approximation, the incident intensity at $\tau = 0$ is given by

$$I_{\text{in}} = I_\nu^+(\tau = 0) = \left(J_\nu - \frac{1}{\sqrt{3}} \partial_\tau J_\nu \right)_{\tau=0}, \quad (2.314)$$

and at the right boundary, we have

$$I_\nu^- = \left(J_\nu + \frac{1}{\sqrt{3}} \partial_\tau J_\nu \right)_{\tau=\tau_0} = 0. \quad (2.315)$$

These two equations give

$$\begin{aligned} c_1 + c_2 - b(c_1 - c_2) &= I_{\text{in}} - B_\nu, \\ c_1 \mathcal{E} + c_2 \mathcal{E}^{-1} + b(c_1 \mathcal{E} - c_2 \mathcal{E}^{-1}) &= -B_\nu, \end{aligned} \quad (2.316)$$

where we have defined

$$b \equiv \sqrt{\epsilon_\nu}, \quad \mathcal{E} \equiv \exp(\sqrt{3\epsilon_\nu}\tau_0) = \exp(\tau_{\text{eff},0}).$$

The solutions to the two equations above are

$$\begin{aligned} c_1 &= \frac{-I_{\text{in}}\mathcal{E}^{-1}(1-b) - B_\nu [(1+b) - \mathcal{E}^{-1}(1-b)]}{\mathcal{E}(1+b)^2 - \mathcal{E}^{-1}(1-b)^2}, \\ c_2 &= \frac{I_{\text{in}}\mathcal{E}(1+b) + B_\nu [(1-b) - \mathcal{E}(1+b)]}{\mathcal{E}(1+b)^2 - \mathcal{E}^{-1}(1-b)^2}, \end{aligned} \quad (2.317)$$

Then the full solution at arbitrary optical depth $0 < \tau < \tau_0$ is obtained by plugging these two constants c_1 and c_2 into the general solution in eq. (2.306). In particular, the mean intensities at the boundaries are

$$\begin{aligned} J_\nu(\tau = 0) &= c_1 + c_2 + B_\nu \\ &= \frac{I_{\text{in}} [\mathcal{E}(1+b) - \mathcal{E}^{-1}(1-b)] + bB_\nu [\mathcal{E}(1+b) + \mathcal{E}^{-1}(1-b) - 2]}{\mathcal{E}(1+b)^2 - \mathcal{E}^{-1}(1-b)^2}, \end{aligned} \quad (2.318)$$

and

$$\begin{aligned} J_\nu(\tau_0) &= c_1 \mathcal{E} + c_2 \mathcal{E}^{-1} + B_\nu \\ &= b \frac{2I_{\text{in}} + B_\nu [\mathcal{E}(1+b) + \mathcal{E}^{-1}(1-b) - 2]}{\mathcal{E}(1+b)^2 - \mathcal{E}^{-1}(1-b)^2}. \end{aligned} \quad (2.319)$$

If we ignore the thermal emission from the slab (setting $B_\nu = 0$), then the *transmitted and reflected fractions* can be defined as

$$f_{\text{tran}} = \frac{I^+(\tau_0)}{I_{\text{in}}} = \frac{2J_\nu(\tau_0)}{I_{\text{in}}} = \frac{4b}{\mathcal{E}(1+b)^2 - \mathcal{E}^{-1}(1-b)^2}, \quad (2.320)$$

and

$$f_{\text{refl}} = \frac{I^-(\tau = 0)}{I_{\text{in}}} = \frac{2J_\nu(\tau = 0)}{I_{\text{in}}} - 1 = \frac{(\mathcal{E} - \mathcal{E}^{-1})(1-b^2)}{\mathcal{E}(1+b)^2 - \mathcal{E}^{-1}(1-b)^2}, \quad (2.321)$$

which are shown in Fig. 2.20. The $\tau_{\text{eff},0} = 1$ case (in right panel) agrees well with the result from Monte Carlo simulation by [Metzger et al. \(2014\)](#) (their Fig. 3).

If we only consider thermal emission (setting $I_{\text{in}} = 0$), then the outward thermal flux on either side of the slab is given by

$$F_\nu^{\text{out}}(\text{thermal}) = \frac{4\pi}{\sqrt{3}} J_\nu(0 \text{ or } \tau_0) = \frac{4\pi b B_\nu}{\sqrt{3}} \frac{\mathcal{E}(1+b) + \mathcal{E}^{-1}(1-b) - 2}{\mathcal{E}(1+b)^2 - \mathcal{E}^{-1}(1-b)^2}. \quad (2.322)$$

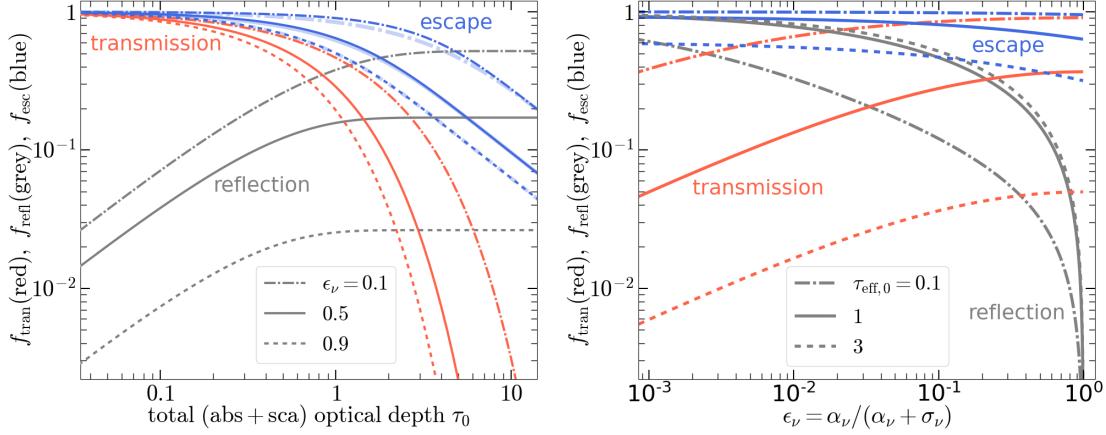


Figure 2.20: Left panel: transmitted (f_{tran} , red), reflected (f_{refl} , grey), and escape (f_{esc} , blue) fractions for a uniform slab of total optical depth τ_0 , for different choices of albedo parameter $\epsilon_\nu \equiv \alpha_\nu / (\alpha_\nu + \sigma_\nu)$. The escape fraction is for the thermal emission from the matter in the slab (thin blue lines = eq. 2.323, and thick blue lines = eq. 2.324). Right panel: The dependences of the three fractions, f_{tran} , f_{refl} , and f_{esc} on the albedo parameter ϵ_ν , for different choices of effective optical depths $\tau_{\text{eff}} = \sqrt{3}\epsilon_\nu\tau_0$. The slab is scattering (or absorption) dominated if $\epsilon_\nu \ll 1$ (or ≈ 1).

The total thermal emission per unit surface area generated within the entire slab is given by $4\pi j_\nu z_0 = 4\pi\alpha_\nu B_\nu z_0 = 4\pi\tau_{a,0}B_\nu$ (z_0 being the geometric thickness), so the escape fraction of the thermal emission is given by (and shown by the thin blue lines in Fig. 2.20)

$$f_{\text{esc}}(\text{thermal}) = \frac{2F_\nu^{\text{out}}(\text{thermal})}{4\pi\tau_{a,0}B_\nu} = \frac{2}{\tau_{\text{eff},0}} \frac{\mathcal{E}(1+b) + \mathcal{E}^{-1}(1-b) - 2}{\mathcal{E}(1+b)^2 - \mathcal{E}^{-1}(1-b)^2}, \quad (2.323)$$

where we have made use of $\sqrt{3}\tau_{a,0}/b = \tau_{\text{eff},0}$. A good analytic approximation for the thermal escape fraction is as follows

$$f_{\text{esc}}(\text{thermal}) \simeq \frac{1 - e^{\tilde{\tau}}}{\tilde{\tau}}, \quad \tilde{\tau} \equiv \frac{1 + \sqrt{\epsilon_\nu}}{2}\tau_{\text{eff},0}, \quad (2.324)$$

which is shown by the thick blue lines on the left panel of Fig. 2.20.

Let us then look at the limiting cases including thermal emission ($B_\nu \neq 0$) and incident radiation. If $\tau_{\text{eff},0} \gg 1$ (and hence $\mathcal{E} \gg 1$), we have

$$J_\nu(\tau = 0) \approx \frac{I_{\text{in}} + \sqrt{\epsilon_\nu}B_\nu}{1 + \sqrt{\epsilon_\nu}}, \quad J_\nu(\tau_0) \approx \frac{\sqrt{\epsilon_\nu}B_\nu}{1 + \sqrt{\epsilon_\nu}}, \quad \text{for } \tau_{\text{eff},0} \gg 1, \quad (2.325)$$

At $\tau = 0$, the radiation field consists of the incident intensity plus the thermal emission from a semi-infinite slab (eq. 2.310). At the right boundary of the slab, the radiation

field only contains the thermal emission from the slab, since the incident photons will be absorbed before they propagate through the slab. Thermal emission contributes an outward flux on either side of the slab

$$F_\nu^{\text{out}}(\text{thermal}) = \frac{4\pi}{\sqrt{3}} J_\nu^{\text{thermal}}(0 \text{ or } \tau_0) \approx \frac{4\pi}{\sqrt{3}} \frac{\sqrt{\epsilon_\nu}}{1 + \sqrt{\epsilon_\nu}} B_\nu, \text{ for } \tau_{\text{eff},0} \gg 1. \quad (2.326)$$

The escape fraction of the thermal emission is given by

$$f_{\text{esc}}(\text{thermal}) \approx \frac{2F_\nu^{\text{out}}(\text{thermal})}{4\pi\tau_{a,0}B_\nu} = \frac{2}{\sqrt{3}} \frac{\sqrt{\epsilon_\nu}}{1 + \sqrt{\epsilon_\nu}} \frac{1}{\tau_{a,0}} = \frac{2}{1 + \sqrt{\epsilon_\nu}} \frac{1}{\tau_{\text{eff},0}}, \quad (2.327)$$

which is in agreement with the analytic approximation in eq. (2.324). The transmitted and reflected fractions of the incident light are

$$f_{\text{tran}} \approx 0, \quad f_{\text{refl}} \approx \frac{1 - \sqrt{\epsilon_\nu}}{1 + \sqrt{\epsilon_\nu}}, \text{ for } \tau_{\text{eff},0} \gg 1, \quad (2.328)$$

and the reflected fraction is in agreement with the case of a semi-infinite slab (eq. 2.313).

In the opposite limit of $\tau_{\text{eff},0} \ll 1$, we have $\mathcal{E}^{\pm 1} \approx 1 \pm \tau_{\text{eff},0} + \tau_{\text{eff},0}^2/2$ (the 2nd-order term is important), and we obtain the following approximations

$$\begin{aligned} \mathcal{E}(1+b) + \mathcal{E}^{-1}(1-b) &\approx 2 + (2\sqrt{3} + 3\tau_0)\epsilon_\nu\tau_0, \\ \mathcal{E}(1+b) - \mathcal{E}^{-1}(1-b) &\approx 2\sqrt{\epsilon_\nu} (1 + \sqrt{3}\tau_0), \\ \mathcal{E}(1+b)^2 - \mathcal{E}^{-1}(1-b)^2 &\approx 2\sqrt{\epsilon_\nu} (2 + \sqrt{3}\tau_0). \end{aligned}$$

Therefore, the two mean intensities at the boundaries are

$$J_\nu(0) \approx \frac{I_{\text{in}}}{1 + (1 + \sqrt{3}\tau_0)^{-1}} + \frac{\sqrt{3}}{2} \tau_{a,0} B_\nu, \quad J_\nu(\tau_0) \approx \frac{I_{\text{in}}}{2 + \sqrt{3}\tau_0} + \frac{\sqrt{3}}{2} \tau_{a,0} B_\nu, \text{ for } \tau_{\text{eff},0} \ll 1, \quad (2.329)$$

where $\tau_{a,0} = \epsilon_\nu\tau_0$ is the absorption optical depth of the whole slab. The thermal emission contributes to an outward flux on either side of the slab

$$F_\nu^{\text{out}}(\text{thermal}) = \frac{4\pi}{\sqrt{3}} J_\nu^{\text{thermal}}(0 \text{ or } \tau_0) \approx 2\pi\tau_{a,0}B_\nu, \text{ for } \tau_{\text{eff},0} \ll 1. \quad (2.330)$$

This is half of the total thermal emission per unit area produced by the entire slab $4\pi\tau_{a,0}B_\nu$, so we find that 100% of the thermal emission has escaped — in agreement with our expectation from the absence of absorption. The transmitted and reflected fractions of the incident light are

$$f_{\text{tran}} \approx \frac{2}{2 + \sqrt{3}\tau_0}, \quad f_{\text{refl}} \approx \frac{\sqrt{3}\tau_0}{2 + \sqrt{3}\tau_0}, \text{ for } \tau_{\text{eff},0} \ll 1. \quad (2.331)$$

2.6.7 *Moments of radiative transfer equation

Numerically solving the full radiative transfer equation is extremely computationally expensive as the intensity function involves 7 dimensions ($t, \mathbf{x}, \nu, \hat{\mathbf{n}}$), so a common approach is to only keep track of the frequency-integrated, 0th, 1st, and 2nd moments of the radiation field $U = \int U_\nu d\nu$, $\mathbf{F} = \int \mathbf{F}_\nu d\nu$ (eq. 1.77), and $\overset{\leftrightarrow}{P} = \int \overset{\leftrightarrow}{P}_\nu d\nu$ (eq. 1.79). The biggest drawback of this approach is that the information about the angular dependence of the specific intensity function $I_\nu(\theta, \phi)$ is lost. This is less of an issue in optically thick regions where we expect the intensity to be only weakly anisotropic, but the problem becomes serious in optically thin regions²¹. Another drawback of the frequency-integrated moments-based approach is the usage of frequency-averaged opacities (see below). Since we are not keeping track of the radiation spectrum, it is non-trivial (or even impossible) to accurately calculate the frequency-averaged opacities. For this reason, frequency-integrated radiation transport methods only strictly apply when either (1) the radiation field and matter are nearly in LTE or (2) the absorption and scattering opacities are nearly gray (i.e., frequency-independent).

In this subsection, we consider the moments of the radiative transfer equation including scattering and absorption

$$\frac{dI_\nu}{ds} = \frac{1}{c} \partial_t I_\nu + \hat{\mathbf{n}} \cdot \nabla I_\nu = -(\alpha_\nu + \sigma_\nu) I_\nu + (\alpha_\nu B_\nu + \sigma_\nu J_\nu), \quad (2.332)$$

where $B_\nu(T_g)$ is the Planck function evaluated at the gas temperature T_g and $\alpha_\nu B_\nu$ stands for the gas emissivity (spontaneous emission) obtained from Kirchhoff's law. The intensity I_ν is evaluated in the comoving frame of the bulk motion of gas where we assume that the gas distribution function is isotropic and thermal (so T_g is well defined). If we wish to consider the intensity I_ν and its moments in the lab frame where the gas has non-negligible bulk motion, one must carry out a Lorentz transformation to obtain the intensity in the gas comoving frame before applying the radiative transfer equation above. Even when the gas motion is non-relativistic, such a Lorentz transformation is required to capture the effect of photon trapping when the effective speed of photon diffusion is slower than the gas speed.

To compute the 0th moment of the radiative transfer equation, we directly integrate both sides over all solid angles and the result is

$$\partial_t U_\nu + \nabla \cdot \mathbf{F}_\nu = -(\alpha_\nu + \sigma_\nu) U_\nu c + (4\pi \alpha_\nu B_\nu + \sigma_\nu U_\nu c) = \alpha_\nu (4\pi B_\nu - U_\nu c), \quad (2.333)$$

where we have made use of $\hat{\mathbf{n}} \cdot \nabla I_\nu = \nabla \cdot (I_\nu \hat{\mathbf{n}})$ as $\hat{\mathbf{n}}$ does not depend on spatial coordinates (\mathbf{x}) and hence $\int d\Omega (\hat{\mathbf{n}} \cdot \nabla I_\nu) = \nabla \cdot (\int d\Omega I_\nu \hat{\mathbf{n}}) = \nabla \cdot \mathbf{F}_\nu$. Note that the radiation energy

²¹For this reason, it is more accurate to directly integrate the radiative transfer equation along a large number of angles (see e.g., [Jiang 2021](#)).

density U_ν is related to the mean intensity J_ν by $U_\nu c = \int d\Omega I_\nu = 4\pi J_\nu$. We then integrate the 0th moment equation over frequency and obtain

$$c^{-1} \partial_t U + \nabla \cdot (\mathbf{F}/c) = -\alpha_E U + \alpha_P a T^4, \quad (2.334)$$

where $a = 4\sigma_{\text{SB}}/c$ is the radiation density constant and we have defined the *energy-mean absorption coefficient* weighted by the spectrum of the local radiation energy density

$$\alpha_E \equiv \frac{\int \alpha_\nu U_\nu d\nu}{\int U_\nu d\nu} = \frac{\int \alpha_\nu U_\nu d\nu}{U}, \quad (2.335)$$

and the Planck-mean absorption coefficient weighted by the Planck function at the gas temperature T_g

$$\alpha_P \equiv \frac{\int \alpha_\nu B_\nu(T_g) d\nu}{\int B_\nu(T_g) d\nu} = \frac{\int \alpha_\nu B_\nu(T_g) d\nu}{\sigma_{\text{SB}} T_g^4 / \pi}. \quad (2.336)$$

Note that the Planck-mean $\alpha_P(\rho, T_g)$ is evaluated at the gas temperature T_g and it only depends on the gas density ρ and gas temperature T_g . Extensive tables of Planck-mean opacity $\kappa_P(\rho, T_g) = \alpha_P/\rho$ and Rosseland-mean opacity $\kappa_R(\rho, T_g)$ have been computed under the LTE assumption, e.g., OPAL (Iglesias & Rogers 1996) and OPLIB (Farag et al. 2024). If the radiation spectrum has a blackbody shape at the radiation temperature T_r (which may be different from the gas temperature T_g) such that $U_\nu \propto B_\nu(T_r)$, then the energy-mean absorption coefficient is given by

$$\alpha_E(\rho, T_g, T_r) \equiv \frac{\int \alpha_\nu B_\nu(T_r) d\nu}{\int B_\nu(T_r) d\nu} = \frac{\int \alpha_\nu B_\nu(T_r) d\nu}{\sigma_{\text{SB}} T_r^4 / \pi}, \quad (2.337)$$

which depend on the gas density ρ , gas temperature T_g (through α_ν), and also the radiation temperature T_r (although in this case one might question the validity of the assumption of thermal distribution function for the gas at temperature T_g). In this case, $\alpha_E(\rho, T_g, T_r)$ is also called the *two temperature Planck-mean absorption coefficient*. In the simpler case where the radiation field and gas are in LTE, then it is appropriate to take

$$\alpha_E(\rho, T_g, T_r) \approx \alpha_P(\rho, T_g), \text{ if LTE}, \quad (2.338)$$

and the hope is that the physical context of the problem in hand provides justification for LTE, as non-LTE radiative transport requires a frequency-resolved treatment.

We then compute the 1st moment of the radiative transfer equation (eq. 2.332) by integrating both sides over all solid angles with a weight of $\hat{\mathbf{n}}$, and the result is

$$c^{-2} \partial_t \mathbf{F}_\nu + \nabla \cdot \overset{\leftrightarrow}{P}_\nu = -(\alpha_\nu + \sigma_\nu) \mathbf{F}_\nu / c. \quad (2.339)$$

The corresponding frequency-integrated version of the 1st-moment equation is

$$c^{-2} \partial_t \mathbf{F} + \nabla \cdot \overset{\leftrightarrow}{P} = -(\alpha_F + \sigma_F) \mathbf{F} / c, \quad (2.340)$$

where we have defined the *flux-mean absorption coefficient*

$$\alpha_F \equiv \frac{\int \alpha_\nu F_\nu d\nu}{\int F_\nu d\nu} = \frac{\int \alpha_\nu F_\nu d\nu}{F}, \quad F_\nu = |\mathbf{F}_\nu|, \quad F = |\mathbf{F}|, \quad (2.341)$$

and similarly the *flux-mean scattering coefficient*²² $\sigma_F \equiv F^{-1} \int \sigma_\nu F_\nu d\nu$. In the optically thin limit, the flux spectrum may be considered as the same as the energy spectrum $F_\nu \propto U_\nu$, so we expect $\alpha_F \approx \alpha_E$.

However, in the optically thick limit, the flux spectrum may be more complicated. Looking at the 1st-moment equation (2.339), the time-derivative term $c^{-2} \partial_t \mathbf{F}_\nu$ describes the variation of radiative flux on the dynamical timescale t_{dy} , which is generally much longer than the mean free time for photon interaction with matter $t_{mfp} = (\alpha_\nu + \sigma_\nu)^{-1}/c$ for mean free path $\ell_{mfp,\nu} = (\alpha_\nu + \sigma_\nu)^{-1}$. This is because the dynamical time is longer than the light-crossing time of the system $t_{dy} > L/c$ whereas the photon mean free time is much shorter than the light-crossing time $t_{mfp} = \ell_{mfp,\nu}/c \ll L/c$, where L is the size of the system. Thus, one can ignore the time-derivative term in eq. (2.339) and obtain

$$\mathbf{F}_\nu \approx -\frac{c}{\alpha_\nu + \sigma_\nu} \nabla \cdot \overset{\leftrightarrow}{P}_\nu = -\ell_{mfp,\nu} c \nabla \cdot \overset{\leftrightarrow}{P}_\nu, \quad \text{if } \tau_\nu \gg 1. \quad (2.342)$$

Since the radiation field is nearly isotropic in optically thick regions, we further approximate²³ $\overset{\leftrightarrow}{P}_\nu \approx \overset{\leftrightarrow}{I} U_\nu / 3$, where $\overset{\leftrightarrow}{I} = \text{diag}(1, 1, 1)$ is the identity matrix. This leads to the well-known result from the Eddington approximation for a nearly isotropic radiation field

$$\mathbf{F}_\nu \approx -\frac{c}{3(\alpha_\nu + \sigma_\nu)} \nabla U_\nu = -\frac{\ell_{mfp,\nu} c}{3} \nabla U_\nu, \quad \text{if } \tau_\nu \gg 1. \quad (2.343)$$

We then integrate the flux density over frequency and obtain

$$\mathbf{F} = \int \mathbf{F}_\nu d\nu = -\frac{c}{3} \int \frac{\nabla U_\nu}{\alpha_\nu + \sigma_\nu} d\nu \approx -\frac{c}{3\rho\kappa_R} \nabla U, \quad \text{if LTE.} \quad (2.344)$$

where in the final expression we have used the Rosseland-mean opacity $\kappa_R(\rho, T_g)$ under the assumption that the radiation field and matter is in LTE such that $T_r = T_g$ and $U_\nu \propto B_\nu(T_g)$ — LTE is achieved when the effective optical depth $\tau_{\text{eff}} \simeq \sqrt{\tau_a(\tau_a + \tau_s)} \gtrsim 1$ at all frequencies. If we do not wish to keep track of the radiation spectrum (the model will be a crude one), the flux-averaged opacity can be estimated under the LTE assumption

$$(\rho\kappa_R)^{-1} = \frac{1}{|\nabla U|} \int (\alpha_\nu + \sigma_\nu)^{-1} |\nabla U_\nu| d\nu \approx \frac{\int (\alpha_\nu + \sigma_\nu)^{-1} \partial_T B_\nu d\nu}{\int \partial_T B_\nu d\nu}. \quad (2.345)$$

²²If scattering coefficient σ_ν is dominated by Thomson scattering by free electrons, then it is nearly frequency-independent and it is appropriate to take $\sigma_F \approx \sigma_T$.

²³The (dimensionless) Eddington tensor for an isotropic radiation field is $\overset{\leftrightarrow}{D} = \overset{\leftrightarrow}{P}/U = \overset{\leftrightarrow}{I}/3$.

or

$$\alpha_F + \sigma_F \approx \rho\kappa_R, \text{ if LTE.} \quad (2.346)$$

Note that the Rosseland-mean opacity κ_R includes both absorption and scattering (they cannot be separated as their contributions to κ_R are not additive).

After formulating the frequency-integrated moment equations (2.334, 2.340) and defining the frequency-averaged absorption/scattering coefficients, we then discuss how the radiation field may be approximately solved using the moment method.

The frequency-integrated 0th-moment equation (2.334) keeps track of how the radiation energy density U changes with time (for given initial conditions), provided that the flux \mathbf{F} is known. The 1st-moment equation (2.340) provides the flux \mathbf{F} at each time, provided that the pressure tensor $\overset{\leftrightarrow}{P}$ is known. The time evolution of the pressure tensor can only be obtained from the 2nd moment of the radiative transfer equation, but that would involve the 3rd moment of the radiation field (which is beyond our scope here). We see that a *closure relation* must be introduced to close the series of moment equations. Many closure relations have been proposed in the literature of radiation hydrodynamics. In the following, we will discuss two popular classes of methods.

The simplest class of closure relations is to truncate the series of moment equations at the 0th order — this is called the *M0 closure method*. In this approach, one introduces an artificial relation to compute the flux vector $\mathbf{F}(t, \mathbf{x})$ from $U(t, \mathbf{x})$. Thus, we only keep track of the radiation energy density $U(t, \mathbf{x})$ as a function of the spacetime coordinates. One might think that this is absurd, because it is impossible to get three numbers (F_x, F_y, F_z) from only one (U). It turns out that this approach is reasonably accurate in optically thick regions where the flux vector is physically determined by the spatial gradient of the radiation energy density, and when it fails in optically thin regions, we hope that the method does not give terribly unphysical results.

Motivated by the Fick's law (which only applies to optically thick regions), we introduce the following closure relation in a framework of the *flux-limited diffusion theory*

$$\mathbf{F}_\nu = -\lambda \frac{c}{\alpha_\nu + \sigma_\nu} \nabla U_\nu \Rightarrow \mathbf{F} = -\lambda \frac{c}{\alpha_F + \sigma_F} \nabla U \approx -\lambda \frac{c}{\rho\kappa_R} \nabla U, \quad (2.347)$$

where the Rosseland-mean approximation only holds under LTE and the dimensionless factor $0 < \lambda < \infty$ is called the *flux limiter* that makes it possible for the above relation to be applied in the optically thick and thin limits (see below).

In the optically thick limit, the diffusion coefficient is given by $\ell_{\text{mfp},R} c / 3 = c / (3\rho\kappa_R)$ (under the Rosseland approximation), so the flux limiter must have the asymptotic limit of $\lambda \rightarrow 1/3$. In the optically thin limit, the mean free path becomes extremely large and our diffusion picture would predict an unphysically large flux. This is because the maximum physical flux is given by $F_{\max} = Uc$, which is achieved when all photons are moving in the

same direction. Motivated by this maximum flux constraint, we expect the flux limiter to have an asymptotic limit of $\lambda \rightarrow \rho\kappa_R U / |\nabla U|$ in the optically thin limit.

There are many choices for the functional form of the flux limiter λ , which, by construction, only depends on the radiation and matter properties locally near position \mathbf{x} where we are trying to solve for the flux vector \mathbf{F} . The most popular choice that satisfies the requirements in the optically thick/thin limits is as follows ([Levermore & Pomraning 1981](#))

$$\lambda(\xi) = \frac{2 + \xi}{6 + 3\xi + \xi^2}, \quad \xi \equiv \frac{|\nabla U|}{\rho\kappa_R U}. \quad (2.348)$$

One can easily verify the following asymptotic behaviors

$$\lambda \rightarrow \begin{cases} 1/3, & \text{for } \xi \ll 1, \text{ (optically thick)} \\ 1/\xi, & \text{for } \xi \gg 1. \text{ (optically thin)} \end{cases} \quad (2.349)$$

Under the flux-limited diffusion approximation, the 0th moment of the radiative transfer equation (2.334) can be written as

$$c^{-1}\partial_t U - \nabla \cdot \left(\frac{\lambda(\xi)}{\rho\kappa_R} \nabla U \right) = -\alpha_E U + \alpha_P a T_g^4 \approx \rho\kappa_P (a T_g^4 - U), \quad (2.350)$$

where in the final expression we have approximated $\alpha_E \approx \alpha_P$ assuming the radiation field and matter to be close to LTE. The above equation allows us to solve for the radiation energy density $U(t, \mathbf{x})$ as a function of time everywhere in the computational domain (for given boundary conditions). We note that the main drawback of the flux-limited diffusion theory is the inaccurate treatment of radiation transport in the optically thin regions, because of the potentially inaccurate assumption that all photons freely stream along the same direction parallel to ∇U .

The next class of closure relations is to truncate the moment equations at the 1st order — this is called the *M1 closure method*. In this approach, one keeps track of U and \mathbf{F} as functions of the spacetime coordinates, whereas an artificial relation is introduced to compute the radiation pressure tensor $\overset{\leftrightarrow}{P}$ from U and \mathbf{F} . The flux vector \mathbf{F} defines the direction $\hat{\mathbf{f}} \equiv \mathbf{F}/F$ for the net energy flow, and we also define a dimensionless flux ratio $f \equiv F/F_{\max} = F/(Uc)$ such that $0 \leq f \leq 1$. Our goal is to construct the (dimensionless) Eddington tensor $\overset{\leftrightarrow}{D} = \overset{\leftrightarrow}{P}/U$ based on the dimensionless flux vector \mathbf{f} which contains the magnitude $f \in [0, 1]$ and direction $\hat{\mathbf{f}}$.

Let us temporarily assume that $\hat{\mathbf{f}} = \hat{\mathbf{z}}$ (one can always orient the local coordinate axes to achieve this). In the optically thin limit, if photons are free-streaming along the same direction (along $\hat{\mathbf{z}}$), we obtain $f = 1$ and the Eddington tensor is $\overset{\leftrightarrow}{D} = \text{diag}(0, 0, 1)$. The assumption of unidirectional photon streaming is in fact problematic, because photons may

not be free-streaming along the same direction if there are more than two radiation sources, e.g., binary stars. Unfortunately, we do not have more information (besides U and \mathbf{F}) to reconstruct the full structure of the pressure tensor $\overset{\leftrightarrow}{P} = c^{-1} \int d\Omega I \hat{\mathbf{n}} \hat{\mathbf{n}}$ (the full angular dependence of the intensity function is needed), so we must let go with the inaccurate results from this method in the optically thin regions. On the other hand, in the optically thick limit, we expect the radiation field to be nearly isotropic, so the Eddington tensor is $\overset{\leftrightarrow}{D} \approx \overset{\leftrightarrow}{I}/3$.

A popular choice for the Eddington tensor that satisfies the optically thin/thick limits as well as the physical requirement of $\text{Tr}(\overset{\leftrightarrow}{D}) = 1$ is ([Levermore 1984](#))

$$\overset{\leftrightarrow}{D} = \text{diag} \left(\frac{1-\chi}{2}, \frac{1-\chi}{2}, \chi \right), \quad (\text{for } \hat{\mathbf{f}} = \mathbf{F}/F = \hat{\mathbf{z}}) \quad (2.351)$$

where the function $\chi(f)$ is taken to be

$$\chi = \frac{3 + 4f^2}{5 + 2\sqrt{4 - 3f^2}}. \quad (2.352)$$

It is easy to verify the following asymptotic limits

$$\chi \rightarrow \begin{cases} 1, & \text{for } f = 0, \text{ (optically thick)} \\ 1/3, & \text{for } f = 1. \text{ (optically thin)} \end{cases} \quad (2.353)$$

If we relax the choice of $\hat{\mathbf{f}} = \mathbf{F}/F = \hat{\mathbf{z}}$, then the corresponding Eddington tensor is generally not diagonal and is given by

$$\overset{\leftrightarrow}{D} = \frac{1-\chi}{2} \delta_{ij} + \frac{3\chi-1}{2} f_i f_j, \quad (\text{for arbitrary } \hat{\mathbf{f}} = \mathbf{F}/F) \quad (2.354)$$

where δ_{ij} ($i, j = x, y, z$) is the Kronecker delta ($\delta_{ii} = 1$ and $\delta_{i \neq j} = 0$) and $f_{i/j}$ are the Cartesian components of $\hat{\mathbf{f}} = \mathbf{F}/F$. In the M1-closure method, the function $\chi(f)$ plays a similar role as the flux limiter $\lambda(\xi)$ in the M0-closure method discussed before, so it is appropriate to consider the M1-closure method to be part of the general flux-limited diffusion theory.

With the approximate Eddington tensor under the M1 closure, one can then write the 0th and 1st moments of the radiative transfer equation as

$$\begin{aligned} c^{-1} \partial_t U + \nabla \cdot (\mathbf{F}/c) &= \alpha_P (a T_g^4 - U), \\ c^{-2} \partial_t \mathbf{F} + \nabla \cdot (\overset{\leftrightarrow}{D} U) &= -\rho \kappa_R \mathbf{F}/c, \end{aligned} \quad (2.355)$$

where we have taken $\alpha_E \approx \alpha_P$ and $\alpha_F + \sigma_F \approx \rho \kappa_R$ under the assumption of LTE (or that the absorption and scattering opacities are nearly gray). It should be noted that the RHS

of the above moment equations are evaluated in the comoving frame of the matter bulk motion. Often, it is desirable to calculate the RHS in the lab frame where the moments of the radiation field are defined. The readers are referred to detailed discussions by [Krumholz et al. \(2007\)](#).

2.7 Radiative transfer in a spherical dense wind

As illustrated in Fig. 2.21, we consider a wind with constant outflowing velocity v launched from an inner radius r_{in} . The wind mass-loss rate is denoted as \dot{M} , and the density profile is given by

$$\rho(r) = \frac{\dot{M}}{4\pi r^2 v} = \rho_{\text{in}}(r/r_{\text{in}})^{-2}, \quad (2.356)$$

where $\rho_{\text{in}} = \dot{M}/(4\pi r_{\text{in}}^2 v)$ is the density at the inner boundary. Let us assume that the wind has a constant total (absorption + scattering) opacity which is roughly equal to the Rosseland-mean opacity, i.e.,

$$\kappa = \kappa_a + \kappa_s \approx \kappa_R \approx \text{const.} \quad (2.357)$$

The corresponding optical depth from r to infinity is

$$\tau(r) = \int_r^\infty \rho \kappa dr = \rho \kappa r = \frac{\kappa \dot{M}}{4\pi r v}. \quad (2.358)$$

The photospheric radius r_{ph} is defined as where $\tau(r_{\text{ph}}) = 1$, which means

$$r_{\text{ph}} = \frac{\kappa \dot{M}}{4\pi v}. \quad (2.359)$$

A larger mass-loss rate and a slower velocity leads to a larger photospheric radius.

At the inner boundary the wind is loaded with radiation field with energy density U_{in} , which may be of the order $(1/2)\rho_{\text{in}}v^2$ if the wind is radiatively driven (but we keep U_{in} a free parameter). We assume that the pressure and internal energy density at all radii are dominated by radiation (ignoring gas pressure). Radiative diffusion produces an energy flux in the radial direction (according to Fick's law, eq. 2.280)

$$F_{\text{dif}}(r) = -\frac{c}{3\rho\kappa} \frac{dU}{dr}. \quad (2.360)$$

The diffusive flux should be understood as in the comoving frame of the fluid's motion. In the lab frame that is at rest with the wind-launching object at the center, the outflowing gas motion produces an advective radiation flux in the outward radial direction

$$F_{\text{adv}}(r) = Uv. \quad (2.361)$$

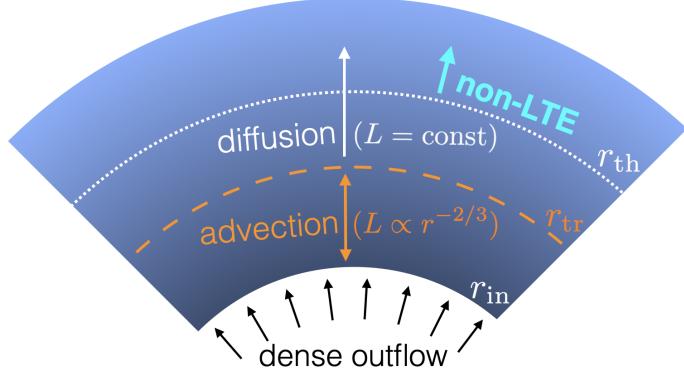


Figure 2.21: Radiative trapping, diffusion, and thermalization in a spherical dense wind.

We consider a non-relativistic wind, so the total flux in the lab frame is the sum of these two, $F = F_{\text{dif}} + F_{\text{adv}}$. The ratio between these two flux components is

$$\frac{F_{\text{dif}}}{F_{\text{adv}}} = \frac{c/v}{3\tau} \left| \frac{dU}{dr} \right| \frac{r}{U}. \quad (2.362)$$

The critical radius where $F_{\text{dif}} = F_{\text{adv}}$ is called the *photon trapping radius* r_{tr} .

Let us assume that the wind is very dense at the inner boundary such that $\tau(r_{\text{in}}) = \rho_{\text{in}}\kappa r_{\text{in}} \gg c/v$. Since the photon diffusion speed is $v_{\text{dif}} \sim c/v$, we know that the advective flux dominates over diffusion at small radii. This means that radiation is trapped by the fluid and hence undergoes adiabatic expansion

$$U(r < r_{\text{tr}}) = U_{\text{in}}(\rho/\rho_{\text{in}})^{\gamma_{\text{ad}}} = U_{\text{in}}(r/r_{\text{in}})^{-8/3}, \quad (2.363)$$

where $\gamma_{\text{ad}} = 4/3$ is the adiabatic index for radiation-dominated gas $P \propto \rho^{\gamma_{\text{ad}}}$ and we have used $P = U/3$. The power-law radiation energy density profile leads to $dU/dr = -(8/3)U/r$, which can be plugged into eq. (2.362) to obtain

$$\frac{F_{\text{dif}}}{F_{\text{adv}}} = \frac{8}{9} \frac{c}{v\tau} \approx \frac{c}{v\tau}. \quad (2.364)$$

Thus, we find the photon trapping radius, where $F_{\text{dif}}/F_{\text{adv}} = 1$, to be given by

$$\tau(r_{\text{tr}}) \approx c/v,$$

and this means

$$r_{\text{tr}} \approx \frac{\kappa \dot{M}}{4\pi c}, \quad (2.365)$$

which is independent of the wind speed v .

At $r < r_{\text{tr}}$, the radiative flux is dominated by advection and the luminosity is

$$L(r < r_{\text{tr}}) = 4\pi r^2 F_{\text{adv}}(r) = L_{\text{in}}(r/r_{\text{in}})^{-2/3}, \quad (2.366)$$

where we have defined $L_{\text{in}} \equiv 4\pi r_{\text{in}}^2 U_{\text{in}} v$ and the decreasing behavior $L \propto r^{-2/3}$ is due to the effect of *adiabatic loss*. At larger radii $r > r_{\text{tr}}$, diffusion is more important than advection. Since radiative diffusion conserves energy (adiabatic loss is negligible), we must have $F_{\text{dif}}(r > r_{\text{tr}}) \propto r^{-2}$ such that $L(r > r_{\text{tr}}) = 4\pi r^2 F_{\text{dif}} = \text{const}$. This means that the radiation energy density decreases with radius more steeply with radius as

$$U(r_{\text{tr}} < r < r_{\text{ph}}) = U(r_{\text{tr}})(r/r_{\text{tr}})^{-3} = U_{\text{in}}(r_{\text{tr}}/r_{\text{in}})^{-8/3}(r/r_{\text{tr}})^{-3}. \quad (2.367)$$

The power-law radiation density profile of $U \propto r^{-3}$ gives $dU/dr = -3U/r$, and hence the diffusive flux is given by

$$F_{\text{dif}}(r_{\text{tr}} < r < r_{\text{ph}}) = \frac{Uc}{\rho\kappa r} = \frac{Uc}{\tau}. \quad (2.368)$$

With the correct boundary condition at r_{tr} , we obtain the radiative luminosity

$$L(r > r_{\text{tr}}) = 4\pi r_{\text{tr}}^2 U(r_{\text{tr}})v = L_{\text{in}}(r_{\text{tr}}/r_{\text{in}})^{-2/3}, \quad (2.369)$$

which is the emerging bolometric luminosity from the wind. To summarize, the radiation energy density profile is given by

$$U(r)/U_{\text{in}} = \begin{cases} (r/r_{\text{in}})^{-8/3}, & \text{for } r_{\text{in}} < r < r_{\text{tr}}, \\ (r_{\text{tr}}/r_{\text{in}})^{-8/3}(r/r_{\text{tr}})^{-3}, & \text{for } r_{\text{tr}} < r < r_{\text{ph}}, \\ (r_{\text{tr}}/r_{\text{in}})^{-8/3}(r_{\text{ph}}/r_{\text{tr}})^{-3}(r/r_{\text{ph}})^{-2}, & \text{for } r > r_{\text{ph}}. \end{cases} \quad (2.370)$$

and the luminosity profile is

$$L(r)/L_{\text{in}} = \begin{cases} (r/r_{\text{in}})^{-2/3}, & \text{for } r_{\text{in}} < r < r_{\text{tr}}, \\ (r_{\text{tr}}/r_{\text{in}})^{-2/3} = \text{const}, & \text{for } r > r_{\text{tr}}. \end{cases} \quad (2.371)$$

In the following, we discuss the radiation temperature profile.

In regions where the effective optical depth $\tau_{\text{eff}} \approx \sqrt{3\tau_a(\tau_a + \tau_s)} \gtrsim 1$ (eq. 2.271), radiation will be in LTE and the temperature is given by $T = (U/a)^{1/4}$, where $a = 4\sigma_{\text{SB}}/c$ is the radiation density constant. The boundary between LTE and non-LTE is called us the *thermalization radius*, which is defined by

$$\tau_{\text{eff}}(r_{\text{th}}) = 1.$$

Then the radiation temperature profile is given by the the following broken power-law

$$T(r)/T_{\text{in}} = \begin{cases} (r/r_{\text{in}})^{-2/3}, & \text{for } r_{\text{in}} < r < r_{\text{tr}}, \\ (r_{\text{tr}}/r_{\text{in}})^{-2/3}(r/r_{\text{tr}})^{-3/4}, & \text{for } r_{\text{tr}} < r < r_{\text{th}}, \\ (r_{\text{tr}}/r_{\text{in}})^{-2/3}(r_{\text{th}}/r_{\text{tr}})^{-3/4} = \text{const}, & \text{for } r > r_{\text{th}}. \end{cases} \quad (2.372)$$

The observed *color temperature* T_c at $r \rightarrow \infty$ is determined at $r = r_{\text{th}}$ and we obtain

$$T_c = T_{\text{in}}(r_{\text{tr}}/r_{\text{in}})^{-2/3}(r_{\text{th}}/r_{\text{tr}})^{-3/4}. \quad (2.373)$$

The flux at the photosphere $L(r_{\text{ph}})/(4\pi r_{\text{ph}}^2) = \sigma_{\text{SB}} T_{\text{eff}}^4$ defines the *effective temperature*

$$T_{\text{eff}} = \left[\frac{L_{\text{in}}(r_{\text{tr}}/r_{\text{in}})^{-2/3}}{4\pi r_{\text{ph}}^2 \sigma_{\text{SB}}} \right]^{1/4}. \quad (2.374)$$

It is interesting to compare T_{eff} to the color temperature T_c . To make this comparison, we use the diffusive flux at the thermalization radius

$$F_{\text{dif}}(r_{\text{th}}) = \frac{U(r_{\text{th}})c}{\tau(r_{\text{th}})} = \frac{acT_c^4}{\tau(r_{\text{th}})} = acT_c^4 \frac{r_{\text{th}}}{r_{\text{ph}}}, \quad (2.375)$$

where we have used the optical depth depth at the thermalization radius $\tau(r_{\text{th}}) = r_{\text{ph}}/r_{\text{th}}$. The radiative flux profile between r_{th} and r_{ph} satisfies the inverse square law, so we write $F_{\text{dif}}(r_{\text{th}}) = \sigma_{\text{SB}} T_{\text{eff}}^4 (r_{\text{ph}}/r_{\text{th}})^2$ and obtain

$$\frac{T_c}{T_{\text{eff}}} \simeq \frac{1}{4^{1/4}} \left(\frac{r_{\text{ph}}}{r_{\text{th}}} \right)^{3/4} \gtrsim 1. \quad (2.376)$$

We see that the color temperature of an optically thick wind is generally higher than the effective temperature, unless the wind opacity is dominated by absorption (with negligible scattering).

The above definition of thermalization radius is based on the assumption of “grey” or frequency-independent opacities and the consequence is that the observed spectrum has a blackbody shape at the color temperature T_c . Scattering by free electrons can usually be described as grey, because the scattering cross-section does not depend on the photon frequency (as long as $h\nu \ll m_e c^2$). However, the absorption opacity is in general not grey and this causes the thermalization radius to be frequency-dependent, so the emerging spectrum generally deviates from the blackbody shape, and hence T_c only provides a roughly estimate for the average photon energy ($\sim 3k_B T_c$). A physical case of the free-free absorption opacity will be discussed in Chapter 6.

2.8 Homework

Prob. 4. Consider spherical dust grains of radius $a = 0.1 \mu\text{m}$ and temperature $T = 100 \text{ K}$. Take the average density of the grain material to be $\rho = 2 \text{ g cm}^{-3}$. The grains in a given cloud has a mass column density of $\Sigma = 10^{-2} M_\odot \text{ pc}^{-2}$, which equals to the mass per unit area if we project the cloud onto a plane perpendicular to our line of sight. [In fact, the gas-to-dust mass ratio is about 100, so the gas column density is $1 M_\odot \text{ pc}^{-2}$, but here we are only interested in the dust grains.]

- (i) If the grain's surface radiates like a blackbody, what are the (frequency-integrated) intensity and luminosity produced by a single grain? Then consider a cloud of such dust grains.
- (ii) Assuming that the grains do not significantly shadow each other (i.e., we are in the optically thin regime) and that the individual grains are unresolved by our telescope (but the cloud is resolved), calculate the average intensity from the cloud.
- (iii) What is the intensity of the cloud if the cloud is optically thick (i.e., the grains do strongly shadow each other)?

Prob. 5. In this exercise, we take a more careful look at the emission from dust grains of radius a and temperature T . For photons with energy near $k_B T$, the wavelength $\lambda \sim hc/(k_B T) \sim 10 \mu\text{m} (T/10^3 \text{ K})^{-1}$ is typically much longer than the grain size, as long as the grain temperature is $T \lesssim 10^3 \text{ K}$. In this limit, the grains are poor emitters/absorbers in that the absorption cross-section is a factor of $\sim a/\lambda \ll 1$ smaller than the geometric cross-section. This means that the emission spectrum takes the form of a so-called “modified blackbody” $j_\nu \propto \nu^q B_\nu$ and $q \approx 1$.

- (i) For the “modified blackbody” with $q = 1$, compute the peak frequency $h\nu_{\text{pk}}$ (as a function of grain temperature T) where j_ν is maximized.
- (ii) What about $q = 2$ (just for curiosity)?

Prob. 6. The Sun has mass M_\odot , radius R_\odot , and luminosity L_\odot . From the Virial Theorem, one can show that the total thermal energy in the solar interior is roughly given by

$$E_{\text{th}} \simeq GM_\odot^2/(2R_\odot),$$

which is comparable to the absolute value of the Sun’s gravitational potential energy. The “average” temperature T in the solar interior is given by

$$(3/2)k_B TN = E_{\text{th}},$$

where $N = M_\odot/(\mu m_p)$ is the total number of particles, $\mu \approx 0.6$ is the mean molecular weight for ionized gas with solar composition, and m_p is the proton mass. The Sun has a radiative core where the energy transport is dominated by radiative diffusion.

- (i) For an average interior temperature of T from above and an average density of $\rho = 3M_\odot/(4\pi R_\odot^3)$, look up (by eye) the corresponding Rosseland-mean opacity κ_R in Fig. 2.17. Use κ_R to estimate the mean-free path for photon diffusion $\ell_{\text{mfp},R} = (\rho\kappa_R)^{-1}$. Then, estimate the photon diffusion timescale t_{dif} from the center to the surface using eq. (2.199) and express your answer in years.

(After you are done, compare your answer with the result in [this paper](#). Note that the authors in this paper *incorrectly* wrote the mean-squared distance after N scatterings as $\langle \mathbf{r}_N^2 \rangle = N\ell_{\text{mfp}}^2/3$, which is smaller than the correct result by a factor of 6. In the end, their diffusion time is too long by a factor of 6.)

(ii) The radiative flux in eq. (2.280) depends on the *radiation* energy density $U_{\text{rad}} = aT^4$. Use the average temperature T (and assume LTE) to estimate the typical radiation energy density in the solar interior. Compare your U_{rad} with the thermal energy density $U_{\text{th}} = 3E_{\text{th}}/(4\pi R_\odot^3)$. Is the thermal energy in the solar interior dominated by radiation?

(iii) The Kelvin-Helmholtz timescale (also known as the thermal timescale) is defined as the time it takes for the Sun to radiate away the thermal energy E_{th} in its interior. Compare your earlier result of t_{dif} to the Kelvin-Helmholtz timescale for the Sun

$$t_{\text{KH}} = \frac{E_{\text{th}}}{L_\odot} \simeq \frac{GM_\odot^2}{2R_\odot L_\odot}.$$

Compare t_{dif} with t_{KH} and comment on their difference at the order-of-magnitude level.

Prob. 7. Consider a homologously expanding spherical supernova ejecta of mass $M = 10M_\odot$ and total energy $E = 10^{51}$ erg. The initial radius of the star is $R = 100R_\odot$. Initially, about half of the total energy is in the form of thermal energy, which is dominated by radiation $E_{\text{rad},0} \sim E/2$. After the ejecta expands by a factor of a few, most of the thermal energy goes into kinetic form, so the average expanding velocity is $v \simeq \sqrt{2E/M}$. Let us take a constant opacity of $\kappa = 0.1 \text{ cm}^2 \text{ g}^{-1}$. For such an expanding ejecta, roughly estimate the luminosity, and duration, and effective temperature near peak luminosity.

P.S., the actual lightcurve of a supernova depends on the effects of hydrogen recombination (affecting the opacity) and radioactive decay of unstable nuclei (affecting the energy). Here, we are ignoring these complications. If you are interested, [Popov 1993](#) discussed the effects of hydrogen recombination.

References

- Rybicki & Lightman, *Radiative processes in astrophysics* (A Wiley-Interscience Publication, New York, 1979).
- Thorne & Blandford, *Modern Classical Physics* (Princeton University Press, New Jersey, 2017).
- Iglesias C. A., Rogers F. J., 1996, ApJ, 464, 943
- Ferguson J. W., Alexander D. R., Allard F., Barman T., Bodnarik J. G., Hauschildt P. H., Heffner-Wong A., Tamai A., 2005, ApJ, 623, 585

- Metzger B. D., Vurm I., Hascoët R., Beloborodov A. M., 2014, MNRAS, 437, 703
Popov D. V., 1993, ApJ, 414, 712

Chapter 3

Special Relativity

This chapter discusses the concept and applications of special relativity.

3.1 Events and Minkowski spacetime

In 3D Euclidean space, we identify a point \mathcal{P} and its position as a vector $\mathbf{x}_{\mathcal{P}}$. The unambiguous meaning of this vector can be specified in two ways: (i) a coordinate-dependent way, e.g., $\mathbf{x}_{\mathcal{P}} = (x_{\mathcal{P}}, y_{\mathcal{P}}, z_{\mathcal{P}}) = (r_{\mathcal{P}}, \theta_{\mathcal{P}}, \phi_{\mathcal{P}})$ in Cartesian and spherical coordinates, respectively; (ii) a coordinate-free way, e.g., $\mathbf{x}_{\mathcal{P}}$ simply standing for the position of Paris, France in the Universe. In the former way, one must also specify the origin and base vectors of the chosen coordinate system. The latter way is called a *geometric* (acronym for “coordinate-free”) description in that the vector does not depend on the choice of coordinate system. This allows us to carry out vectorial analysis (e.g., dot/cross products, divergence, curl) without referring to the coordinates.

When \mathcal{P} and \mathcal{Q} stand for the positions of objects at *two different times*, specifying the spatial positions $\mathbf{x}_{\mathcal{P}}$ and $\mathbf{x}_{\mathcal{Q}}$ does not fully describe what happened — an accurate description of an *event* requires an extra time dimension, so we must deal with a 4D spacetime. In any spacetime, the meaning of an event may also be understood in either the coordinate-dependent or geometric way.

As a concrete example, let us consider two events \mathcal{P} and \mathcal{Q} that are connected by a direct light signal, through the emission and absorption of a given photon. These two events are studied by two observers independently, but one observer in the \mathcal{O}' frame which is moving at a constant velocity \mathbf{v} with respect to the other one in the \mathcal{O} frame. We assume that neither observers are undergoing accelerations, meaning that each sits in their own inertial frames — such a non-accelerating frame is called a *Lorentz frame*. There can be many observers in the same inertial/Lorentz frame, as long as they are not moving with respect

to each other. This can be arranged if they constantly bounce light signals off each other — if the round-trip light-travel times between two observers stay constant, they confirm that there is no relative motion with respect to each other. Using the same method, they can calibrate their own clocks such that they all agree with a *universal coordinate time* in a given Lorentz frame, despite that the observers are located at different positions.

These two observers both agree that \mathcal{P} stands for the emission of the photon and \mathcal{Q} stands for the absorption — this is the geometric way of describing these two events. One may complain that the above geometric descriptions sound a little vague, but it is possible to avoid confusion with other events in the Universe by adding details like the frequency and the nature of the emitter/receiver of the given photon. Here, we treat the photon as a classical particle and ignore the effects due to the Uncertainty Principle as we are interested in the properties of the spacetime.

On the other hand, one may find it more satisfying to describe these two events by specifying the spacetime coordinates in a coordinate system established by a given observer — we will work with this description in this Chapter (while noting that the geometric description has advantages in other contexts). However, a potential drawback is that the coordinate-dependent description is generally different for different observers who might be moving relative to each other and may choose different coordinate systems (e.g., Cartesian, spherical coordinates). This drawback is not too bad if the two observers are in the same Lorentz frame with the same universal coordinate time, so we may only need to transform from spherical to Cartesian coordinates or from one Cartesian system to another rotated/translationally shifted Cartesian coordinate system — these are trivial tasks based on our Euclidean geometry. The difficulty is that, when two observers are moving relative to each other, they do not share the same coordinate time (due to the effect of “time dilation”, see below) and the difference in the time interval between the two events also mixes into the difference in spatial coordinates due to the relative motion between the two observers. In the following, we will exclusively use Cartesian coordinate system for all observers, which has the nice property that the base vectors $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})$ are independent of the position (x, y, z) and time t . We further assume that the x, y, z-axes are parallel between the two Lorentz frames \mathcal{O} and \mathcal{O}' .

Suppose an observer in the \mathcal{O} frame measures the spacetime position of \mathcal{P} (the photon emission event) to be¹

$$(t_{\mathcal{P}}, \mathbf{x}_{\mathcal{P}}),$$

and the same for \mathcal{Q} (the absorption event) at $(t_{\mathcal{Q}}, \mathbf{x}_{\mathcal{Q}})$. Another observer in the a different frame \mathcal{O}' carries out her own time and position measurements and writes down $(t'_{\mathcal{P}}, \mathbf{x}'_{\mathcal{P}})$ and $(t'_{\mathcal{Q}}, \mathbf{x}'_{\mathcal{Q}})$ for the same two events.

¹The spacetime coordinate of an event is generally not a 4-vector (to be defined later).

In Special Relativity, we work with the Minkowski spacetime that has three defining properties: spatial + temporal homogeneity, spatial isotropy, and constancy of the speed of light c . In General Relativity, we know that the spacetime near a massive body is curved and hence not homogeneous or isotropic, but Special Relativity still applies *locally* in a region of size much smaller than the local radius of the spacetime curvature. In fact, the entire Universe can be considered as a mosaic of an infinite number of small patches of locally homogeneous and isotropic Minkowski spacetime. Regarding the constancy of speed of light, the *principle of special relativity* states that, in homogeneous and isotropic spacetime, all physics laws are independent of the choice of Lorentz frames. From the Maxwell equations, which is frame-independent under the above principle, we derived the wave equation in Ch. 1 and showed that the group and phase speeds of EM waves in vacuum are equal to c everywhere in the spacetime. This means that the speed of light is the same in any inertial frame, as confirmed by the [Michelson–Morley experiment](#).

For the two events connected by a light signal, the constancy of the speed of light in both frames \mathcal{O} and \mathcal{O}' means

$$-(c\Delta t)^2 + (\Delta \mathbf{x})^2 = -(c\Delta t')^2 + (\Delta \mathbf{x}')^2 = 0, \quad (3.1)$$

where $\Delta \mathbf{x} = \mathbf{x}_{\mathcal{Q}} - \mathbf{x}_{\mathcal{P}}$ is the spatial separation, $(\Delta \mathbf{x})^2 = \Delta x^2 + \Delta y^2 + \Delta z^2$, Δt is the time separation in frame \mathcal{Q} , and similarly for $\Delta \mathbf{x}'$ and $\Delta t'$ in frame \mathcal{O}' . The above example demonstrates that the 4-dimensional “squared distance” between the two events \mathcal{P} and \mathcal{Q} is independent of the choice of Lorentz frames, and we say that the *spacetime interval* (=“squared distance”) between the two events \mathcal{P} and \mathcal{Q} is *Lorentz invariant* (an acronym for “Lorentz frame-independent”).

For *any* two events \mathcal{P} and \mathcal{Q} in Minkowski spacetime, which are not necessarily connected by a direct light signal, one can generally write their spacetime interval in the two Lorentz frames \mathcal{O} and \mathcal{O}' as follows

$$\begin{aligned} (\Delta s)^2 &= A(c\Delta t)^2 + B_1\Delta x^2 + B_2\Delta y^2 + B_3\Delta z^2 \\ &= A'(c\Delta t')^2 + B'_1(\Delta x')^2 + B'_2(\Delta y')^2 + B'_3(\Delta z')^2, \end{aligned} \quad (3.2)$$

where the squared sum is based on the fact that the four axes are orthogonal to each other. In the following, we first use physical arguments to determine the coefficients A, B_1, B_2, B_3 in frame \mathcal{O} and then similarly in frame \mathcal{O}' . Since the Minkowski spacetime is homogeneous (in both space and time), we know that these four coefficients A, B_1, B_2, B_3 do not depend on the spacetime coordinates. The spatial isotropy of the Minkowski spacetime means $B_1 = B_2 = B_3 = B$, where B is a constant that does not depend on spacetime coordinates. The spacial case of eq. (3.1) tells us $A = -B$, so we arrive at (using similar arguments for the coefficients in frame \mathcal{O}')

$$(\Delta s)^2 = B [-(c\Delta t)^2 + \Delta \mathbf{x}^2] = B' [(c\Delta t')^2 + (\Delta \mathbf{x}')^2]. \quad (3.3)$$

Since the Minkowski spacetime is spatially isotropic, the ratio of the two B-coefficients $\lambda(v) = B/B'$, which describes a possible change of scale, can only depends on the *magnitude* of the relative velocity $v = |\mathbf{v}|$ between the two frames. Additionally, a forward transformation and then an inverse transformation between the two Lorentz frames must give the same physical result, so we obtain $\lambda^2(v) = 1$. The physical solution corresponds to a real and positive scale factor between the two frames, and hence we choose² $\lambda(v) = 1$ and hence $B = B'$.

We choose $B = 1$ so as to be consistent with the Euclidean squared distance at a constant coordinate time $\Delta s^2 = \Delta x^2 + \Delta y^2 + \Delta z^2$ (for $\Delta t = 0$). Some other textbooks may take $B = -1$ or even $B = 1/c^2$, and in fact, the choice of the constant B only reset the units of the spacetime interval but has no physical significance.

Therefore, we conclude that the spacetime interval between any two events is given by

$$(\Delta s)^2 = -(c\Delta t)^2 + \Delta x^2 + \Delta y^2 + \Delta z^2 = (c\Delta t')^2 + (\Delta x')^2 + (\Delta y')^2 + (\Delta z')^2. \quad (3.4)$$

which is Lorentz invariant — this allows us to compute $(\Delta s)^2$ in any frame and then use it in any other frames.

Instead of two widely separated events, it is often more convenient to consider an infinitesimal separation (for reasons below) and write

$$ds^2 = -c^2 dt^2 + dx^2 + dy^2 + dz^2. \quad (3.5)$$

An equivalent way of expressing the above spacetime interval is to first define a *4-vector*,

$$d\vec{x} = (c dt, d\mathbf{x}), \quad (3.6)$$

which is called the *4-displacement vector*. Hereafter, we may use the coordinate form (A^μ) interchangeably with the vector form \vec{A} to denote a 4-vector, whereas the usual 3-vector in Euclidean space is denoted as \mathbf{A} . Then, we define the rule for the dot-product of the 4-displacement vector with itself as

$$ds^2 = d\vec{x} \cdot d\vec{x} = \eta_{\mu\nu} dx^\mu dx^\nu, \quad (3.7)$$

where the spacetime indices³ μ, ν go from 0 to 3. Hereafter, we follow the Einstein summation convention — Greek indices are summed over 0 to 3 when they are repeated in a single term with one up and one down. Such a summation is called *index contraction* —

²The choice of $\lambda(v) = -1$ would lead to the following unphysical result: even if two Lorentz frames \mathcal{O} and \mathcal{O}' only has an arbitrarily small relative motion $v \approx 0$, the spacetime interval $(\Delta s)^2$ obtained in these two (nearly identical) frames would flip sign.

³We following the convention that Greek indices (α, β , etc) represent a number from 0 to 3, whereas Latin indices (i, j, k , etc) only stand for 1, 2, 3 — the spatial components.

a term with all indices contracted is a Lorentz invariant *scalar* (simply a number). Here $(\eta_{\mu\nu})$ is called the *Minkowski metric*, which may be understood in a geometric way as a (rank-2) tensor or in a coordinate-dependent way as a 4×4 matrix. All the properties of the spacetime (e.g., Lorentz transformation, the effects of time dilation and Doppler beaming, see later) is contained in eq. (3.7) and hence ds^2 is called a *line element* of the Minkowski spacetime as it describes the “squared length” of the 4-displacement vector $d\vec{x}$.

Full-scale tensor analysis is not needed for special relativity. Since the Minkowski metric is particularly simple, when lowering and raising indices, we only need to apply the simple “sign flip if temporal” rule (the name following Thorne & Blandford’s book): when one raises/lowers a temporal (0) index, the sign changes; whereas raising/lowering a spatial (1, 2, 3) index does not change the numerical value of a component. For instance, $T^{ij} = T^i_j = T_{ij}$, $T^{0j} = -T_0^j = -T_{0j}$, where Latin indices only stand for the spatial components. When dealing with curved spacetime in general relativity, one will recognize that $(\eta^{\mu\nu})$ is simply the inverse-matrix of $(\eta_{\mu\nu})$, but for our purpose here, the sign-flip-if-temporal rule is sufficient.

In Cartesian coordinates, the components of $(\eta_{\mu\nu})$ are

$$(\eta_{\mu\nu}) = \text{diag}(-1, 1, 1, 1) = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (3.8)$$

For spherical coordinates $(x^1, x^2, x^3) = (r, \theta, \phi)$ (where θ and ϕ are the polar and azimuthal angles), the components of the Minkowski metric are $(\eta_{\mu\nu}) = \text{diag}(-1, 1, r, r \sin \theta)$. One must be careful with curvilinear (e.g., spherical or cylindrical) coordinates, because the base vectors are only defined *locally* — unlike the Cartesian basis which are global. This is why it is more convenient to use the infinitesimal 4-displacement vector $d\vec{x}$ to define the spacetime interval ds^2 . This chapter will mainly be based on Cartesian coordinates, so do not worry about the curvilinear form for now.

Similar to eq. (3.7), the dot-product between any two 4-vectors \vec{A} and \vec{B} is defined as

$$\vec{A} \cdot \vec{B} = \eta_{\mu\nu} A^\mu B^\nu. \quad (3.9)$$

We will later show that all other 4-vectors can be defined based on the 4-displacement vector $d\vec{x}$ and that the components of 4-vectors (A^μ) satisfy the Lorentz transformation from one Lorentz frame to another (which ensures the laws of physics are the same in all frames). A given 4-vector \vec{A} is called *time-like* if $\vec{A} \cdot \vec{A} < 0$, *space-like* if $\vec{A} \cdot \vec{A} > 0$, and *null* if $\vec{A} \cdot \vec{A} = 0$.

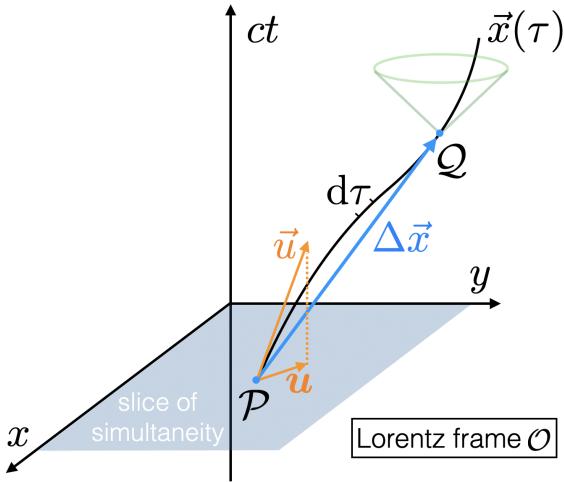


Figure 3.1: In a Lorentz frame \mathcal{O} , the worldline (or spacetime trajectory) of a particle is described $\vec{x}(\tau)$, where τ is the proper time if the particle has non-zero mass. We have omitted the third spatial dimension z , which is okay as long as the particle does not move in that direction (i.e., $z = \text{const}$). The 4-velocity is defined as the tangent of the worldline $\vec{u} = d\vec{x}/d\tau$, the spatial part of which is $\mathbf{u} = d\mathbf{x}/d\tau$. A massive particle's future worldline is confined within the lightcone (shown in green). The law of causality requires that physical motions of particles must go forward in time in any Lorentz frame. For a massless particle, which moves on the surface of the lightcone, instead of 4-velocity, one should work with its 4-momentum $\vec{p} = d\vec{x}/d\lambda$, where λ is a renormalized version of the proper time — called an *affine parameter*. The slice of simultaneity (3-space at the same time coordinate t) is shown as a blue plane — all events on this plane are simultaneous as measured by all observers in this Lorentz frame. The coordinate separation between two events \mathcal{P} and \mathcal{Q} along the worldline is $\Delta\vec{x} = \vec{x}_{\mathcal{Q}} - \vec{x}_{\mathcal{P}} = (c\Delta t, \Delta\mathbf{x})$, but the physical spacetime interval is given by $(\Delta s)^2 = -(c\Delta t)^2 + (\Delta\mathbf{x})^2 = -(c\Delta\tau)^2 = -c^2(\tau_{\mathcal{Q}} - \tau_{\mathcal{P}})^2$.

3.2 Proper time, 4-velocity, and Lorentz factor

A particle's trajectory in the Minkowski spacetime is called its *worldline*, which is illustrated in Fig. 3.1 along with other concepts to be discussed in this section: proper time, 4-velocity, lightcone, slice of simultaneity, spacetime interval, and Lorentz frame.

Let us consider a particular event \mathcal{P} and another adjacent event at a displacement $d\vec{x}$ from \mathcal{P} along the worldline a massive particle (the spacial case of a massless particle will be discussed later). The spacetime interval ds^2 between these two events is given by eq. (3.5). As ds^2 is Lorentz invariant, we can calculate it in any Lorentz frame — a special one is the *momentary rest frame* (or *comoving frame*) of the particle near event \mathcal{P} . In this momentary rest frame denoted as \mathcal{O}' , we write the 4-displacement vector as $d\vec{x}' = (c dt', dx', dy', dz')$ and immediately realize that $dx' = dy' = dz' = 0$ as the particle is at rest in frame \mathcal{O}' . Since the momentary is a special frame (and also the simplest), we call the time coordinate in this frame the *proper time* and denote it as $\tau = t'$ — the time recorded by a clock that a particle carries along its worldline. Therefore, the spacetime interval of the 4-displacement vector is given by

$$ds^2 = -c^2 d\tau^2. \quad (3.10)$$

This shows that the proper time of a particle can be defined by the spacetime interval along its worldline

$$d\tau = c^{-1} \sqrt{-ds^2}, \quad \text{and} \quad \tau = c^{-1} \int \sqrt{-ds^2}, \quad (3.11)$$

provided that the particle moves at a sub-luminal speed $v < c$ in any Lorentz frame with $(dx)^2 < (cdt)^2$ and hence τ is real. It can be shown (by studying the relativistic equation of motion) that a non-zero rest mass⁴ $m \neq 0$ is a sufficient and necessary condition for sub-luminal speeds $v < c$ in all Lorentz frames. We see that the proper time of a given massive particle is Lorentz invariant. Note that the proper time is well defined even for an object undergoing acceleration, because at each moment, one can always go to its momentary rest frame in which the object is temporarily at rest and its velocity only increases quadratically with time $\propto d\tau^2$ and quadratic terms can be dropped when taking infinitesimal changes $d\tau$ in the integral in eq. (3.11).

The *4-velocity* of a particle with non-zero rest mass is defined as

$$\vec{u} \equiv d\vec{x}/d\tau = (u^0, \mathbf{u}), \quad (3.12)$$

⁴The *rest mass* of a given particle is the mass measured in its momentary rest frame. The rest mass may be measured by the ratio between a force (due to e.g., an electric field) and the corresponding acceleration, or by the gravitational potential produced by the particle. Einstein's Equivalence Principle states that the effects of acceleration and gravity (due to spacetime curvature in General Relativity) are indistinguishable and hence the two mass measurements are identical. As long as the particle does not spontaneously lose mass by decaying to a lower energy level in its comoving frame, its rest mass is Lorentz invariant.

where the temporal component is $u^0 = dt/d\tau$ and the spatial components are $\mathbf{u} = d\mathbf{x}/d\tau$. Since $d\tau$ is Lorentz invariant, we know that \vec{u} is a 4-vector which differs from the 4-displacement vector $d\vec{x}$ only by a scalar factor. The usual 3-velocity is given by

$$\mathbf{v} = \frac{d\mathbf{x}}{dt} = \frac{d\mathbf{x}}{d\tau} \frac{d\tau}{dt} = \frac{\mathbf{u}}{u^0} c. \quad (3.13)$$

The dot-product of 4-velocity with itself is given by

$$\vec{u}^2 \equiv \vec{u} \cdot \vec{u} = -(u^0)^2 + \mathbf{u}^2 = \frac{\eta_{\mu\nu} dx^\mu dx^\nu}{d\tau^2} = \frac{ds^2}{d\tau^2} = -c^2. \quad (3.14)$$

We find that $\vec{u}^2 < 0$ so the 4-velocity of a massive particle is always time-like. We then insert $\mathbf{u}^2 = v^2 c^2 / (u^0)^2$ into the above equation and solve for the temporal component of the 4-velocity

$$\frac{dt}{d\tau} = u^0/c = \gamma = \frac{1}{\sqrt{1 - \beta^2}}, \quad (3.15)$$

where $\beta \equiv v/c$ is the dimensionless 3-velocity. The γ factor above is called the *Lorentz factor* of the particle, which is equal to the ratio between the coordinate time interval and the proper time interval. Since $\gamma \geq 1$, the coordinate time interval dt is always longer or equal to the corresponding proper time interval $d\tau$ — this is the effect of *time dilation*. Finally, the time and spatial components of the 4-velocity can be written as

$$\vec{u} = \gamma(c, \mathbf{v}), \quad \text{or} \quad u^0 = \gamma c, \mathbf{u} = \gamma \mathbf{v}, \quad (3.16)$$

where we have made use of the Lorentz factor γ .

3.3 Lorentz transformation

A given 4-vector, e.g., the 4-displacement vector $d\vec{x}$ or 4-velocity, is a geometric object whose nature does not depend on the choice of Lorentz frame or coordinate system in a given frame. If we would like to describe a 4-vector in the matrix form, e.g., $(dx^\mu) = (cdt, dx, dy, dz)$, its components would be frame-dependent (and also coordinate-dependent, but we always adopt Cartesian coordinates and do not worry about the coordinate transformation here). For a given 4-vector, its components in different Lorentz frames must be related in a way that only depends on the properties of the Minkowski spacetime. *Lorentz transformation* tells us how the components of any given 4-vector transform from one Lorentz frame to another.

Without losing generality, let us consider that the frame \mathcal{O}' is moving at a velocity $\mathbf{v} = \beta c = \beta c \hat{\mathbf{x}}$ along the $\hat{\mathbf{x}}$ direction as viewed in the frame \mathcal{O} . The Cartesian axes in both frames are parallel to each other. In the following, we make use of the Lorentz invariance

of the Minkowski spacetime interval ds^2 to derive the Lorentz transformation for the 4-displacement vector $d\vec{x}$ between the \mathcal{O} and \mathcal{O}' frames.

First, since the relative motion between the two frames is along the \hat{x} direction, the other two orthogonal directions \hat{y} and \hat{z} are unaffected (as the Minkowski spacetime is spatially isotropic), so we must have

$$dy' = dy, \quad dz' = dz. \quad (3.17)$$

The Lorentz transformation is linear so as to allow a forward transformation followed by an inverse transformation to recover the original 4-vector components. Thus, we write the transformation in the following general form

$$\begin{aligned} c dt' &= A_0 c dt + A_1 dx, \\ dx' &= B_0 c dt + B_1 dx, \end{aligned} \quad (3.18)$$

where the unknown coefficients A_0, A_1, B_0, B_1 only depend on the relative velocity β but not on the coordinates (as the Minkowski spacetime is homogeneous).

Let us first consider a simple case. For a particle that is at rest in frame \mathcal{O}' , we write

$$dt' = d\tau, \quad dx' = 0, \quad (3.19)$$

where $d\tau$ is the proper time of the particle. From the rule of time dilation (eq. 3.15), we know that the time interval in frame \mathcal{O} is given by $dt = \gamma d\tau$, and the particle travels a distance $dx = \beta c dt = \gamma \beta c d\tau$ over this time interval. Plugging all these into the general linear form for Lorentz transformation, we obtain

$$\begin{aligned} cd\tau &= A_0 c \gamma d\tau + A_1 \gamma \beta c d\tau, \\ 0 &= B_0 c \gamma d\tau + B_1 \gamma \beta c d\tau. \end{aligned} \quad (3.20)$$

These lead to

$$A_0 = \gamma^{-1} - \beta A_1, \quad B_0 = -\beta B_1, \quad (3.21)$$

and hence our Lorentz transformation becomes

$$\begin{aligned} c dt' &= (\gamma^{-1} - \beta A_1) c dt + A_1 dx, \\ dx' &= B_1 (-\beta c dt + dx), \end{aligned} \quad (3.22)$$

We need two more equations to determine A_1 and B_1 . We then plug these into the Lorentz invariant spacetime interval ds^2 and obtain

$$\begin{aligned} ds^2 &= -c^2 dt^2 + dx^2 + dy^2 + dz^2 = -c^2 (dt')^2 + (dx')^2 + (dy')^2 + (dz')^2 \\ &= -[(\gamma^{-1} - \beta A_1) c dt + A_1 dx]^2 + B_1^2 (-\beta c dt + dx)^2 + dy^2 + dz^2 \\ &= [-(\gamma^{-1} - \beta A_1)^2 + \beta^2 B_1^2] c^2 dt^2 + (-A_1^2 + B_1^2) dx^2 \\ &\quad - 2 [(\gamma^{-1} - \beta A_1) A_1 + \beta B_1^2] c dt dx + dy^2 + dz^2. \end{aligned} \quad (3.23)$$

The coefficients in front of each of the terms (dt^2 , dx^2 , dy^2 and dz^2) must be identical, so we obtain (one of the following is redundant)

$$\begin{aligned} -1 &= -(\gamma^{-1} - \beta A_1)^2 + \beta^2 B_1^2, \\ 1 &= -A_1^2 + B_1^2, \\ 0 &= (\gamma^{-1} - \beta A_1)A_1 + \beta B_1^2. \end{aligned} \quad (3.24)$$

and then we can solve for A_1, B_1 and hence A_0, B_0 as follows

$$A_0 = \gamma, \quad A_1 = -\gamma\beta, \quad B_0 = -\gamma\beta, \quad B_1 = \gamma, \quad (3.25)$$

where we have used the identity $\gamma^2\beta^2 = \gamma^2 - 1$ in an intermediate step.

With the coefficients determined, we then write down the Lorentz transformation for *any* 4-vector $\vec{x} = (x^\mu)$ between the two frames \mathcal{O} and \mathcal{O}'

$$\begin{pmatrix} x^{0'} \\ x^{1'} \\ x^{2'} \\ x^{3'} \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix}, \text{ or more concisely } x^{\mu'} = L^{\mu'}_\nu x^\nu, \quad (3.26)$$

or explicitly

$$\begin{aligned} x^{0'} &= \gamma(x^0 - \beta x^1), \\ x^{1'} &= \gamma(-\beta x^0 + x^1), \\ x^{2'} &= x^2, \\ x^{3'} &= x^3. \end{aligned} \quad (3.27)$$

The *Lorentz transformation matrix* $L^{\mu'}_\nu$ only depends on the orientation and relative motion between the two frames but does not depend on what is happening in each frame.

The most important property of the Lorentz transformation is that all laws of physics in Minkowski spacetime is independent of the choice of inertial frames as long as all 4-vectors in each frame are transformed from each other according to eq. (3.26) — the fundamental reason is that the Lorentz invariance of the spacetime interval ds^2 is preserved. The negative sign in front of the $\gamma\beta$ terms in eq. (3.26) comes from the fact that a passing train runs slower when you are in a frame that tries to catch up with it. Note that the *inverse transformation* would simply reverse the direction of the relative velocity — this is because for observer's in the \mathcal{O}' frame, the \mathcal{O} frame is moving at velocity $-\mathbf{v}$. A forward Lorentz transformation and then an inverse one would return the original 4-vector, and this means

$$L^\alpha_{\mu'} L^{\mu'}_\beta = \delta^\alpha_\beta, \quad (3.28)$$

where $\delta_\beta^\alpha = 0$ when $\alpha \neq \beta$ and $\delta_\beta^\alpha = 1$ for $\alpha = \beta$.

In the next section, we discuss various 4-vectors, whose components in different Lorentz frames satisfy the Lorentz transformation.

3.4 Four-vectors

All 4-vectors in Minkowski spacetime are derived from the 4-displacement vector $d\vec{x}$, which can be considered as the most fundamental 4-vector of all.

In §3.2, we have already introduced the 4-velocity $\vec{u} = d\vec{x}/d\tau$ based on the proper time τ of a massive particle, and its spacetime components are $(u^\mu) = \gamma(c, \mathbf{v})$ with $\mathbf{v} = dx/dt = \beta c$ being the usual 3-velocity and $\gamma = (1 - \beta^2)^{-1/2}$ being the Lorentz factor. The 4-velocity is always time-like as its squared length is given by $\vec{u} \cdot \vec{u} = -c^2 < 0$. We can further define the particle's *4-momentum* as its 4-velocity multiplied by the rest mass m ,

$$\vec{p} = m\vec{u} = \gamma m(c, \mathbf{v}). \quad (3.29)$$

Since the rest mass m is Lorentz invariant, we know that \vec{p} is also a 4-vector. It is also straightforward to show $\vec{p}^2 \equiv \vec{p} \cdot \vec{p} = -m^2 c^2$, which is also time-like for $m \neq 0$.

In the following, we use a simple example to demonstrate that the time component of the 4-momentum p^0 is related to the total energy E and the spatial components $\mathbf{p} = (p^1, p^2, p^3)$ corresponds to the usual 3-momentum, as follows

$$\vec{p} = (E/c, \mathbf{p}), \quad \text{or} \quad E = \gamma mc^2, \quad \mathbf{p} = \gamma m\mathbf{v}. \quad (3.30)$$

Let us consider that the particle in consideration has charge q and is moving in a uniform electric field $\mathbf{E} = E\hat{\mathbf{x}}$ (one should not confuse the two E 's) that is parallel to the particle's velocity \mathbf{v} , i.e., $\mathbf{E} \parallel \mathbf{v}$. In the momentary rest frame of the particle, the electric field is given by $\mathbf{E}' = E'\hat{\mathbf{x}}' = E\hat{\mathbf{x}}$ (we have used $E' = E$ according to the Lorentz transformation of electric field, see §3.6.2). Over a time interval $d\tau$ in the comoving frame \mathcal{O}' (or $dt = \gamma d\tau$ in the lab-frame \mathcal{O}), the particle gains velocity

$$dv' = \frac{Eq}{m} d\tau$$

according to acceleration $a' = dv'/d\tau = Eq/m$. This corresponds to a 4-velocity

$$d\vec{u}' \approx (c, dv', 0, 0)$$

in Cartesian components in the comoving frame, where we have ignored 2nd-order terms $\mathcal{O}(d\tau^2)$ and taken the Lorentz factor to be 1 in the comoving frame. We then (inverse)

Lorentz transform $d\vec{u}'$ to the lab-frame \mathcal{O} where the particle was originally moving with 4-velocity $\vec{u} = \gamma(c, v, 0, 0)$ at coordinate time t , and the 4-velocity at time $t + dt$ is

$$\vec{u} + d\vec{u} = (\gamma c + \gamma \beta dv', \gamma \beta c + \gamma dv', 0, 0),$$

so we obtain the change in 4-momentum within the time interval of dt

$$d\vec{p} = m d\vec{u} = Eq dt(\beta, 1, 0, 0), \quad (3.31)$$

where we have used $dt = \gamma d\tau$. The spatial part of eq. (3.31) gives the equation of motion $d\mathbf{p}/dt = \mathbf{E}q$ in frame \mathcal{O} . If we then take the dot product with \mathbf{v} and make use of the following identity $d(\gamma\mathbf{v}) \cdot \mathbf{v} = v^2 d\gamma + (1/2)\gamma dv^2 = c^2 d\gamma$, this leads to

$$\frac{d\mathbf{p}}{dt} \cdot \mathbf{v} = m \frac{d(\gamma\mathbf{v})}{dt} \cdot \mathbf{v} = \frac{d\gamma}{dt} mc^2,$$

which reproduces the temporal part of eq. (3.31).

The above is a special case with the electric field along the $\hat{\mathbf{x}}$ direction. It is possible to show that, for a general electric field \mathbf{E} along an arbitrary direction in frame \mathcal{O} , the relativistic equation of motion is given by (see §3.6.2)

$$\frac{d\mathbf{p}}{dt} = q\mathbf{E}, \quad (3.32)$$

and the corresponding energy change can be obtained by dot-producing the above equation of motion with \mathbf{v} , and the result is an energy equation

$$\frac{d\gamma}{dt} mc^2 = q\mathbf{E} \cdot \mathbf{v}. \quad (3.33)$$

This clearly shows that γmc^2 is the energy of a particle, because $d\mathbf{x} = \mathbf{v}dt$ is the positional change of the particle and $q\mathbf{E} \cdot \mathbf{v}dt = q\nabla\phi \cdot d\mathbf{x} = q d\phi$ is the change in potential energy, where the electric potential ϕ is defined as $\mathbf{E} = \nabla\phi$.

We conclude that the temporal part of the 4-momentum is related to the total energy of a particle given by Einstein's famous mass-energy equation $E = \gamma mc^2$, where γm is the total relativistic mass, mc^2 is the *rest-mass energy*, and $(\gamma - 1)mc^2$ is the *kinetic energy*, and the spatial part of the 4-momentum is the relativistic momentum $\mathbf{p} = \gamma m\mathbf{v}$. Using the identity $\gamma^2\beta^2 = \gamma^2 - 1$, we obtain the *energy-momentum relation*

$$E^2 = p^2 c^2 + m^2 c^4. \quad (3.34)$$

The above discussion is for a massive particle with non-zero rest mass.

A massless particle can be considered as a quantum (e.g., a photon) associated with a wave packet with energy $\hbar\omega$ and momentum $\hbar\mathbf{k}$, where ω is the angular frequency and \mathbf{k} is the wavevector. Thus, its 4-momentum is given by

$$\vec{p} = \hbar(\omega/c, \mathbf{k}), \quad (3.35)$$

from which we can define the *4-wavevector*

$$\vec{k} = (\omega/c, \mathbf{k}). \quad (3.36)$$

The physical meaning of a 4-wavevector is a plane EM wave with angular frequency ω , wavenumber $k = 2\pi/\lambda$ (λ = wavelength), propagating along the direction of $\hat{\mathbf{n}} = \mathbf{k}/k$. The fact that $\vec{p} \cdot \vec{p} = -m^2c^2 = 0$ for a massless particle means that the angular frequency is related to the wavenumber by $\omega^2 = k^2c^2$ (which we already know as the dispersion relation for vacuum EM waves).

In the following, we discuss a few more 4-vectors.

For massive particles the *4-acceleration* is defined as

$$\vec{a} = d\vec{u}/d\tau = d^2\vec{x}/d\tau^2, \quad (3.37)$$

and the *4-force* is defined similar to Newton's 2nd law as

$$\vec{F} = m\vec{a} = d\vec{p}/d\tau = m d^2\vec{x}/d\tau^2. \quad (3.38)$$

Since $d\tau$ is Lorentz invariant, it is easy to understand that \vec{F} and \vec{a} are 4-vectors and obey the Lorentz transformation.

Instead of a single particle, let us consider a large number of particles in a fluid. Suppose the particles in a local region of a given fluid body has an average velocity \mathbf{v} in the lab-frame \mathcal{O} . In the frame \mathcal{O}' that is comoving with the average velocity \mathbf{v} , the particles have random motions with a zero mean value. For simplicity, suppose the random motions are sufficiently slow such that the number density of particles $n' = n_0$ in the comoving frame \mathcal{O}' remains constant in the time interval of interest to us, where n_0 is known as the *comoving number density*. Note that random motions may cause the fluid to expand or contract, causing the comoving number density n_0 to change on the dynamical timescale t'_{dy} in frame \mathcal{O}' , but here, we are focusing on the relativistic kinematics regarding the bulk motion of the system on timescales much shorter than the dynamical timescale t'_{dy} .

We would like to know the particle number density n in the lab frame. To calculate this, let us consider dN particles that are contained in a volume $dV' = dx'dy'dz'$ in the comoving frame \mathcal{O}' . Note that the positions of the dN particles are obtained *simultaneously* in frame \mathcal{O}' at the same coordinate time t' . If we project all dN particles onto the $\hat{\mathbf{x}}'$ -axis, the separation between rightmost particle and the leftmost particle is given by dx' at a given

global coordinate time t' — the particles have negligible motions in frame \mathcal{O}' anyway, so the condition of simultaneity does not make a difference here. Our goal is to obtain the volume occupied by the dN particles in frame \mathcal{O} . Again, we project all dN particles onto the \hat{x} -axis, the separation between rightmost particle and the leftmost particle is given by dx at a given coordinate time t . The potentially confusing point is that, the number density in any given frame is defined by the positions of all particles *simultaneously* in that frame. To compute dx , we must use $dt = 0$ as we are considering the positions of particles simultaneously in frame \mathcal{O} . According to Lorentz transformation

$$dt' = \gamma(\gamma c dt - \beta dx), \quad dx' = \gamma(-\beta c dt + dx) \quad (3.39)$$

and using $dt = 0$, we obtain

$$dx' = \gamma dx. \quad (3.40)$$

This is the effect of *length contraction*. Since $dy' = dy$ and $dz' = dz$, we see that the number density in frame \mathcal{O} is a factor of γ higher than the comoving number density,

$$n = \gamma n' = \gamma n_0. \quad (3.41)$$

Based on this conclusion, we can define the *4-number flux* as

$$\vec{N} = n(c, \mathbf{v}) = \gamma n_0(c, \mathbf{v}) = n_0 \vec{u}, \quad (3.42)$$

where $n\mathbf{v}$ is the 3-number flux vector (the number of particles per unit area per unit time passing through a given surface whose normal direction is along \mathbf{v}), n_0 is the comoving number density (the Lorentz invariant number density of particles in the rest-frame of their average velocity), and \vec{u} is their average 4-velocity.

Analogous to the 4-number flux, if the particles each has charge q , we can define the *4-current density* as

$$\vec{J} = \vec{N}q = (\rho c, \mathbf{J}) = n_0 q \vec{u}, \quad (3.43)$$

where $\rho = nq$ is the charge density and $\mathbf{J} = nq\mathbf{v}$ is the 3-current density. Note that the charge q of a given particle is Lorentz invariant as it is an intrinsic property similar to the rest-mass. The above 4-current density is based on a given particle species, and it can be extended to the more general case consisting of multiple species using the total charge density $\rho = \sum_j n_j q_j$ and total current density $\mathbf{J} = \sum_j n_j q_j \mathbf{v}_j$, so we write

$$\vec{J} = \sum_j n_{0,j} q_j \vec{u}_j, \quad (3.44)$$

where the sum accounts for all charge species j .

The 4-vectors in the above discussion will be frequently used in the rest of this book, so the readers are encouraged to gain a good understanding of them. There are many more 4-vectors that we have not discussed, but a general theme is that they are descendants of the 4-displacement vector $d\vec{x}$ and that they all obey the Lorentz transformation.

3.5 Various relativistic effects

In this section, we discuss some very interesting relativistic effects in the Minkowski space-time. Before doing that, let us motivate ourselves by looking at a class of astrophysical sources showing clear evidence of relativistic motion.

3.5.1 Apparent motion of a moving source

Superluminal apparent motion is a counter-intuitive effect that has been observed in some astrophysical sources and they provide the first evidence for collimated relativistic jets pointing near our line of sight. A famous example is 3C 279 as shown in the left panel of Fig. 3.2. In fact, no knowledge about special relativity is needed to explain the superluminal apparent motion in these sources, but it demonstrates relativistic physical motion.

As shown in Fig. 3.3, let us consider a distant emitting point source that is moving at velocity $\mathbf{v} = \beta c$ at an angle θ away from our line of sight. The source emitted two photons in a row along our line of sight at time t_1 and t_2 in the lab frame. In this time interval, the source has moved a distance $v(t_2 - t_1)$ in physical space, which can be decomposed into $\Delta\ell_{\parallel} = v(t_2 - t_1) \cos \theta$ along our line of sight and $\Delta\ell_{\perp} = v(t_2 - t_1) \sin \theta$ perpendicular to our line of sight. By looking at the separation between the two photons along the line of sight $c(t_2 - t_1) - v(t_2 - t_1) \cos \theta$, we see that the two photons will arrive at the observer with a time separation

$$\Delta t_{\text{obs}} = (1 - \beta \cos \theta)(t_2 - t_1). \quad (3.45)$$

Often it is only possible to observe the transverse motion of the source — this is known as the *proper motion*⁵. An observer would measure an apparent speed of proper motion

$$\beta_{\text{app}} = \frac{\Delta\ell_{\perp}}{c\Delta t_{\text{obs}}} = \frac{\beta \sin \theta}{1 - \beta \cos \theta}. \quad (3.46)$$

It is easy to show that the above apparent motion can be faster than the speed of light (meaning $\beta_{\text{app}} > 1$) when the source is moving at an ultra-relativistic speed. For a given true speed β , the apparent speed is maximized at a critical viewing angle θ_{crit} given by

$$\cos \theta_{\text{crit}} = \beta, \quad \text{or} \quad \sin \theta_{\text{crit}} = \gamma^{-1}, \quad (3.47)$$

which is obtained by solving $\partial \beta_{\text{app}} / \partial \theta = 0$. Here, $\gamma = (1 - \beta^2)^{-1/2}$ is the Lorentz factor. At this critical angle, the maximum apparent speed is

$$\beta_{\text{app,max}} = \beta_{\text{app}}(\theta = \theta_{\text{crit}}) = \gamma\beta, \quad (3.48)$$

⁵A celestial object's apparent position on the sky shifts due to its intrinsic motion as well as Earth's rotation and orbital motion around the Sun. When one subtracts the effects due to observer's motion, the leftover is the intrinsic motion of a celestial object projected on the sky called the “proper motion”. It is unrelated to the proper time τ (eq. 3.11).

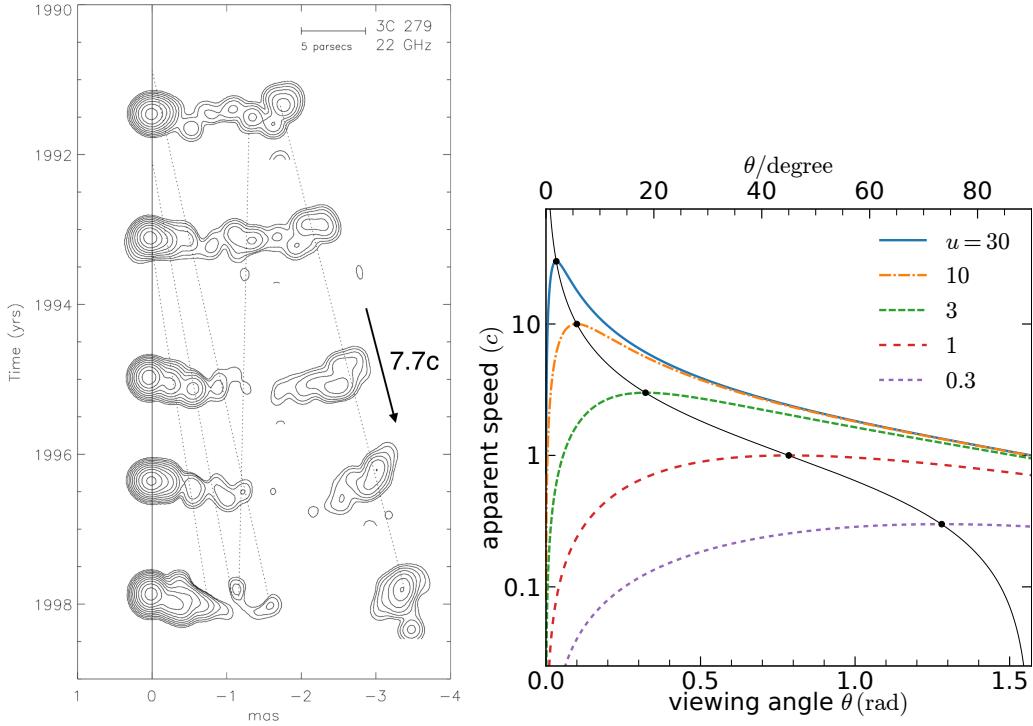


Figure 3.2: *Left panel:* Wehrle et al. 2001 presented 22 GHz radio images of 3C279, which is an active galactic nuclei (i.e., an accreting supermassive black hole) with a relativistic jet pointing near our line of sight. The jet has a stationary core that is near the position of the black hole marked by “0” in the horizontal axis. The location of the rightmost knot moved about 1.5 milliarcseconds (mas) in about 6.5 years in observer’s time. The source is located at redshift of 0.54. According to the standard Λ CDM cosmology given by the Planck satellite, 1.5 mas corresponds to a projected distance of 32 light years in the local frame of the host galaxy, and 6.5 years of time span corresponds to 4.2 years in the host galaxy’s rest-frame. One concludes that the emitting gas has an apparent superluminal speed of about $7.7c$. *Right panel:* Apparent speed as a function of viewing angle θ for different source 4-speed $u = \gamma\beta$. The black dots show the maximum apparent speed for each curve, and the maximum value of $\beta_{app,max} = u$ is achieved at a critical viewing angle $\theta_{crit} = \arcsin(\gamma^{-1})$.

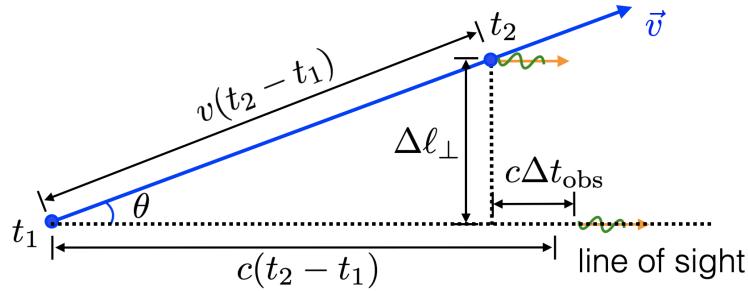


Figure 3.3: Geometry for analyzing the apparent motion of an emitter (blue dot).

We conclude the maximum apparent speed of an emitting source is given by its 4-speed $u = \gamma\beta$. Turning the argument around, if an observer measures an apparent speed of β_{app} , she immediately knows that the true 4-speed of the source must be greater than β_{app} . On the other hand, if we somehow know the true speed of the source, the apparent speed can be used to measure the viewing angle of the source.

Let us consider the ultra-relativistic limit of $\gamma \gg 1$ and the observer can only see bright emission at viewing angles $\theta \lesssim \gamma^{-1} \ll 1$ as a result of relativistic beaming (see Fig. 3.14 later). In this limit, we adopt the Taylor expansions of $\sin \theta \approx \theta$, $\cos \theta \approx 1 - \theta^2/2$, and $\beta \approx 1 - 1/(2\gamma^2)$, and the apparent speed becomes

$$\beta_{\text{app}} \approx \frac{2\gamma}{\gamma\theta + (\gamma\theta)^{-1}}, \text{ for } \gamma \gg 1 \text{ and } \theta \ll 1. \quad (3.49)$$

We see that if the observer is located at an angle $\theta = \gamma^{-1} \approx \theta_{\text{crit}}$, then the apparent speed is $\beta_{\text{app}} \approx \gamma$, which is indeed superluminal.

3.5.2 Doppler factor

Consider a moving emitter and an observer at rest in the lab-frame. The ratio between the observed frequency (in the observer's frame = lab-frame) and the emitted frequency (in the source's comoving frame) is called *the Doppler factor* and denoted as \mathcal{D} . As we will show later in §3.7, the Doppler factor is highly useful when one studies the properties (e.g., the specific intensity I_ν) of a radiation field in different Lorentz frames.

Going back to the geometry shown in Fig. 3.3 where an emitter is moving with velocity $v = \beta c$ at an angle θ away from the observer's line of sight. Two photons emitted with a lab-frame time interval $\Delta t = t_2 - t_1$ will arrive at the observer separated by $\Delta t_{\text{obs}} = (1 - \beta \cos \theta)\Delta t$ (eq. 3.45). From the effect of time dilation (eq. 3.15), we know that Δt is related to the proper time interval $\Delta\tau = \tau_2 - \tau_1$ by $\Delta t = \gamma\Delta\tau$, so we obtain

$$\Delta t_{\text{obs}} = \Delta\tau/\mathcal{D}, \quad (3.50)$$

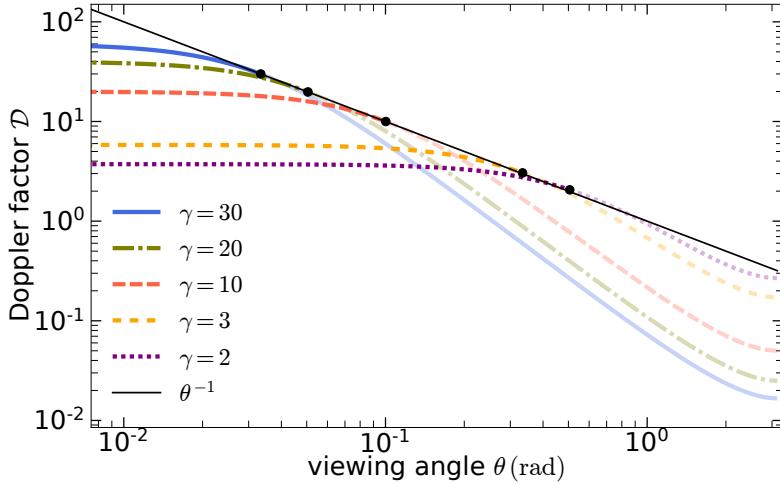


Figure 3.4: Doppler factor as a function of viewing angles, for different Lorentz factors (different line colors). The black dots mark the edge of the beaming cone $\theta = \gamma^{-1}$. The thin black line shows $D = \theta^{-1}$, which is the approximate result for $\theta = \gamma^{-1}$ in the ultra-relativistic limit.

where \mathcal{D} is the Doppler factor (as shown in Fig. 3.4)

$$\mathcal{D} = \frac{1}{\gamma(1 - \beta \cos \theta)}. \quad (3.51)$$

A physically intuitive way of understanding eq. (3.50) is as follows. If a wave train emitted between t_1 and t_2 consists of N pulses (and N may be any positive real number), the frequency of pulsations in the emitter's comoving frame is $\nu' = N/\Delta\tau$ and the frequency in the lab-frame is $\nu = N/\Delta t_{\text{obs}}$, where N is Lorentz invariant as it is based on simply counting. Therefore, we see that the two frequencies are related by

$$\nu = \mathcal{D}\nu', \quad (3.52)$$

so \mathcal{D} indeed describes the Doppler shift.

In the non-relativistic limit of $\beta \ll 1$, the Doppler factor is given by

$$\mathcal{D} \approx 1 + \beta \cos \theta, \quad \text{or} \quad \nu/\nu' \approx 1 + \frac{v_{\parallel}}{c}, \quad \text{for } \gamma \approx 1, \quad (3.53)$$

where $v_{\parallel} = v \cos \theta$ is the projected velocity along the line of sight. The above result is in agreement with the non-relativistic rule of Doppler shift — the frequency of a car's sound increases as it approaches the observer, resulting in a higher perceived pitch. In the

ultra-relativistic limit of $\gamma \gg 1$, the Doppler factor is the highest at small viewing angles $\theta \lesssim \gamma^{-1} \ll 1$ and we obtain

$$\mathcal{D} \approx \frac{2\gamma}{1 + (\gamma\theta)^2}, \text{ for } \gamma \gg 1 \& \theta \ll 1. \quad (3.54)$$

For $\theta = \gamma^{-1}$ (roughly the edge of the beaming cone), we obtain $\mathcal{D} \approx \gamma = \theta^{-1}$ — the Doppler factor can be much greater than unity in the ultra-relativistic limit.

3.5.3 Time dilation and length contraction

Let us consider the Lorentz transformation of the 4-displacement vector $d\vec{x} = (cdt, d\vec{x})$, which describes the spacetime separation between two nearby events

$$\begin{aligned} cdt' &= \gamma(cdt - \beta dx), \\ dx' &= \gamma(-\beta cdt + dx), \\ dy' &= dy, \quad dz' = dz. \end{aligned} \quad (3.55)$$

If we consider frame \mathcal{O}' to be the comoving with the particle, then $d\vec{x}' = (dx', dy', dz') = 0$ and dt' represents the proper time $d\tau$ of the particle. According to the second expression in eq. (3.55), we see that $dx' = 0$ means $dx = \beta cdt$, which makes sense as the frame \mathcal{O}' is moving at velocity $\mathbf{v} = \beta c\hat{\mathbf{x}}$ as viewed in frame \mathcal{O} . Let us then plug $dx = \beta cdt$ into the first expression in eq. (3.55) and obtain $dt' = d\tau = \gamma(1 - \beta^2)dt$, or

$$dt = \gamma d\tau, \quad (3.56)$$

which is just the definition of the Lorentz factor (eq. 3.15). The fact that lab-frame time is longer than the proper time is a relativistic effect called *time dilation*.

A famous example of time dilation is the observations of atmospheric muons. Hadronic cosmic ray particles (protons and heavier nuclei) collide with the Nitrogen and Oxygen nuclei in the Earth's upper atmosphere at a height ~ 20 km above the ground, and the most common secondary hadrons are charged (π^\pm) and neutral (π^0) pions with roughly 1/3 probability for each of the three outcomes. In the rest frame of a neutral pion, it quickly decays into two photons with a lifetime of $\tau_{\pi^0} \approx 25 \text{ nm}/c$. The charge pions have a much longer rest-frame lifetime of $\tau_{\pi^\pm} \approx 7.8 \text{ m}/c$, so they undergo additional hadronic interactions, which produce lower-energy daughter pions. In each generation of hadronic interactions, 1/3 of the energy is taken away by π^0 , and 2/3 remains in π^\pm . Eventually, the charged pions are not energetic enough to undergo further hadronic interactions and will decay into muons and (anti-)neutrinos following $\pi^\pm \rightarrow \mu^\pm + \nu_\mu/\bar{\nu}_\mu$. A muon particle in its rest frame spontaneous decays into an electron/positron and a neutrino on a lifetime of $\tau \approx 0.66 \text{ km}/c$. In the Earth's frame, a muon with Lorentz factor γ can travel a distance of $\gamma c\tau$. Thus, muons with $\gamma \gtrsim 30$ would be able to reach the ground. The rest mass of

muons is roughly $207m_e \approx 105.8$ MeV, so experiments at ground level only detect muons above a threshold energy of about 3 GeV.

Another example of time dilation is related to the so-called “twin paradox”. Consider a pair of twins who are born at the same time and same position in space. One of them always stays at rest in the frame \mathcal{O} , so her biological age or proper time is equal to the coordinate time $\tau_1 = t$ (and the twins’ birthday is denoted as $t = 0$). The other one wanders around the universe at relativistic speeds as measured in the frame \mathcal{O} and then comes back to the birthplace and meets with her sister at time t . Due to relativistic motion, the latter twin’s biological age/proper time is given by $\tau_2 = \int_0^t dt/\gamma(t)$. Since $\gamma(t) \geq 1$, we conclude that $\tau_2 \leq t = \tau_1$, so the one who wandered around the Universe is younger.

The second law of thermodynamics (“total entropy always increases with time”) defines the direction of time such that $d\tau > 0$ for any particle with a non-zero rest-mass, which means $dt > 0$ in any Lorentz frame — the coordinate time must go forward into the future but not the past. It can be shown that this conclusion also applies to photons (whose $d\tau \rightarrow 0$ and $\gamma \rightarrow \infty$). As a result, the *cause* must precede its *effect* according to all inertial observers — this is the *law of causality*. Another equivalent way of stating this law is: an event can only affect things within its future lightcone (the green region in Fig. 3.1).

On the other hand, let us consider a rigid rod of length L_0 as measured in its comoving frame. If the rod is moving at speed $\mathbf{v} = \beta c \hat{\mathbf{x}}$ in frame \mathcal{O} and if an observer measures the positions of both ends of the rod *simultaneously* (i.e., at the same coordinate time) in the frame \mathcal{O} . A simultaneous measurement in frame \mathcal{O} means $dt = 0$, and using the second expression in eq. (3.55), we obtain

$$dx = dx'/\gamma, \text{ for } dt = 0. \quad (3.57)$$

This means that a rigid rod of length L_0 measured in its comoving frame would appear to have a shortened length of $L = L_0/\gamma$ if it is moving at a high speed in frame \mathcal{O} . Note that simultaneity in frame \mathcal{O} ($dt = 0$) does not mean simultaneity in the comoving frame \mathcal{O}' , since $c dt' = -\gamma \beta dx \neq 0$. It is not possible to achieve simultaneity at different spatial locations in both frames. Due to the effect of length contraction, the number density of a group of particles in frame \mathcal{O} appears higher than the comoving number density n_0 in frame \mathcal{O}' , i.e., $n = \gamma n_0$, as was already discussed earlier near eq. (3.41).

The spacetime described by special relativity has been incredibly well tested by numerous experiments. Perhaps the one closest to our everyday life is the Global Position System (GPS), which takes both special and general relativistic effects into account. Since GPS relies on satellites that move with respect to the ground at a typical speed $v/c \simeq 2.6 \times 10^{-5}$ (or Lorentz factor $\gamma \simeq 1 + 3.5 \times 10^{-10}$), if we fail to correct for time dilation or length contraction, then the cumulative error would be $30 \mu\text{s}$ or 9 km per day — this is unacceptable for navigation. The effects of general relativity (i.e., spacetime curvature due to the Earth’s mass) are of the same order.

3.5.4 *Relative velocity

Consider two particles moving at different 4-velocities \vec{u}_a and \vec{u}_b as measured in the lab frame. Our goal is to calculate the speed of particle a in the comoving frame of particle b — this is because many collisional cross-sections are measured for targets that are held at rest. We discuss two ways of doing this, and then discuss the implications.

The first way is to carry out a Lorentz transformation. Since the relative velocity between the comoving frame of b and the lab frame is $\mathbf{v}_b = \boldsymbol{\beta}_b c$, the 4-velocity of particle a in the comoving frame of particle b is given by

$$\begin{aligned} u_a^{0'} &= \gamma_b (u_a^0 - \beta_b u_a^1), \\ u_a^{1'} &= \gamma_b (-\beta_b u_a^0 + u_a^1), \\ u_a^{2'} &= u_a^2, \quad u_a^{3'} = u_a^3, \end{aligned} \tag{3.58}$$

which means the *relative Lorentz factor* is given by

$$\gamma_{\text{rel}} \equiv \gamma'_a = u_a^{0'}/c = \gamma_a \gamma_b (1 - \boldsymbol{\beta}_a \cdot \boldsymbol{\beta}_b). \tag{3.59}$$

One can show that this is also the Lorentz factor of particle b in the comoving frame of particle a .

The second way is to use a Lorentz invariant quantity to skip Lorentz transformation. The sum of the two 4-velocities $\vec{u} = \vec{u}_a + \vec{u}_b$ is a 4-vector, so $\vec{u}^2 \equiv \vec{u} \cdot \vec{u}$ is a scalar and is Lorentz invariant. Expressing \vec{u}_a and \vec{u}_b in the lab frame (hereafter taking $c = 1$), we find

$$\vec{u}^2 = \vec{u}_a^2 + \vec{u}_b^2 + 2\vec{u}_a \cdot \vec{u}_b = -2 - 2\gamma_a \gamma_b (1 - \boldsymbol{\beta}_a \cdot \boldsymbol{\beta}_b).$$

In the comoving frame of particle b , we have $\vec{u}'_b = (1, 0, 0, 0)$ and hence

$$\vec{u}'_a \cdot \vec{u}'_b = -\gamma'_a \equiv -\gamma_{\text{rel}}.$$

Therefore, we obtain $(\vec{u}')^2 = (\vec{u}'_a)^2 + (\vec{u}'_b)^2 + 2\vec{u}'_a \cdot \vec{u}'_b = -2 - 2\gamma_{\text{rel}}$. We then make use of the Lorentz invariance $\vec{u}^2 = (\vec{u}')^2$ and obtain the Lorentz factor of particle a in the comoving frame of particle b , $\gamma_{\text{rel}} = \gamma_a \gamma_b (1 - \boldsymbol{\beta}_a \cdot \boldsymbol{\beta}_b)$, same as in eq. (3.59).

In the following, we discuss some properties of the relative velocity. The parallel ($1'$) component of the relative velocity is

$$\beta_{\text{rel},\parallel} \equiv \beta'_{a,\parallel} = \frac{u_a^{1'}}{u_a^{0'}} = \frac{\beta_{a,\parallel} - \beta_b}{1 - \boldsymbol{\beta}_a \cdot \boldsymbol{\beta}_b}, \tag{3.60}$$

and the perpendicular component (either $2'$ or $3'$) is

$$\beta_{\text{rel},\perp} \equiv \beta'_{a,\perp} = \frac{u_a^{2'}}{u_a^{0'}} = \frac{\beta_{a,\perp}}{\gamma_b (1 - \boldsymbol{\beta}_a \cdot \boldsymbol{\beta}_b)}. \tag{3.61}$$

For the spacial case of aligned velocities $\beta_a \cdot \beta_b = \beta_a \beta_b$, and in the ultra-relativistic limit, the relative Lorentz factor is

$$\gamma_{\text{rel}} \approx \frac{1}{2} \left(\frac{\gamma_a}{\gamma_b} + \frac{\gamma_b}{\gamma_a} \right), \text{ for aligned velocities, } \gamma_a \gg 1 \text{ and } \gamma_b \gg 1. \quad (3.62)$$

For the anti-aligned case $\beta_a \cdot \beta_b = -\beta_a \beta_b$, we obtain

$$\gamma_{\text{rel}} \approx 2\gamma_a \gamma_b, \text{ for anti-aligned velocities, } \gamma_a \gg 1 \text{ and } \gamma_b \gg 1. \quad (3.63)$$

To achieve a maximum *relative* Lorentz factor, experiments at hadronic colliders are carried out with anti-aligned particle velocities.

3.5.5 *Threshold energy for inelastic collisions

In this subsection, we discuss the necessary condition for inelastic collision

$$a + b \rightarrow p + q, \quad (3.64)$$

where we require that the total rest mass increases, $m_p + m_q > m_a + m_b$, so the reaction is endothermic. From energy conservation, we know that some kinetic energy must be cost during the collision, and our goal is to calculate the threshold energy for such an endothermic reaction to occur. If the reaction is exothermic, there is no threshold energy.

Suppose the 4-momenta of the reactants are $\vec{p}_a = (E_a/c, \mathbf{p}_a)$ and $\vec{p}_b = (E_b/c, \mathbf{p}_b)$ in the lab frame. Let us denote the 4-momenta of the two particles in the *center-of-momentum frame* as $\vec{p}'_a = (E'_a, \mathbf{p}'_a)$ and $\vec{p}'_b = (E'_b, \mathbf{p}'_b)$, and the sum of the 3-momenta vanishes in this frame (where the two particles collide head-on with opposite momenta)

$$\mathbf{p}'_a + \mathbf{p}'_b = 0.$$

We make use of the fact that $\vec{p}^2 = (\vec{p}_a + \vec{p}_b)^2$ is Lorentz invariant and evaluate \vec{p}^2 in both frames. From $(\vec{p}')^2 = -(E'_a + E'_b)^2/c^2$, so the total energy in the center-of-momentum frame is given by

$$E' = E'_a + E'_b = \sqrt{-(\vec{p}')^2 c^2} = \sqrt{-\vec{p}^2 c^2}. \quad (3.65)$$

The minimum requirement for the inelastic collision to occur is that all the energy E' goes to the rest-mass energies of the two products p and q . Thus, we require

$$E' \geq (m_p + m_q)c^2, \quad (3.66)$$

which means

$$-\vec{p}^2 = -\vec{p}_a^2 - \vec{p}_b^2 - 2\vec{p}_a \cdot \vec{p}_b \geq (m_p + m_q)^2 c^2. \quad (3.67)$$

Making use of $\vec{p}_a^2 = -m_a^2 c^2$ and $\vec{p}_b^2 = -m_b^2 c^2$ (where m_a and m_b are the rest masses of the two reactants), we find

$$-2\vec{p}_a \cdot \vec{p}_b \geq (m_p + m_q)^2 c^2 - (m_a^2 + m_b^2)c^2, \quad (3.68)$$

or

$$2E_a E_b - 2\mathbf{p}_a \cdot \mathbf{p}_b c^2 \geq (m_p + m_q)^2 c^2 - (m_a^2 + m_b^2) c^2. \quad (3.69)$$

Let θ be the angle between the two momenta \mathbf{p}_a and \mathbf{p}_b such that $\mathbf{p}_a \cdot \mathbf{p}_b = p_a p_b \cos \theta$, and we obtain

$$\mathbf{p}_a \cdot \mathbf{p}_b c^2 = p_a p_b \cos \theta c^2 = \cos \theta \sqrt{(E_a^2 - m_a^2 c^4)(E_b^2 - m_b^2 c^4)}. \quad (3.70)$$

Therefore, we obtain the threshold for the inelastic collision

$$2E_a E_b - 2\sqrt{(E_a^2 - m_a^2 c^4)(E_b^2 - m_b^2 c^4)} \cos \theta \geq (m_p + m_q)^2 c^4 - (m_a^2 + m_b^2) c^4, \quad (3.71)$$

where no approximation has been made.

Let us then consider the limit where the two reactant particles a and b are ultra-relativistic in the lab frame such that $E_a \gg m_a^2 c^2$ and $E_b \gg m_b^2 c^2$, and in this limit the condition for inelastic collision becomes

$$E_a E_b \geq \frac{(m_p + m_q)^2 c^4 - (m_a^2 + m_b^2) c^4}{2(1 - \cos \theta)}, \text{ in the ultra-relativistic limit.} \quad (3.72)$$

Since the RHS of the above equation is positive (as $m_p + m_q > m_a + m_b$ by construction), the energy requirement is minimized when the two particles collide head-on such that $\cos \theta = -1$ (\mathbf{p}_a and \mathbf{p}_b are anti-parallel), so we obtain the *threshold condition for inelastic collision*

$$E_a E_b \geq \frac{(m_p + m_q)^2 c^4 - (m_a^2 + m_b^2) c^4}{4}, \text{ for ultra-relativistic head-on collision.} \quad (3.73)$$

In the following, we consider some examples of such collisions.

- For pair production from the collision between two photons

$$\gamma_1 + \gamma_2 \rightarrow e^\pm, \quad (3.74)$$

the threshold condition in eq. (3.73) becomes

$$E_{\gamma 1} E_{\gamma 2} \geq m_e^2 c^4, \quad (3.75)$$

where we have used $m_{\gamma 1} = m_{\gamma 2} = 0$ for photons and the rest mass for electron/positron m_e . An important application is that the intergalactic gamma-ray background (IGB) spectrum has a high-energy cut-off caused by pair production collisions with the cosmic microwave background (CMB). In the local universe at low redshifts ($z \ll 1$), the typical CMB photon energy of $3k_B T_{\text{CMB}} \approx 7 \times 10^{-4}$ eV (for $T_{\text{CMB}} = 2.7$ K), so we find that gamma-ray photons above an energy $m_e^2 c^4 / (3k_B T_{\text{CMB}}) \approx 4 \times 10^{14}$ eV would undergo pair production while propagating in the intergalactic medium. The mean-free path for very high energy gamma-ray photons propagating in the intergalactic medium has been calculated by [Coppi & Aharonian \(1997\)](#), and their results are shown in Fig. 3.5.

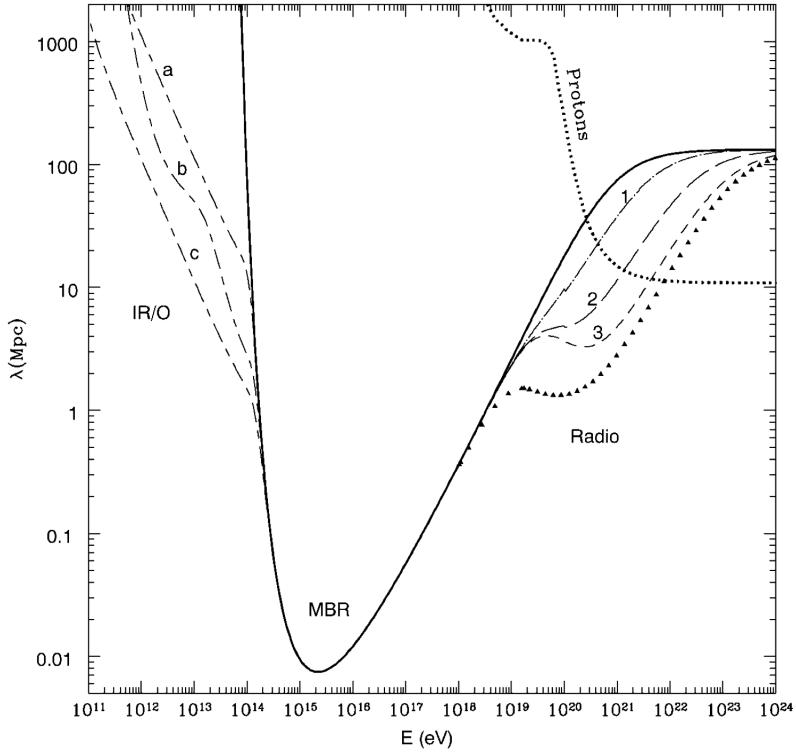


Figure 3.5: Pair production mean free path for gamma-ray photons propagating in the intergalactic medium (ignoring cosmic redshift). Below 10^{14} eV, gamma-ray photons interact with infrared and optical photons (denoted by “IR/O” under three different models); between 10^{14} and 10^{19} eV, they interact with the CMB; and above 10^{19} eV, they interact with radio photons (using four different models). The dotted line marked by “Protons” denote the energy loss mean free path for cosmic ray protons, which interact with the CMB by pair production ($p + \gamma \rightarrow p + e^\pm$) and photo-meson production (eqs. 3.76, 3.78).

- For the inelastic collision between a proton and a photon

$$p + \gamma \rightarrow n + \pi^+, \quad (3.76)$$

In the limit of an ultra-relativistic cosmic ray proton, the threshold condition in eq. (3.73) becomes

$$E_p \geq \frac{(m_n + m_{\pi^+})^2 c^4 - m_p^2 c^4}{4E_\gamma} = 1.0 \times 10^{20} \text{ eV} \frac{7 \times 10^{-4} \text{ eV}}{E_\gamma}, \text{ for } \pi^+, \quad (3.77)$$

where we have plugged in the rest-mass energies of neutron $m_n c^2 = 939.56 \text{ MeV}$, charged pion $m_{\pi^+} c^2 = 139.57 \text{ MeV}$, proton $m_p c^2 = 938.27 \text{ MeV}$, and the typical energy of CMB photons $E_\gamma = 3k_B T_{\text{CMB}} \approx 7 \times 10^{-4} \text{ eV}$ in the local universe. The $p\gamma$ inelastic collision can also produce a neutral pion via

$$p + \gamma \rightarrow p + \pi^0, \quad (3.78)$$

for which the threshold energy is only slightly lower

$$E_p \geq \frac{(m_p + m_{\pi^0})^2 c^4 - m_p^2 c^4}{4E_\gamma} = 9.7 \times 10^{19} \text{ eV} \frac{7 \times 10^{-4} \text{ eV}}{E_\gamma}, \text{ for } \pi^0, \quad (3.79)$$

where we have used the rest-mass energy of neutral pion $m_{\pi^0} c^2 = 134.98 \text{ MeV}$. Therefore, we expect that photo-meson production via eqs. (3.76) and (3.78) will lead to a cut-off in the energy spectrum of ultra-high energy cosmic rays (UHECRs) above $\sim 10^{20} \text{ eV}$. This cut-off has been predicted by Greisen, Zatsepin, and Kuzmi (GZK), and the GZK cut-off is the most spectacular feature of the UHECR spectrum⁶. Our estimate for the cut-off energy is only approximate in that it does not consider the energy spread in CMB photons. It turns out that more energetic CMB photons (on the Wien tail) with energies $E_\gamma \simeq 10k_B T_{\text{CMB}}$ are sufficiently abundant to interact with UHECRs, and this produces a cutoff that starts around $E_{\text{GZK}} \simeq 3 \times 10^{19} \text{ eV}$.

- For the reactions in eqs. (3.76, 3.78), we can also consider the interactions in the proton comoving frame where the proton's 3-momentum vanishes (and hence $\mathbf{p}_a \cdot \mathbf{p}_b = 0$). In the proton comoving frame, we obtain a threshold energy for the gamma-ray photon from eq. (3.71)

$$E_\gamma \geq \frac{(m_n + m_{\pi^+})^2 c^4 - m_p^2 c^4}{2m_p c^2} = 151.4 \text{ MeV}, \text{ for } \pi^+, \quad (3.80)$$

⁶The existence of the GZK cut-off has been confirmed by the [HiRes Experiment](#) and [Pierre Auger Observatory](#). However, the detailed UHECR spectral modeling is complicated by the uncertain elemental composition — it is possible that UHECR is dominated by heavier ions than protons and photo-disintegration of heavier ions also plays an important role in shaping the UHECR energy spectrum.

and

$$E_\gamma \geq \frac{(m_p + m_{\pi^0})^2 c^4 - m_p^2 c^4}{2m_p c^2} = 144.7 \text{ MeV, for } \pi^0. \quad (3.81)$$

We find that the energy threshold for neutral pion production is slightly lower.

Finally, we briefly mention the important pion decaying products following the photo-meson interactions (3.76, 3.78) as well as $n + \gamma \rightarrow p + \pi^-$ and $n + \gamma \rightarrow n + \pi^0$ for heavier ions that contain neutrons. The pions rapidly decay as follows

$$\begin{aligned} \pi^0 &\rightarrow 2\gamma, \quad t_{1/2}(\pi^0) = 5.8 \times 10^{-17} \text{ s}, \\ \pi^+ &\rightarrow \mu^+ + \nu_\mu, \quad t_{1/2}(\pi^+) = 1.80 \times 10^{-8} \text{ s}, \\ \pi^- &\rightarrow \mu^- + \bar{\nu}_\mu, \quad t_{1/2}(\pi^-) = 1.80 \times 10^{-8} \text{ s}, \end{aligned} \quad (3.82)$$

where $t_{1/2}$'s are the corresponding half-life times. The muons then decay as follows

$$\mu^+ \rightarrow e^+ + \bar{\nu}_\mu + \nu_e, \quad \mu^- \rightarrow e^- + \nu_\mu + \bar{\nu}_e, \quad (3.83)$$

with half-life time $t_{1/2}(\mu^\pm) = 1.5 \times 10^{-6} \text{ s}$. The end result is that each photo-pion production gives rise to two muon neutrinos and one electron neutrino (during propagation, neutrino oscillation then redistributes flavors such that $\nu_e : \nu_m u : \nu_\tau = 1 : 1 : 1$). When cosmic ray protons propagate in very dense gas, pions may be produced by $p + p \rightarrow p + p + \pi^0$, $p + p \rightarrow p + n + \pi^+$, and $p + p \rightarrow p + p + \pi^+ + \pi^-$, etc. These neutrinos are important messengers in high-energy astrophysics.

3.5.6 *Extreme elastic and inelastic collisions

Consider a projectile particle of rest mass m_a and velocity $\beta_a = \beta_a \hat{x}$ colliding with a target particle of rest mass $m_b \leq m_a$ that is initially at rest in the lab frame. We restrict ourselves to the ultra-relativistic limit of $\gamma_a \gg 1$. As one can always define the “lab frame” as the comoving frame of the less massive particle, so $m_b \leq m_a$ is not a restriction. An elastic collision conserves the total energy and momentum of the system, whereas an inelastic collision would dissipate some energy.

The question is best analyzed in the center-of-momentum frame (denoted by a prime ') where $\vec{p}'_a + \vec{p}'_b = 0$. Let us denote the Lorentz factor of center-of-momentum frame wrt. the lab frame as γ . We first write the 4-momenta of the two particles in the lab frame

$$\vec{p}_a = \gamma_a m_a (1, \beta_a, 0, 0), \quad \vec{p}_b = m_b (1, 0, 0, 0), \quad (3.84)$$

and then Lorentz transformation gives the 4-momenta in the center-of-momentum frame

$$\vec{p}'_a = \gamma \gamma_a m_a (1 - \beta \beta_a, \beta_a - \beta, 0, 0), \quad \vec{p}'_b = \gamma m_b (1, -\beta, 0, 0). \quad (3.85)$$

We then solve for the relative velocity between the two frames based on the condition of $\mathbf{p}'_a + \mathbf{p}'_b = 0$,

$$\beta = \frac{\gamma_a m_a}{\gamma_a m_a + m_b} \beta_a \approx 1 - \frac{m_b}{\gamma_a m_a} - \frac{1}{2\gamma_a^2}, \quad (3.86)$$

where we used Taylor expansion for small quantities $m_b/(\gamma_a m_a) \ll 1$ and $1/\gamma_a \ll 1$. The corresponding Lorentz factor is

$$\gamma = \frac{1}{\sqrt{1 - \beta^2}} \approx \frac{\gamma_a}{\sqrt{1 + \xi}}, \quad \xi \equiv 2\gamma_a m_b / m_a, \quad (3.87)$$

where we have defined a dimensionless quantity ξ that may or may not be much less (or greater) than unity.

There are two asymptotic limits: if $\xi \ll 1$ (i.e., the projectile particle is extremely heavy), then we obtain $\gamma \approx \gamma_a$ as the center of momentum of the system almost coincides with the projectile particle; in the opposite limit of $\xi \gg 1$, we obtain $\gamma \approx \sqrt{\gamma_a m_a / (2m_b)}$ as the center of momentum lies largely in between the two particles. A special case is when $m_a = m_b$, then we obtain $\gamma \approx \sqrt{\gamma_a / 2}$.

Let us then consider two extreme outcomes of the collision: (i) an elastic collision that maximizes the energy received by the target particle; (ii) a complete inelastic collision that maximizes the energy dissipation of the system. The results are as follows.

- (i) In an elastic collision, the magnitude of each momentum is preserved in the center-of-momentum frame, and the directions of \mathbf{p}'_a and \mathbf{p}'_b are deflected. A physical example is Coulomb collision in the absence of radiative losses. The quantities after the collision will be denoted with a tilde. From momentum conservation, we obtain

$$|\tilde{\mathbf{p}}'_a| = |\mathbf{p}'_a| = |\tilde{\mathbf{p}}'_b| = |\mathbf{p}'_b|. \quad (3.88)$$

However, the direction of $\tilde{\mathbf{p}}'_a$ is unspecified (and $\tilde{\mathbf{p}}'_b = -\tilde{\mathbf{p}}'_a$). It can be shown that the energy gain by the target particle (b) in the lab frame is maximized when the deflection angle is $\theta = 180^\circ$. In this case, we obtain $\tilde{\mathbf{p}}'_b = \gamma \beta m_b \hat{\mathbf{x}}$. Hence, the maximum energy gain by the target particle in the lab frame is given by

$$\tilde{E}_b - m_b = 2\gamma^2 \beta^2 m_b = \frac{2\gamma_a^2 \beta_a^2 m_b}{1 + 2\gamma_a m_b / m_a + (m_b / m_a)^2} \approx \frac{\xi \gamma_a m_a}{1 + \xi}, \quad \xi = \frac{2\gamma_a m_b}{m_a}. \quad (3.89)$$

The post-collision energy of the projectile particle in the lab frame is then

$$\tilde{E}_a = \gamma_a m_a - (\tilde{E}_b - m_b) \approx \frac{\gamma_a m_a}{1 + \xi}. \quad (3.90)$$

We see that the parameter ξ controls how the two particles share their energies.

- (ii) In the case of a complete inelastic collision, both particles dissipate all their kinetic energies in the center-of-momentum frame, and hence $\tilde{p}'_a = \tilde{p}'_b = 0$. Then, the post-collision energies of the two particles in the lab frame are

$$\tilde{E}_{a/b} = \gamma m_{a/b} \approx \frac{\gamma_a}{\sqrt{1+\xi}} m_{a/b}. \quad (3.91)$$

The fractional energy dissipation during the collision is given by

$$f_{\text{loss}} = 1 - \frac{\tilde{E}_a + \tilde{E}_b}{\gamma_a m_a + m_b} \approx 1 - \frac{1 + m_b/m_a}{\sqrt{1+\xi}} \quad (3.92)$$

If $\xi \ll 1$, we obtain $f_{\text{loss}} \approx \xi/2 \ll 1$, so only a small fraction of the energy is dissipated as the projectile particle is extremely heavy in this limit. In the opposite limit of $\xi \gg 1$, we obtain $f_{\text{loss}} \approx 1$, so nearly 100% of the kinetic energy is dissipated in a very strong collision. We also note that the post-collision total momentum in the lab frame is $\tilde{p} = \gamma\beta(m_a + m_b)$ whereas the initial total momentum is $p = \gamma_a\beta_a m_a$, so we obtain

$$\tilde{p}/p \approx \frac{1 + m_b/m_a}{\sqrt{1+\xi}}. \quad (3.93)$$

The fact that $\tilde{p} \neq p$ does not violate momentum conservation, because it is assumed that the dissipated kinetic energy has been radiated away and the radiation carries away some momentum along \hat{x} .

3.5.7 *A rigid rod moving in both frames

It is easy to get confused about (i) time simultaneity in different frames and (ii) when should the time dilation and length contraction formulae be used. Carrying out the following exercises should get rid of such confusion. We take $c = 1$ in this subsection.

Let us consider a rigid rod of length L' as measured simultaneously in frame \mathcal{O}' , meaning that at $t' = 0$, we define two events \mathcal{P} and \mathcal{Q} corresponding to the left and right ends of the rod. Since \mathcal{P} and \mathcal{Q} are just two spacetime events, the rod can be massless with its two ends moving even at speed of light. Let the rod lie along the \hat{x}' axis. Without losing generality, we choose $x'_P = 0$ and $x'_Q = L'$ at $t' = 0$, so these two events have spacetime coordinates

$$\mathcal{P} : (t' = 0, x' = 0), \quad \mathcal{Q} : (t' = 0, x' = L'), \quad (3.94)$$

where the $y'z'$ dimensions have been omitted. By Lorentz transforming the separation 4-vector between these two events, one obtains their coordinates in frame \mathcal{O}

$$\mathcal{P} : (t = 0, x = 0), \quad \mathcal{Q} : (t = \gamma\beta L', x = \gamma L'). \quad (3.95)$$

In most textbooks (e.g., Rybicki & Lightman), the rod is considered not moving in frame \mathcal{O}' , and in that case L' is equal to its proper lengths L_0 , which is only well-defined for

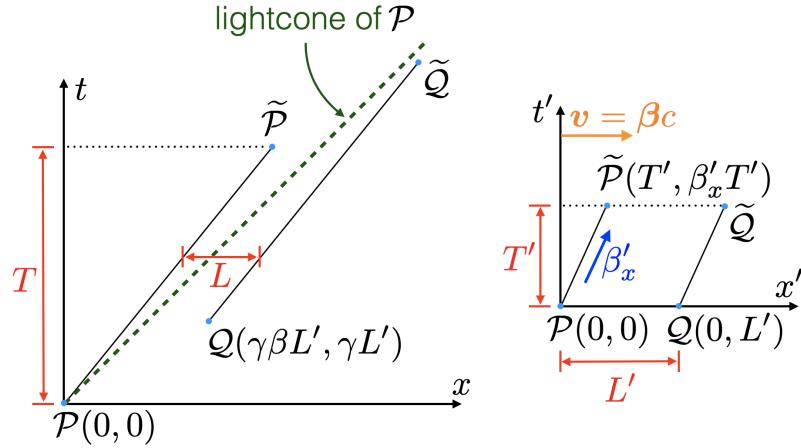


Figure 3.6: Spacetime diagram of a rigid rod. It has length L' as measured simultaneously (meaning at the same coordinate time t' on both ends) in frame \mathcal{O}' . The rod moves in frame \mathcal{O}' at velocity β' (with an x -component β'_x) for a duration T' . In frame \mathcal{O} , its length contracts to L , and we have $L/L' = 1/[\gamma(1 + \beta\beta'_x)]$ and $T/T' = \gamma(1 + \beta\beta'_x)$, which means that LT = Lorentz invariant.

non-zero mass. However, here we consider a more general case where the rod is moving with a velocity component $v'_x = \beta'_x c$ in frame \mathcal{O}' (the other two components v'_y and v'_z are arbitrary). The spacetime diagram is shown in Fig. 3.6.

After time T' in frame \mathcal{O}' , the spacetime position of the left end of the rod is denoted as event $\tilde{\mathcal{P}}$, and the right end is denoted as event $\tilde{\mathcal{Q}}$. The coordinates of these two new events are

$$\begin{aligned} \text{frame } \mathcal{O}' &: \tilde{\mathcal{P}} = (T', \beta'_x T'), \quad \tilde{\mathcal{Q}} = (T', L' + \beta'_x T'); \\ \text{frame } \mathcal{O} &: \tilde{\mathcal{P}} = [\gamma(1 + \beta\beta'_x)T', \gamma(\beta + \beta'_x)T'] ; \\ &\quad \tilde{\mathcal{Q}} = [\gamma\beta L' + (\gamma(1 + \beta\beta'_x)T', \gamma L' + \gamma(\beta + \beta'_x)T'] . \end{aligned} \quad (3.96)$$

The key result is that the slope of $\mathcal{P}\tilde{\mathcal{P}}$ (and also $\mathcal{Q}\tilde{\mathcal{Q}}$) is

$$\frac{dx}{dt} = v_{\text{rod}} = \frac{\beta + \beta'_x}{1 + \beta\beta'_x}, \quad (3.97)$$

which is simply the relativistic addition of velocities (see eq. 3.60). At $t = t_{\mathcal{Q}} \equiv \gamma\beta L'$, the position of the left end of the rod is $t_{\mathcal{Q}}v_{\text{rod}}$, and hence the length of the rod as measured simultaneously in the lab frame \mathcal{O} is

$$L = \gamma L' \left[1 - \frac{\beta(\beta + \beta'_x)}{1 + \beta\beta'_x} \right] = \frac{L'}{\gamma(1 + \beta\beta'_x)}, \quad (3.98)$$

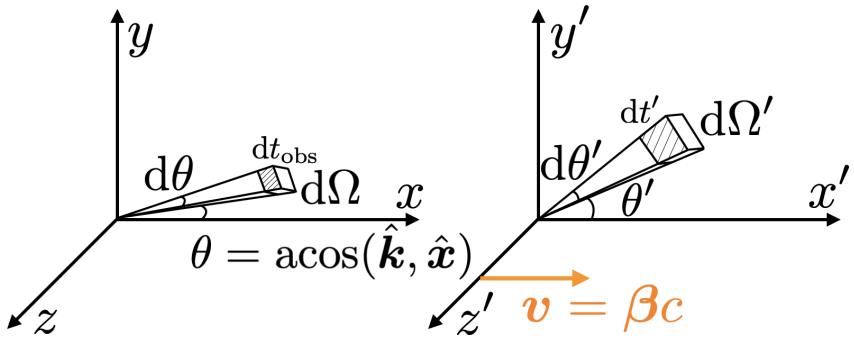


Figure 3.7: Lorentz transformation of a plane EM wave. A source is moving in the lab frame (xyz) at a velocity $\mathbf{v} = \beta c \hat{x}$ direction. The source produces plane EM waves into a solid angle $d\Omega'$ along the \mathbf{k}' direction in its comoving frame ($x'y'z'$) and the angle between the wavevector \mathbf{k}' and \hat{x}' is θ' . In the lab frame, the waves are beamed into a solid angle $d\Omega = d\Omega'/D^2$ (D being the Doppler factor defined in eq. 3.104) along the \mathbf{k} direction at an angle θ away from the \hat{x} axis. The two inclination angles are related by $\sin \theta = \sin \theta' / D$. A pulse of duration dt' in the comoving frame corresponds to a duration of $dt_{\text{obs}} = dt'/D$ as seen by a fixed observer in the lab frame (whose line of sight is along \mathbf{k}). The received power per solid angle by the observer is $dP_{\text{obs}}/d\Omega = D^4 dP'/d\Omega'$.

which we call the *generalized length contraction*. For the special case of a massive rod that is not moving in frame \mathcal{O}' (i.e. $\beta' = 0$), we know that L' equals its proper length L_0 and hence $L = L_0/\gamma$ — this is the usual length contraction. However, for a massive rod moving in a general direction (not necessarily along \hat{x}'), then the relation between lab-frame length L and the proper length L_0 is more complicated, because only the length component parallel to the relative velocity contracts whereas the perpendicular length components are the same in different frames.

3.6 Lorentz transformation of EM fields

In this section, we first look at the Lorentz transformation of a plan EM wave as described by the corresponding 4-wavevector \vec{k} . Then, we will study the how an arbitrary EM field transforms between different Lorentz frames.

3.6.1 Monochromatic plane EM waves

Consider a source emitting a monochromatic plane EM wave, which is specified by the 4-wavevector $\vec{k}' = (\omega'/c, \mathbf{k}')$ in the source's comoving frame \mathcal{O}' . As measured by observers in the lab frame \mathcal{O} , the source is moving at a constant velocity $\mathbf{v} = \beta c \hat{x}$. The geometry is shown in Fig. 3.7. Let us calculate the 4-wavevector $\vec{k} = (\omega/c, \mathbf{k})$ in the lab frame.

Without losing generality, we orient the Cartesian y/y' -axis in both frames such that \mathbf{k}' is in the $\hat{\mathbf{x}}'/\hat{\mathbf{y}}'$ plane. Thus, the 4-wavevector in the comoving frame can be written as

$$\vec{k}' = (\omega'/c, k'_x, k'_y, 0), \quad (3.99)$$

where $k'_x = \mathbf{k}' \cdot \hat{\mathbf{x}}'$ and $k'_y = \mathbf{k}' \cdot \hat{\mathbf{y}}'$. Instead of using the Lorentz transformation in eq. (3.26) which goes from \mathcal{O} to \mathcal{O}' frames, here we need the inverse transformation.

We consider the waves to be propagating in vacuum, so the dispersion relation⁷ gives

$$(\mathbf{k}')^2 = (k')^2 = (k'_x)^2 + (k'_y)^2 = (\omega'/c)^2. \quad (3.100)$$

Let us denote the angle between \mathbf{k}' and $\hat{\mathbf{x}}'$ as θ' , so $\sin \theta' = k'_y/k'$ and $\cos \theta' = k'_x/k'$.

Using the inverse Lorentz transform (eq. 3.26 with a flipped sign for β), we obtain

$$\begin{aligned} \omega/c &= \gamma(\omega'/c + \beta k'_x) \Rightarrow \omega = \gamma \omega'(1 + \beta \cos \theta'), \\ k_x &= \gamma(\beta \omega'/c + k'_x) = k' \gamma(\beta + \cos \theta'), \\ k_y &= k'_y = k' \sin \theta', \\ k_z &= k'_z = 0. \end{aligned} \quad (3.101)$$

Then, in the \mathcal{O} frame, the angle between \mathbf{k} and $\hat{\mathbf{x}}$ is given by

$$\sin \theta = \frac{k_y}{k} = \frac{\omega' \sin \theta'}{\omega} = \frac{\sin \theta'}{\gamma(1 + \beta \cos \theta')}, \quad (3.102)$$

and

$$\cos \theta = \frac{k_x}{k} = \frac{\cos \theta' + \beta}{1 + \beta \cos \theta'}, \quad \cos \theta' = \frac{\cos \theta - \beta}{1 - \beta \cos \theta}. \quad (3.103)$$

It can be easily shown that $k_x^2 + k_y^2 = k^2 = (\omega/c)^2$ — Lorentz transformation preserves the vacuum dispersion relation (in accordance with the principle of relativity).

In fact, since we know that the Lorentz transformation must obey the principle of relativity, the above algebra can be simplified in the following way. The first step is to do the frequency transformation, which gives the well-known Doppler factor

$$\mathcal{D} = \frac{\omega}{\omega'} = \gamma(1 + \beta \cos \theta') = \frac{1}{\gamma(1 - \beta \cos \theta)}. \quad (3.104)$$

This means that the amplitude of the wavevector in the lab frame is given by

$$k = \omega/c = \mathcal{D} k'. \quad (3.105)$$

⁷It might seem odd that $k = \omega/c$ is called the “dispersion relation” here, because it indicates that waves at all frequencies propagate at the same speed of c , i.e., there is no wave dispersion in vacuum. When considering EM waves in a plasma in Ch. 5, we do find that waves at different frequencies propagate at different speeds. The concept of dispersion relation will make more sense there.

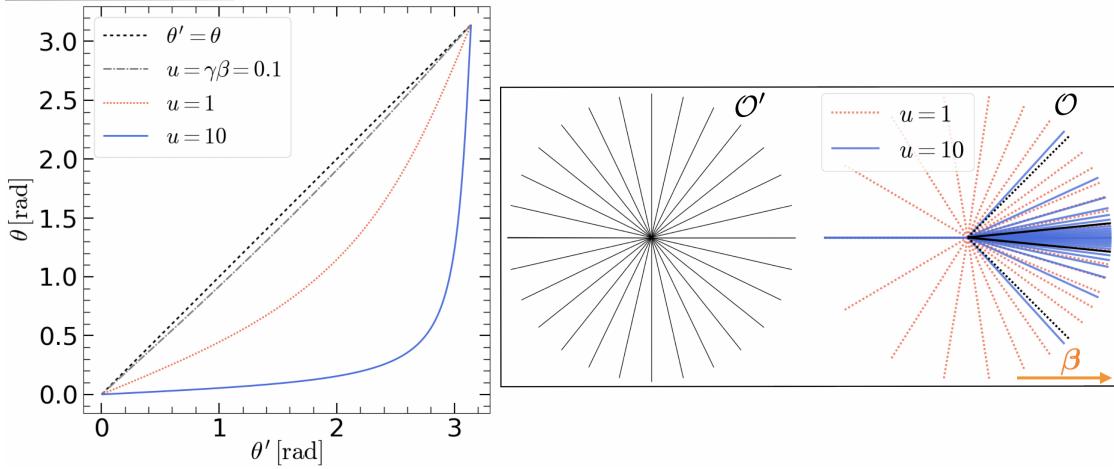


Figure 3.8: Left panel: the monotonic correspondence between the wave propagation angles θ' and θ in the two Lorentz frames \mathcal{O}' and \mathcal{O} , for a few velocities as indicated by $u = \gamma\beta$. Right panel: A source emitting isotropic rays in the comoving frame \mathcal{O}' and Doppler beaming cause the rays to bundle up near the direction of the velocity vector β in the lab frame \mathcal{O} . The black rays in the lab frame \mathcal{O} correspond to $\theta' = \pi/2$ or $\theta = \text{acos}(\beta)$.

The wavevector component in the transverse direction is given by $k_y = k'_y = k' \sin \theta' = (k/\mathcal{D}) \sin \theta'$, which means

$$\sin \theta = \frac{\sin \theta'}{\mathcal{D}}. \quad (3.106)$$

Knowing ω and θ , the problem is essentially solved — the only remaining issue is the sign of $k_x = \pm \sqrt{k^2 - k_y^2}$ or $\cos \theta = \pm \sqrt{1 - (\sin \theta'/\mathcal{D})^2}$, which can be determined by finding a critical angle θ'_c at which $\sin \theta = 1$ and hence $\cos \theta$ flips sign here. This critical angle in the comoving frame, if needed, is given by $\cos \theta'_c = -\beta$.

The correspondence between θ' and θ is shown in Fig. 3.8. For a highly relativistic source with $\gamma \gg 1$, the emission along most angles θ' in the comoving frame will be beamed in the forward direction with an angle $\theta \ll 1$ from the velocity vector β . At the same time, each photon along the forward direction gains energy by a factor of the Doppler factor $\mathcal{D} \gg 1$. Moreover, the arrival times of the emitted photons are compressed to be within a time $\Delta t_{\text{obs}} = \Delta\tau/\mathcal{D} \ll \Delta\tau$. All these effects make the apparent power of the source much higher than that in the source's comoving frame. These effects are called *relativistic beaming* and they can be described by the Doppler factor. Below is a summary of some interesting facts about the Doppler factor.

- In the limit $\gamma \gg 1$ such that $\beta \approx 1 - 1/(2\gamma^2)$, for most angles in the comoving frame (θ' not extremely close to 0 or π), the Doppler factor is of the order $\mathcal{D} \sim \gamma$ and most

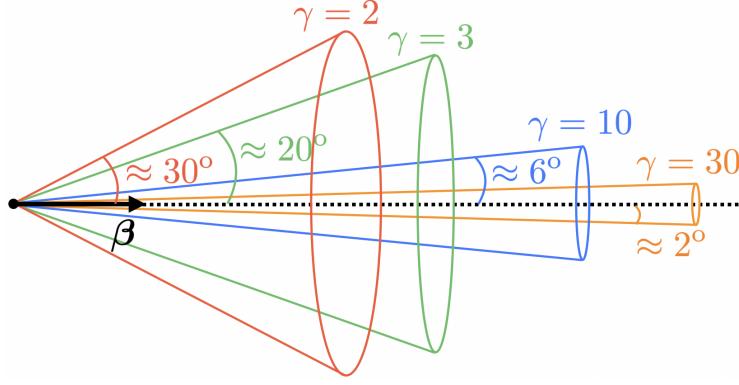


Figure 3.9: Approximate beaming cones ($\theta \approx 1/\gamma$) for an emitter moving at different Lorentz factors γ .

photons are beamed into a narrow angle $\theta \sim \mathcal{D}^{-1} \sim \gamma^{-1}$ in the lab frame. However, when considering extreme angles ($\theta' \approx 0$ or π), the Doppler factor can span a wide range $(2\gamma)^{-1} < \mathcal{D} < 2\gamma$.

- From the 4-momentum whose time component is the total energy (including rest mass if non-zero), we know that the energies E and E' of *any* particle in the two frames are related by

$$E = \mathcal{D}E'. \quad (3.107)$$

- The differential solid angle in the lab frame is $d\Omega = d(\cos \theta)d\phi$, and similarly for $d\Omega'$ in the comoving frame. By differentiating eq. (3.103), one can easily show that $d(\cos \theta) = d(\cos \theta')/\mathcal{D}^2$, and since $d\phi = d\phi'$, we obtain

$$d\Omega = d\Omega'/\mathcal{D}^2. \quad (3.108)$$

- The energy emitted per solid angle is

$$dE/d\Omega = \mathcal{D}^3 dE'/d\Omega'. \quad (3.109)$$

- The *observed* power per solid angle for an observer in the lab frame is given by

$$\frac{dP_{\text{obs}}}{d\Omega} = \frac{dE}{dt_{\text{obs}} d\Omega} = \mathcal{D}^4 \frac{dE'}{dt' d\Omega'} = \mathcal{D}^4 \frac{dP'}{d\Omega'}, \quad (3.110)$$

where $dP'/d\Omega'$ is the emitted power per solid angle in the comoving frame. If the observer *assumes* (in the absence of more information) that the source emits in an isotropic manner, then she infers an *isotropic-equivalent luminosity*

$$L_{\text{iso}} = 4\pi (dP_{\text{obs}}/d\Omega). \quad (3.111)$$

For a highly relativistic source with $\gamma \gg 1$, the factor of \mathcal{D}^4 boost in L_{iso} is so strong that an on-axis observer located within an angle $\theta \lesssim \gamma^{-1}$ from the velocity vector would see extremely bright emission whereas an off-axis observer at $\theta \gg \gamma^{-1}$ usually does not detect the source. The relativistic beaming cones for a relativistic emitter with different Lorentz factors are schematically shown in Fig. 3.9.

3.6.2 EM fields

Unfortunately, the EM fields themselves do not form a nice 4-vector that follows the simple Lorentz transformation rule. The components of \mathbf{E} and \mathbf{B} forms an antisymmetric rank-2 *field strength tensor*

$$F^{\mu\nu} = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & B_z & -B_y \\ -E_y & -B_z & 0 & B_z \\ -E_z & B_y & -B_x & 0 \end{pmatrix}. \quad (3.112)$$

This tensor undergoes Lorentz transformation in the following way

$$F^{\mu'\nu'} = L_{\alpha}^{\mu'} L_{\beta}^{\nu'} F^{\alpha\beta}. \quad (3.113)$$

The result is

$$\begin{aligned} \mathbf{E}'_{\parallel} &= \mathbf{E}_{\parallel}, & \mathbf{B}'_{\parallel} &= \mathbf{B}_{\parallel}, \\ \mathbf{E}'_{\perp} &= \gamma(\mathbf{E}_{\perp} + \boldsymbol{\beta} \times \mathbf{B}), & \mathbf{B}'_{\perp} &= \gamma(\mathbf{B}_{\perp} - \boldsymbol{\beta} \times \mathbf{E}), \end{aligned} \quad (3.114)$$

where subscripts \parallel and \perp stand for field components parallel and perpendicular to the relative velocity vector $\mathbf{v} = \boldsymbol{\beta}c$ between the two frames. Note that “parallel” and “perpendicular” on the LHS and RHS of eq. (3.114) should be understood as in the comoving frame and lab frame for primed and unprimed quantities, respectively. The inverse transformations for the \mathbf{E} and \mathbf{B} fields are given by reversing the sign of $\boldsymbol{\beta}$,

$$\begin{aligned} \mathbf{E}_{\parallel} &= \mathbf{E}'_{\parallel}, & \mathbf{B}_{\parallel} &= \mathbf{B}'_{\parallel}, \\ \mathbf{E}_{\perp} &= \gamma(\mathbf{E}'_{\perp} - \boldsymbol{\beta} \times \mathbf{B}'), & \mathbf{B}_{\perp} &= \gamma(\mathbf{B}'_{\perp} + \boldsymbol{\beta} \times \mathbf{E}'), \end{aligned} \quad (3.115)$$

In the following, we provide an intuitive understanding the field transformation above.

Let us consider a very thin, uniformly charged slab inside the x' - y' plane, with charge density ρ' and thickness h' in its comoving frame \mathcal{O}' . We can calculate the electric field strengths in the comoving frame, using the Gauss' theorem (Maxwells equations apply in all Lorentz frames)

$$\nabla' \cdot \mathbf{E}' = 4\pi\rho' \Rightarrow \int (\nabla' \cdot \mathbf{E}') dV' = \oint \mathbf{E}' \cdot d\mathbf{S}' = 4\pi \int \rho' dV'. \quad (3.116)$$

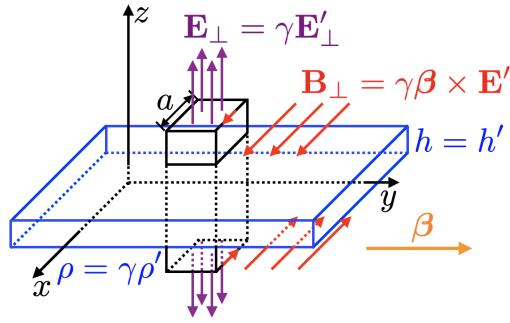


Figure 3.10: EM fields from a uniformly charged slab moving with velocity $\beta = \beta \hat{y}$.

Due to the symmetry of the problem, the electric fields must be perpendicular to the slab, along the $+\hat{z}$ and $-\hat{z}$ direction above and below the slab, respectively. For a cubic volume that extends across the slab with a projected area of A' (see Fig. 3.10), we obtain the area integral $\oint \mathbf{E} \cdot d\mathbf{S}' = 2E'A'$ and the volume integral $\int \rho' dV' = \rho'A'h'$, so the electric field strength outside the slab in the comoving frame \mathcal{O}' is

$$\mathbf{E}' = 2\pi\rho'h \frac{z}{|z|} \hat{z}. \quad (3.117)$$

There is no magnetic field in the comoving frame $\mathbf{B}' = 0$.

Now consider that, in the lab frame \mathcal{O} , the slab is moving with velocity $\mathbf{v} = \beta c \hat{y}$. Because of this motion, the charge density in the lab frame is given by $\rho = \gamma\rho'c$, because the number density of charged particles transforms as $n = \gamma n'$ as a result of length contraction. Again, the electric fields in the lab frame must be perpendicular to the slab, parallel to the \hat{z} axis. Again, we consider a cubic volume that extends across the slab with a projected area A . Applying Gauss' theorem in the lab frame,

$$\nabla \cdot \mathbf{E} = 4\pi\rho \Rightarrow \int (\nabla \cdot \mathbf{E}) dV = \oint \mathbf{E} \cdot d\mathbf{S} = 2EA = 4\pi \int \rho dV = 4\pi\rho Ah, \quad (3.118)$$

and using $h = h'$, we obtain

$$\mathbf{E} = 2\pi\rho h \frac{z}{|z|} \hat{z} = \gamma \mathbf{E}'. \quad (3.119)$$

As the relative motion is perpendicular to the electric field in both frames, we conclude $\mathbf{E}_\perp = \gamma \mathbf{E}'_\perp$ (in agreement with eq. 3.115 for $\mathbf{B}' = 0$).

In the lab frame, there is also a current density $\mathbf{J} = \rho\mathbf{v} = \gamma\rho'\beta c \hat{y}$ near the x-y plane, which means that there will be magnetic fields that are parallel to the slab. Let us consider a rectangular loop that is in a plane parallel to the x-z plane and extends across the slab. The length of the segment parallel to the x-axis is a . We then apply the Stokes theorem

as follows (note the vanishing displacement current as $\partial_t \mathbf{E} = 0$)

$$\nabla \times \mathbf{B} = 4\pi \mathbf{J} \Rightarrow \int (\nabla \times \mathbf{B}) \cdot d\mathbf{A} = \oint \mathbf{B} \cdot d\ell = 4\pi \int \mathbf{J} \cdot \mathbf{A}. \quad (3.120)$$

For our geometry, we obtain the loop integral $\oint \mathbf{B} \cdot d\ell = 2Ba$ and the area integral $\int \mathbf{J} \cdot \mathbf{A} = J ah = \gamma\beta\rho'cah$, so the magnetic field strength outside the slab in the lab frame is

$$\mathbf{B} = 2\pi\gamma\beta\rho'h \frac{z}{|z|} \hat{x} = \gamma\beta \times \mathbf{E}', \quad (3.121)$$

which is in agreement with eq. (3.115) for $\mathbf{B}' = 0$.

Although we have only considered a special case in the above discussion, one can generalize the consideration to a slab moving in an arbitrary direction in the lab frame and verify the correctness of the Lorentz transformation of the EM fields. For instance, if the slab is moving with velocity $\beta = \beta\hat{z}$, which is along the normal direction of the slab. This motion does not change the charge column density of the slab, so we obtain $\mathbf{E}_{||} = \mathbf{E}'_{||} = 2\pi\rho'h'$ outside the slab. There will be no magnetic field outside the slab.

3.6.3 Coulomb field of a moving charge

Consider a point charge moving at a constant velocity $\mathbf{v} = \beta c \hat{x}$ relative to an observer, as shown in Fig. 3.11. The impact parameter, i.e. the minimum separation between the particle and observer, for the relative motion is b . In the observer's frame \mathcal{O} , at a given coordinate time t (simultaneous for the charge and the observer), the 3D Cartesian position of the charge is $\mathbf{x}_q = (\Delta x, -b, 0)$ and the observer is at the coordinate origin $\mathbf{x}_{\text{obs}} = (0, 0, 0)$. Without loss of generality, we take $t = 0$ to be the moment when the charge crosses the y -axis in the observer's frame, and this corresponds to

$$\Delta x = \beta ct. \quad (3.122)$$

In the charge's comoving frame \mathcal{O}' , the observer is moving at velocity $\mathbf{v}' = -\beta c \hat{x}'$. We place the charge at the (fixed) 3D Cartesian position $\mathbf{x}'_q = (0, -b, 0)$ on the y' -axis, and the observer's position is denoted as $\mathbf{x}'_{\text{obs}} = (\Delta x', 0, 0)$. Our plan is to first calculate $\Delta x'$, and then the Coulomb field \mathbf{E}' at the observer's position \mathbf{x}'_{obs} in the charge's comoving frame \mathcal{O}' can be easily obtained, and finally we use eq. (3.115) to Lorentz transform the electric field from frame \mathcal{O}' to frame \mathcal{O} . Our final goal is to obtain the electric field \mathcal{E} and magnetic field \mathbf{B} at the observer's position at time t in frame \mathcal{O} .

At each coordinate time t in the observer's frame \mathcal{O} , we consider the charge and observer to be *simultaneous* (i.e., $\Delta t = 0$), and hence, according to Lorentz transformation, we obtain the separation between the charge and observer along the x' -axis

$$\Delta x' = -\gamma\beta c\Delta t + \gamma\Delta x = \gamma\Delta x. \quad (3.123)$$

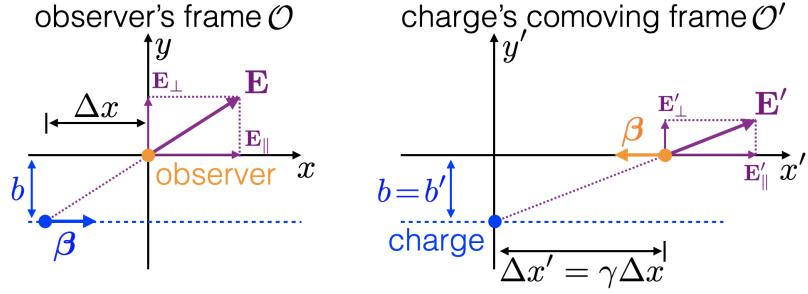


Figure 3.11: Geometry for a charge undergoing constant motion relative to an observer, with an impact parameter of b .

It should be noted that simultaneity in frame \mathcal{O} does not mean the same in frame \mathcal{O}' , because Lorentz transformation gives $c\Delta t' = \gamma c\Delta t - \gamma\beta\Delta x = -\gamma\beta\Delta x \neq 0$ (although our following discussion will not use $\Delta t'$). In the transverse direction, the impact parameter stays unchanged in both frames $b' = b$.

Since we have switched the perspective between the two frames and defined $\Delta x = x_q^1 - x_{\text{obs}}^1$ and $\Delta x' = x_{\text{obs}}^1 - x_q^1$ (where “1” means the x-component), there is a sign change

$$\Delta x' = -\gamma\Delta x = -\gamma\beta ct, \quad (3.124)$$

which means that for $t < 0$ we have $\Delta x < 0$ but $\Delta x' > 0$ (as the charge is on the left side of the observer). Although, this sign change purely comes from our notation and has no physical significance, this might be a bit confusing to some readers. It is possible to demonstrate this more explicitly by Lorentz transforming the 4-displacements of the charge and observer (from the spacetime origin) from frame \mathcal{O}

$$\vec{x}_q = (t, \beta ct, -b, 0), \quad \vec{x}_{\text{obs}} = (t, 0, 0, 0), \quad (3.125)$$

to frame \mathcal{O}'

$$\vec{x}'_q = (ct/\gamma, 0, -b, 0), \quad \vec{x}'_{\text{obs}} = (\gamma ct, -\gamma\beta ct, 0, 0). \quad (3.126)$$

Thus, one can see that $\Delta x' = -\gamma\Delta x$.

At observer's time t , the separation between the charge and observer in the charge's comoving frame is

$$R'(t) = \sqrt{(\Delta x')^2 + b^2} = \sqrt{(\gamma\beta ct)^2 + b^2}. \quad (3.127)$$

Then, we write down the Coulomb field at the observer's position in the charge's comoving frame \mathcal{O}' ,

$$\mathbf{E}' = \frac{q}{(R')^2} \left[\frac{\Delta x'}{R'} \hat{x}' + \frac{b}{R'} \hat{y}' \right], \quad (3.128)$$

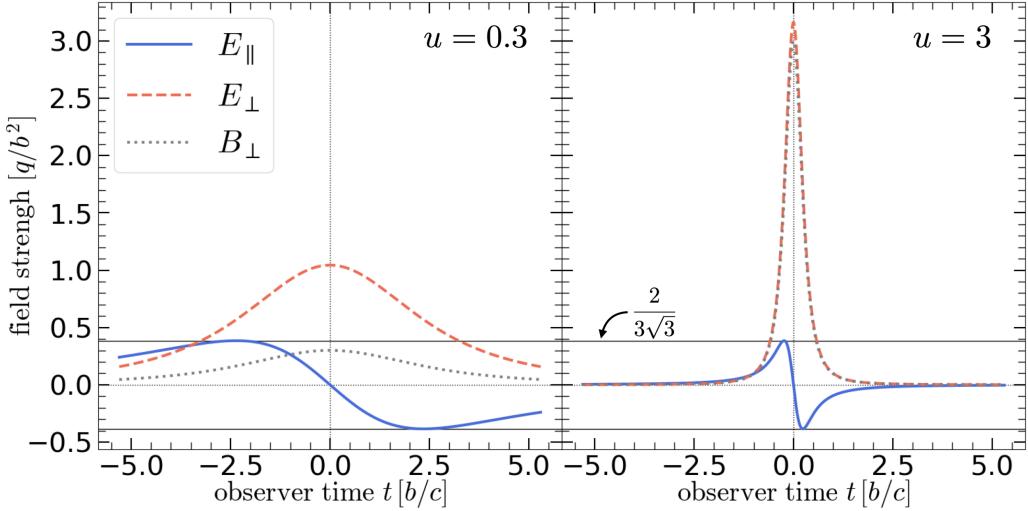


Figure 3.12: The electric and magnetic fields from a moving charge q with constant velocity β at impact parameter b . The two panels show different velocities $u = \gamma\beta = 0.3$ and 3 . In the relativistic case ($u = 3$), the transverse magnetic field becomes comparable to the transverse electric field $B_\perp \approx E_\perp \gg E_\parallel$.

where the first and second terms are E'_\parallel and E'_\perp respectively. Then, we Lorentz transform the parallel and perpendicular components of the electric field to the observer's frame and obtain (note that $\mathbf{B}' = 0$ in frame \mathcal{O}')

$$\begin{aligned} \mathbf{E}_\parallel &= \mathbf{E}'_\parallel = -\frac{q}{(R')^3}\gamma\beta ct\hat{x}, \quad \mathbf{E}_\perp = \gamma\mathbf{E}'_\perp = \frac{q}{(R')^3}\gamma b\hat{y}; \\ \mathbf{B}_\parallel &= \mathbf{B}'_\parallel = 0, \quad \mathbf{B}_\perp = \gamma\beta \times \mathbf{E}' = \frac{q}{(R')^3}\gamma\beta b\hat{z}, \end{aligned} \quad (3.129)$$

or

$$\mathbf{E} = \frac{\gamma q}{[(\gamma\beta ct)^2 + b^2]^{3/2}}(-\beta ct\hat{x} + b\hat{y}); \quad \mathbf{B} = \frac{\gamma q}{[(\gamma\beta ct)^2 + b^2]^{3/2}}\beta b\hat{z}. \quad (3.130)$$

The three non-zero components of the EM fields are shown in Fig. 3.12.

The maximum amplitude of E_\parallel is achieved at $t = \pm b/(\sqrt{2}\gamma\beta c)$ when $E_\parallel = \pm 2q/(3\sqrt{3}b^2)$, which is of the same order as the Coulomb field of a non-relativistic charge at separation b . The maximum amplitude of E_\perp and B_\perp are achieved at $t = 0$ when $E_\perp = \gamma q/b^2$ and $B_\perp = \gamma\beta q/b^2 = \beta E_\perp$. In the ultra-relativistic limit $\gamma \gg 1$, the fields are mostly transverse with $E_\perp \approx B_\perp \gg E_\parallel$. This means that the field lines are bundled up near the directions perpendicular to the velocity, as schematically shown in Fig. 3.13. The field strength only stays high for a short amount of time $\Delta t \sim b/(\gamma\beta c)$, which corresponds to a Fourier frequency width of $\Delta\omega \simeq 1/\Delta t \sim \gamma\beta c/b$. In the non-relativistic limit ($\beta \ll 1$), the two

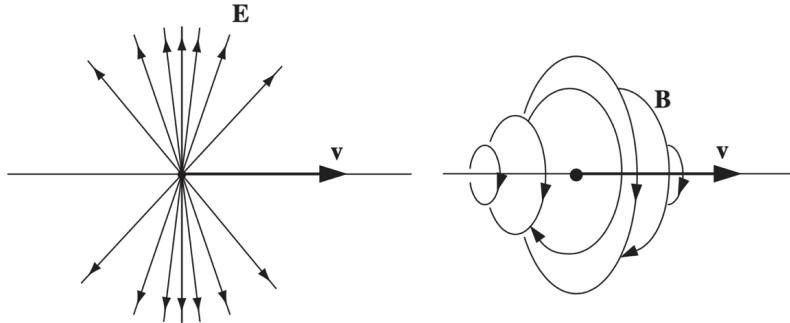


Figure 3.13: The lab-frame fields from a uniformly moving positive charge (at a mildly relativistic speed). Taken from Figs. 10.10 and 10.11 in Griffiths' book.

electric field components E_{\parallel} and E_{\perp} have comparable maximum values (of the order q/b^2) but the magnetic field is negligible as it is smaller than the electric field by a factor of $\beta \ll 1$.

Finally, we note that

$$\frac{E_{\parallel}}{E_{\perp}} = \frac{\Delta x}{b} = \frac{\beta ct}{b}, \quad (3.131)$$

which means that the electric field vector always points from *current* position of the particle (at t) to the observer, regardless of how fast the particle is moving at. This result is counter-intuitive, as one might naively and incorrectly expect that, based on light-travel delay (or the *principle of causality*), that the E-field should point from the particle position at the retarded time (when a causal signal reaching the observer at t was emitted). This issue will be discussed in the next chapter where the EM fields from a particle undergoing acceleration will be presented.

3.7 Lorentz invariant quantities of radiation field

Often we want to study a given radiation field in different Lorentz frames. A labor-efficient way is to use the Lorentz invariant quantities.

3.7.1 Phase-space distribution function

Let us consider a radiation field described the phase-space photon density in a Lorentz frame \mathcal{O} ,

$$f(\mathbf{x}, \mathbf{p}, t) = \frac{dN}{d^3x d^3p}, \quad (3.132)$$

where dN is the number of particles existing in the phase volume element $dV = d^3x d^3p$ at the same coordinate time t . The spacetime locations of all the dN particles can be

considered as individual events, and each of them has a 4-momentum indicating their instantaneous motion. Our goal is to obtain the Lorentz transformation of the distribution function f . The following approach is similar to the [lecture notes](#) by Prof. van Hees.

Let us consider another frame \mathcal{O}' moving at velocity $\mathbf{v} = \beta c \hat{\mathbf{x}}$ with respect to the \mathcal{O} frame. We would like to know the phase volume $d\mathcal{V}' = d^3\mathbf{x}'d^3\mathbf{p}'$ occupied by the same dN particles (or events) in the \mathcal{O}' frame. The 4-momentum of a particle transforms as

$$\begin{aligned} E'/c &= \gamma(E/c - \beta p_x), \\ p'_x &= \gamma(-\beta E/c + p_x), \\ p'_y &= p_y, \quad p'_z = p_z. \end{aligned} \tag{3.133}$$

We obtain the following ratio from the Jacobian (noticing that $p'_y = p_y$ and $p'_z = p_z$)

$$\frac{d^3\mathbf{p}'}{d^3\mathbf{p}} = \frac{\partial p'_x}{\partial p_x} = \gamma \left(1 - \frac{\beta}{c} \frac{\partial E}{\partial p_x} \right) = \frac{\gamma}{E/c} (E/c - \beta p_x) = \frac{E'}{E}, \tag{3.134}$$

where we have made use of the Lorentz invariant $\vec{p} \cdot \vec{p} = -(E/c)^2 + \mathbf{p}^2 = -m^2c^2$ to obtain $\partial E/\partial p_x = p_x/E$. The above argument shows

$$\frac{d^3\mathbf{p}}{E} = \text{Lorentz invariant.} \tag{3.135}$$

The spatial volume element $d^3\mathbf{x}'$ containing the dN events can be obtained based on the picture that any two particles/events can be considered as the ends of a rigid rod, because they have the same 4-momentum in all frames (since we are considering an infinitesimal spread in momentum space). The perpendicular length components stay the same in that $dy' = dy$ and $dz' = dz$, and the parallel component undergoes contraction. Here, it should be noted that in frame \mathcal{O}' , the particles are moving with momentum $\mathbf{p}' \neq 0$, so the length contraction is given by eq. (3.98),

$$\frac{\partial x'}{\partial x} = \gamma(1 + \beta\beta'_x), \tag{3.136}$$

where $\beta'_x = p'_x/(E'/c)$. Then, we obtain, again from the Jacobian,

$$\frac{d^3\mathbf{x}'}{d^3\mathbf{x}} = \frac{\partial x'}{\partial x} = \gamma \left(1 + \beta \frac{-\beta E/c + p_x}{E/c - \beta p_x} \right) = \frac{E}{E'}, \tag{3.137}$$

which means

$$Ed^3\mathbf{x} = \text{Lorentz invariant.} \tag{3.138}$$

Finally, we arrive at the conclusion $d\mathcal{V}' = d\mathcal{V}$ — meaning that

$$f(\mathbf{x}, \mathbf{p}, t) = f'(\mathbf{x}', \mathbf{p}', t') = \text{Lorentz invariant,} \tag{3.139}$$

which holds for photons and massive particles.

A way of testing the Lorentz-invariance of the distribution function is to look at the special case of the Planck distribution, which can be expressed in the center of momentum frame of all photons,

$$f_{\text{Planck}} = \frac{2/h^3}{e^{-h\nu/k_B T} - 1} = \frac{2/h^3}{e^{-\vec{p}\cdot\vec{u}/k_B T} - 1}, \quad (3.140)$$

where in the second expression we have used $h\nu = \vec{p}\cdot\vec{u}$ using particle 4-momentum $\vec{p} = (h\nu/c, \mathbf{p})$ and the 4-velocity of the frame $\vec{u} = (c, \mathbf{0})$ (no bulk motion). Since $\vec{p}\cdot\vec{u}$ is a Lorentz scalar, we know that the distribution function must be Lorentz invariant.

Prof. J. Goodman's [lecture notes](#) include the following (brilliant) thought by his fellow graduate student T. McGlynn. The phase-space volume $dV = \mathbf{x}^3 d\mathbf{p}^3$ involves three factors of $dx^i dp^i$ ($i = x, y, z$). The Uncertainty Principle tells us that $\Delta x^i \Delta p^i \sim h$ (h being the Planck constant) holds in any Lorentz frame, so we know that the Lorentz transformations of the momentum interval Δp_x and spatial separation Δx (and similarly for the y and z components) at a fixed coordinate time must preserve the product of $\Delta x \Delta p_x$. This demonstrates the Lorentz invariance of the phase-space volume dV and hence the phase-space distribution function f .

Another way of thinking is that, since the photon occupation number η_γ (see Ch. 1) is given by $f/(h^3/g_s)$ (where $g_s = 2$ is the degeneracy factor due to spin) and η_γ describes the number of stimulated photons for each spontaneously emitted photon, we see that η_γ only involves photon counting and hence must be Lorentz invariant.

3.7.2 Optical depth, I_ν/ν^3 , j_ν/ν^2 , emitting power, etc.

Since the intensity is related to the distribution function by $I_\nu/\nu^3 \propto f$, we obtain

$$I_\nu/\nu^3 = \text{Lorentz invariant.} \quad (3.141)$$

Since the frequencies in two different frames are related by a Doppler factor $\nu = \mathcal{D}\nu'$, we have $I_\nu = \mathcal{D}^3 I'_\nu$. The frequency-integrated intensities in the two frames are related by $I = \mathcal{D}^4 I'$. The angular dependence of the intensity for an isotropic emitter in the comoving frame is shown in Fig. 3.14.

Below, we summarize a few other quantities that are Lorentz invariant (all observers in different Lorentz frames must agree on the same number).

- Total number of particles in a given system, total rest mass, and total charge⁸ are Lorentz invariant.

⁸In a steady state, a box containing N protons and N electrons all moving at different velocities is overall charge neutral.

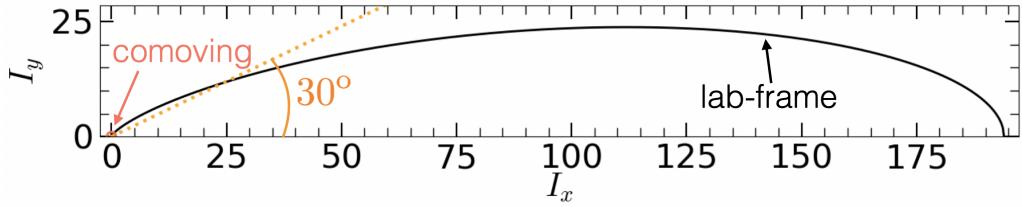


Figure 3.14: Lab-frame angular distribution of intensity, $I(\theta) = \sqrt{I_x^2 + I_y^2} = \mathcal{D}^4(\theta)I'$ at viewing angle $\theta = \arccos(I_x/I)$, for an isotropic emitter in its comoving frame, normalized such that $I' = 1$. Here, the emitter is moving along the \hat{x} axis with a modest Lorentz factor $\gamma = 2$.

- Optical depth τ = Lorentz invariant, since attenuation can be considered as the reduction of the number of photons in the beam by a factor of $e^{-\tau}$.
- Column density $\int n ds$ along any direction = Lorentz invariant, as long as the direction (which can be considered as a ray of photons) is correctly Lorentz transformed between different frames.
- j_ν/ν^2 = Lorentz invariant. This can be shown by carrying out the following “Lorentz transformation gymnastics”

$$j_\nu = n \frac{dE}{dtd\Omega d\nu} = (\gamma n') \frac{(\mathcal{D}dE')}{(\gamma dt')(\mathcal{D}^{-2}d\Omega')(\mathcal{D}\nu')} = \mathcal{D}^2 j'_{\nu'} = \left(\frac{\nu}{\nu'}\right)^2 j'_{\nu'}. \quad (3.142)$$

- The product of the frequency and absorption coefficient $\nu\alpha_\nu$ = Lorentz invariant. This is because $j_\nu = \alpha_\nu S_\nu$ and the source function S_ν transforms like the specific intensity I_ν (an example of this is to consider a thermal distribution for which Kirchhoff’s law gives $S_\nu = B_\nu$).
- Proper time = Lorentz invariant.
- Emitting power $P = dE/dt$ = Lorentz invariant, provided that the radiation does not carry net linear momentum in the comoving frame of the emitter. This is demonstrated by combining time dilation $dt/d\tau = \gamma$ and Lorentz transformation of the total 4-momentum of the emitted radiation $dE = \gamma(dE' + \beta p^1')$ (“1” is along the relative motion). In classical electrodynamics, since the emitting power from an accelerating charged particle has a dipolar angular pattern in the comoving frame (i.e., photons do not carry net linear momentum⁹, see §4.1.2), we have $dp' = 0$ and

⁹Note that there are sources where the emitted photons do carry net linear momentum. For instance, the Earth is hotter on the day side than the night side, so the thermal emission (as well as scattered radiation) from the Earth carries non-zero net linear momentum, so the total emitting power from the Earth is in fact not Lorentz invariant.

hence $dE = \gamma dE'$. Note that the statement of $dE = \gamma dE'$ is not contradictory with eq. (3.109) where we have $dE = \mathcal{D}dE'$ along *one viewing angle*. Here, we are considering the total emitted energy over all angles. We conclude that the emitting power is Lorentz invariant: $P = dE/dt = dE'/d\tau = P'$. As we will show in §4.1.2, for a charge q undergoing acceleration $\mathbf{a}' = d\mathbf{v}'/d\tau$ in its comoving frame, the Larmor formula gives the emitting power $P' = (2q^2/3c^3)(\mathbf{a}')^2$. In the comoving frame, the 4-velocity is $\vec{u} = (c, \mathbf{0})$, so we obtain the 4-acceleration $\vec{a} = (0, d\mathbf{v}'/d\tau)$ and hence $(\mathbf{a}')^2 = \vec{a} \cdot \vec{a}$. Thus, the emitting power is

$$P = \frac{2q^2}{3c^3} \vec{a} \cdot \vec{a} = \text{Lorentz invariant.} \quad (3.143)$$

- $E^2 - B^2$ and $\mathbf{E} \cdot \mathbf{B}$ are both Lorentz invariant.
- Polarization state of a monochromatic plane EM wave (see §3.7.4).

3.7.3 Lorentz transformation of radiation energy density

The transformation of radiation energy density between different Lorentz frames cannot be described by a single Doppler factor. It is important to take into account the anisotropy of the radiation field, as we show below.

Consider an object (or a fluid element) moving at velocity $\beta = \beta \hat{x}$ and Lorentz factor γ in the lab frame. Suppose the radiation field is described by the frequency-integrated intensity functions I in the lab frame and I' in the comoving frame of the object.

The radiation energy density in either of two frames is given by the 0th moment of the intensity function

$$U_{\text{rad}} = c^{-1} \int d\Omega I, \quad U'_{\text{rad}} = c^{-1} \int d\Omega' I' = c^{-1} \int d\Omega I \mathcal{D}^{-2}, \quad (3.144)$$

where the intensities and differential solid angles are related to each other through the Doppler factor \mathcal{D} ,

$$I' = I/\mathcal{D}^4, \quad d\Omega' = \mathcal{D}^2 d\Omega, \quad \mathcal{D} = \gamma(1 + \beta\mu') = \frac{1}{\gamma(1 - \beta\mu)}, \quad (3.145)$$

and $\mu = \cos\theta$, $\mu' = \cos\theta'$ are for the angles θ and θ' between a given ray and the particle's velocity β in the lab and comoving frame, respectively.

In the following, we discuss a few examples.

- A single beam at an angle θ from β . In this case, we obtain

$$\frac{U'_{\text{rad}}}{U_{\text{rad}}} = \frac{1}{\mathcal{D}^2(\theta)} = \gamma^2(1 - \beta\mu)^2, \quad \text{for a single beam.} \quad (3.146)$$

This applies to the case where we would like to know the radiation energy density from a point-like star in the comoving frame of a fast spacecraft moving in the interstellar medium. For an ultra-relativistic object ($\beta \approx 1$), the energy density ratio between the two frames spans a wide range from $(4\gamma^2)^{-1}$ (for $\mu = 1$) and $4\gamma^2$ (for $\mu = -1$).

- Isotropic radiation field in the lab frame. In this case, we obtain $U_{\text{rad}} = 4\pi I/c$ and

$$U'_{\text{rad}} = \frac{2\pi I}{c} \gamma^2 \int_{-1}^1 (1 - 2\beta\mu + \beta^2\mu^2) d\mu = \frac{4\pi I}{c} \gamma^2 (1 + \beta^2/3), \quad (3.147)$$

so we obtain

$$\frac{U'_{\text{rad}}}{U_{\text{rad}}} = \gamma^2 (1 + \beta^2/3), \quad \text{for isotropic radiation field in lab frame.} \quad (3.148)$$

For example, the Earth is moving through the cosmic microwave background at a velocity of roughly 600 km/s or $\beta \approx 0.002$, so we obtain $U'_{\text{rad}}/U_{\text{rad}} \approx 1 + 4\beta^2/3 \approx 1 + 5 \times 10^{-6}$ — this is a very small difference. In the other case where the radiation field is isotropic in the object's comoving frame, one can show that $U_{\text{rad}}/U'_{\text{rad}} = \gamma^2(1 + \beta^2/3)$.

- Isotropic radiation field in one hemisphere in the lab frame. We consider two hemispheres with $-1 < \mu < 0$ and $0 < \mu < 1$ separately. In these two cases, it is easy to show that

$$\frac{U'_{\text{rad}}}{U_{\text{rad}}} = \begin{cases} \gamma^2 \int_{-1}^0 (1 - 2\beta\mu + \beta^2\mu^2) d\mu = \gamma^2(1 + \beta + \beta^2/3), & \text{for } -1 < \mu < 0, \\ \gamma^2 \int_0^1 (1 - 2\beta\mu + \beta^2\mu^2) d\mu = \gamma^2(1 - \beta + \beta^2/3), & \text{for } 0 < \mu < 1. \end{cases} \quad (3.149)$$

This two cases apply to a particle near the surface of a star (ignoring limb darkening) moving radially away or towards the stellar center.

3.7.4 *Lorentz invariance of polarization state

Here we show that the polarization state of a given wave is Lorentz invariant, even though the wave amplitude is not. For any elliptically polarized wave, the electric vector rotates in the same manner in all inertial frames, and the azimuthal angle describing the rotation is invariant just as the wave phase. The simplest way of understanding is that each photon is either left- or right-circularly polarized and the *helicity* is Lorentz invariant, since it is not possible to find an inertial frame where the photon moves backwards.

Let us consider a general elliptically polarized wave described by $E_j(t) = E_{j,0} \cos(\omega t - \phi_j)$ ($j = x, y$) in frame \mathcal{O} . The two Cartesian components of the electric vector are in the x - y plane, and the wave vector \mathbf{k} is along the $\hat{\mathbf{z}}$ axis. A different inertial frame \mathcal{O}' is moving at velocity $\mathbf{v} = \beta c$ with respect to \mathcal{O} , and all three Cartesian axes are aligned with each other in both frames. The angle between the velocity vector and wave vector is $\theta = \arccos(\mathbf{v} \cdot \mathbf{k})$. Without losing generality, the relative velocity can be put in the x - z plane.

The Lorentz transformation (eq. 3.114) of component $E_x \hat{\mathbf{x}}$ and the associated magnetic field $B_y \hat{\mathbf{y}} = E_x \hat{\mathbf{y}}$ gives

$$\mathbf{E}'_1 = [\gamma(1 - \beta \cos \theta) - (\gamma - 1) \sin^2 \theta] E_x \hat{\mathbf{x}}' + \sin \theta [\gamma \beta - (\gamma - 1) \cos \theta] E_x \hat{\mathbf{z}}'. \quad (3.150)$$

Similarly for the component $E_y \hat{\mathbf{y}}$ and the associated magnetic field $B_x \hat{\mathbf{x}} = -E_y \hat{\mathbf{x}}$, one obtains

$$\mathbf{E}'_2 = \gamma(1 - \beta \cos \theta) E_y \hat{\mathbf{y}}'. \quad (3.151)$$

The total electric field in the \mathcal{O}' frame is given by $\mathbf{E}' = \mathbf{E}'_1 + \mathbf{E}'_2$. The wave vector in the \mathcal{O}' frame may be obtained directly by a transformation of the four-vector $\vec{k} = (\omega/c, \mathbf{k})$, but a simpler way is to make use of the fact that the wave is transverse in any frame $\mathbf{k}' \perp \mathbf{E}'_1$ and $\mathbf{k}' \perp \mathbf{E}'_2$. Thus, the direction of the wave vector $\hat{\mathbf{k}}'$ in the \mathcal{O}' frame is given by $k'_x/k'_z = -E'_z/E'_x$ and $k'_y = 0$. The magnitude of \mathbf{k}' is given by a Doppler shift $k' = \gamma(1 - \beta \cos \theta)k$, because $\omega' = k'$ in vacuum. Thus, in the \mathcal{O} frame, the wave is still elliptically polarized, but the components have been stretched — in the $\hat{\mathbf{y}}'$ direction by a factor of $|\mathbf{E}'_2|/E_y$ and in the $\hat{\mathbf{y}}' \times \hat{\mathbf{k}}$ direction by a factor of $|\mathbf{E}'_1|/E_x$. It is easy to show that the two stretching factors are identical $|\mathbf{E}'_1|/E_x = |\mathbf{E}'_2|/E_y = \gamma(1 - \beta \cos \theta)$, and therefore the shape and orientation of the polarization ellipse (and hence the polarization location on the Poincaré sphere) stay unchanged between different frames.

3.8 Energy-momentum tensor (or stress-energy tensor)

In this section, we discuss the energy-momentum tensors for radiation field and ideal gas (the energy-momentum tensor for classical EM fields will be discussed in §4.2.8). The *energy-momentum tensor* (also called *stress-energy tensor*) of a given system ($T^{\mu\nu}$) is defined as

$$T^{\mu\nu} = \begin{cases} \text{flux of } \mu \text{ component of 4-momentum} \\ \text{across a surface of constant } x^\nu, \end{cases} \quad (3.152)$$

where $(x^\nu) = (ct, x, y, z)$ are the spacetime coordinates. More specifically,

- (1) T^{00} = energy density.
- (2) T^{0i} = energy flux (divided by c — to have the same dimension as the energy density).
- (3) T^{i0} = momentum density (multiplied by c) = energy flux (divided by c). Note that momentum density = relativistic mass \times velocity divided by volume and that energy flux = energy divided by volume \times velocity. Since energy = relativistic mass $\times c^2$, we see that the momentum density has the same physical meaning as the energy flux, apart from factors of c . This is why $T^{0i} = T^{i0}$.
- (4) T^{ij} = momentum flux = pressure/stress. We usually refer to the diagonal terms T^{ii} as pressure on a surface of constant x^i and the off-diagonal terms T^{ij} (for $i \neq j$) as

the *shear stress*. The 3×3 tensor (T^{ij}) is called the *pressure tensor* or *stress tensor*. It is possible to show that (T^{ij}) is always symmetric (we omit the details here).

In special relativity¹⁰, the importance of $(T^{\mu\nu})$ is that it captures all energies, momenta, pressures, and stresses in one single form and it allows us to easily write the energy and momentum conservation laws. In the following, we will show that the interactions between the radiation field and gas is captured by the 4-force density that describes the energy and momentum exchange rate between the two components.

Let us start with the 0th, 1st, 2nd moments of the radiation field (see §1.4.3)

$$U = c^{-1} \int d\Omega I, \quad \mathbf{F} = \int d\Omega I \hat{\mathbf{n}}, \quad \overset{\leftrightarrow}{P} = c^{-1} \int d\Omega I \hat{\mathbf{n}} \hat{\mathbf{n}}, \quad (3.153)$$

which are the energy density, energy flux vector, and pressure tensor. These three moments form a 4×4 , symmetric *energy-momentum tensor*

$$(R^{\mu\nu}) = \begin{pmatrix} U & \mathbf{F}^T/c \\ \mathbf{F}/c & \overset{\leftrightarrow}{P} \end{pmatrix}, \quad (3.154)$$

where the Greek indices μ and ν stands for the four spacetime dimensions. The energy-momentum tensor $(R^{\mu\nu})$ has 9 independent components given the its symmetry and the fact that $\text{Tr}(\overset{\leftrightarrow}{P}) = U$. In the spacial case of a perfectly isotropic radiation field (with vanishing net flux), the energy-momentum tensor takes the following diagonal form

$$(R^{\mu\nu}) = \begin{pmatrix} U & 0 & 0 & 0 \\ 0 & U/3 & 0 & 0 \\ 0 & 0 & U/3 & 0 \\ 0 & 0 & 0 & U/3 \end{pmatrix}. \quad (\text{for isotropic radiation field}) \quad (3.155)$$

In the absence of interactions with matter, the energy and momentum conservation laws for the radiation field are given by (see §2.6.7)

$$\partial_t U + \nabla \cdot \mathbf{F} = 0, \quad c^{-2} \partial_t \mathbf{F} + \nabla \cdot \overset{\leftrightarrow}{P} = 0. \quad (\text{for radiation in vacuum}) \quad (3.156)$$

The four equations above can be written in a more compact way as follows

$$R_{,\mu}^{\mu\nu} \equiv \sum_{\mu} \frac{\partial R^{\mu\nu}}{\partial x^{\mu}} = 0, \quad (\text{for radiation in vacuum}) \quad (3.157)$$

where $(x^{\mu}) = (ct, x, y, z)$ stands for the four spacetime coordinates.

¹⁰In general relativity, the supreme importance of $(T^{\mu\nu})$ lies in that it is the source of gravitational field or spacetime curvature.

Let us then study the energy-momentum tensor for ideal gas — we ignore e.g., Coulomb interactions between particles. In the comoving frame of the fluid's bulk motion, we know that (1) the total gas energy density U_g only involves rest-mass energy density $\rho_0 c^2$ and thermal energy density e due to microscopic (thermal or degeneracy) particle motions,

$$U_g = \rho_0 c^2 + e, \quad (3.158)$$

and (2) the gas pressure tensor is isotropic

$$\overset{\leftrightarrow}{P}_g = \begin{pmatrix} P_g & 0 & 0 \\ 0 & P_g & 0 \\ 0 & 0 & P_g \end{pmatrix}. \text{ (in the comoving frame for ideal gas)} \quad (3.159)$$

Depending on the equation of state (whether microscopic motions are relativistic and whether particles are degenerate), the thermal energy density e generally depends on the rest-mass density ρ_0 and pressure P_g in the comoving frame. For instance, one can show that, if the microscopic motions are non-relativistic (or ultra-relativistic), we have $e = 2P_g/3$ (or $e = P_g/3$).

Since microscopic particle motions are isotropic and hence do not carry a net energy flux in the comoving frame of the fluid's bulk motion, we can then write the energy-momentum tensor in this frame (denoted by a prime)

$$(T^{\mu'\nu'}) = \begin{pmatrix} \rho_0 c^2 + e & 0 & 0 & 0 \\ 0 & P_g & 0 & 0 \\ 0 & 0 & P_g & 0 \\ 0 & 0 & 0 & P_g \end{pmatrix}. \text{ (in the comoving frame for ideal gas)} \quad (3.160)$$

Our next task is to calculate the energy-momentum $(T^{\mu\nu})$ in the lab frame where the fluid has a bulk motion described by a 4-velocity (u^μ) . Such a task can be done with Lorentz transformation $(T^{\mu'\nu'}) \rightarrow (T^{\mu\nu})$, which involves tedious tensor multiplications. A simpler way is to directly write $(T^{\mu'\nu'})$ in a tensor form using the fact that $(u^{\mu'}) = c(1, \mathbf{0})$ in the comoving frame, and one can show that eq. (3.160) is identical to the following

$$T^{\mu'\nu'} = (\rho_0 + e/c^2 + P_g/c^2)u^{\mu'}u^{\nu'} + P_g\eta^{\mu'\nu'}, \quad (3.161)$$

where $(\eta^{\mu\nu})$ is the Minkowski metric. For instance, $T^{0'0'} = (\rho_0 + e/c^2 + P_g/c^2)c^2 - P_g = \rho c^2 + e$ (as $\eta^{0'0'} = -1$), and one can easily verify other components. The above expression uniquely implies that the energy-momentum tensor in the lab frame is given by

$$T^{\mu\nu} = (\rho_0 + e/c^2 + P_g/c^2)u^\mu u^\nu + P_g\eta^{\mu\nu}, \quad (3.162)$$

where ρ_0 is the gas rest-mass density, e is the thermal energy density, P_g is the gas pressure (all these three are defined in the comoving frame), $(u^\mu) = \gamma(c, \mathbf{v})$ is the four-velocity for

the gas bulk motion in the lab frame (γ = Lorentz factor), and $\eta^{\mu\nu} = \text{diag}(-1, 1, 1, 1)$ is the Minkowski metric.

In the absence of interactions with radiation (or any other fields), the energy and momentum conservation laws can be written as (the same form as eq. 3.157)

$$T_{,\mu}^{\mu\nu} \equiv \sum_{\mu} \frac{\partial T^{\mu\nu}}{\partial x^{\mu}} = 0, \quad (\text{for matter only}) \quad (3.163)$$

which holds for all four components $\nu = 0, 1, 2, 3$ (or t, x, y, z).

Let us then allow energy and momentum exchange between the radiation field and matter, and we denote the coupling between the two by the *radiation 4-force density* (G^{ν}), and the law of total energy and momentum conservation can be written as

$$R_{,\mu}^{\mu\nu} = -G^{\nu} = -T_{,\mu}^{\mu\nu}. \quad (3.164)$$

Back in §2.6.7, we studied the 0th and 1st moments of the radiation transfer equation in the comoving frame of the gas motion. From eqs. (2.334) and (2.340), we obtain the radiation 4-force density in the gas comoving frame (denoted with a prime)

$$\begin{aligned} G^{0'} &= (\alpha_E U' - \alpha_P a T_g^4), \\ G^{1'} &= (\alpha_F + \sigma_F) \mathbf{F}' / c, \end{aligned} \quad (3.165)$$

where U' and \mathbf{F}' are the radiation energy density and flux vectors in the gas comoving frame, and four frequency-averaged attenuation coefficients (related to the gas opacity) are also evaluated in the comoving frame. Note that $G^{\nu'}$ is defined in the gas comoving frame where the emissivity, absorption/scattering coefficients may be considered as isotropic. However, $G^{\nu'}$ is somewhat inconvenient to use because we must perform a Lorentz transformation $R^{\mu\nu} \rightarrow R^{\mu'\nu'}$ to obtain U' and \mathbf{F}' in the gas comoving frame — the readers are referred to detailed discussions by [Mihalas & Auer \(2001\)](#), [Krumholz et al. \(2007\)](#), and [McKinney et al. \(2015\)](#).

Before closing this section, we verify the temporal ($\nu = 0$) and spatial ($\nu = i = 1, 2, 3$) components of the gas-only conservation law in eq. (3.163) in the non-relativistic limit. The i th spatial component of eq. (3.163) is given by

$$\begin{aligned} 0 &= \partial_t [\gamma^2(\rho_0 + e/c^2 + P_g/c^2)v^i] + \nabla \cdot [\gamma^2(\rho_0 + e/c^2 + P_g/c^2)v^i \mathbf{v}] + \partial P_g / \partial x^i \\ &\approx \partial_t (\rho v^i) + \nabla \cdot [\rho v^i \mathbf{v}] + \partial P_g / \partial x^i \\ &= \partial_t (\rho v^i) + \mathbf{v} \cdot \nabla(\rho v^i) + (\nabla \cdot \mathbf{v})\rho v^i + \partial P_g / \partial x^i \\ &= d(\rho v^i)/dt - v^i d\rho/dt + \partial P_g / \partial x^i \\ &= \rho dv^i / dt + \partial P_g / \partial x^i, \end{aligned} \quad (3.166)$$

where we have used approximations for non-relativistic gas ($\rho_0 \approx \rho$ and $e \ll \rho c^2$) and the result $\nabla \cdot \mathbf{v} = -\rho d\rho/dt$ from the mass continuity equation (not included in $T_{,\mu}^{\mu\nu} = 0$)

$$0 = \partial_t \rho + \nabla \cdot (\rho \mathbf{v}) = \partial_t \rho + \mathbf{v} \cdot \nabla \rho + \rho (\nabla \cdot \mathbf{v}) = d\rho/dt + \rho (\nabla \cdot \mathbf{v}). \quad (3.167)$$

Note that $d/dt = \partial_t + \mathbf{v} \cdot \nabla$ stands for the Lagrangian derivative = describing temporal variations along with the gas motion. The above result shows that $T_{,\mu}^{\mu i} = 0$ is equivalent to momentum conservation in hydrodynamics (as $d\mathbf{v}/dt = -\rho^{-1} \nabla P$).

The temporal component of eq. (3.163) is given by

$$\begin{aligned} 0 &= \partial_t [\gamma^2(\rho_0 c^2 + e + P_g) - P_g] + \nabla \cdot [\gamma^2(\rho_0 c^2 + e + P_g) \mathbf{v}] \\ &\approx \partial_t [\gamma \rho c^2 + e] + \nabla \cdot [(\gamma \rho c^2 + e + P_g) \mathbf{v}] \\ &\approx [\partial_t \rho + \nabla \cdot (\rho \mathbf{v})] + \partial_t (\rho v^2/2 + e) + \nabla \cdot [\rho v^2/2 + e + P_g \mathbf{v}] \\ &= \partial_t (\rho v^2/2 + e) + \nabla \cdot [(\rho v^2/2 + e + P_g) \mathbf{v}], \end{aligned} \quad (3.168)$$

where we have used various approximations for non-relativistic gas ($\rho = \gamma \rho_0$, $\gamma^2 e \approx e$, $\gamma^2 P_g \approx P_g$, $\gamma \rho c^2 \approx \rho c^2 + \rho v^2/2$) and in the final expression we have used the mass continuity equation (3.167). The above result shows that that $T_{,\mu}^{\mu 0} = 0$ is equivalent to energy conservation in hydrodynamics — if it is not obvious, one can expand the divergence term and show that the final equation is equivalent to

$$\frac{de}{dt} - \frac{e + P_g}{\rho} \frac{d\rho}{dt} = \rho T_g \frac{ds}{dt} = 0, \quad (3.169)$$

where s is the gas entropy per unit mass, T_g is the gas temperature, and $T_g ds$ is the heat gain/loss per unit mass (which vanishes).

3.9 *Pair production at relativistic temperatures (LTE)

At extremely high temperatures $k_B T \gtrsim m_e c^2$ in LTE, electron-positron pairs co-exist with radiation. The total energy density (or pressure) is no longer aT^4 (or $aT^4/3$), because one must include the contribution from pairs. From the pair creation process $\gamma + \gamma \leftrightarrow e^\pm$, we know the chemical potentials of electrons μ_- and positrons μ_+ satisfy

$$\mu_- + \mu_+ = 0. \quad (3.170)$$

We need one more equation to solve for the chemical potentials, and that is given by charge conservation

$$n_- - n_+ = n_e, \quad (3.171)$$

where n_e is the net electron number density (not to be confused with the electron occupation number η_e in Ch. 2) and the lepton number densities are given by

$$n_\pm = \int \frac{2/h^3}{e^{(\epsilon - \mu_\pm)/(k_B T)} + 1} d^3 p, \quad d^3 p = 4\pi p^2 dp, \quad (3.172)$$

and the relativistic particle energy is related to momentum by

$$\epsilon(p) = \sqrt{m_e^2 c^4 + p^2 c^2}. \quad (3.173)$$

The net electron number density is given by

$$n_e \approx \frac{Y_e \rho}{m_p}, \quad (3.174)$$

where Y_e is the fraction of nucleons that are protons (known as the electron fraction), ρ is the mass density of nuclei, and m_p is the proton mass (ignoring binding energy of nuclei). If the chemical potentials are solved, one can then calculate the pressure and energy density of electrons or positrons

$$P_{\pm} = \int \frac{2/h^3}{e^{(\epsilon - \mu_{\pm})/(k_B T)} + 1} \frac{4\pi p v}{3} p^2 dp, \quad v = pc^2/\epsilon, \quad (3.175)$$

$$U_{\pm} = \int \frac{2/h^3}{e^{(\epsilon - \mu_{\pm})/(k_B T)} + 1} 4\pi \epsilon p^2 dp. \quad (3.176)$$

These integrals can be carried out by changing variables, and it is convenient to define the following Fermi-Dirac integrals (Cox & Giuli 1968)

$$F_k(\eta, \Theta) \equiv \int_0^\infty \frac{x^k (1 + 0.5\Theta x)^{1/2}}{\exp(x - \eta) + 1}, \quad \Theta = k_B T / m_e c^2, \quad x = (\epsilon / m_e c^2 - 1) / \Theta, \quad (3.177)$$

This leads to

$$\begin{aligned} n_{\pm} &= \frac{\sqrt{2}}{\pi^2 V} \Theta^{3/2} \left(F_{\frac{1}{2}} + \Theta F_{\frac{3}{2}} \right), \\ P_{\pm} &= \frac{2^{3/2} m_e c^2}{3\pi^2 V} \Theta^{5/2} \left(F_{\frac{3}{2}} + \frac{\Theta}{2} F_{\frac{5}{2}} \right), \\ U_{\pm} &= n_{\pm} m_e c^2 + \frac{\sqrt{2} m_e c^2}{\pi^2 V} \Theta^{5/2} (F_{\frac{3}{2}} + \Theta F_{\frac{5}{2}}), \end{aligned} \quad (3.178)$$

where the two arguments of the integral functions F_k are $\eta_{\pm} \equiv (\mu_{\pm} - 1) / \Theta$ and Θ , and V is a volume unit defined as

$$V = \frac{(m_e c)^3}{\hbar^3}. \quad (3.179)$$

The full equation of state for a radiation+gas mixture is shown in Fig. 3.15. For more extensive discussion of a plasma consisting of protons, neutrons, helium nuclei, electron-positron pairs, and radiation, see [pagen.pdf](#) at this [Github link](#).

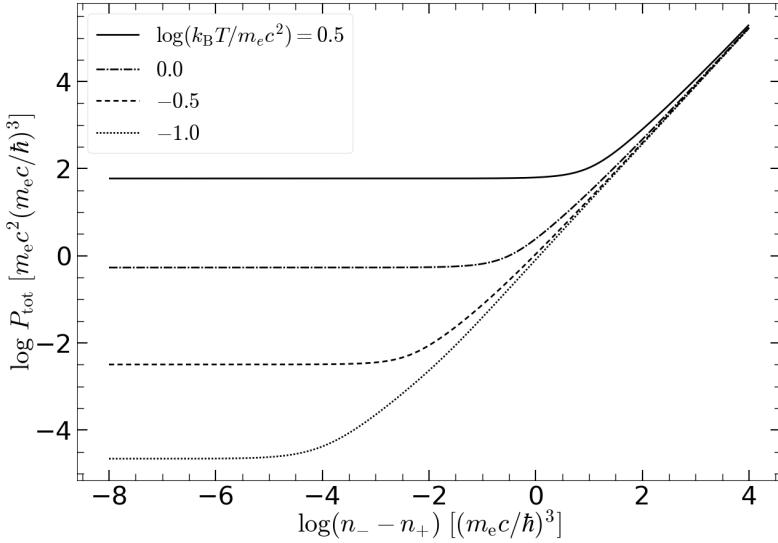


Figure 3.15: Equation of state for a mixture of radiation and matter both at LTE — total pressure $P_{\text{tot}} = P_{\text{nuclei}} + P_{\pm} + P_{\gamma}$ as a function of net electron density $n_e = n_- - n_+$, at different temperatures. Matter is assumed to be protons and electrons (i.e., ionized hydrogen) plus pairs.

In the limit of zero net electron density $n_e \rightarrow 0$ (the radiation pressure-dominated regime), we have $\mu_{\pm} \rightarrow 0$ and hence the pair pressure is given by

$$P_{\pm} = \frac{m_e c^2}{3\pi^2 V} \int_0^\infty \frac{u^3 \beta du}{e^{\gamma/\Theta} + 1}. \quad (3.180)$$

When $\Theta \ll 1$, there are no pairs and $P_{\pm} \rightarrow 0$. At ultra-relativistic temperatures $\Theta \gg 1$, we have

$$P_- + P_+ \approx \frac{7\pi^2 \Theta^4}{180} \frac{m_e c^2}{V} = \frac{7}{12} a T^4, \text{ for } k_B T / m_e c^2 \gg 1, \quad (3.181)$$

and hence the total pressure is given by

$$P_{\text{tot}} \approx P_- + P_+ + P_{\gamma} \approx \frac{11}{12} a T^4, \text{ for } k_B T / m_e c^2 \gg 1. \quad (3.182)$$

3.10 Homework

Prob. 8. Consider a jet that is moving in the lab frame at velocity βc and Lorentz factor $\gamma = (1 - \beta^2)^{-1/2}$ that are time-dependent. The jet is narrowly beamed such that we can approximately consider all fluid elements to be moving at the same velocity. The velocity vector β is misaligned with the observer's line of sight.

The jet is interacting with surrounding gas (the interstellar medium). A famous solution for the hydrodynamic evolution of an ultra-relativistic ($\gamma \gg 1$) spherical blastwave is provided by [Blandford & McKee \(1976\)](#). The basic result is that, for blastwave running into a constant-density medium, energy conservation gives $\gamma^2 r^3 \propto \text{const}$ and r is the distance to the center of explosion. Since $r \simeq ct$, we obtain

$$\gamma(t) \approx \gamma_0 (t/t_0)^{-3/2}, \text{ for } t > t_0, \quad (3.183)$$

where t is the lab-frame time, γ_0 is the initial Lorentz factor of the jet, and t_0 is called the (lab-frame) *deceleration timescale*. Before the deceleration time, the blastwave moves at a constant Lorentz factor $\gamma \simeq \gamma_0$, since not enough material has been swept up yet. This solution also applies for a narrowly beamed jet so long as its opening angle is more than $1/\gamma$, since no information can propagate much beyond the $1/\gamma$ cone due to causality — the fluid elements inside the jet do not know/care if the entire explosion is spherical or not. This means that the radial position of the jet is given by

$$r(t) = c \int_0^t \beta(t') dt' \approx c \int_0^t \left(1 - \frac{1}{2\gamma^2(t')}\right) dt' = \left(1 - \frac{t^3}{8\gamma_0^2 t_0^3}\right) ct = \left(1 - \frac{1}{8\gamma^2(t)}\right) ct, \text{ for } t > t_0. \quad (3.184)$$

Let us approximate the jet as a point source whose trajectory is given by the equation above. An observer located at an angle $\theta = 20^\circ$ away from the jet axis took a high-resolution (radio) image of the jet at observer's time $t_{\text{obs}} = 230$ day since the explosion. The result shows that the angular position of the jet is located at 5 milliarcsec away from that of the center of explosion. We know that the source is located at a distance of about 40 Mpc, so the angular shift corresponds to a distance of ≈ 1 pc projected on the sky.

Note that the observer's time starts at the arrival of the first causal signal (e.g., gravitational waves) that was emitted when the explosion occurred. This means

$$ct_{\text{obs}} = ct - r(t) \cos \theta, \quad (3.185)$$

where t is the lab-frame time when the light detected by the observer was emitted (here t is known as the *retarded time*).

- (i) Calculate the apparent speed of the jet: the projected motion (i.e., proper motion) on the sky divided by observer's time.
- (ii) How far away was the jet from the center of explosion at the retarded time corresponding to $t_{\text{obs}} = 230$ d?
- (iii) What was the jet Lorentz factor at the retarded time that corresponds to $t_{\text{obs}} = 230$ d?
- (iv) Suppose the viewing angle θ is unknown, what additional observation(s) can tell us θ ?

(p.s., the above problem is a simplified version of the famous source GW170817. For earlier epochs $t_{\text{obs}} \lesssim 160$ d, the observer sees *different* emitting fluid elements inside the jet which has some non-trivial internal structure, so the Blandford-McKee solution does not directly apply between different epochs.)

Prob. 9. A *thin* spherical shell expanding at a constant Lorentz factor $\gamma \gg 1$.

- (i) If the material in the shell radiates isotropically in its own local rest frame, what fraction of the shell is observable for a distant observer? (Hint: nearly all the emission is beamed within a cone of angular size $1/\gamma$ near the velocity vector of each fluid element)
- (ii) If the shell only produces radiation when its radius is between r and $r + \Delta r$ (where $\Delta r \ll r$), how long will the observed flare last for a distant observer? (please provide a rough estimate)
- ^{*}(iii) [not required] Suppose the total volume of the shell is V in the lab frame meaning that the radial thickness of the shell is $V/(4\pi r^2)$. If the frequency-integrated volume emissivity is $j' = \int j'_{\nu'} d\nu'$ in the comoving frame of each fluid element, estimate the bolometric flux $F = \int F_{\nu} d\nu$ measured by an observer at distance D .

References

- Jackson, *Classical Electrodynamics* (John Wiley & Sons, New York, 1975).
- Rybicki & Lightman, *Radiative processes in astrophysics* (A Wiley-Interscience Publication, New York, 1979).
- Griffiths, *Introduction to Electrodynamics* (Pearson Education, Illinois, 2013).
- Thorne & Blandford, *Modern Classical Physics* (Princeton University Press, New Jersey, 2017).
- Cox & Giuli, *Principles of stellar structure* (Gordon and Breach, New York, 1968)
- Blandford & McKee, 1976, Physics of Fluids, 19, 1130

Chapter 4

Moving Charges

This chapter discusses (1) the EM fields produced by a moving charged particle with a given trajectory and (2) the motion of charged particles for given EM fields. These two topics are treated separately, for the following reason. The motion of a charged particle is mainly controlled by *external* EM fields that are produced by other particles. It is often the case that the *radiative cooling timescale* $t_{\text{cool}} = (\gamma - 1)mc^2/P$ (P being the emitting power given by the Larmor formula, see §4.1.2) over which the particle loses a large fraction of its kinetic energy is much longer than the *dynamical timescale* $t_{\text{dy}} = \gamma|\mathbf{u}|/|\mathbf{a}|$ (\mathbf{u} and \mathbf{a} being the spatial components of the 4-velocity and 4-acceleration, respectively) over which the particle undergoes significant acceleration. When $t_{\text{cool}} \gg t_{\text{dy}}$, it is safe to ignore the drag force due to radiative losses and only consider the effects of external EM fields. This is usually an excellent assumption for particles that are only modestly relativistic¹.

4.1 EM fields of moving charges

In this section, we discuss the EM fields of an isolated charged particle or an isolated system of charged particles in arbitrary motions.

4.1.1 Causality and retarded time

A fundamental principle of physics is *causality*: the effect can not precede the cause. For a moving charged particle in vacuum, the EM fields at the observer's spacetime position (t, \mathbf{r}) must only depend on the information of the particle (its position, velocity, and acceleration) at the *retarded time*

$$t_{\text{ret}} = t - c^{-1}|\mathbf{r} - \mathbf{x}(t_{\text{ret}})|, \quad (4.1)$$

¹A counter example of an electron with an extremely high Lorentz factor will be discussed in §8.3.4.

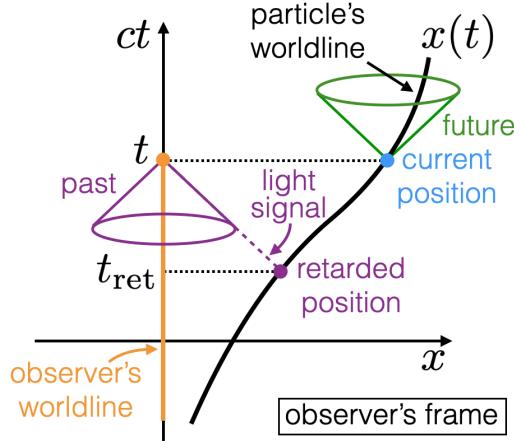


Figure 4.1: Illustration of the concepts of retarded time and retarded position in the observer’s frame (for an observer who stays at rest at $x = 0$). For each observer’s time t , a given emitting particle has a unique retarded position $\mathbf{x}(t_{\text{ret}})$ that is obtained by the intersection between the observer’s past lightcone and the particle’s worldline.

where $\mathbf{x}(t_{\text{ret}})$ is the position of the particle at the retarded time. This is because the information about the particle propagates at the speed of light in vacuum in between the particle and the observer, and it takes time $c^{-1}|\mathbf{r} - \mathbf{x}(t_{\text{ret}})|$ for the information to arrive at the observer. The relationship between current position $\mathbf{x}(t)$ and retarded position $\mathbf{x}(t_{\text{ret}})$ is shown in Fig. 4.1.

Differentiating eq. (4.1) with respect to the retarded time while keeping the observer’s position fixed, we obtain

$$\frac{dt}{dt_{\text{ret}}} = 1 - \hat{\mathbf{n}} \cdot \boldsymbol{\beta}, \quad (4.2)$$

where

$$\hat{\mathbf{n}} = \frac{\mathbf{r} - \mathbf{x}(t_{\text{ret}})}{|\mathbf{r} - \mathbf{x}(t_{\text{ret}})|}$$

is the unit vector pointing from the particle’s retarded position to the observer and

$$\boldsymbol{\beta}(t_{\text{ret}}) = d\mathbf{x}/dt_{\text{ret}}$$

is the velocity of the particle at the retarded time. Since $dt/dt_{\text{ret}} > 0$ for any physical speed $\beta < 1$, the function $t(t_{\text{ret}})$ (at a fixed observer’s position) is *monotonically increasing*, which means that there is a one-to-one map between t and t_{ret} , so there is only one retarded time for a given particle’s trajectory. The factor dt/dt_{ret} is due to “geometrical compression” of the information sent out from the particle — in the limit $\hat{\mathbf{n}} \parallel \boldsymbol{\beta}$ and $\beta \approx 1$ (ultra-relativistic), the emission within a time interval Δt_{ret} is received by the observer over a much shorter duration of $\Delta t \approx \Delta t_{\text{ret}}/(2\gamma^2)$.

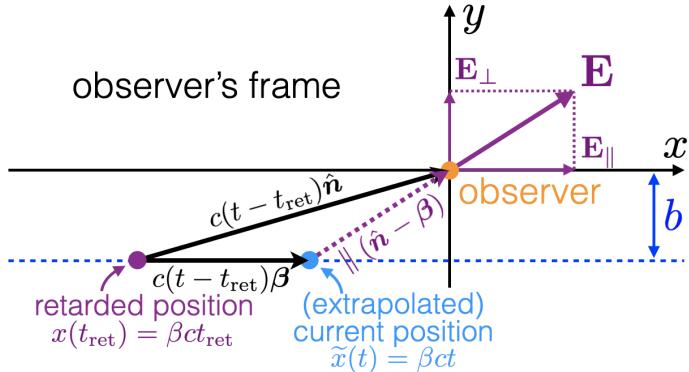


Figure 4.2: The Coulomb field of a moving particle is determined by the information of the charged particle at the retarded time.

The EM fields at the observer's spacetime position does not depend on the current position of the particle $\mathbf{x}(t)$. In fact, the observer has no way of knowing where the current position is. All she could do is to extrapolate the trajectory of the particle based on the information she had before it reached the retarded position $\mathbf{x}(t_{\text{ret}}(t, \mathbf{r}))$ — such an extrapolation can be precisely made if the particle is moving with a constant velocity but in case of a non-zero and time-dependent acceleration, it is generally difficult to make an accurate extrapolation.

Earlier in §3.6.3, we have shown that a charge q with trajectory $(x = \beta ct, y = -b, z = 0)$ produces a Coulomb field at the observer's location $(x = 0, y = 0, z = 0)$ as given by

$$\mathbf{E} = \frac{\gamma q}{(R')^3} (-\beta ct \hat{\mathbf{x}} + b \hat{\mathbf{y}}), \quad R' = \sqrt{\gamma \beta ct)^2 + b^2}, \quad (4.3)$$

where b is the impact parameter, $\gamma = (1 - \beta^2)^{-1/2}$ is the Lorentz factor of the particle, and the physical meaning of R' is the separation between the charge's *current* position $\mathbf{x}(t)$ and observer in the charge's *comoving frame*. Surprisingly, the Coulomb field is always in the direction that points from the particle's *current* position $\mathbf{x}(t)$ to the observer, as one can see from the fact that $E_x/E_y = -\beta ct/b$. This seems to be in contradiction with causality. In fact, this is just a coincidence, as one can cast the above result in another form that only depends on the information about the particle at the retarded time. Below, we show that the separation R' based on the current position of the charge is indeed physically irrelevant; what matters is the separation between the particle's *retarded* position and the observer. The relevant geometry of the system is shown in Fig. 4.2.

From the retarded position $\mathbf{x}(t_{\text{ret}}) = (x = \beta ct_{\text{ret}}, y = -b, z = 0)$, one can solve for the retarded time,

$$c(t - t_{\text{ret}}) = |\mathbf{x}(t_{\text{ret}})| = \sqrt{(\beta ct_{\text{ret}})^2 + b^2} \Rightarrow (t_{\text{ret}}/\gamma)^2 - 2t_{\text{ret}}t = (b/c)^2 - t^2. \quad (4.4)$$

We then add $(\gamma t)^2$ to both sides of the equation so as to complete the square,

$$\begin{aligned} \text{LHS} + (\gamma t)^2 &= (t_{\text{ret}}/\gamma)^2 - 2t_{\text{ret}}t + (\gamma t)^2 = (\gamma t - t_{\text{ret}}/\gamma)^2, \\ \text{RHS} + (\gamma t)^2 &= (b/c)^2 - t^2 + (\gamma t)^2 = (b/c)^2 + (\gamma \beta t)^2, \text{ using } \gamma^2 - 1 = (\gamma \beta)^2, \end{aligned} \quad (4.5)$$

so we obtain $R' = \sqrt{\gamma \beta ct)^2 + b^2} = c(\gamma t - t_{\text{ret}}/\gamma)$. Note that we have used the fact $\gamma t > t_{\text{ret}}/\gamma$ as constrained by causality.

Let us denote the separation between the particle's retarded position $\mathbf{x}(t_{\text{ret}})$ and the observer as

$$R = |\mathbf{R}| = c(t - t_{\text{ret}}) = \sqrt{(\beta ct_{\text{ret}})^2 + b^2},$$

and the unit vector from the retarded position $\mathbf{x}(t_{\text{ret}})$ to the observer $\mathbf{r} = \mathbf{0}$ as

$$\hat{\mathbf{n}} = \mathbf{R}/R = -\beta ct_{\text{ret}}\hat{\mathbf{x}} + b\hat{\mathbf{y}}.$$

Note that R and R' are not connected by Lorentz transformation (as they mean different things) but they are related by a Doppler factor $\mathcal{D} = [\gamma(1 - \hat{\mathbf{n}} \cdot \boldsymbol{\beta})]_{\text{ret}}$ as follows

$$\begin{aligned} R/\mathcal{D} &= \gamma(1 - \hat{\mathbf{n}} \cdot \boldsymbol{\beta})R = \gamma(R + \beta^2 ct_{\text{ret}}) = \gamma [c(t - t_{\text{ret}} + \beta^2 ct_{\text{ret}})] \\ &= \gamma c(t - t_{\text{ret}}/\gamma^2) = c(\gamma t - t_{\text{ret}}/\gamma) = R'. \end{aligned} \quad (4.6)$$

We see that the denominator in the Coulomb field expression (eq. 4.3) is given by $(R')^{3/2} = (R/\mathcal{D})^{3/2}$, which depends on the position (and velocity) of the particle at the retarded time. Similarly, the vectorial factor in the numerator can also be re-written as follows

$$\begin{aligned} (-\beta ct\hat{\mathbf{x}} + b\hat{\mathbf{y}}) &= -\beta ct_{\text{ret}}\hat{\mathbf{x}} - \beta c(t - t_{\text{ret}})\hat{\mathbf{x}} + b\hat{\mathbf{y}} \\ &= -\beta ct_{\text{ret}}\hat{\mathbf{x}} - \beta R\hat{\mathbf{x}} + b\hat{\mathbf{y}} = R(\hat{\mathbf{n}} - \boldsymbol{\beta}). \end{aligned} \quad (4.7)$$

Putting them together, we re-write the observed Coulomb field as

$$\mathbf{E}(t, \mathbf{r}) = \left[\frac{\gamma \mathcal{D}^3 q}{R^2} (\hat{\mathbf{n}} - \boldsymbol{\beta}) \right]_{\text{ret}}, \quad (4.8)$$

where all quantities on the RHS are evaluated at the retarded time. The vector pointing from the particle's *current* position $\mathbf{x}(t)$ and the observer $\mathbf{r} = \mathbf{0}$ is given by

$$\mathbf{r} - \mathbf{x}(t) = [\mathbf{r} - \mathbf{x}(t_{\text{ret}})] - [\mathbf{x}(t) - \mathbf{x}(t_{\text{ret}})] = c(t - t_{\text{ret}})\hat{\mathbf{n}} - c(t - t_{\text{ret}})\boldsymbol{\beta}, \quad (4.9)$$

which is indeed parallel to $(\hat{\mathbf{n}} - \boldsymbol{\beta})$ and hence to the Coulomb field \mathbf{E} .

We conclude that the Coulomb field is indeed physically determined by the information of the particle at the retarded time *only*, as there is no dependence on the properties before or after the retarded time. If the particle suddenly undergoes some arbitrary acceleration

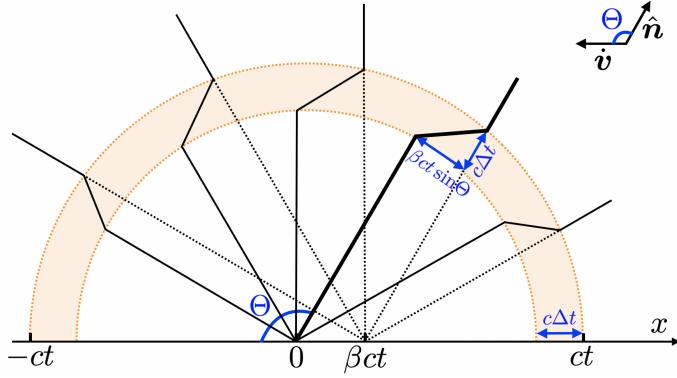


Figure 4.3: A charged particle with initial speed $\beta \ll 1$ along \hat{x} suddenly decelerated to $\beta = 0$ within a short time interval Δt right after it reached $x = 0$ at $t = 0$. The E-field at $t (\gg \Delta t)$ is shown by the black lines. The E-field line for a viewing angle Θ is highlighted by a thick black line. At time t , the E-field outside a sphere of radius ct centered at $x = 0$ would point to the extrapolated current position of $\tilde{x} = \beta ct$. Since the particle has been staying at $x \approx 0$ (as $\Delta t \ll 1$) since the deceleration, the E-field inside radius $c(t - \Delta t)$ is “informed” and hence points to the true position of $x \approx 0$. The field lines in between the two spheres of radii ct and $c(t - \Delta t)$ are connected (as required by Gauss’s law in vacuum) in a way that they are almost perpendicular to the radial direction — they correspond to the radiation field E_{rad} .

after the retarded time, the observed electric field is still given by eq. (4.8) and it is along the direction pointing from the *extrapolated current position*

$$\tilde{\mathbf{x}}(t) = \mathbf{x}(t_{\text{ret}}) + c(t - t_{\text{ret}})\boldsymbol{\beta}$$

to the observer, i.e.,

$$\mathbf{E} \parallel (\hat{\mathbf{n}} - \boldsymbol{\beta}) \parallel (\mathbf{r} - \tilde{\mathbf{x}}(t)).$$

For the case of a uniform motion, the extrapolated current position $\tilde{\mathbf{x}}(t)$ is identical to the actual current position $\mathbf{x}(t)$, so $\mathbf{E} \parallel (\mathbf{r} - \mathbf{x}(t))$ is simply a coincident. For an accelerated trajectory, these two positions $\mathbf{x}(t)$ and $\tilde{\mathbf{x}}(t)$ are generally not the same, but the Coulomb field is always parallel to $\mathbf{r} - \tilde{\mathbf{x}}(t)$.

However, an accelerated charge produces not only the Coulomb field but also another field component that is related to the non-zero acceleration at the retarded time. This will be discussed in the next section.

4.1.2 Larmor formula

The Coulomb field of a particle undergoing uniform motion does not carry energy away to infinity. This can be seen from the fact that the field strength decreases rapidly with

distance to the charge $E \propto 1/R^2$, and hence the energy density scales as $E^2 \propto R^{-4}$. In this section, we show that a particle that undergoes acceleration generates EM waves that can carry energy to infinity — an accelerating particle radiates its kinetic energy away. We will first consider the radiation from a non-relativistic particle and then generalize our results to the arbitrarily relativistic case (as one can always go to the momentary rest frame of a particle where it is non-relativistic).

Consider a particle moving along the x -axis with a constant velocity $\beta \ll 1$ suddenly comes to a stop during a very short time interval from $t = 0$ to Δt . Suppose the deceleration process is linear with a constant $\dot{\beta} = \ddot{x} = -\beta/\Delta t$, the trajectory of the particle is

$$x(t) = \begin{cases} \beta ct, & \text{for } t < 0, \\ (\beta t/2)(2 - t/\Delta t) \approx 0, & \text{for } 0 < t < \Delta t, \\ \beta \Delta t/2 \approx 0, & \text{for } t > \Delta t. \end{cases} \quad (4.10)$$

For observers who are studying the EM fields of the system at a much later time $t \gg \Delta t$, we can ignore higher order terms $\mathcal{O}(\Delta t^2)$ and take $x(t > 0) \approx 0$. The electric field at time $t \gg \Delta t$ is shown in Fig. 4.3.

At a given time t , let us consider two observers located at different positions: the first one is at a radius $r_1 > ct$ and the second one located at radius $r_2 < c(t - \Delta t)$. The observer at the larger radius does not know that the particle has come to a stop, as it takes a light travel time $r_1/c > t$ for the information to propagate to her. Thus, the first observer measures a Coulomb field in the direction pointing from the particle's extrapolated current position $\tilde{x} = \beta ct$ to the observer. For the second observer at the smaller radius is informed that the particle has come to a stop, so she measures an electric field along the direction pointing from $x \approx 0$ to the observer. Both observers only detect the Coulomb field of a uniformly moving particle, as given by eq. (4.8).

The electric field lines between two spheres of radii $c(t - \Delta t)$ and ct continuously connect the interior to the exterior. This region is called the “radiation zone” where the electric field has two components: a radial component, and a transverse component that is perpendicular to the line of sight \hat{n} (which is nearly aligned with the radial vector \hat{r} for a non-relativistic system). The radial component is identified as the Coulomb field that only depends on the velocity of the particle at the retarded time for a given observer’s position — this component is commonly called the *velocity field* and denoted as \mathbf{E}_{vel} . The transverse component is caused by the non-zero acceleration of the particle and is hence called the *acceleration field*. We will see later that the acceleration field carries energy to infinity and comes from the radiative energy loss from the particle, so this field component is also called the *radiation field* and denoted as \mathbf{E}_{rad} .

Based on the geometry in Fig. 4.3, we find the ratio between the magnitudes of these two

components to be

$$\frac{E_{\text{rad}}}{E_{\text{vel}}} = \frac{\beta ct \sin \Theta}{c\Delta t}, \quad (4.11)$$

where Θ is the angle between the line of sight and the acceleration vector. Note that the acceleration vector in this problem is given by

$$\dot{\mathbf{v}} = -(\beta c/\Delta t)\hat{\mathbf{x}},$$

where the negative sign comes from the deceleration. The velocity field near radius $R \approx ct$ is equal to the Coulomb field in the non-relativistic limit

$$E_{\text{vel}} = q/R^2 = q/(ct)^2, \quad (4.12)$$

so we obtain

$$E_{\text{rad}} = \frac{q}{Rc^2} |\dot{\mathbf{v}}| \sin \Theta, \quad (4.13)$$

where we have used the magnitude of the acceleration $|\dot{\mathbf{v}}| = \beta c/\Delta t$. The direction of the radiative E-field can be obtained based on the fact that \mathbf{E}_{rad} is inside the plane containing $\dot{\mathbf{v}}$ and $\hat{\mathbf{n}}$ but perpendicular to $\hat{\mathbf{n}}$. The vectorial form of the radiative E-field at the observer's spacetime position (t, \mathbf{r}) is given by

$$\mathbf{E}_{\text{rad}}(t, \mathbf{r}) = \frac{q}{Rc^2} \hat{\mathbf{n}} \times (\hat{\mathbf{n}} \times \dot{\mathbf{v}}(t_{\text{ret}})) = -\frac{q}{Rc^2} \dot{\mathbf{v}}_{\perp}(t_{\text{ret}}), \quad (4.14)$$

where $\dot{\mathbf{v}}_{\perp} \equiv \dot{\mathbf{v}} - (\hat{\mathbf{n}} \cdot \dot{\mathbf{v}})\hat{\mathbf{n}}$ is the component of the acceleration $\dot{\mathbf{v}}$ perpendicular² to the line of sight $\hat{\mathbf{n}}$ and the retarded time t_{ret} is given by

$$t_{\text{ret}} = t - R/c. \quad (4.15)$$

The corresponding radiative B-field must be given by $\mathbf{B}_{\text{rad}} = \hat{\mathbf{n}} \times \mathbf{E}_{\text{rad}}$ so as to make transverse EM waves in vacuum.

The Poynting vector points outwards along the line of sight $\hat{\mathbf{n}}$ ($\parallel \hat{\mathbf{r}}$) and is given by

$$\mathbf{S} = \frac{E_{\text{rad}}^2 c}{4\pi} \hat{\mathbf{n}} \propto R^{-2}, \quad (4.16)$$

which obeys the inverse-square law so we know that the energy of the radiation fields will be carried away to infinity. Then, the emitted power per solid angle along a given viewing angle is given by

$$\frac{dP}{d\Omega} = S(\Theta)R^2 = \frac{q^2 \dot{v}^2}{4\pi c^3} \sin^2 \Theta, \quad (4.17)$$

which has the characteristic *dipole pattern* $dP/d\Omega \propto \sin^2 \Theta$ as shown in Fig. 4.4. [this video](#) provides a nice 3D demonstration of the dipole pattern. Another nice Jupyter-Notebook

²This can be obtained from the vector identity $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}$.

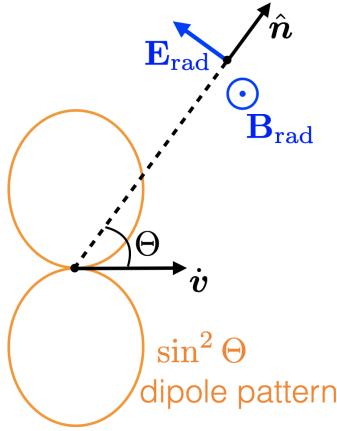


Figure 4.4: The radiation fields of a non-relativistic particle undergoing acceleration $\dot{\mathbf{v}}$. The unit vector points from the particle's position to the observer. The observed E-field is inside the $\hat{\mathbf{n}} \cdot \dot{\mathbf{v}}$ plane, and the B-field is along $\hat{\mathbf{n}} \times \dot{\mathbf{v}}$. The two orange lobes shows the angular distribution of the emitting power $dP/d\Omega \propto \sin^2 \Theta$, where Θ is the observer's viewing angle = the angle between $\hat{\mathbf{n}}$ and $\dot{\mathbf{v}}$.

demonstration of the radiation fields is [here](#) (provided by Prof. A. Parsons). The *total emitting power* of this particle is given by the angular integral

$$P = \int \frac{dP}{d\Omega} d\Omega = \frac{2q^2 \dot{v}^2}{3c^3}, \quad (4.18)$$

which is the *Larmor formula* — arguably the most important formula of this course.

In the following, we generalize the Larmor formula to the relativistic case by noticing that eq. (4.18) holds in the momentary rest frame of an accelerated charge. In the momentary rest frame \mathcal{O}' , the particle's 4-velocity is given by $\vec{u}'/c = (1, \mathbf{0})$, and the 4-acceleration is $\vec{a}' = d\vec{u}'/d\tau = (0, \dot{\mathbf{v}}')$. Therefore, we obtain $(\dot{\mathbf{v}}')^2 = \vec{a} \cdot \vec{a} = \eta_{\mu\nu} a^\mu a^\nu = a_\mu a^\mu$, where $(\eta_{\mu\nu})$ is the Minkowski metric. Since $a_\mu a^\mu$ is Lorentz invariant, we can write the total emitting power as

$$P = \frac{2q^2}{3c^3} a_\mu a^\mu, \quad (4.19)$$

which is called the *generalized Larmor formula* and can be evaluated in any Lorentz frame. The above expression shows that the total emitting power is Lorentz invariant. The Lorentz invariance of an emitting particle can also be demonstrated using Lorentz transformation (see §3.7). For instance, for an arbitrary particle trajectory $\mathbf{x}(t)$ measured in the lab frame, we can calculate the 4-velocity $\vec{u} = (\gamma c, d\mathbf{x}/d\tau)$ (note: $d\tau = dt/\gamma$) and then 4-acceleration $\vec{a} = d\vec{u}/d\tau$, and hence the total emitting power at lab-frame time t is given by $P(t) = (2q^2/3c^3) [-(a^0)^2 + \sum_i (a^i)^2]$ for $i = 1, 2, 3$.

It is possible to Lorentz transform the radiation field (eq. 4.14) from the particle's momentary rest frame to the lab frame, but it is a tedious task. We will discuss the radiation field in an arbitrary Lorentz frame based on the Liénard-Wiechart potential later.

Finally, we mention an important caveat that the Larmor formula is only applicable for incoherent emission — the EM waves produced by different particles have no phase coherence such that the total power is given by the incoherent sum from each individual particles. In most systems where particles' (e.g., thermal) motions are randomized, this is an excellent approximation. When a large number ($N \gg 1$) of particles emit coherently like electrons in a short antenna with oscillating currents, we would need to add up the amplitude of radiated electric and magnetic fields from individual particles. In this case, since the total field amplitude becomes N times the field produced by individual particles, we obtain a total power that is N^2 times that from an isolated particle! In this case, the radiative back-reaction forces may be strong, which would lead to a relatively large resistance in an otherwise nearly perfectly conducting antenna.

4.1.3 Retarded 4-potential

Let us consider a system with charge density $\rho(t, \mathbf{x})$ and current $\mathbf{J}(t, \mathbf{x})$ as given functions of spacetime coordinates. As we have seen in Ch. 3, these two quantities form a 4-vector that obey the Lorentz transformation — the 4-current density $\vec{J}(t, \mathbf{x}) = (\rho c, \mathbf{J})$.

The divergence-free property of the B-field ($\nabla \cdot \mathbf{B} = 0$) allows us to define the 3-vector potential \mathbf{A} such that

$$\mathbf{B} \equiv \nabla \times \mathbf{A}. \quad (4.20)$$

Then, from $\nabla \times \mathbf{E} = -c^{-1}\partial_t \mathbf{B}$, we obtain

$$\nabla \times (\mathbf{E} + c^{-1}\partial_t \mathbf{A}) = 0, \quad (4.21)$$

which means that $\mathbf{E} + c^{-1}\partial_t \mathbf{A}$ can be expressed as the (negative) gradient of the scalar potential ϕ , i.e.,

$$\mathbf{E} \equiv -\nabla\phi - c^{-1}\partial_t \mathbf{A}. \quad (4.22)$$

The first term above is the potential component (think about a capacitor) and the second term is the inductive component (think about an inductor).

The potentials ϕ and \mathbf{A} are not uniquely specified by the conditions (eqs. 4.20 and 4.22) imposed above. The same (\mathbf{E}, \mathbf{B}) could correspond to many sets of (ϕ, \mathbf{A}) — this is called *Gauge freedom*. In fact, for any scalar function $\Lambda(t, \mathbf{x})$, the following two sets of potentials

$$(\phi, \mathbf{A}) \text{ and } (\phi - c^{-1}\partial_t \Lambda, \mathbf{A} + \nabla \Lambda), \quad (4.23)$$

give the same (\mathbf{E}, \mathbf{B}) , as can be easily verified. There is in fact a functional degree of freedom — the scalar function $\Lambda(t, \mathbf{x})$ can be arbitrarily chosen for *our purpose* so as to

simplified the problem at hand. A common choice is the *Lorenz gauge* to be used later in this section

$$\nabla \cdot \mathbf{A} + c^{-1} \partial_t \phi = 0, \quad (4.24)$$

which constrains the scalar function by $\nabla^2 \Lambda - c^{-2} \partial_t^2 \Lambda = 0$ but still does not fully specify it — nor do we need to, as it is the EM fields (not the potentials) that are measurable and hence concern us anyway.

Another popular choice is the *Coulomb gauge*

$$\nabla \cdot \mathbf{A} = 0, \quad (4.25)$$

which makes it easier to solve for the scalar potential ϕ from a simple Poisson equation (see eq. 4.26) but the vector potential \mathbf{A} is much more difficult to solve under this gauge choice. In the special situation that the sources are time-independent or in the absence of charge sources, we have $\partial_t \phi = 0$, so the Lorenz gauge turns into the Coulomb gauge.

Let us then consider how the potentials are determined by the sources which are specified by the 4-current density $\vec{J} = (\rho c, \mathbf{J})$. The $\nabla \cdot \mathbf{E} = 4\pi\rho$ equation gives

$$\nabla^2 \phi + c^{-1} \partial_t (\nabla \cdot \mathbf{A}) = -4\pi\rho. \quad (4.26)$$

The $\nabla \times \mathbf{B} = 4\pi c^{-1} \mathbf{J} + c^{-1} \partial_t \mathbf{E}$ equation gives

$$\nabla^2 \mathbf{A} - c^{-2} \partial_t^2 \mathbf{A} - \nabla (\nabla \cdot \mathbf{A} + c^{-1} \partial_t \phi) = -4\pi \mathbf{J}/c. \quad (4.27)$$

Under the Lorenz gauge (eq. 4.24), the above two equations can be simplified into the following form of a set of four wave equations

$$(-c^2 \partial_t^2 + \nabla^2) \phi = -4\pi\rho, \quad (-c^2 \partial_t^2 + \nabla^2) \mathbf{A} = -4\pi \mathbf{J}/c, \quad (4.28)$$

These four wave equations are equivalent to the set of four Maxwell equations.

The above form of wave equations, together with the 4-current $\vec{J} = (\rho c, \mathbf{J})$, motivates us to define the electromagnetic 4-potential

$$\vec{A} = (\phi, \mathbf{A}), \quad (4.29)$$

and then write

$$(-c^2 \partial_t^2 + \nabla^2) \vec{A} = \eta^{\mu\nu} \partial_\mu \partial_\nu \vec{A} = -4\pi \vec{J}/c, \quad (4.30)$$

where we have used the compact notation for the *wave-equation operator* $\square \equiv \eta^{\mu\nu} \partial_\mu \partial_\nu = \partial^\nu \partial_\nu = -c^{-2} \partial_t^2 + \nabla^2$ together with the inverse-matrix of the Minkowski metric ($\eta^{\mu\nu}$). Note that the “sign-flip-if-temporal” rule of the Minkowski spacetime, together with Einstein summation, gives $\partial^\nu \partial_\nu = -\partial_0^2 + \sum_{i=1,2,3} \partial_i^2$, where $\partial_\nu = \partial/(\partial x^\nu)$ means partial derivative along coordinate x^ν . We see that the wave-equation operator \square is Lorentz invariant, as

it has undergone two index contractions (for both μ and ν). An important conclusion is that, under the Lorenz gauge, $\vec{A} = (\phi, \mathbf{A})$ is a 4-vector as it scales linearly with the 4-current \vec{J} . We also note that the Lorenz gauge condition (eq. 4.24) can be written in a frame-independent form of $\partial_\mu A^\mu = 0$, which means that the 4-potential is divergence free under the Lorenz gauge.

The four wave equations (4.30) can be formally solved using Fourier transform and Green's function methods. We skip the tedious details here (see Jackson's book). The result is the *retarded 4-potential* at the observer's spacetime position (t, \mathbf{r})

$$\vec{A}(t, \mathbf{r}) = c^{-1} \int \frac{\vec{J}(t_{\text{ret}}(\mathbf{x}), \mathbf{x}) d^3x}{|\mathbf{r} - \mathbf{x}|}, \quad (4.31)$$

where the 4-current density \vec{J} is evaluated at the *retarded time*

$$t_{\text{ret}}(\mathbf{x}) = t - c^{-1} |\mathbf{r} - \mathbf{x}|, \quad (4.32)$$

for any source position \mathbf{x} that we will integrate over. The meaning of the above solution is that the potential $\vec{A}(t, \mathbf{r})$ is given by the charge and current density distribution at different positions \mathbf{x} at the local retarded time t_{ret} , which depends on \mathbf{x} for given observer's time t and position \mathbf{r} . To carry out the integral, the procedure would be: for each position \mathbf{x} , one first calculates the retarded time t_{ret} (from eq. 4.32), then evaluate the 4-current density $\vec{J}(t_{\text{ret}}, \mathbf{x})$, then add the contribution from the volume d^3x to the potential, and finally sum over the entire volume. As already mentioned in §4.1.1, the concept of retarded time is the consequence of *causality*: the information at position \mathbf{x} can only affect the fields at the observer's position \mathbf{r} after a light-travel time of $c^{-1} |\mathbf{r} - \mathbf{x}|$. Here, the retarded 4-potential solution to the wave equation shows that the EM fields of any current system indeed preserve causality.

We note that, in the limit of a static charge distribution, eq. (4.31) corresponds to the Coulomb potential $\phi(\mathbf{r}) = \int \rho d^3x / |\mathbf{r} - \mathbf{x}|$, so it is appropriate to call the retarded 4-potential $\vec{A}(t, \mathbf{r})$ the *generalized Coulomb potential*.

4.1.4 Liénard-Wiechart 4-potential for a point charge

In this section, we consider the EM fields of a relativistic point charge moving along a known trajectory $\mathbf{x}(t)$ with arbitrary acceleration. For given observer's time t and position \mathbf{r} , we can solve the retarded time t_{ret} from eq. (4.32), and we know that the EM fields at the observer's spacetime position only depend on the position $\mathbf{x}(t_{\text{ret}})$, velocity $\mathbf{v} = \dot{\mathbf{x}} = d\mathbf{x}/dt_{\text{ret}}$, and acceleration $\ddot{\mathbf{v}} = \ddot{\mathbf{x}} = d^2\mathbf{x}/dt_{\text{ret}}^2$ of the particle at the retarded time. Since we are ignoring radiation back-reaction force (see 4.2.9), the changing rate of the acceleration $\ddot{\mathbf{v}} = \ddot{\mathbf{x}}$ does not play a role in our results.

Surprisingly, not matter how small the charged particle is, it is important to carefully account for the fact that the retarded time for different parts of the charge distribution is slightly different! This is demonstrated by the following example of a uniformly charged moving rod (see Fig. 4.5) — an arbitrarily shaped object can be decomposed into a collection of infinitesimal rods.

First, let us define things in the comoving frame of the rod: the charge density is ρ' , the proper length is L_0 , and the cross-sectional area is S . The total charge is

$$q = \rho' L_0 S. \quad (4.33)$$

In the comoving frame, the 4-current density is $(\rho' c, \mathbf{0})$ where there is matter. In the lab frame, the rod is moving with velocity βc and Lorentz factor γ along its longitudinal direction. We Lorentz transform the 4-current density into the lab frame and obtain

$$\vec{J} = \rho' \vec{u} = (\gamma \rho' c, \gamma \beta \rho' c, 0, 0), \quad (4.34)$$

where \vec{u} is the 4-velocity. Thus, in the lab frame, the charge density is $\rho = \gamma \rho'$ and current density is $\mathbf{J} = \gamma \beta \rho' c \hat{x}$ where there is matter (and $\rho = J = 0$ where there is no matter). The longitudinal length is L_0/γ due to length contraction, so the total charge is $\rho(L_0/\gamma)S = q$, confirming that total charge is Lorentz invariant.

For an distant observer located at a viewing angle θ wrt. the velocity vector, the retarded time for different parts of the rod is different: one can geometrically show that the retarded positions back and front ends of the rod are separated by $\mathcal{D}L_0$ along the x -axis (see Fig. 4.5), where the Doppler factor is given by

$$\mathcal{D} = \frac{1}{\gamma(1 - \beta \cos \theta)}. \quad (4.35)$$

Thus, one can say that, at fixed observer's time and position, the *retarded length* of the rod is $\mathcal{D}L_0$. This factor of \mathcal{D} cannot be ignored no matter how small the rod physically is. Thus, when carrying out the volume integral in eq. (4.31), one is forced to conclude that the charge density at the local retarded time $[\rho]_{\text{ret}}$ spans the whole apparent length of $\mathcal{D}L_0$ and equals to $\rho = \gamma \rho'$. Similarly, the current density at the local retarded time is given by $[\mathbf{J}]_{\text{ret}} = \gamma \beta \rho' c \hat{x}$ throughout the retarded length of $\mathcal{D}L_0$.

If the distance to the observer, R , is much larger than the size of the emitting rod, the denominator gives $|\mathbf{r} - \mathbf{x}| \approx R$, and then we obtain the 4-potential at the observer's spacetime position (t, \mathbf{r}) in the limit of a point charge

$$\vec{A}(t, \mathbf{r}) = \frac{\int \vec{J}(t_{\text{ret}}, \mathbf{x}) d^3x}{R(t_{\text{ret}})c} = \left[\frac{\rho' \mathcal{D}L_0 S}{Rc} \vec{u} \right]_{\text{ret}} = \left[\frac{\mathcal{D}q}{Rc} \vec{u} \right]_{\text{ret}}, \quad (4.36)$$

where we have made use of the (Lorentz invariant) total charge $q = \rho' L_0 S$. The above is called the *Liénard-Wiechart 4-potential* for a point charge.

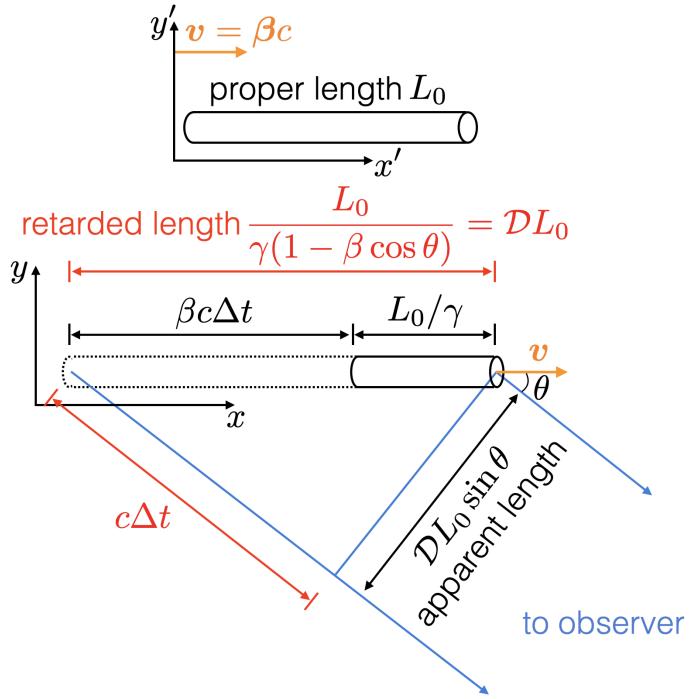


Figure 4.5: A moving rigid rod with *proper length* L_0 (as measured simultaneously in its comoving or $x'y'$ frame) is length-contracted to L_0/γ as measured simultaneously in the lab or xy frame. From the viewpoint of a distant observer, the photons from the back end of the rod takes a path that is longer by $c\Delta t$ than photons from the front end, and during this time, the rod has moved a distance $\beta c\Delta t$. Denoting the angle between the observer's line of sight and the velocity vector as θ , one can show that $c\Delta t = \mathcal{D}L_0 \cos \theta$, where $\mathcal{D} = [\gamma(1 - \beta \cos \theta)]^{-1}$ is the Doppler factor. Therefore, the *retarded length* of the rod is $\beta c\Delta t + L_0/\gamma = \mathcal{D}L_0$. If the observer takes a picture of the rod with a high-speed camera (with an infinitely short exposure time), it shows an *apparent length* of $\mathcal{D}L_0 \sin \theta$ in the image (since this is the size transverse to the line of sight). But if the geometric projection is corrected for, I conclude that, at a given observer's time, the back and front ends of the rod are separated by $\mathcal{D}L_0$ along the x -axis. (Sometimes, without the knowledge of the viewing angle, one would naively conclude that the rod has length $\mathcal{D}L_0 \sin \theta$ based on the image in hand)

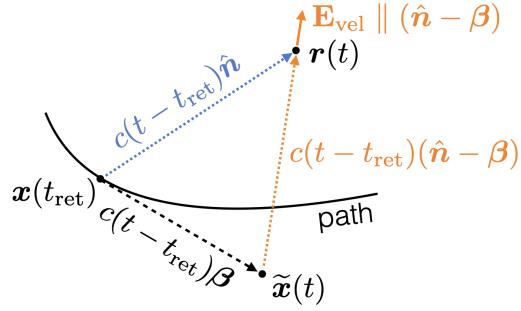


Figure 4.6: The velocity field $\mathbf{E}_{\text{vel}}(t, \mathbf{r})$ is parallel to the vector pointing from the particle's extrapolated current position $\tilde{\mathbf{x}}(t)$ to the observer \mathbf{r} .

In the process of obtaining the above results, we did not assume the velocity of the rod to be constant. The comoving ($x'y'$) frame we chose can be considered the momentary rest frame of the rod, in which the length of the rod in the longitudinal direction is measured to be L_0 . Therefore, the above results apply as long as \vec{u} is evaluated at the retarded time (the moment when the radiation is produced), θ is defined as the angle between the observer's line of sight and the velocity vector at the retarded time, and R is defined as $|\mathbf{r} - \mathbf{x}(t_{\text{ret}})|$, where $\mathbf{x}(t_{\text{ret}})$ is the source position at the retarded time.

The four components of the Liénard-Wiechart are then

$$\phi(t, \mathbf{r}) = \left[\frac{q}{(1 - \hat{\mathbf{n}} \cdot \beta) |\mathbf{r} - \mathbf{x}|} \right]_{\text{ret}}, \quad (4.37)$$

$$\mathbf{A}(t, \mathbf{r}) = \left[\frac{q\beta}{(1 - \hat{\mathbf{n}} \cdot \beta) |\mathbf{r} - \mathbf{x}|} \right]_{\text{ret}}, \quad (4.38)$$

which are evaluated at the retarded time t_{ret} .

It is a lengthy task to differentiate the above potentials to obtain the EM fields (see Jackson's book), because t_{ret} depends on the observer's time and position in a complicated way (involving the particle's trajectory) as specified by eq. (4.1). The result is

$$\begin{aligned} \mathbf{E}(t, \mathbf{r}) &= q \left[\frac{\gamma \mathcal{D}^3}{R^2} (\hat{\mathbf{n}} - \beta) \right]_{\text{ret}} + \frac{q}{c} \left[\frac{(\gamma \mathcal{D})^3}{R} \hat{\mathbf{n}} \times ((\hat{\mathbf{n}} - \beta) \times \dot{\beta}) \right]_{\text{ret}}, \\ \mathbf{B}(t, \mathbf{r}) &= [\hat{\mathbf{n}}]_{\text{ret}} \times \mathbf{E}(t, \mathbf{r}), \end{aligned} \quad (4.39)$$

where

$$\mathcal{D} \equiv \frac{1}{\gamma(1 - \hat{\mathbf{n}} \cdot \beta)}, \quad \hat{\mathbf{n}} = \frac{\mathbf{R}}{R}, \quad \mathbf{R} = \mathbf{r} - \mathbf{x}(t_{\text{ret}}). \quad (4.40)$$

Note that $\dot{\beta} = d\beta/dt_{\text{ret}} = \ddot{\mathbf{x}} = d^2\mathbf{x}/dt_{\text{ret}}^2$ is the acceleration that is evaluated at the retarded time. The strong dependence on the Doppler factor means that the EM fields

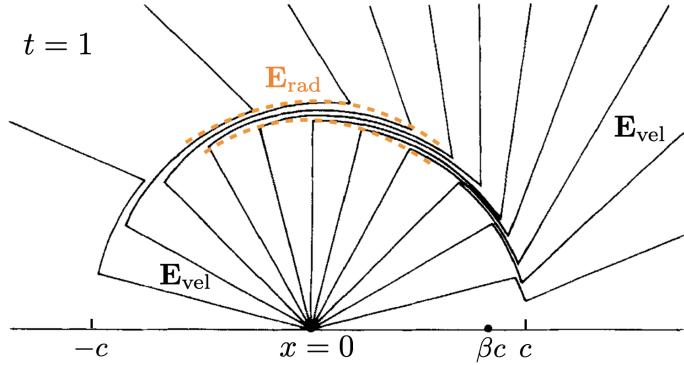


Figure 4.7: The E-field from a particle at time $t = 1$. The particle originally moving at a mildly relativistic uniform velocity β along the x -axis is suddenly stopped at $t = 0$ when it reaches $x = 0$. The deceleration occurs in a short time interval $\Delta t \ll 1$.

from an ultra-relativistic particle are strongest when $\hat{\mathbf{n}}$ is nearly parallel to β to within an angle of the order γ^{-1} .

The first term in the total E-field is proportional to R^{-2} is called the *velocity field*, which is in agreement with the Coulomb field of a uniformly moving charge (eq. 4.8). Note that the velocity field \mathbf{E}_{vel} is in the direction of $\hat{\mathbf{n}} - \beta$. This can be geometrically understood using Fig. 4.6. If we *assume* that the particle moves with a constant velocity $\beta(t_{\text{ret}})$, then by time t the particle would reach the *extrapolated current position* $\tilde{\mathbf{x}}(t) = \mathbf{x}(t_{\text{ret}}) + c(t-t_{\text{ret}})\beta$. The observer position can be written as $\mathbf{r} = \mathbf{x}(t_{\text{ret}}) + c(t-t_{\text{ret}})\hat{\mathbf{n}}$. Therefore, the vector from the particle's extrapolated current position to the observer is $\mathbf{r} - \tilde{\mathbf{x}}(t) = c(t-t_{\text{ret}})(\hat{\mathbf{n}} - \beta)$, which is parallel to the velocity E-field \mathbf{E}_{vel} .

The second term in the total E-field is linearly proportional to the acceleration $\dot{\beta}$, and due to its much shallower scaling with distance R^{-1} , this term carries energy away to large distances — it is called the *radiation field* or *acceleration field* \mathbf{E}_{rad} . Since \mathbf{E}_{rad} is perpendicular to $[\hat{\mathbf{n}}]_{\text{ret}}$, we know that the corresponding magnetic field $\mathbf{B}_{\text{rad}} = [\hat{\mathbf{n}}]_{\text{ret}} \times \mathbf{E}_{\text{rad}}$ has the same magnitude as \mathbf{E}_{rad} . The three vectors \mathbf{E}_{rad} , \mathbf{B}_{rad} and $[\hat{\mathbf{n}}]_{\text{ret}}$ are mutually perpendicular to each other, and they form transverse vacuum EM waves propagating along $[\hat{\mathbf{n}}]_{\text{ret}}$, as can be seen from the Poynting flux $\mathbf{S}_{\text{rad}} = \mathbf{E}_{\text{rad}} \times \mathbf{B}_{\text{rad}} \propto [\hat{\mathbf{n}}]_{\text{ret}}$.

A nice demonstration of the velocity field \mathbf{E}_{vel} and radiation field \mathbf{E}_{rad} for a mildly relativistic particle that has undergone rapid deceleration is provided by Fig. 3.2 in Rybicki & Lightman's book, which is adapted in Fig. 4.7.

4.2 Motion of charged particles in EM fields

In this section, we ignore radiative losses and consider particle motion under external EM fields as described by known functions $\mathbf{E}(\mathbf{x}, t)$ and $\mathbf{B}(\mathbf{x}, t)$.

4.2.1 Equation of motion

Let us consider the equation of motion of a relativistic particle of charge q and rest-mass m . In the lab frame \mathcal{O} , the particle has velocity $\mathbf{v} = \beta c \hat{\mathbf{x}}$ at time t and the EM fields near the particle's location are given by \mathbf{E} and \mathbf{B} . In the particle's comoving frame, the components of the electric field are given by (cf. eq. 3.114)

$$E'_x = E_x, \quad \mathbf{E}'_{\perp} = \gamma(\mathbf{E}_{\perp} + \boldsymbol{\beta} \times \mathbf{B}), \quad (4.41)$$

where $\mathbf{E}_{\perp} = \mathbf{E} - (\mathbf{E} \cdot \hat{\mathbf{x}})\hat{\mathbf{x}}$ and $\mathbf{E}'_{\perp} = \mathbf{E}' - (\mathbf{E}' \cdot \hat{\mathbf{x}})\hat{\mathbf{x}}$ contain the y and z components that are perpendicular to the particle's motion. For the purpose of the equation of motion in the particle's comoving frame, we do not care about the magnetic field \mathbf{B}' as its contribution to the Lorentz force vanishes for a particle at rest $\mathbf{v}' = 0$.

The components of the particle's acceleration in the comoving frame parallel and perpendicular to the x -axis are given by

$$a'_x = \frac{dv'_x}{d\tau} = \frac{qE'_x}{m}, \quad \text{and} \quad \mathbf{a}'_{\perp} = \frac{d\mathbf{v}'_{\perp}}{d\tau} = \frac{q\mathbf{E}'_{\perp}}{m}. \quad (4.42)$$

After a proper time interval $d\tau$ (which corresponds to a lab-frame time interval $dt = \gamma d\tau$), the particle's 4-velocity in the comoving frame becomes

$$d\vec{u}' \approx (c, dv'_x, dv'_y, dv'_z), \quad (4.43)$$

where $dv'_x = (qE'_x/m)d\tau$, the perpendicular components are given by $d\mathbf{v}'_{\perp} = (q\mathbf{E}'_{\perp}/m)d\tau$, and we have ignored higher order terms $\mathcal{O}(dv'^2)$. We then (inverse) Lorentz transform $d\vec{u}'$ into the lab-frame and obtain the updated 4-velocity at time $t + dt$

$$\vec{u} + d\vec{u} = (\gamma c + \gamma \beta dv'_x, \gamma \beta c + \gamma dv'_x, dv'_y, dv'_z), \quad (4.44)$$

where $\vec{u} = (\gamma c, \gamma \beta c, 0, 0)$ is the original 4-velocity at time t . Thus, we obtain the change in 4-velocity in the lab-frame within time interval dt

$$d\vec{u} = (\gamma \beta dv'_x, \gamma dv'_x, dv'_y, dv'_z). \quad (4.45)$$

The temporal component of $d\vec{u}$ is

$$du^0 = d\gamma c = \gamma \beta dv'_x = \gamma \beta \frac{qE'_x}{m} d\tau = \frac{qE_x}{m} \beta = \frac{q}{mc} \mathbf{E} \cdot \mathbf{v}, \quad (4.46)$$

which means that the particle gains energy at a rate

$$\frac{d\gamma}{dt} mc^2 = q \mathbf{E} \cdot \mathbf{v}, \quad (4.47)$$

which is expected as the Lorentz force $\propto \mathbf{v} \times \mathbf{B}$ does no work. The spatial components of $d\vec{u}$ are given by

$$du_x = \gamma dv'_x = \gamma \frac{qE'_x}{m} d\tau = \frac{qE_x}{m} dt, \quad du_{\perp} = dv'_{\perp} = \frac{qE'_{\perp}}{m} d\tau = \frac{q(\mathbf{E}_{\perp} + \boldsymbol{\beta} \times \mathbf{B})}{m} dt. \quad (4.48)$$

Using the 4-momentum $\vec{p} = m\vec{u} = (\gamma mc, \mathbf{p})$, we then write the *equation of motion* in terms of the rate of changing for the spatial components of the 4-momentum

$$\frac{dp}{dt} = q(\mathbf{E} + \boldsymbol{\beta} \times \mathbf{B}), \quad (4.49)$$

where the RHS includes the electric force $q\mathbf{E}$ and the Lorentz force $\boldsymbol{\beta} \times \mathbf{B}$, and dt in the denominator of the LHS is the coordinate time in frame \mathcal{O} (not proper time). It is also possible to derive the energy equation (4.47) by dot-producing both sides of the equation of motion (4.49) by \mathbf{v} and noticing $d(\gamma\mathbf{v}) \cdot \mathbf{v} = v^2 d\gamma + (1/2)\gamma dv^2 = c^2 d\gamma$ (using $\beta^2 = 1 - \gamma^{-2}$), so we know that the energy equation is in fact redundant.

4.2.2 Uniform B (helical orbit, cyclotron frequency, Larmor radius)

We consider a charged particle's motion in a uniform, time-independent magnetic field $\mathbf{B} = B\hat{z}$. The equation of motion is given by

$$\frac{dp}{dt} = q\boldsymbol{\beta} \times \mathbf{B}. \quad (4.50)$$

Since there is no E-field, the particle's Lorentz factor $\gamma = (1 - \beta^2)^{-1/2}$ stays constant. Since $\boldsymbol{\beta} \times \mathbf{B}$ is perpendicular to \mathbf{B} , the component of the momentum \mathbf{p}_{\parallel} (and hence velocity $\boldsymbol{\beta}_{\parallel}$) parallel to the B-field remains conserved.

We define the *non-relativistic cyclotron frequency* for a particle of charge q and rest-mass m as

$$\omega_B = \frac{qB}{mc} \Rightarrow \nu_B = \frac{\omega_B}{2\pi} = 2.80 \times 10^6 \text{ Hz} (B/\text{G}) \frac{m_e/e}{m/q}. \quad (4.51)$$

which will be the frequency unit in our description of the particle motion. Eq. (4.50) can then be re-written in terms of the velocity components perpendicular and parallel to the B-field

$$\frac{d\boldsymbol{\beta}_{\perp}}{dt} = -\frac{\omega_B}{\gamma} \boldsymbol{\beta}_{\perp} \times \hat{z}, \quad \beta_{\parallel} = |\boldsymbol{\beta}_{\parallel}| = \text{const.} \quad (4.52)$$

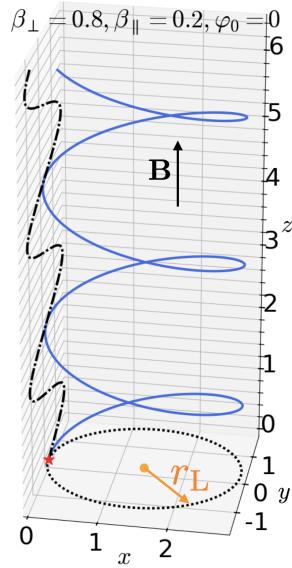


Figure 4.8: Helical orbit of a particle in a uniform B -field (blue line). Black dash-dotted and dotted lines show the projections. The coordinates are in units of c/ω_B . The chosen parameters correspond to a Larmor radius of $r_L = \gamma\beta_{\perp}c/\omega_B = 1.41c/\omega_B$.

In terms of Cartesian components, we have

$$\frac{d\beta_y}{dt} = -\frac{\omega_B}{\gamma}\beta_x, \quad \frac{d\beta_x}{dt} = \frac{\omega_B}{\gamma}\beta_y, \quad \beta_z = \text{const.} \quad (4.53)$$

The two 1st-order differential equations can be combined into 2nd-order ones

$$\frac{d^2\beta_i}{dt^2} = -\left(\frac{\omega_B}{\gamma}\right)^2\beta_i, \quad i = x, y, \quad (4.54)$$

It can be easily shown that the solutions for $\beta_{x/y}(t)$ describe circular motion at the *relativistic cyclotron frequency* ω_B/γ , i.e.

$$\beta_x = \beta_{\perp} \sin(\gamma^{-1}\omega_B t + \varphi_0), \quad \beta_y = \beta_{\perp} \cos(\gamma\omega_B t + \varphi_0), \quad (4.55)$$

where $\beta_{\perp} = \sqrt{\beta_x^2 + \beta_y^2} = \text{const}$ and φ_0 are determined by the initial conditions: $\beta_x(t=0), \beta_y(t=0)$. The circular motion inside the x - y plane combined with a uniform motion along the z -axis describes a *helical orbit*. The ratio between β_{\perp} and β_{\parallel} defines the *pitch angle* α of the helical orbit

$$\tan \alpha = \frac{\beta_{\perp}}{\beta_{\parallel}} = \frac{p_{\perp}}{p_{\parallel}}. \quad (4.56)$$

For a uniform B-field, the magnitudes of the parallel (p_{\parallel}) and perpendicular (p_{\perp}) momentum components, as well as the pitch angle remain constant. The angular frequency of the gyro-motion ω_B/γ is called the *relativistic cyclotron frequency*. This can be understood because the total mass is γm and a heavier particle gyrates slower. An example of such a helical orbit is shown in Fig. 4.8.

The projected trajectory in the x - y plane is a circle whose radius is called the *Larmor radius*, as given by the velocity in the x - y plane, $\beta_{\perp} c$, divided by the relativistic cyclotron frequency ω_B/γ ,

$$r_L = \frac{\beta_{\perp} c}{\omega_B/\gamma} = \frac{\gamma \beta_{\perp} mc^2}{qB} = \frac{p_{\perp} c}{qB} = 1.70 \times 10^3 \text{ cm} \gamma \beta_{\perp} (B/G)^{-1} \frac{m/q}{m_e/e}. \quad (4.57)$$

The factor of $p_{\perp} c/q$, where $p_{\perp} = \gamma \beta_{\perp} mc$ is the perpendicular momentum component, is called the *rigidity* of a particle — this concept is used in the cosmic ray community as it quantifies how easy/hard it is for a given B-field to bend the trajectory of a particle. In the non-relativistic limit $\gamma \approx 1$, the frequency of gyro-motion is independent of the particle's speed and pitch angle and is equal to ω_B — a faster particle will have a larger Larmor radius but the ratio between r_L and $\beta_{\perp} c$ stays constant.

Another important concept related to the particle's helical motion is the *orbital magnetic dipole moment*. The gyro-motion produces a time-averaged loop current $I = q\omega_B/(2\pi\gamma)$ (charge per unit time). Since the Lorentz force is in the direction of $qv \times \mathbf{B}$, the direction of current density (along qv) is such that the magnetic dipole moment of the current loop $\boldsymbol{\mu}$ (along the right-hand-rule direction) is *always anti-parallel* to the B-field, regardless of the charge sign. This can be understood because an electron (with charge $q = -e$) gyrates in the direction opposite to a proton (with charge $q = e$), so they create the same sense of current flow. Since the area of the current loop is $A = \pi r_L^2$, we obtain the *orbital magnetic dipole moment* of the particle

$$\boldsymbol{\mu} = -\frac{IA}{c}\hat{\mathbf{z}} = -\frac{\gamma \beta_{\perp}^2 mc^2}{2B}\hat{\mathbf{z}}. \quad (4.58)$$

The gyro-motions of *all* charged particles, regardless of the charge sign, each generate a secondary B-field that tries to cancel with the external B-field. Such a cancellation *reduces* the total B-field strength and hence lowers the total energy of the system as compared to the unphysical case where the secondary B-field associated with the particle's motion adds to the external field. For this reason, plasmas are *diamagnetic*.

It can be easily shown that the orbital angular momentum of the particle is given by

$$\mathbf{L} = \mathbf{r}_{\perp} \times \mathbf{p}_{\perp} = -\frac{\gamma^2 \beta_{\perp}^2 m_e c^3}{qB} \hat{\mathbf{z}} = \frac{2\gamma mc}{q} \boldsymbol{\mu}, \quad (4.59)$$

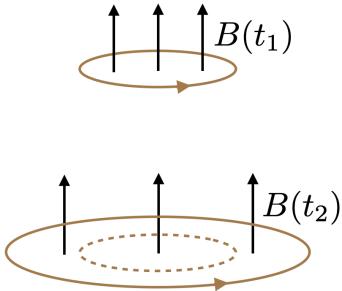


Figure 4.9: The orbit of an electron adiabatically adjusting to the slowly varying B-field to conserve the enclosed magnetic flux $r_L^2 B$.

where \mathbf{r}_\perp is the projected position of the particle in the x - y plane and $|\mathbf{r}_\perp| = r_L$. The ratio between orbital magnetic dipole moment and angular momentum is uniquely determined by a particle's charge-to-mass ratio as

$$\frac{\boldsymbol{\mu}}{\mathbf{L}} = \frac{q}{2\gamma mc}. \quad (4.60)$$

Analogous to the above result, a charged particle with a non-zero spin angular momentum \mathbf{S} (quantized) also has a non-zero magnetic moment given by

$$\boldsymbol{\mu}_s = g \frac{e}{2mc} \mathbf{S}, \quad (4.61)$$

where *spin g-factor* depends on the nature of the particle (experimentally, $g = -2.0023$ for electron, $g = 5.586$ for proton, and $g = -3.826$ for neutron).

In most astrophysical situations, the Larmor radius in eq. (4.57) is much smaller than the curvature radii of the local B-field lines. In this limit, the particle will always gyrate around the *same number of field lines*³ — the ones enclosed by the circle of area πr_L^2 . In the situation that the strength of the B-field changes smoothly on lengthscales much longer than r_L or timescales much longer than $(\omega_B/\gamma)^{-1}$, the magnetic flux through the area of the Larmor radius is conserved, i.e.

$$\frac{d}{dt}(r_L^2 B) = \frac{d}{dt} \left(\frac{p_\perp^2}{B} \right) = 0. \quad (4.62)$$

This can be demonstrated in the following way. Let us go to the comoving frame of the parallel component of the velocity β_{\parallel} . We ignore drift motions and hence the center of the gyro-orbit in this frame stays unchanged. Let the B-field strength $B'(t')$ near the gyro-center be time-dependent, which could be due to non-zero spatial gradient of the lab-frame

³Due to drift motions caused by spatial or temporal variations of the B-field, the center of gyration may slowly move from one field line to another (see §2 of F. F. Chen's book).

field strength along the direction of the β_{\parallel} direction ($\beta_{\parallel} \cdot \nabla B \neq 0$) or due to the time dependence of the B-field strength ($\partial_t B \neq 0$). We consider a time interval of a gyro-period in the comoving frame

$$\Delta t' = \frac{\Delta t}{\gamma_{\parallel}} = \frac{2\pi\gamma/\omega_B}{\gamma_{\parallel}}, \quad (4.63)$$

where $\gamma_{\parallel} = (1 - \beta_{\parallel}^2)^{-1/2}$ is the relative Lorentz factor between the two frames. Suppose after a gyro-period $\Delta t'$ (this could also be any integer times the gyro-period), the B-field has changed by $\Delta\mathbf{B}'$, which is equal to the change in the lab-frame $\Delta\mathbf{B}$ because the field line is parallel to the direction of the relative velocity between two frames. In the comoving frame, we write $\Delta\mathbf{B}'/\Delta t' = \partial_{t'}\mathbf{B}'$ because the gyro-center does not move, and the right-hand side of this expression is given by $-c\nabla' \times \mathbf{E}'$ from the Maxwell-Faraday equation, so we obtain the changing rate of the magnetic flux through the current loop of area A'

$$\int \frac{\Delta\mathbf{B}'}{\Delta t'} \cdot \mathbf{A}' = -c \int (\nabla' \times \mathbf{E}') \cdot \mathbf{A}' = -c \oint_C \mathbf{E}' \cdot d\ell', \quad (4.64)$$

where we have used the Stokes' theorem to convert the integral over an area A' into a loop integral around its boundary C . Since $A' = A = \pi r_L^2$ (the area is perpendicular to β_{\parallel}) and $\Delta B' = \Delta B$, the left-hand side equals to

$$\Delta B \pi r_L^2 \frac{\gamma_{\parallel} \omega_B}{2\pi\gamma}, \quad (4.65)$$

Since $d\ell' = \mathbf{v}' dt'$ and $\mathbf{E}' \cdot \mathbf{v}' = (d\gamma'/dt')(mc/q)$, the right-hand side equals to

$$-\Delta\gamma' mc^3/q, \quad (4.66)$$

where $\Delta\gamma' = \int_0^{\Delta t'} (d\gamma'/dt') dt'$. Using $\gamma' = \gamma/\gamma_{\parallel}$, $p'_{\perp} = p_{\perp}$, and $\Delta p'_{\perp} = \Delta p_{\perp}$, we further obtain

$$\Delta\gamma' = \frac{p'_{\perp} \Delta p'_{\perp}}{\gamma' m^2 c^2} = \frac{\gamma_{\parallel} p_{\perp} \Delta p_{\perp}}{\gamma m^2 c^2}. \quad (4.67)$$

Finally,

$$\frac{\Delta B}{B} = -2 \frac{\Delta p_{\perp}}{p_{\perp}} \rightarrow \frac{p_{\perp}^2}{B} = \text{const, or } r_L^2 B = \text{const.} \quad (4.68)$$

We see that smooth spatial or temporal variations of the B-field can gradually change the perpendicular momentum $p_{\perp} = \gamma\beta_{\perp}m_e c$ of the electron, and this can lead to the change in pitch angle α .

On the other hand, it is generally possible that the B-field lines have structures on length-scales comparable or smaller than the particle's Larmor radius. Such small-scale irregularities may be due to turbulence cascade, which can transport energy carried by the fluctuations of fluid quantities (e.g., density, pressure, B-fields) from large scales to small

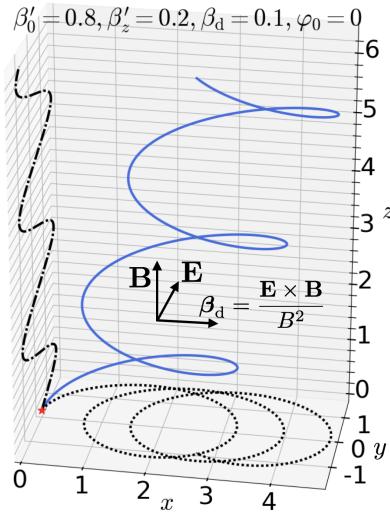


Figure 4.10: Blue solid line shows the lab-frame trajectory of a charge particle in uniform EM fields given by $\mathbf{E} = E\hat{\mathbf{y}}$ and $\mathbf{B} = B\hat{\mathbf{z}}$. Black dotted line shows the projection in the x - y plane, a superposition of a constant drift motion of $\beta_d = \beta_d\hat{\mathbf{x}}$ and gyro-motion. Black dash-dotted line shows the projection in the y - z plane, a superposition of a constant velocity $\beta_z\hat{\mathbf{z}}$ and sinusoidal oscillation along $\hat{\mathbf{y}}$. The coordinates are in units of c/ω_B .

scales. Magnetic fluctuations are also generated by streaming motions of high-energy particles (e.g., cosmic rays), which excite short-wavelength waves due to plasma instabilities. We will not discuss these in detail, but the important result is that, a particle's pitch angle is strongly modified when it interacts with magnetic irregularities on scales comparable to its Larmor radius — this is called *pitch-angle scattering* analogous to the momentum deflection caused by Coulomb scattering. The net effect of scatterings off magnetic fluctuations is to randomize the pitch angle distribution such that the plasma's entropy is maximized.

4.2.3 Uniform \mathbf{E} & \mathbf{B} (\mathbf{E} -cross- \mathbf{B} drift plus helical orbit)

Consider a charged particle moving in spatially uniform EM fields in the lab-frame,

$$\mathbf{E} = E\hat{\mathbf{y}}, \quad \mathbf{B} = B\hat{\mathbf{z}}. \quad (4.69)$$

The choice of $\mathbf{E} \perp \mathbf{B}$ here is motivated by the fact that the E-field component parallel to the B-field cannot be Lorentz transformed away (as $\mathbf{E} \cdot \mathbf{B}$ = Lorentz invariant). If $\mathbf{E} \cdot \mathbf{B} \neq 0$, we know that the particle will be accelerated indefinitely and other pieces of physics (e.g., radiation drag, pair production) must come in play before the Lorentz factor goes to infinity.

We will focus on the magnetically dominated case with $E < B$, because otherwise the particle will be accelerated indefinitely. If $E < B$, then it is possible to find a Lorentz frame where the E-field vanishes. For the choice of EM fields in eq. (4.69), let us consider an other inertial frame that is moving wrt. the lab-frame at velocity

$$\beta_d = \frac{\mathbf{E} \times \mathbf{B}}{B^2} = \frac{E}{B} \hat{\mathbf{x}}, \quad (4.70)$$

where the subscript d stands for drift and β_d describes the “E-cross-B drift”. Note that the E-cross-B drift speed⁴ does not depend on the mass or charge of the particle — a proton will have the same drift velocity as an electron. In the comoving frame of velocity β_d , we denote all quantities with a prime ('') and the EM fields are given by

$$\mathbf{E}' = 0, \quad \mathbf{B}' = (B/\gamma_d) \hat{\mathbf{z}}, \quad (4.71)$$

where $\gamma_d = (1 - \beta_d^2)^{-1/2}$ is the Lorentz factor of the drift.

Let us define the non-relativistic cyclotron frequency in the comoving frame

$$\omega'_B = \frac{qB'}{mc} = \frac{\omega_B}{\gamma_d}, \quad (4.72)$$

where $\omega_B = qB/(mc)$ is the cyclotron frequency in the lab frame.

In the comoving frame of the E-cross-B drift, we expect the particle to take a helical orbit with a relativistic gyro-frequency of ω'_B/γ' , where γ' is a constant Lorentz factor. Following the case of a uniform B-field as discussed in §4.2.2, we write the solutions for the velocity components in the comoving frame of the E-cross-B drift

$$\beta'_x = \beta'_0 \sin(\gamma'^{-1}\omega'_B t' + \varphi_0), \quad \beta'_y = \beta'_0 \cos(\gamma'^{-1}\omega'_B t' + \varphi_0), \quad \beta'_z = \text{const}, \quad (4.73)$$

where $\beta'_0 = \sqrt{(\beta'_x)^2 + (\beta'_y)^2}$, $\gamma' = [1 - (\beta'_0)^2 - (\beta'_z)^2]^{-1/2}$, and φ_0 are determined by the initial conditions of $\beta'_x(t' = 0), \beta'_y(t' = 0)$ and the constant β'_z .

Having the 4-velocity in the comoving frame $\vec{u}'/c = \gamma'(1, \beta'_x, \beta'_y, \beta'_z)$, we then go back to the lab frame by doing a Lorentz transformation

$$\gamma = \gamma_d \gamma'(1 + \beta_d \beta'_x), \quad u_x/c = \gamma_d \gamma'(\beta_d + \beta'_x), \quad u_y/c = \gamma' \beta'_y, \quad u_z/c = \gamma' \beta'_z = \text{const}. \quad (4.74)$$

An example of the particle’s trajectory in the lab frame is shown in Fig. 4.10. The above solution can be understood as the superposition of a helical motion (for a uniform B-field) and the E-cross-B drift motion $\beta_d = \mathbf{E} \times \mathbf{B}/B^2$.

⁴There are many other types of drifts when the fields are not uniform or there is a gravitational acceleration, etc (see F. F. Chen’s *Introduction to Plasma Physics*). We will not discuss those in detail.

4.2.4 Lagrangian and Hamiltonian

The *Lagrangian* describes the current state of the particle based on its coordinate \mathbf{x} , time derivative of the coordinate $\mathbf{v} = \dot{\mathbf{x}} = d\mathbf{x}/dt$, and the coordinate time t . For a charge particle interacting with the EM field, the Lagrangian is (see Jackson's book)

$$\mathcal{L}(\mathbf{x}, \mathbf{v}, t) = -mc^2/\gamma + q(\mathbf{A} \cdot \mathbf{v}/c - \phi), \quad (4.75)$$

where the 1st term is the contribution from a free particle, the 2nd and 3rd terms are the interaction terms, and $\vec{A} = (\phi, \mathbf{A})$ is the 4-potential related to the electric field $\mathbf{E} = -\nabla\phi - c^{-1}\partial_t\mathbf{A}$ and magnetic field $\mathbf{B} = \nabla \times \mathbf{A}$ (see §4.1.3). It is easy to show that $\gamma\mathcal{L} = \vec{p} \cdot \vec{p}/m + (q/c)\vec{A} \cdot \vec{u}$, where $\vec{p} = m\vec{u}$ is the 4-momentum, is Lorentz invariant. This comes from the fact that the action integral between two different coordinate times t_1 and t_2 or proper times τ_1 and τ_2 ,

$$S \equiv \int_{t_1}^{t_2} \mathcal{L} dt = \int_{\tau_1}^{\tau_2} = \gamma\mathcal{L} d\tau, \quad (4.76)$$

must be Lorentz invariant, because the equations of motion are determined by the *principle of least action* or *Hamilton's principle*, $\delta S = 0$, where δ means path variation.

The extremum condition $\delta S = 0$ leads to the *Euler-Lagrange equation* for each coordinate x^i ($i = 1, 2, 3$)

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \mathbf{v}} \right) - \frac{\partial \mathcal{L}}{\partial \mathbf{x}} = 0, \quad (4.77)$$

which reproduce eq. (4.49) if the Lagrangian derivative is understood as $d/dt = \partial_t + \mathbf{v} \cdot \nabla$, i.e., the derivative along the particle's trajectory. Using $\partial\gamma/\partial\mathbf{v} = \gamma^3\mathbf{v}/c^2$, we obtain the *canonical momentum* conjugate to the position coordinate

$$\mathbf{P} \equiv \frac{\partial \mathcal{L}}{\partial \mathbf{v}} = \mathbf{p} + q\mathbf{A}/c. \quad (4.78)$$

The *Hamiltonian*, as a function of the coordinate \mathbf{x} , its conjugate momentum \mathbf{P} , and time t , is given by

$$\mathcal{H}(\mathbf{x}, \mathbf{P}, t) \equiv \mathbf{P} \cdot \mathbf{v} - \mathcal{L}(\mathbf{x}, \mathbf{v}, t) = \gamma mc^2 + q\phi = \sqrt{(c\mathbf{P} - q\mathbf{A})^2 + m^2c^4} + q\phi. \quad (4.79)$$

We see that the Hamiltonian equals to the total (kinetic+potential) energy of the particle. In fact, the Hamiltonian and canonical momentum forms the *4-canonical momentum*

$$\vec{P} = (\mathcal{H}/c, \mathbf{P}), \quad (4.80)$$

which is similar to the case of the kinetic 4-momentum $\vec{p} = (E/c, \mathbf{p})$ (here $E = \gamma mc^2$ means the kinetic energy including rest mass). It is easy to verify that these two 4-momenta differ by an interaction term that is proportional to the charge q and the 4-potential \vec{A} ,

$$\vec{P} = \vec{p} + q\vec{A}/c. \quad (4.81)$$

Taking the total derivative of the above Hamiltonian (eq. 4.79) and using $\mathbf{P} = \partial\mathcal{L}/\partial\mathbf{v}$ (eq. 4.78) and $\dot{\mathbf{P}} = \partial\mathcal{L}/\partial\mathbf{x}$ (eq. 4.77), we obtain

$$d\mathcal{H} = \mathbf{v} \cdot d\mathbf{P} - \dot{\mathbf{P}} \cdot d\mathbf{x} - \partial_t\mathcal{L} dt. \quad (4.82)$$

We then identify \mathbf{v} as $\partial_{\mathbf{P}}\mathcal{H}$, $-\dot{\mathbf{P}}$ as $\partial_{\mathbf{x}}\mathcal{H}$, and $-\partial_t\mathcal{L}$ as $\partial_t\mathcal{H}$. These give rise to the *canonical Hamilton equations*

$$\frac{d\mathbf{x}}{dt} = \frac{\partial\mathcal{H}}{\partial\mathbf{P}}, \quad \frac{d\mathbf{P}}{dt} = -\frac{\partial\mathcal{H}}{\partial\mathbf{x}}, \quad (4.83)$$

and we also find the third equation

$$\frac{d\mathcal{H}}{dt} = \frac{\partial\mathcal{H}}{\partial t} = -\frac{\partial\mathcal{L}}{\partial t}, \quad (4.84)$$

which is useful only if the 4-potential \vec{A} is time-independent (and hence the Hamiltonian/Lagrangian depends explicitly on the coordinate time). When the 4-potential is independent, then the Hamiltonian is conserved.

It is useful to understand Hamiltonian formalism for at least two reasons: (i) it is easy to identify the *conserved quantities* in the system — if the Hamiltonian (or Lagrangian) does not explicitly depend on a certain coordinate x_i , then we know that its conjugate momentum P_i is conserved along the trajectory ($dP_i/dt = 0$); (ii) when we transition to quantum theory, the Hamiltonian \mathcal{H} determines the energies of stable states and the transition rates between them. Another advantage of Hamiltonian mechanics is the usage of generalized coordinates, but for our purposes in this Chapter, Cartesian coordinates (t, x, y, z) will be used as we did for Special Relativity (Ch. 3).

Sometimes, it might be hard to remember which of the two Hamilton equations has a minus sign. A trick is to consider a simple, special case. For instance, for particle motion in a static electric field, the canonical momentum is equal to the kinetic momentum, $\mathbf{P} = \mathbf{p}$, and the equation of motion is $\dot{\mathbf{p}} = -q\nabla\phi = -\partial_{\mathbf{x}}\mathcal{H}$ (note the negative sign), as the Hamiltonian is the sum of kinetic (γmc^2) and potential ($q\phi$) energies. Another simple case is a non-relativistic free particle for which $\mathcal{H} = \mathbf{p}^2/(2m)$, so we obtain $\partial_{\mathbf{p}}\mathcal{H} = \mathbf{p}/m = \mathbf{v} = d\mathbf{x}/dt$ (note the absence of negative sign).

In the non-relativistic (NR) limit, the Lagrangian and Hamiltonian are given by (without including the rest mass)

$$\mathcal{L}^{(\text{NR})} = \frac{\mathbf{p}^2}{2m} + q\mathbf{A} \cdot \mathbf{v}/c - q\phi, \quad (4.85)$$

$$\mathcal{H}^{(\text{NR})} = \frac{\mathbf{p}^2}{2m} + q\phi = \frac{(\mathbf{P} - q\mathbf{A}/c)^2}{2m} + q\phi. \quad (4.86)$$

In the following, we use the Lagrangian formulism to study the motion of a charged particle under a strong EM wave.

4.2.5 *Strong EM wave (“figure-8” motion)

Let us consider a monochromatic plane EM wave which can be described by the following vector potential

$$\mathbf{A} = \frac{cE_0}{\omega} \hat{\mathbf{x}} \sin \psi, \quad (4.87)$$

where $\psi \equiv kz - \omega t$ and $\mathbf{k} = (\omega/c)\hat{\mathbf{z}}$ is the wavevector in vacuum. We do not use the complex notation, because for the general case considered here, the results depend on the strength of the EM wave. The scalar potential of the wave field is taken to be $\phi_w = 0$. The wave EM fields are

$$\begin{aligned} \mathbf{E} &= -\frac{1}{c} \partial_t \mathbf{A} = E_0 \hat{\mathbf{x}} \cos \psi, \\ \mathbf{B} &= \nabla \times \mathbf{A} = E_0 \hat{\mathbf{y}} \cos \psi, \end{aligned} \quad (4.88)$$

The Lagrangian of a particle with charge q and rest mass m is

$$\mathcal{L} = -mc^2/\gamma + \frac{q}{c} \mathbf{A} \cdot \mathbf{v}, \quad (4.89)$$

and the canonical momentum is

$$\mathbf{P} = \gamma m \mathbf{v} + \frac{q}{c} \mathbf{A}. \quad (4.90)$$

It is convenient to define the dimensionless strength of the wave

$$a_0 = \frac{qE_0}{m\omega c}, \quad (4.91)$$

which is called the *non-linearity parameter*. The wave is linear in the limit $a_0 \ll 1$, since the particle’s motion is non-relativistic. Our consideration here is general.

Since \mathcal{L} does not explicitly depend on coordinate x and y , the canonical momenta in these two directions are conserved

$$\begin{aligned} P_x/mc &= \gamma \beta_x + a_0 \sin \psi = \text{const}, \\ P_y/mc &= \gamma \beta_y = \text{const}. \end{aligned} \quad (4.92)$$

The Euler-Lagrange equation for the z -coordinate is

$$\frac{d}{dt}(\gamma \beta_z) = \omega a_0 \beta_x \cos \psi = \frac{d\gamma}{dt}, \quad (4.93)$$

where the second equality comes from the fact that only the wave electric field does work (see eq. 4.47). This means $\gamma(1 - \beta_z) = \text{const}$. We consider the case that the particle is initially at rest, so the 4-velocity is given by

$$\gamma = 1 + (1/2)a_0^2 \sin^2 \psi, \quad u_x = \gamma \beta_x = -a_0 \sin \psi, \quad u_z = \gamma \beta_z = (1/2)a_0^2 \sin^2 \psi, \quad (4.94)$$

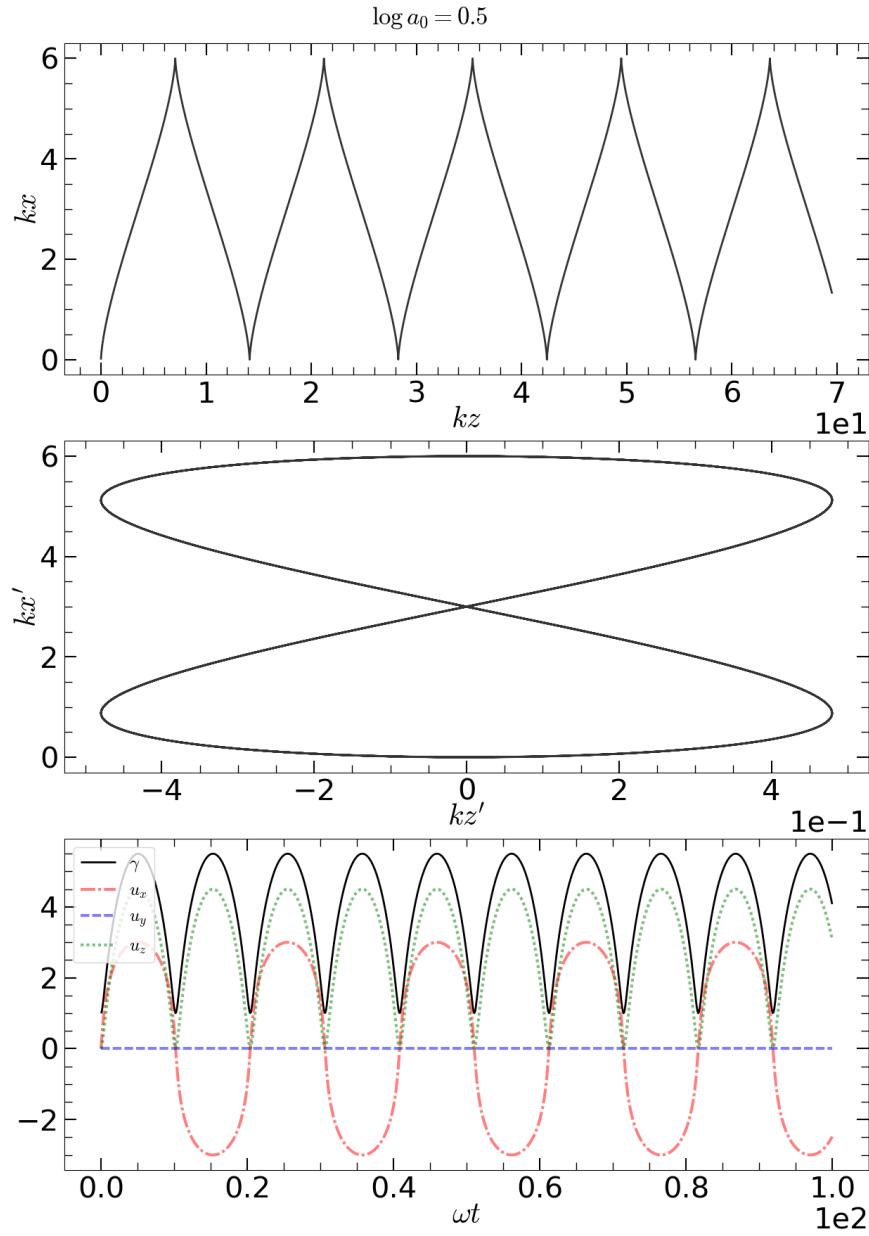


Figure 4.11: Motion of a charge particle under a strong EM wave with dimensionless strength $a_0 = qE_0/(m\omega c) \simeq 3$. The upper panel shows the trajectory projected onto the z - x plane in the lab frame (length unit: $k^{-1} = c/\omega$). The middle panel shows the “figure-8” trajectory in the comoving frame of the mean velocity $\langle \beta_z \rangle = 1/(4/a_0^2 + 1)$. The bottom panel shows the 4-velocity as a function of time in the lab frame.

Since $d\psi/dt = \omega(\beta_z - 1)$, we can eliminate β_z and solve for the phase

$$\psi = -\frac{\omega t}{1+\eta} + \frac{\eta}{2} \frac{\sin 2\psi}{1+\eta}, \quad \eta \equiv a_0^2/4, \quad (4.95)$$

and then the position of the particle is given by

$$\begin{aligned} kx &= a_0(1 - \cos \psi), \\ kz &= \frac{a_0^2}{4} \left(-\psi + \frac{\sin 2\psi}{2} \right) = \frac{\eta}{1+\eta} \omega t + \frac{\eta}{2} \frac{\sin 2\psi}{1+\eta}, \end{aligned} \quad (4.96)$$

where we have used

$$\gamma \frac{dx}{dt} = \gamma \frac{d\psi}{dt} \frac{dx}{d\psi} = \omega \gamma (\beta_z - 1) \frac{dx}{d\psi} = \omega \frac{dx}{d\psi} = -a_0 c \sin \psi. \quad (4.97)$$

This solution is a homework problem in Landau & Lifshitz's *Classical Theory of Fields*. The maximum Lorentz factor is $\gamma_{\max} = 1 + a_0^2/2$. The particle moves in the z direction (parallel to the wavevector) at a mean speed of

$$\langle \beta_z \rangle = c^{-1} \langle dz/dt \rangle = \frac{\eta}{1+\eta}, \quad (4.98)$$

which corresponds to a Lorentz factor of $\langle \gamma \rangle \approx 2^{-3/2} a_0$ in the ultra-relativistic limit ($a_0 \gg 1$). If we go to the comoving frame of this mean velocity, then the trajectory looks like a "figure-8", as shown in Fig. 4.11. In the non-relativistic limit ($a_0 \ll 1$), the particle oscillates mainly in the transverse (x) direction with an amplitude of $\Delta x = a_0/k = a_0 \lambda/2\pi \ll \lambda$ away from the time-averaged position, but there is also an oscillatory motion in the longitudinal direction with a much smaller amplitude of $\Delta z = a_0^2/(8k) = a_0^2 \lambda/(16\pi) \ll \lambda$ away from the time-averaged position.

4.2.6 *Strong EM wave plus uniform B-field (stochastic acceleration)

Let us add a static background B-field in the x - y plane⁵ given by

$$\mathbf{A}_B = (B_y \hat{x} - B_x \hat{y})z, \text{ or } \mathbf{B}_{bg} = B_x \hat{x} + B_y \hat{y}, \quad (4.99)$$

where $B_x = B_{bg} \sin \theta_B$ and $B_y = B_{bg} \cos \theta_B$. The total vector potential is $\mathbf{A} = \mathbf{A}_{wave} + \mathbf{A}_B$ and the scalar potential is $\phi = 0$. We define the (non-relativistic) cyclotron frequency for the static B-field

$$\omega_B = \frac{qB_{bg}}{mc}. \quad (4.100)$$

⁵Note that placing the static B-field in the x - y plane does not lose the generality of the setup. This is because for an arbitrary angle between \mathbf{B}_{bg} and the wavevector $\mathbf{k} = k\hat{z}$ in one inertial frame \mathcal{O} , one can go to another inertial frame \mathcal{O}' that is moving at a velocity $\mathbf{v}_{rel} = \hat{\mathbf{B}}_{bg} c \cos \theta$ wrt. frame \mathcal{O} , where $\cos \theta = \hat{\mathbf{B}}_{bg} \cdot \hat{\mathbf{z}}$. In frame \mathcal{O}' , the wavevector is in the x' - y' plane and the angular frequency is $\omega \sin \theta$. And since the relative velocity is along $\hat{\mathbf{B}}_{bg}$, there is no static electric field in frame \mathcal{O}' .

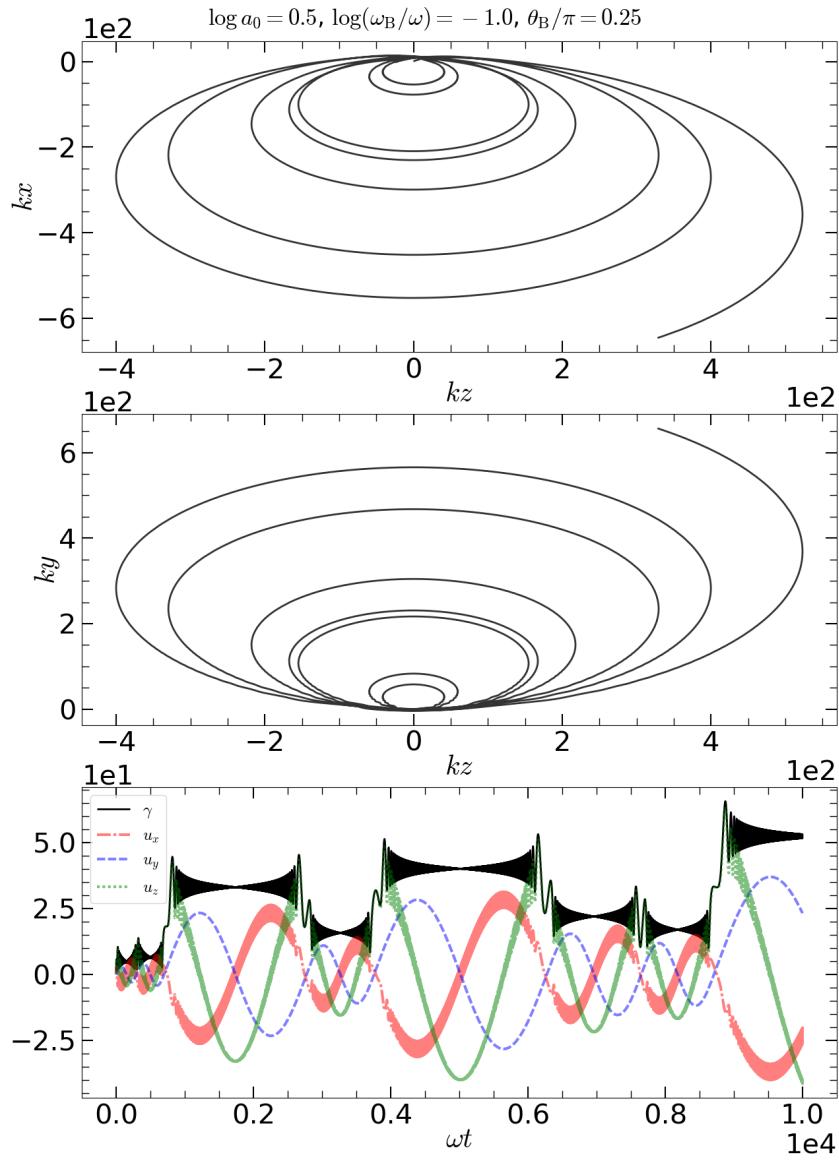


Figure 4.12: Motion of a charge particle under a strong EM wave $\mathbf{E} = E_0 \hat{x} \cos(kz - \omega t)$ plus a static background B-field \mathbf{B}_{bg} which is in the plane perpendicular to the wavevector $\mathbf{k} = k \hat{z}$. The orientation of the static B-field is specified by the angle $\theta_B = \text{acos}(\hat{x}, \hat{\mathbf{B}}_{\text{bg}})$. The EM wave has dimensionless strength $a_0 = qE_0/(m\omega c) = 3$ and the static B-field corresponds to a non-relativistic cyclotron frequency $\omega_B/\omega = 0.1$ (such that $B_{\text{bg}}/E_0 = 1/30$). The upper and middle panels show the projected trajectory (length unit: $k^{-1} = c/\omega$). The bottom panel shows the 4-velocity as a function of time, all in the lab frame.

The conservation of canonical momenta in the x and y directions give

$$\begin{aligned} P_x/mc &= \gamma\beta_x + \frac{\omega_B z}{c} \sin \theta_B + a_0 \sin \psi = \text{const}, \\ P_y/mc &= \gamma\beta_y - \frac{\omega_B z}{c} \cos \theta_B = \text{const}. \end{aligned} \quad (4.101)$$

The Euler-Lagrange equation for the z -coordinate is

$$\frac{d}{dt}(\gamma\beta_z) = \omega a_0 \beta_x \cos \psi + \omega_B (\sin \theta_B \beta_x - \cos \theta_B \beta_y). \quad (4.102)$$

Again, the first term on the RHS equals to $d\gamma/dt$ (using eq. 4.47), so the above equation can be written as

$$\frac{d}{dt} \left[\gamma(1 - \beta_z) + \frac{\omega_B}{c} (x \sin \theta_B - y \cos \theta_B) \right] = 0, \quad (4.103)$$

which gives another integral of motion.

We then specify the initial conditions as $\mathbf{v}(t = 0) = \mathbf{0}$ and $\mathbf{x}(t = 0) = \mathbf{0}$. It is possible to workout the solution for the general case with an arbitrary initial position z_0 (or phase ϕ_0) and velocity \mathbf{v}_0 , but in the following we only consider the simplest case.

Under the initial conditions that the particle is at rest at the coordinate origin, we obtain from the integrals of motion

$$\begin{aligned} \gamma\beta_x &= -a_0 \sin \psi - \frac{\omega_B z}{c} \sin \theta_B, \\ \gamma\beta_y &= \frac{\omega_B z}{c} \cos \theta_B, \\ \gamma\beta_z &= \gamma - 1 + \frac{\omega_B}{c} (x \sin \theta_B - y \cos \theta_B). \end{aligned} \quad (4.104)$$

Then, the Lorentz factor can be obtained as a function of spacetime coordinates (t, \mathbf{x}) ,

$$\gamma = 1 + \frac{a_0^2 \sin^2 \psi + (\omega_B/c)^2 [z^2 + 2a_0 z \sin \theta_B \sin \psi (c/\omega_B) + (x \sin \theta_B - y \cos \theta_B)^2]}{2 [1 - (\omega_B/c) (x \sin \theta_B - y \cos \theta_B)]}. \quad (4.105)$$

We must numerically integrate an ordinary differential equation, which is of the form $d\mathbf{x}/dt = f(t, \mathbf{x})$, to obtain the full trajectory of the particle. The results from a 4th-order Runge-Kutta integration are shown in Fig. 4.12 for a case with non-linearity parameter $a_0 = 3$ and weak static B-field $\omega_B/\omega = 0.1$. We see that the static B-field confines the particle's motion in the z -direction (recall that in the EM wave-only case, the particle can move to $z \rightarrow \infty$). The particle's trajectory is not periodic. It gains a much higher peak Lorentz factor $\gamma \sim 50$ than in the EM wave-only case (where $\gamma_{\max} = 5.5$), as a result of stochastic acceleration every once it returns near the original position.

4.2.7 *EM wave and Coulomb potential (under construction)

Instead of a static B-field, we would like to consider the case of an EM wave plus Coulomb potential. An example is an electron moving near a nucleus while being illuminated by radiation — this is the case of *inverse-Bremsstrahlung* or *free-free absorption*.

The Coulomb field is time-independent (ignoring the motion of the nucleus)

$$\phi = Zq/|\mathbf{r}|, \quad (4.106)$$

where $|\mathbf{r}| = \sqrt{x^2 + y^2 + z^2}$ is the distance between the electron and nucleus. Since the system is rather complicated, we only consider the non-relativistic case with $a_0 \ll 1$ but keep the wave amplitude finite. The (non-relativistic) Lagrangian becomes

$$\mathcal{L} = \frac{\mathbf{p}^2}{2m} + q\mathbf{A} \cdot \mathbf{v}/c - q\phi, \quad (4.107)$$

where \mathbf{A} for describes the EM wave. For an electron with initial velocity v and impact parameter b , the most important part of the Coulomb interaction occurs on a timescale of b/v . The maximum impact parameter we consider is

$$b_{\max} = v/\omega, \quad (4.108)$$

because for larger $b > b_{\max}$, the scattering occurs over multiple wave periods and the electron adiabatically adjusts to the slowly evolving combined electromagnetic+Coulomb fields and hence there is little energy exchange. The minimum impact parameter is

$$b_{\min} = \max \left(\frac{q^2}{mv^2}, \frac{\hbar}{mv} \right), \quad (4.109)$$

where the first limit corresponds to a large-angle ($\sim 90^\circ$) deflection and the second limit comes from Heisenberg's uncertainty principle.

In the limit $b \ll b_{\max}$, the scattering occurs on a timescale much shorter than the wave period (this is the *impuse approximation*), so the local EM wave fields can be considered as time-independent. This leads to $\partial_t \mathcal{L} \approx 0$ and hence the Hamiltonian is approximately conserved, i.e.

$$\mathcal{H} = \frac{(\mathbf{P} - q\mathbf{A}/c)^2}{2m} + q\phi = \frac{P^2}{2m} - \frac{q}{mc}\mathbf{P} \cdot \mathbf{A} + \frac{q^2 A^2}{2mc^2} + q\phi = \text{const.} \quad (4.110)$$

The key here is that long before and long after the Coulomb collision, the electron is only affected by the EM wave and hence its canonical momenta \mathbf{P}_1 (before) and \mathbf{P}_2 (after the collision) are conserved. It is the Coulomb collision that causes the $\mathbf{P}_1 \rightarrow \mathbf{P}_2$ energy transition. The particle approaches the nucleus from (effectively) infinity and goes back to

infinity after the collision, so the potential energy $q\phi$ does not change. Under the impulse approximation, the term $q^2 A^2 / (2mc^2)$ stays unchanged, since the interaction occurs within a short lengthscale $b \ll b_{\max} \ll c/\omega$ and a short timescale $b/v \ll \omega^{-1}$. Therefore, we obtain the energy change of the particle

$$\Delta E = \frac{\mathbf{P}_1^2 - \mathbf{P}_2^2}{2m} = -\frac{q}{mc} (\mathbf{P}_2 - \mathbf{P}_1) \cdot \mathbf{A}. \quad (4.111)$$

The following thought experiment help to understand why the energy change ΔE is given by the difference in canonical momenta $(\mathbf{P}_1^2 - \mathbf{P}_2^2)/2m$, instead of kinetic momenta.

Consider an electron with kinetic energy $p_1^2/2m$ in the absence of EM wave, so the canonical momentum is $\mathbf{P}_1 = \mathbf{p}_1$ initially. Now we slowly ramp up the amplitude of the EM wave and the electron starts oscillating. For non-relativistic amplitudes $a_0 \ll 1$, the oscillation is mainly in the direction perpendicular to the wavevector (the transverse velocity is $\sim a_0 c$ whereas the longitudinal velocity is $\sim a_0^2 c$, see eq. 4.94). Since the Lagrangian of the wave-particle interaction does not explicitly depend on the transverse coordinates (x and y in §4.2.5), the canonical momenta in these two directions (which dominate the total canonical momentum) are conserved. Therefore, we conclude that as we slowly ramp up the amplitude of the EM wave, the canonical momentum must stay as $\mathbf{P}_1 = \mathbf{p}_1$. The same applies as we slowly decreases the EM wave amplitude back to zero. We see that although the particle might be oscillating at a high speed (potentially greater than its initial thermal speed) when the EM wave is on, the kinetic energy $p^2/2m$ does not all “belong” to the particle — when the EM wave retreats, the particle must give the oscillating kinetic energy back to the wave and retain its original share of $p_1^2/2m$. In this sense, the canonical momentum (rather than the kinetic momentum) better describes the actual energy of the particle moving in an EM wave.

Let us then briefly discuss the Coulomb collision part of the story, which involves the $\mathbf{p}_1 \rightarrow \mathbf{p}_2$ transition in terms of the physical kinetic momenta. Under the impulse approximation, the EM wave phase does not change during the collision, so we have $\Delta \mathbf{p} = \mathbf{p}_2 - \mathbf{p}_1 = \mathbf{P}_2 - \mathbf{P}_1$, and hence

$$\Delta E = -\frac{q}{mc} \Delta \mathbf{p} \cdot \mathbf{A}. \quad (4.112)$$

We denote the deflection angle as $\theta = \text{acos}(\hat{\mathbf{p}}_1 \cdot \hat{\mathbf{p}}_2)$. Since the scattering is elastic (ignoring nucleus motion), the momentum does not change in magnitude, i.e. $p_2 = p_1 = p$, and hence one can geometrically show that $\sin(\theta/2) = \Delta p/p$. Since most collisions are weak ($b \gg b_{\min}$) with a small deflection angle, the interaction roughly gives a momentum change of the order $\Delta p \sim Z q^2 / b^2 \times b/v$ (force multiplied by the interaction time, Z being the charge number of the nucleus) in the direction perpendicular to \mathbf{p}_1 . It is the dot-product between the momentum change $\Delta \mathbf{p}$ and the potential vector \mathbf{A} that gives the energy exchange between the particle and EM wave.

(under construction)

4.2.8 *Energy-momentum tensor for EM fields

The equation of motion $d\mathbf{p}/dt = q(E + \mathbf{v}/c \times \mathbf{B})$ (eq. 4.49) can be expressed in a Lorentz-invariant (frame-independent) way that better describes the energy and momentum exchange between EM fields and plasma particles.

We start from the law of energy conservation. Suppose a plasma has charge density $\rho = nq$ where n is the lab-frame number density⁶ of particles of charge q . As a result of the electric force $q\mathbf{E}$, each particle moving at velocity \mathbf{v} gains kinetic energy at a rate given by $q\mathbf{E} \cdot \mathbf{v}$ (eq. 4.47). Thus, the EM fields suffer energy loss at a rate $-nq\mathbf{E} \cdot \mathbf{v} = -\mathbf{J} \cdot \mathbf{E}$ per unit volume, where $\mathbf{J} = nq\mathbf{v}$ is the current density. Since the EM fields have energy density $U_{\text{EM}} = (E^2 + B^2)/(8\pi)$ and energy flux $\mathbf{S} = (\mathbf{E} \times \mathbf{B})c/(4\pi)$, the continuity equation for EM field energy can be written as

$$\partial_t U_{\text{EM}} + \nabla \cdot \mathbf{S} = -\mathbf{J} \cdot \mathbf{E}, \quad (4.113)$$

where $\nabla \cdot \mathbf{S}$ describes the divergence of the energy flux and $-\mathbf{J} \cdot \mathbf{E}$ is a sink term.

Let us then study the law of momentum conservation. For a plasma of charge density $\rho = nq$ and current density $\mathbf{J} = nq\mathbf{v}$, the Lorentz force per unit volume is given by

$$\mathbf{f} = nq(\mathbf{E} + \mathbf{v}/c \times \mathbf{B}) = \rho\mathbf{E} + c^{-1}\mathbf{J} \times \mathbf{B}, \quad (4.114)$$

which is called the *Lorentz force density*. The EM fields suffer momentum loss at a rate $-\mathbf{f}$ per unit volume. The continuity equation for EM field momentum can be written as

$$c^{-2}\partial_t \mathbf{S} + \nabla \cdot \overset{\leftrightarrow}{T} = -\mathbf{f}, \quad (4.115)$$

where $\partial_t \mathbf{S}$ describes the rate of change for the momentum density (identical to the energy flux as one can check the units of the two) and $\nabla \cdot \overset{\leftrightarrow}{T}$ describes the divergence of the *Maxwell pressure tensor* $\overset{\leftrightarrow}{T}$ (also called the *Maxwell stress tensor*). The diagonal components T_{ii} (for $i = 1, 2, 3$) describes the pressure (or momentum flux) across a surface at a constant x_i , and the off-diagonal components T_{ij} (for $i \neq j$) describes a quantity called the *shear stress*. Generally, the component T_{ij} (for any $i, j = 1, 2, 3$) describes the flux of i momentum component along the j direction (or across a surface at constant x_j). In the following, we go through some tedious calculations to demonstrate that the Maxwell pressure tensor is given by

$$T_{ij} = \frac{1}{4\pi} \left(\frac{1}{2} \delta_{ij} (E^2 + B^2) - (E_i E_j + B_i B_j) \right), \quad (4.116)$$

⁶If there are multiple charge species, then $\rho = nq$ denotes the net charge density.

where δ_{ij} is the Kronecker delta ($\delta_{ii} = 1$ and $\delta_{i\neq j} = 0$). Note that some other textbooks (e.g., Griffiths) define the Maxwell pressure tensor with a negative sign (such that $\mathbf{f} = \nabla \cdot \overset{\leftrightarrow}{T} - c^{-2} \partial_t \mathbf{S}$). Our choice in eq. (4.116) is motivated by the simpler expression of the *energy-momentum tensor* (also called *stress-energy tensor*) for EM fields

$$(T^{\mu\nu}) = \begin{pmatrix} U_{\text{EM}} & \mathbf{S}^T/c \\ \mathbf{S}/c & \overset{\leftrightarrow}{T} \end{pmatrix}, \quad (4.117)$$

which is a 4×4 symmetric tensor. Note that $T^{ij} = T_{ij}$ as i and j are both spatial indices (the sign would only flip when we raise or lower a temporal index).

We use the charge density $\rho = \nabla \cdot \mathbf{E}/(4\pi)$ and current density $\mathbf{J} = (c\nabla \times \mathbf{B} - \partial_t \mathbf{E})/(4\pi)$ to re-express the force density as follows

$$\begin{aligned} 4\pi \mathbf{f} &= (\nabla \cdot \mathbf{E}) \mathbf{E} + (\nabla \times \mathbf{B} - c^{-1} \partial_t \mathbf{E}) \times \mathbf{B} \\ &= (\nabla \cdot \mathbf{E}) \mathbf{E} - \mathbf{B} \times (\nabla \times \mathbf{B}) + c^{-1} \mathbf{E} \times \partial_t \mathbf{B} - c^{-1} \partial_t (\mathbf{E} \times \mathbf{B}) \\ &= (\nabla \cdot \mathbf{E}) \mathbf{E} - \mathbf{E} \times (\nabla \times \mathbf{E}) + (\nabla \cdot \mathbf{B}) \mathbf{B} - \mathbf{B} \times (\nabla \times \mathbf{B}) - c^{-1} \partial_t (\mathbf{E} \times \mathbf{B}) \\ &= [(\nabla \cdot \mathbf{E}) \mathbf{E} + (\mathbf{E} \cdot \nabla) \mathbf{E}] + [(\nabla \cdot \mathbf{B}) \mathbf{B} + (\mathbf{B} \cdot \nabla) \mathbf{B}] - \frac{\nabla (E^2 + B^2)}{2} - c^{-1} \partial_t (\mathbf{E} \times \mathbf{B}). \end{aligned} \quad (4.118)$$

In the (tedious) calculations above, we have used the following $\partial_t B = -c\nabla \times \mathbf{E}$, $\nabla(E^2) = 2(\nabla \cdot \mathbf{E}) \mathbf{E} + 2\mathbf{E} \times (\nabla \times \mathbf{E})$ (and similarly for $\nabla(B^2)$), and $\nabla \cdot \mathbf{B} = 0$. The final expression for \mathbf{f} shows that

$$\begin{aligned} f_i &= \frac{1}{4\pi} \sum_{j=1,2,3} \partial_j (E_i E_j + B_i B_j) - (1/2) \partial_i (E^2 + B^2) - c^{-2} \partial_t S_i \\ &= - \sum_{j=1,2,3} \partial_j T_{ij} - c^{-2} \partial_t S_i, \end{aligned} \quad (4.119)$$

where $\partial_i = \partial/\partial x_i$ means the partial derivative wrt. the x_i coordinate and S_i denotes the i component of the Poynting vector. One can then verify that the Maxwell pressure tensor is indeed given by eq. (4.116).

Using the energy-momentum tensor (eq. 4.117), we can then write the conservation laws for energy and momentum as follows

$$T_{,\mu}^{\mu\nu} = \sum_{\mu} \frac{\partial T^{\mu\nu}}{\partial x^{\mu}} = -f^{\nu}, \quad (f^{\nu}) = (\mathbf{J} \cdot \mathbf{E}/c, \rho \mathbf{E} + c^{-1} \mathbf{J} \times \mathbf{B})^T, \quad (4.120)$$

where (f^{ν}) is called the *4-Lorentz force density*, $(x^{\mu}) = (ct, x, y, z)$ denotes the spacetime coordinates, and the notation of $T_{,\mu}^{\mu\nu}$ (4-dimensional divergence) contains the Einstein

summation rule for index μ . Another way of writing the 4-Lorentz force density is to use the field strength tensor ($F^{\mu\nu}$) (eq. 3.112),

$$f^\nu = F_\mu^\nu J^\mu / c, \quad (4.121)$$

where $(J^\mu) = qn_0 (u^\mu) = \rho(c, \mathbf{v})$ is the 4-current density (n_0 being the number density in the comoving frame and ρ being the charge density in the lab frame), and $F_\mu^\nu = F^{\nu\beta} \eta_{\beta\mu}$ is obtained by lowering one of the indices in the field strength tensor using the inverse matrix of the Minkowski metric.

4.2.9 *Abraham-Lorentz force for radiation reaction

In the above discussion, we have ignored the effects of the radiation energy loss from an accelerating particle on its equation of motion. In the momentary rest frame of the particle, we consider a constant acceleration \dot{v} on a dynamical timescale of τ and this gives rise to a kinetic energy of $E_k = m(\dot{v}\tau)^2/2$. Meanwhile, the amount of energy radiated during this time interval is $\Delta E = P\tau = 2q^2\dot{v}^2\tau/(3c^3)$. The ratio between the radiated energy and the kinetic energy is

$$\frac{\Delta E}{E_k} = \frac{P\tau}{m(\dot{v}\tau)^2/2} = \frac{4r_e/c}{3\tau} \sim \frac{10^{-23} \text{ sec}}{\tau}, \quad (4.122)$$

where we have taken $q = e$ and $m = m_e$ for an electron. We see that as long as the dynamical timescale τ is much longer than the light-crossing time of the electron classical radius $r_e/c \sim 10^{-23}$ sec, the effects of radiation loss on the dynamics of the particle are weak. The radiation reaction force can be included in a perturbative way as given by the *Abraham-Lorentz* formula (for the case of an electron)

$$\mathbf{F}_{\text{ext}} = m \left(\dot{\mathbf{v}} - \frac{2r_e}{3c} \ddot{\mathbf{v}} \right), \quad (4.123)$$

where \mathbf{F}_{ext} is the external force (e.g., the Lorentz force, gravity) on the particle, \mathbf{v} is the acceleration, and $\ddot{\mathbf{v}}$ is the changing rate of the acceleration. The above formula only applies to Newtonian speeds⁷ and is derived by requiring energy conservation on over a sufficiently long time interval between t_1 and t_2 .

Let us consider the interaction between a radiation pulse and a particle initially at rest. The amplitude of the wavepacket slowly ramps up and then ramps down on timescales much longer than the typical period of the incident waves, so the particle comes back to rest after the passage of the pulse — quantum mechanically, the particle is at the same energy level before and after the interaction. The total work done by the external force during

⁷This is not a strong restriction as we can always go to the momentary rest frame of the particle.

the interaction must equal to the amount of energy radiated away by the particle,

$$\begin{aligned} \int_{t_1}^{t_2} \mathbf{F}_{\text{ext}} \cdot \mathbf{v} dt &= \int_{t_1}^{t_2} \frac{2q^2}{3c^3} \dot{\mathbf{v}} \cdot \dot{\mathbf{v}} dt = \frac{2q^2}{3c^3} \int_{t_1}^{t_2} \dot{\mathbf{v}} \cdot d\mathbf{v} \\ &= \frac{2q^2}{3c^3} \left[(\dot{\mathbf{v}} \cdot \mathbf{v})|_{t_1}^{t_2} - \int_{t_1}^{t_2} \ddot{\mathbf{v}} \cdot \mathbf{v} dt \right], \end{aligned} \quad (4.124)$$

where we have done integral by parts. We then notice that the boundary term vanishes because $\dot{\mathbf{v}} = 0$ before and after the pulse in the absence of other forces. We then obtain

$$\int_{t_1}^{t_2} \left(\mathbf{F}_{\text{ext}} + \frac{2q^2}{3c^3} \ddot{\mathbf{v}} \right) \cdot \mathbf{v} dt = 0. \quad (4.125)$$

Thus, energy conservation will be satisfied if we take the radiation reaction force to be

$$\mathbf{F}_{\text{rad}} = \frac{2q^2}{3c^3} \ddot{\mathbf{v}}, \quad (4.126)$$

which then leads to eq. (4.123).

4.3 Thomson scattering

Under the action of the incoming EM waves, a charge is set in motion. This motion then produces radiation in all directions — this is called the scattered waves. The observer sees the superposition of the original and scattered waves.

4.3.1 Linearly polarized incident waves

Here we would like to calculate the differential cross-section under the *Born approximation* — taking the incident field as the total field that drives the motion of the charge, which is valid as long as the scattered field is much weaker than the incident field.

Consider an electron that is driven to oscillate in a (low-amplitude) monochromatic linearly polarized plane EM wave,

$$\mathbf{E}(z, t) = E_0 \hat{x} \cos(kz - \omega t). \quad (4.127)$$

For an electron located at $z = 0$, its acceleration is given by

$$\dot{\mathbf{v}} = \frac{qE_0}{m} \hat{x} \cos \omega t. \quad (4.128)$$

From eq. (4.17), we obtain the time-averaged power per solid angle

$$\left\langle \frac{dP}{d\Omega} \right\rangle = \frac{q^4 E_0^2}{8\pi m^2 c^3} \sin^2 \Theta, \quad (4.129)$$

where $\langle \dots \rangle$ stands for time average, $\langle \cos^2 \omega t \rangle$ gives a factor of $1/2$, and Θ is the angle between the line of sight and $\hat{\mathbf{x}}$. The scattered radiation is linearly polarized in the plane that contains the line of sight and the acceleration vector (or the E-field of the linearly polarized incident wave).

Since the time-averaged incident flux is $\langle S \rangle = E_0^2 c / (8\pi)$, the differential cross-section for linearly polarized incident waves is defined as

$$\left(\frac{d\sigma}{d\Omega} \right)_{\text{lin}} = \frac{\langle dP/d\Omega \rangle}{\langle S \rangle} = r_e^2 \sin^2 \Theta, \quad (4.130)$$

where r_e is the *classical electron radius* as defined as

$$r_e = \frac{q^2}{mc^2} \approx 2.82 \times 10^{-13} \text{ cm} \quad (4.131)$$

The above definition can be understood as the Coulomb potential energy q^2/r_e being comparable to the rest-mass energy of the electron. The angle-integral gives the total *Thomson cross-section*

$$\sigma_T = \int \frac{d\sigma}{d\Omega} d\Omega = \frac{8\pi}{3} r_e^2 \approx 6.65 \times 10^{-25} \text{ cm}^2, \quad (4.132)$$

where we have used $\int \sin^2 \Theta d\Omega = 2\pi \int_0^\pi \sin^3 \Theta d\Theta = 8\pi/3$. Note that massive charge particles, e.g. protons, produces much less scattering as it is suppressed by a factor of m^{-2} . For a fully ionized gas with Hydrogen mass fraction X and Helium mass fraction $Y = 1 - X$, the Thomson scattering opacity is given by

$$\kappa_T = \left(X + \frac{Y}{2} \right) \frac{\sigma_T}{m_p} \approx 0.2(1+X) \text{ cm}^2 \text{ g}^{-1}, \quad (4.133)$$

where m_p is the proton mass.

4.3.2 Arbitrarily polarized incident waves

Consider incident waves consisting of a superposition of N polarization states (denoted by index $a = 1, 2, \dots, N$), each described by the components of their Jones vectors

$$E_j^{(a)}(t) = E_{j,0}^{(a)} \cos(\omega^{(a)} t + \phi_j^{(a)}), \quad j = x, y. \quad (4.134)$$

We assume that the wave modes are mutually incoherent — their phases $\phi_{x,y}^{(a)}$ are uncorrelated and they are at different frequencies⁸ $\omega^{(a)}$. All wave modes are propagating along $\hat{\mathbf{z}}$.

⁸If all wave modes have identical frequency $\omega^{(a)} = \omega$, then one can show that the sum is still elliptically polarized $E_{j,0}^{(a)} \cos(\omega t - \phi_j^{(a)}) = A \cos(\omega t + \phi)$, where A and ϕ are constants.

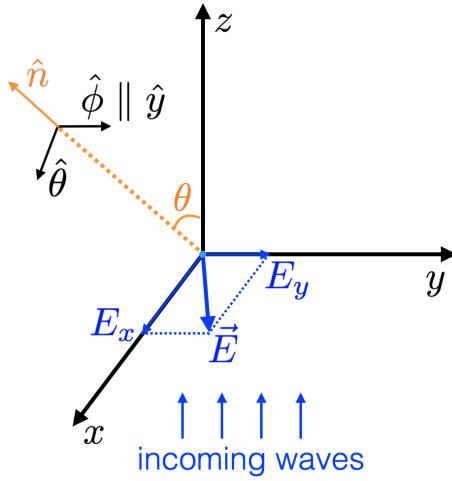


Figure 4.13: Geometry for Thomson scattering of arbitrarily polarized waves.

The observer is located in the direction \hat{n} that is in the x - z plane at an angle θ from the z -axis, i.e.,

$$\hat{n} = \sin \theta \hat{x} + \cos \theta \hat{z}. \quad (4.135)$$

The geometry is shown in Fig. 4.13. An electron is driven into oscillation by the superposition of all wave modes (this is only valid if the incident wave amplitudes are small), so the acceleration along the j -direction is

$$\dot{v}_j(t) = (q/m) \sum_a E_j^{(a)}(t). \quad (4.136)$$

For $j = x$, the angle between \hat{n} and \hat{x} is $\pi/2 - \theta$, and the contribution to the scattered radiation is polarized in the $\hat{\theta}$ direction defined as

$$\hat{\theta} = \hat{y} \times \hat{n} = \cos \theta \hat{x} - \sin \theta \hat{z}. \quad (4.137)$$

For $j = y$, the angle between \hat{n} and \hat{y} is $\pi/2$, and this contributes to scattered radiation that is polarized in the $\hat{\phi} = \hat{y}$ direction. Therefore, we obtain the radiated wave amplitudes along the $\hat{\theta}$ and $\hat{\phi}$ directions (perpendicular to \hat{n})

$$E_{\text{rad},\theta} = \frac{q^2 \cos \theta}{m R c^2} \sum_a E_x^{(a)}(t), \quad E_{\text{rad},\phi} = \frac{q^2}{m R c^2} \sum_a E_y^{(a)}(t). \quad (4.138)$$

Since each individual wave mode (specified by a) is incoherent with the others, we have $\langle (\sum_a E_x^{(a)}(t))^2 \rangle = \sum_a \langle (E_x^{(a)}(t))^2 \rangle$ as the cross terms do not contribute. Therefore, the

time-averaged scattered power per solid angle along $\hat{\mathbf{n}}$ is given by (using eq. 4.17)

$$\begin{aligned}\left\langle \frac{dP}{d\Omega} \right\rangle &= \frac{R^2 c}{4\pi} (\langle E_{\text{rad},\theta}^2 \rangle + \langle E_{\text{rad},\phi}^2 \rangle) \\ &= \frac{q^4}{4\pi m^2 c^3} \sum_a \left[\left\langle (E_x^{(a)}(t))^2 \right\rangle \cos^2 \theta + \left\langle (E_y^{(a)}(t))^2 \right\rangle \right].\end{aligned}\quad (4.139)$$

The time-averaged incident flux is

$$\langle S \rangle = \frac{c}{4\pi} \sum_a \left[\left\langle (E_x^{(a)}(t))^2 \right\rangle + \left\langle (E_y^{(a)}(t))^2 \right\rangle \right]. \quad (4.140)$$

Finally, the differential scattering cross-section is given by $d\sigma/d\Omega = \langle dP/d\Omega \rangle / \langle S \rangle$. For the special cases of a circularly polarized incident wave *or* unpolarized incident waves, we have $\sum_a \langle (E_x^{(a)}(t))^2 \rangle = \sum_a \langle (E_y^{(a)}(t))^2 \rangle$, so the differential cross-section is

$$\left(\frac{d\sigma}{d\Omega} \right)_{\text{cir/unpol}} = \frac{1}{2} r_e^2 (1 + \cos^2 \theta). \quad (4.141)$$

The consequence of the above differential cross-section is that the scattered waves are generally polarized. This is because the power contained in the two orthogonal components $\langle E_{\text{rad},\theta}^2 \rangle$ and $\langle E_{\text{rad},\phi}^2 \rangle$ are in general different and their ratio depends on the scattering angle θ between the wavevector and the line of sight. For the case of unpolarized incident waves, we obtain

$$\langle E_{\text{rad},\theta}^2 \rangle / \langle E_{\text{rad},\phi}^2 \rangle = \cos^2 \theta. \quad (4.142)$$

From the Stokes parameters, we then find the scattered waves to be partially linearly polarized and the polarization fraction to be

$$\Pi = \frac{\sqrt{Q^2 + U^2 + V^2}}{I} = \frac{1 - \cos^2 \theta}{1 + \cos^2 \theta}, \quad (4.143)$$

where $I = \langle E_{\text{rad},\theta}^2 \rangle + \langle E_{\text{rad},\phi}^2 \rangle$ and $Q = \langle E_{\text{rad},\theta}^2 \rangle - \langle E_{\text{rad},\phi}^2 \rangle$ are non-zero but $U = V = 0$ because there is no temporal correlation between $E_{\text{rad},\theta}$ and $E_{\text{rad},\phi}$. Note that, if the observer's (orthogonal) antenna arms are not perfectly oriented along the $\hat{\theta}$ and $\hat{\phi}$ directions, one would still get $V = 0$, but the other three Stokes parameters, I, Q, U , are non-zero. However, the polarization fraction is still given by eq. (4.143). When the observer's line of sight is perpendicular to the wavevector of the incident wave (i.e., for a scattering angle of $\theta = \pi/2$), the scattered waves are 100% linearly polarized in the $\hat{\phi}$ direction. We conclude that electron scatterings produce linear polarization in the direction that is *perpendicular* to the plane that contains the observer, the scatterer (=the

electron here), and the source. The same is true for Rayleigh scattering, which explains the polarization of the Sun-lit and Moon-lit sky (see §4.4.3).

Integrating the above differential cross-section $(d\sigma/d\Omega)_{\text{cir/unpol}}$ (eq. 4.141) over all θ , we then find that the total cross-section is equal to the Thomson value $\sigma_T = 8\pi r_e^2/3$. More generally, one can show that the total cross-section for arbitrary polarization states (correlated or not) is equal to σ_T . The argument goes as follows. Consider an elliptically polarized incident wave with two orthogonal E-field amplitudes $E_{x,0}$ and $E_{y,0}$, which correspond to the amplitudes of oscillating acceleration components

$$\mathbf{a}_{x,0} = (q/m)E_{x,0}\hat{\mathbf{x}}, \quad \mathbf{a}_{y,0} = (q/m)E_{y,0}\hat{\mathbf{y}}. \quad (4.144)$$

These two acceleration components lead to oscillating radiation E-fields with total amplitude (using eq. 4.14)

$$\mathbf{E}_{\text{rad},0} = \frac{q}{Rc^2} \{a_{x,0} [\hat{\mathbf{x}} - (\hat{\mathbf{x}} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}] + a_{y,0} [\hat{\mathbf{y}} - (\hat{\mathbf{y}} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}]\}, \quad (4.145)$$

and the time-averaged emitting power per solid angle along an arbitrary direction $\hat{\mathbf{n}}$ (not necessarily in the x - z plane) is given by

$$\begin{aligned} \left\langle \frac{dP}{d\Omega} \right\rangle &= \frac{1}{2} \frac{\mathbf{E}_{\text{rad},0}^2 c}{4\pi} R^2 = \frac{q^2}{8\pi c^3} \{a_{x,0} [\hat{\mathbf{x}} - (\hat{\mathbf{x}} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}] + a_{y,0} [\hat{\mathbf{y}} - (\hat{\mathbf{y}} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}]\}^2 \\ &= \frac{q^2}{8\pi c^3} \left\{ a_{x,0}^2 [\hat{\mathbf{x}} - (\hat{\mathbf{x}} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}]^2 + a_{y,0}^2 [\hat{\mathbf{y}} - (\hat{\mathbf{y}} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}]^2 - 2a_{x,0}a_{y,0}(\hat{\mathbf{x}} \cdot \hat{\mathbf{n}})(\hat{\mathbf{y}} \cdot \hat{\mathbf{n}}) \right\}, \end{aligned} \quad (4.146)$$

where the factor of 1/2 comes from time averaging of the sinusoidal oscillation. To account for the contributions from acceleration components $\mathbf{a}_{x,0}$ and $\mathbf{a}_{y,0}$, we need to first add up the two corresponding radiation E-field amplitudes and then take the square of the total amplitude $\mathbf{E}_{\text{rad},0}$. The total emitting power not only has the incoherent sum of the two contributions $a_{j,0}^2 [\hat{\mathbf{e}}_j - (\hat{\mathbf{e}}_j \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}]^2$ (for $j = x, y$) but also includes a cross term of $-2a_{x,0}a_{y,0}(\hat{\mathbf{x}} \cdot \hat{\mathbf{n}})(\hat{\mathbf{y}} \cdot \hat{\mathbf{n}})$. Fortunately, when we integrate over all viewing angles, the cross-term does not contribute to the angle-integrated total emitting power, because

$$\int d\Omega (\hat{\mathbf{x}} \cdot \hat{\mathbf{n}})(\hat{\mathbf{y}} \cdot \hat{\mathbf{n}}) = 0, \quad (4.147)$$

which can be easily verified by writing out $\hat{\mathbf{n}}$ with its Cartesian components $\hat{\mathbf{n}} = \sin\theta \cos\phi \hat{\mathbf{x}} + \sin\theta \sin\phi \hat{\mathbf{y}} + \cos\theta \hat{\mathbf{z}}$, for polar angle θ and azimuthal angle ϕ . Therefore, we find that the angle-integrated emitting power is given by the incoherent sum of the contributions from two orthogonal acceleration components, i.e.,

$$\langle P \rangle = \frac{q^2}{8\pi c^3} (a_{x,0}^2 + a_{y,0}^2) \int d\Omega [\hat{\mathbf{x}} - (\hat{\mathbf{x}} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}]^2 = \frac{q^2}{3c^3} (a_{x,0}^2 + a_{y,0}^2), \quad (4.148)$$

where we have used $\int d\Omega [\hat{\mathbf{x}} - (\hat{\mathbf{x}} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}]^2 = \int d\Omega [\hat{\mathbf{y}} - (\hat{\mathbf{y}} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}]^2 = \int d\Omega [\hat{\mathbf{z}} - (\hat{\mathbf{z}} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}]^2 = \int \sin^2 \theta d\Omega = 8\pi/3$ due to symmetry. We then plug in the two acceleration components and obtain

$$\langle P \rangle = \frac{r_e^2 c}{3} (E_{x,0}^2 + E_{y,0}^2) = \frac{8\pi r_e^2}{3} \langle S \rangle \Rightarrow \sigma_{\text{tot}} = \frac{\langle P \rangle}{\langle S \rangle} = \sigma_T, \quad (4.149)$$

where the Poynting flux of the incident wave is $\langle S \rangle = (c/8\pi)(E_{x,0}^2 + E_{y,0}^2)$. We conclude that the total scattering cross-section is given by the incoherent sum of the scattering cross-sections of the two orthogonal components \mathbf{E}_x and \mathbf{E}_y .

Finally, we note that the total cross-section σ_T only applies to low-frequency ($h\nu \ll mc^2$), low-amplitude $E_0 q / (m\omega c^2) \ll 1$ waves interacting with a single particle. At sufficiently high wave frequencies ($h\nu \gtrsim mc^2$), photon momentum impact on the electron is no longer negligible and one must take into account the recoil effect which causes the scattered photons to have lower frequencies than the incident ones. Compton scattering will be discussed in Ch. 7. When the amplitude of the incident waves is large $E_0 q / (m\omega c^2) \gtrsim 1$, the electron undergoes relativistic longitudinal and transverse motions (see §4.2.5). To obtain the scattering cross-section, one can first calculate the electron's trajectory (ignoring radiation losses) and then use the Larmor formula to obtain the emitted power along the known trajectory. Finally, when there are N electrons in a small region of transverse size $\lesssim \lambda/\theta$ (θ being the deflection angle between the incident and scattered beams), we are in the *coherent scattering regime* as total amplitude of the scattered waves by these electrons add up coherently as N times that from each electron (due to negligible phase difference), and therefore the total differential cross-section is $N^2 d\sigma/d\Omega$.

4.3.3 Plasma dispersion by a thin sheet

Consider a monochromatic linearly polarized plane wave $E_0 \hat{\mathbf{x}} e^{i(kz - \omega t)}$ that encounters a thin layer of ionized hydrogen gas at $z = 0$, as shown in Fig. 4.14. The plasma layer is located inside the x - y plane and has a spatially uniform free electron column density $N_{\text{col}} \ll \sigma_T^{-1}$ such that multiple scattering can be ignored. The observer is located at a distance z_0 from the plasma layer. Our goal is to calculate the observed total E-field, which is a superposition of the incident and scattered waves.

The electrons in a small patch of area $dx dy$ near position $(x, y, z = 0)$ have acceleration

$$\dot{\mathbf{v}} = \frac{qE_0 \hat{\mathbf{x}}}{m} e^{-i\omega t_{\text{ret}}}, \quad (4.150)$$

where the retarded time is given by

$$t_{\text{ret}} = t - R/c, \quad (4.151)$$

t is the observer's time and $R = \sqrt{x^2 + y^2 + L^2}$ is the distance to the observer. It will be clear later that the contribution to the scattered waves mainly comes from the circular

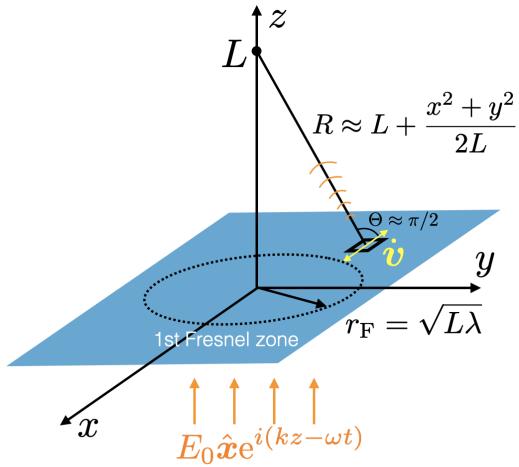


Figure 4.14: Scattering of an incident plane wave (propagating along \hat{z}) by a thin plasma sheet located in the x - y plane. Each small patch of plasma column emits secondary waves. Most contribution to the total amplitude of the scattered waves comes from the first Fresnel zone shown by a dotted circle.

region of radius of the order

$$r_F = \sqrt{L\lambda}, \quad (4.152)$$

which is called the *Fresnel scale* and $\lambda = 2\pi c/\omega$ is the wavelength. According to the above definition for the Fresnel scale (other textbooks may have slightly different definitions), a ray path at a transverse separation $\sqrt{x^2 + y^2} = r_F$ from the center of the screen ($z = 0$) has a geometric phase delay of π .

We consider the limit $L \gg \lambda$ — the observer is in the “far zone”. Thus, $r_F \ll L$ and we have the following approximation

$$R \approx L + \frac{x^2 + y^2}{2L}. \quad (4.153)$$

In this limit, the acceleration vector $\dot{\mathbf{v}}$ is nearly perpendicular to the line of sight $\hat{\mathbf{R}}$, i.e. $\Theta \approx \pi/2$. Therefore, the contribution to the scattered waves at the observer’s position at time t by the $N_{\text{col}} dx dy$ electrons is given by (using eq. 4.14)

$$d\mathbf{E}_{\text{rad}}(t) = -\frac{qN_{\text{col}}\dot{\mathbf{v}}(t_{\text{ret}})}{Rc^2} dx dy \approx -\frac{q^2 E_0 N_{\text{col}} \hat{\mathbf{x}}}{mc^2 L} e^{i\phi_L} \exp\left(ik \frac{x^2 + y^2}{2L}\right) dx dy \quad (4.154)$$

where $\phi_L = kL - \omega t$ is the phase of the original wave at the observer’s location. An integral over the entire x - y plane gives the total amplitude of the scattered waves. We make use of

the following integral ([proof](#))

$$\int_{-\infty}^{\infty} e^{i\xi^2} d\xi = \sqrt{i\pi}, \quad (4.155)$$

which is related to the Fresnel integrals $S(x) = \int_0^x \sin(\xi^2) d\xi$ and $C(x) = \int_0^x \cos(\xi^2) d\xi$ ([see here](#)). The scattered waves from the entire sheet is given by

$$\mathbf{E}_{\text{rad}}(t) = \iint_{\text{xy plane}} d\mathbf{E}_{\text{rad}} \approx -i \frac{2\pi q^2 N_{\text{col}}}{\omega mc} E_0 \hat{\mathbf{x}} e^{i\phi_L}. \quad (4.156)$$

Most of the contribution to the integrated \mathbf{E}_{rad} comes from the region $|x^2 + y^2| \lesssim r_F$, which is called the *first Fresnel zone*. This is because the secondary waves from regions at $x^2 + y^2 \gg r_F^2$ have random, incoherent phases. The total amplitude is given by the superposition of the original ($E_0 \hat{\mathbf{x}} e^{i\phi_L}$) and scattered waves

$$\mathbf{E}_{\text{tot}}(t) \approx E_0 \hat{\mathbf{x}} e^{i\phi_L} (1 - i\Delta\phi), \quad (4.157)$$

where the phase delay is

$$\Delta\phi = \frac{2\pi q^2 N_{\text{col}}}{m\omega c}. \quad (4.158)$$

For a plasma slab with a very small column density such that $\Delta\phi \ll 1$, we have $1 - i\Delta\phi \approx e^{-i\Delta\phi}$, and hence

$$\mathbf{E}_{\text{tot}}(z, t) \approx E_0 \hat{\mathbf{x}} e^{i(kz - \omega t)} \cdot e^{-i\Delta\phi(\omega)}. \quad (4.159)$$

Therefore, the main effect of the plasma slab is to *delay* the phase of the original wave by $\Delta\phi$ (eq. 4.158). In fact, one can always divide all the plasma along the line of sight into a large number of thin slices, so N_{col} can be understood as the *total* column density and the corresponding $\Delta\phi$ is the *total* phase delay.

The above consideration is for a monochromatic wave, and it can be generalized to incident waves with an arbitrary Fourier spectrum $\tilde{E}(\omega) = (2\pi)^{-1} \int E(t) e^{i\omega t} dt$. For each Fourier component, the plasma slab produces a (frequency-dependent) phase delay $\tilde{E}(\omega) \rightarrow \tilde{E}(\omega) e^{-i\Delta\phi(\omega)}$, where $\Delta\phi(\omega)$ is given by eq. (4.158). Thus, the observed waveform at position $z = L$ is given by (dropping the $\hat{\mathbf{x}}$ unit vector in the following)

$$E(t) = \int \tilde{E}(\omega) e^{-i\Delta\phi(\omega)} e^{-i\omega t} d\omega. \quad (4.160)$$

If we consider the waveform in a narrow frequency channel $\omega_0 - \Delta\omega/2 < \omega < \omega_0 + \Delta\omega/2$, it is appropriate to take the following Taylor expansion

$$\Delta\phi(\omega) = \Delta\phi(\omega_0) + \left. \frac{d\Delta\phi}{d\omega} \right|_{\omega_0} (\omega - \omega_0) + \mathcal{O}[(\omega - \omega_0)^2] \quad (4.161)$$

We then plug the above expression (only up to the linear term) into eq. (4.160) and, after some arrangements, obtain the waveform in the frequency channel of width $\Delta\omega$,

$$E_{\Delta\omega}(t) = e^{-i[\omega_0 t + \Delta\phi(\omega_0)]} \int_{\omega_0 - \Delta\omega/2}^{\omega_0 + \Delta\omega/2} \tilde{E}(\omega) e^{-i(\omega - \omega_0)(t + d\Delta\phi/d\omega|_{\omega_0})} d\omega. \quad (4.162)$$

In practice, as long as the Fourier spectrum of the incident waves, $\tilde{E}(\omega)$, is a smooth function of frequency, we can take $\tilde{E}(\omega) \approx \tilde{E}(\omega_0)$ for a sufficiently narrow frequency channel $\Delta\omega \ll \omega_0$. Then, the integral can be carried out analytically and we obtain

$$E_{\Delta\omega}(t) = \Delta\omega \tilde{E}(\omega_0) \frac{\sin \xi}{\xi} e^{-i[\omega_0 t + \Delta\phi(\omega_0)]}, \quad \xi = \left(t + \left. \frac{d\Delta\phi}{d\omega} \right|_{\omega_0} \right) \Delta\omega/2, \quad (4.163)$$

where the factor of $(\sin \xi)/\xi$ is the sinc function $\text{sinc } \xi$. The observed waveform in the frequency channel between $\omega_0 \pm \Delta\omega/2$ reaches the maximum amplitude when $\xi = 0$, and this gives the arrival time of the wavepacket near frequency ω_0 ,

$$t_a(\omega_0) = - \left. \frac{d\Delta\phi}{d\omega} \right|_{\omega_0} = \frac{2\pi q^2 N_{\text{col}}}{m\omega^2 c}. \quad (4.164)$$

We see that the arrival time for the signal in a narrow frequency channel is equal to the time for stationary phase across the channel.

In radio astronomy, we usually call the column density of free electrons, N_{col} , the *dispersion measure*

$$\text{DM} \equiv N_{\text{col}}, \quad (4.165)$$

because it gives rise to an easily measurable *frequency-dependent time delay* for a short-duration radio pulse

$$t_a(\omega) - t_a(\infty) = \frac{2\pi q^2}{m\omega^2 c} \text{DM} = (4.15 \text{ ms}) \frac{\text{DM}}{\text{pc cm}^{-3}} \nu_{\text{GHz}}^{-2}, \quad (4.166)$$

where $\nu_{\text{GHz}} = \nu/\text{GHz}$ (for radio observations). Wave dispersion will be discussed in more detail in Chapter 5.

4.3.4 *Connection to the Optical Theorem

In this subsection, we discuss an interesting application of the Abraham-Lorentz formula for the radiation reaction force in the context of Thomson scattering. Let us consider a plane wave described by $E_0 \hat{x} e^{i(kz - \omega t)}$ interacting with an electron at position $z = 0$. We would like to calculate the scattered wave at the observer's position on the z-axis at $z = L$. We will drop the unit vector \hat{x} because both the incoming and the scattered waves

considered here are both polarized in this direction. Using the Abraham-Lorentz formula (eq. 4.123), we write the equation of motion for the electron

$$E_0 q e^{-i\omega t} = m \left(\dot{v} - \frac{2r_e}{3c} \ddot{v} \right). \quad (4.167)$$

The zeroth-order solution for the acceleration is the usual result

$$\dot{v} \approx \frac{E_0 q}{m} e^{-i\omega t}. \quad (4.168)$$

From the above solution, we obtain the time derivative of the acceleration

$$\ddot{v} \approx -i\omega \frac{E_0 q}{m} e^{-i\omega t}, \quad (4.169)$$

which can then be plugged back to eq. (4.167) to obtain the first-order correction to the acceleration

$$\dot{v} \approx \frac{E_0 q}{m} \left(1 - i\omega \frac{2r_e}{3c} \right) e^{-i\omega t}. \quad (4.170)$$

Thus, the electric field of the scattered wave at the observer's position on the z-axis is given by (after replacing t in the above expression by the retarded time $t_{\text{ret}} = t - L/c$)

$$E_{\text{sca}} = -\frac{q\dot{v}}{Lc^2} = \frac{f_0 E_0 e^{ikL}}{L} e^{-i\omega t}, \quad f_0 = -r_e + ik \frac{2r_e^2}{3}, \quad (4.171)$$

where f_0 is the dimensionless complex amplitude of the scattered waves in the *forward direction* along the z-axis. The complex amplitude (without the trivial $e^{-i\omega t}$) of the total observed electric field is the superposition of incident wave plus scattered wave

$$E_{\text{tot}} = \left(1 + \frac{f_0}{L} \right) E_0 e^{ikL}. \quad (4.172)$$

It is a bit counter-intuitive, but when we consider the total contribution from an entire screen of scatterers inside the xy plane (perpendicular to the wave vector), the real part of f_0 gives rise to a phase delay as discussed in §4.3.3, and the imaginary part of f_0 (due to radiation reaction) corresponds to a reduction in the amplitude of the total observed electric field — and this reduction is due to the fact that the wave energy is scattered into other directions.

The *optical theorem* is a general law of physics, which connects the forward scattering amplitude f_0 to the *total* (absorption+scattering) cross section of the scatterer σ_{tot} by

$$\sigma_{\text{tot}} = \frac{4\pi}{k} \text{Im}(f_0). \quad (4.173)$$

If using the forward scattering amplitude in eq. (4.171), we find that $\sigma_{\text{tot}} = 8\pi r_e^2/3$, which is in agreement with the Thomson scattering cross-section in eq. (4.132). When both absorption and scattering are important in wave propagation through a given medium, the meaning of the “forward scattering amplitude” f_0 in the optical theorem is more general, and it describes the perturbations to the wave complex amplitude due to both absorption and scattering by a given “scatterer” as follows

$$\frac{f_0}{L} \equiv \frac{E_{\text{tot}}}{E_0 e^{ikL}} - 1. \quad (4.174)$$

For the derivation of the optical theorem (eq. 4.173) from energy conservation, we refer to Jackson’s book.

4.3.5 *Scattering in the presence of a strong B-field

In this subsection, we consider the Thomson scattering of EM waves by an electron in a strongly magnetized plasma. Such a consideration is relevant for radiative transfer inside the magnetosphere of neutron stars and white dwarfs where the electron cyclotron frequency $\nu_B = 2.8 \times 10^{14} \text{ Hz}(B/10^8 \text{ G})$ may be comparable to the photon frequency of interest. In this regime, under the combined influence of the incident EM wave and the plasma B-field, the components of the electron’s velocity/acceleration perpendicular to the plasma B-field is strongly reduced. Our goal is to calculate the total scattering cross-section using classical theory. Readers who are interested in the literature (including quantum mechanical calculations) are referred to [Canuto et al. \(1971\)](#), [Blandford & Scharlemann \(1976\)](#), and [Ventura \(1979\)](#).

We restrict our consideration to the limit of $\omega_p \ll \omega \ll m_e c^2/\hbar$ and $\omega_p \ll \omega_B \ll m_e c^2/\hbar$, where ω , $\omega_p = \sqrt{4\pi n e^2/m_e}$ (n = number density of free electrons), $\omega_B = \sqrt{eB/(m_e c)}$ (B = plasma B-field strength) are the EM wave frequency, plasma frequency, and electron cyclotron frequency, respectively. This is because (1) the magnetospheres of neutron stars and white dwarfs typically have $\omega_p \ll \omega_B$ and (2) scattering of low-frequency (e.g., radio) EM waves with $\omega \ll \omega_p \ll \omega_B$ is more complex because the propagation of the scattered waves are strongly affected by plasma effects as the refractive index of the Alfvén mode significantly deviates from unity. We only consider the limit of classical Thomson scattering where the electron motion is non-relativistic, and quantum effects (e.g., Landau quantization, electron spin, and photon momentum) are ignored. The restriction of $\omega_B \ll m_e c^2/\hbar$ corresponds to sub-critical B-field strengths $B \ll B_{\text{QED}} \equiv m_e^2 c^3/(e\hbar) = 4.4 \times 10^{13} \text{ G}$. Readers who are interested in higher energy photons or super-critical B-fields are referred to [Harding & Lai \(2006\)](#) and references therein.

Since we are considering $\omega \gg \omega_p$, the two primary modes of EM waves are called the O-mode and X-mode and their corresponding refractive indices are very close to unity (so they propagate as if in vacuum, see §5.3.5). We consider the plasma to be nearly uniform

such that the variational lengthscale is much longer than the wavelength of the EM wave. The O-mode has electric vector inside the \mathbf{k} - \mathbf{B} plane and perpendicular to \mathbf{k} , and the X-mode has electric vector perpendicular to the \mathbf{k} - \mathbf{B} plane, where \mathbf{k} is the wavevector of the plane EM wave and \mathbf{B} is the plasma B-field.

Let us consider the plasma B-field to be along the z -axis and the wavevector to be inside the x - z plane at an angle θ from the B-field, i.e.,

$$\mathbf{B} = B\hat{\mathbf{z}}, \quad \mathbf{k} = k(\cos\theta\hat{\mathbf{z}} + \sin\theta\hat{\mathbf{x}}), \quad (4.175)$$

where $k \approx \omega/c$ as the refractive indices of both O-mode and X-mode are close to unity. We consider that the incident EM waves contain these two modes

$$\mathbf{E}_O(t, \mathbf{r}) = E_{O,0} (\sin\theta\hat{\mathbf{x}} - \cos\theta\hat{\mathbf{z}}) e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}, \quad (4.176)$$

and

$$\mathbf{E}_X(t, \mathbf{r}) = E_{X,0} \hat{\mathbf{y}} e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}, \quad (4.177)$$

where $E_{O,0}$ and $E_{X,0}$ are the real amplitudes of the two modes.

Under the combined actions of any linearly polarized EM wave with real amplitude \mathbf{E}_0 and the plasma B-field, an electron at position \mathbf{r} is driven into oscillation at velocity (results taken from eqs. 5.71 and 5.72)

$$\mathbf{v}(t, \mathbf{r}) = \mathbf{v}_0 e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}, \quad \mathbf{v}_0 = -\frac{e}{m_e \omega} \left[iE_{0,\parallel}\hat{\mathbf{z}} + \frac{\omega^2}{\omega^2 - \omega_B^2} \left(i\mathbf{E}_{0,\perp} - \frac{\omega_B}{\omega} \mathbf{E}_0 \times \hat{\mathbf{z}} \right) \right], \quad (4.178)$$

where $E_{0,\parallel} = \mathbf{E}_0 \cdot \hat{\mathbf{z}}$ is the component of \mathbf{E}_0 that is parallel to the plasma B-field, and $\mathbf{E}_{0,\perp} = \mathbf{E}_0 - E_{0,\parallel}\hat{\mathbf{z}}$ is the perpendicular component. The above expression was obtained in the non-relativistic limit with $|\mathbf{v}_0| \ll c$, so our conclusions are only valid for low-amplitude incident waves satisfying the following two conditions

$$\frac{eE_{0,\parallel}}{m_e \omega c} \ll 1 \quad \text{and} \quad \frac{eE_{0,\perp}}{m_e \omega_B c} = \frac{E_{0,\perp}}{B} \ll 1. \quad (4.179)$$

Once we find the oscillating velocity $\mathbf{v}(t, \mathbf{r})$, the acceleration is given by $\dot{\mathbf{v}} = -i\omega\mathbf{v}$.

We apply the above result to the O-mode and X-mode and obtain the corresponding acceleration amplitudes

$$\dot{\mathbf{v}}_{O,0} = \frac{ieE_{O,0}}{m_e} \left[-i \sin\theta \hat{\mathbf{z}} + \frac{\omega^2 \cos\theta}{\omega^2 - \omega_B^2} \left(i\hat{\mathbf{x}} - \frac{\omega_B}{\omega} \hat{\mathbf{y}} \right) \right] \quad (4.180)$$

and

$$\dot{\mathbf{v}}_{X,0} = \frac{ieE_{X,0}}{m_e} \frac{\omega^2}{\omega^2 - \omega_B^2} \left(i\hat{\mathbf{y}} - \frac{\omega_B}{\omega} \hat{\mathbf{x}} \right). \quad (4.181)$$

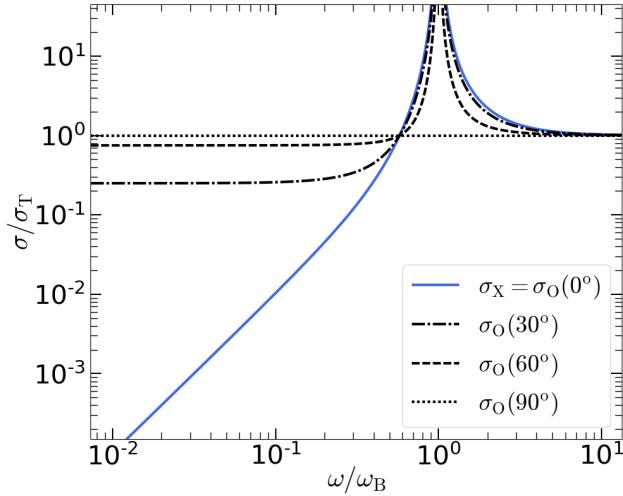


Figure 4.15: Total cross-section (in units of σ_T) for Thomson scattering by a free electron at rest in the presence of a strong B-field \mathbf{B} , as a function of the incident wave frequency ω in units of the cyclotron frequency $\omega_B = eB/(m_ec)$. Two polarization modes for the incident waves are considered: O-mode (σ_O) and X-mode (σ_X). For the O-mode, the dependence on the angle between the wavevector \mathbf{k} and \mathbf{B} is shown by different linestyles.

We then calculate the time-averaged emitting power from each acceleration by adding up the contributions from all orthogonal components incoherently (eq. 4.148)

$$\begin{aligned} \langle P \rangle_O &= \frac{r_e^2 c}{3} E_{O,0}^2 \left[\sin^2 \theta + \frac{\omega^4 \cos^2 \theta}{(\omega^2 - \omega_B^2)^2} \left(1 + \frac{\omega_B^2}{\omega^2} \right) \right] \\ &= \frac{r_e^2 c}{3} E_{O,0}^2 \left[\sin^2 \theta + \frac{\cos^2 \theta}{2} \left(\frac{\omega^2}{(\omega + \omega_B)^2} + \frac{\omega^2}{(\omega - \omega_B)^2} \right) \right], \end{aligned} \quad (4.182)$$

and

$$\langle P \rangle_X = \frac{r_e^2 c}{3} E_{X,0}^2 \frac{\omega^4}{(\omega^2 - \omega_B^2)^2} \left(1 + \frac{\omega_B^2}{\omega^2} \right) = \frac{r_e^2 c}{3} E_{X,0}^2 \left(\frac{\omega^2}{(\omega + \omega_B)^2} + \frac{\omega^2}{(\omega - \omega_B)^2} \right), \quad (4.183)$$

where $r_e = e^2/(m_ec^2)$ is the classical electron radius. For each mode, the incident Poynting flux is $\langle S \rangle = E_0^2 c / (8\pi)$, so we obtain the total scattering cross-sections from the definition $\sigma \equiv \langle P \rangle / \langle S \rangle$, and hence

$$\begin{aligned} \sigma_O/\sigma_T &= \sin^2 \theta + \frac{\cos^2 \theta}{2} \left[\frac{\omega^2}{(\omega + \omega_B)^2} + \frac{\omega^2}{(\omega - \omega_B)^2} \right], \\ \sigma_X/\sigma_T &= \frac{1}{2} \left[\frac{\omega^2}{(\omega + \omega_B)^2} + \frac{\omega^2}{(\omega - \omega_B)^2} \right], \end{aligned} \quad (4.184)$$

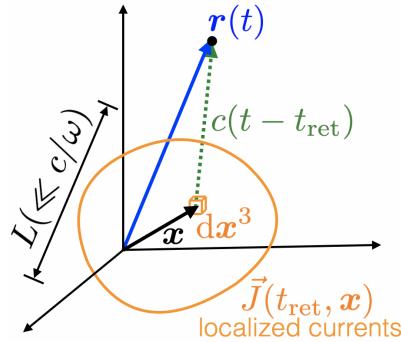


Figure 4.16: Fields at positions far ($|r| \gg L$) from a localized current system with spatial size $L \ll c/\omega$, where ω is the typical angular frequency for the current variations.

where $\sigma_T = 8\pi r_e^2/3$ is the Thomson cross-section.

In the low-frequency limit $\omega \ll \omega_B$, we obtain the following asymptotic results

$$\sigma_O \approx \sigma_T \left[\sin^2 \theta + (\omega/\omega_B)^2 \cos^2 \theta \right], \quad \sigma_X \approx \sigma_T (\omega/\omega_B)^2. \quad (4.185)$$

We find that the scattering cross-section at $\omega \ll \omega_B$ is strongly suppressed by the strong B-field. In the high-frequency limit $\omega \gg \omega_B$, we obtain $\sigma_O \approx \sigma_X \approx \sigma_T$, as long as we are in the Thomson limit of $\hbar\omega \ll m_ec^2$ (otherwise the scattering cross-section is given by the Klein-Nishina result σ_{KN} , see eq. 7.11).

4.4 Multipolar fields

Recall that the source term in the general wave equation for EM fields is proportional to $\partial_t \mathbf{J}$ (see Chapter 1), which means that a *time-dependent* current system will radiate energy to infinity. In this section, we consider localized, slowly varying charge/current distributions that can be analyzed in a perturbative method by multipolar expansion. The emitting power and the angular distribution of the power will be studied.

4.4.1 Spatially confined, slowly varying, non-relativistic current system

Consider a spatially localized, slowly varying current system of size $L \ll r$ (r being the distance to the observer) and variational timescale $t_{\text{var}} \gg L/c$, as shown in Fig. 4.16. If the currents are oscillating at a typical angular frequency $\omega \sim t_{\text{var}}^{-1}$, then the latter condition means $L/\lambda \ll 1$, where $\lambda = 2\pi c/\omega$ is the wavelength of the emission from the changing currents. The above conditions are satisfied by a system of non-relativistic particles that are confined in a sufficiently small volume $\ll \lambda^3$.

Under the above constraints, the retarded potential $\vec{A}(t, \mathbf{r}) = c^{-1} \int d^3x \vec{J}(t_{\text{ret}}, \mathbf{x}) / |\mathbf{r} - \mathbf{x}|$ can be simplified by an expansion series. Note that $\vec{A} = (\phi, \mathbf{A})$ is the 4-potential and that $\vec{J} = (\rho c, \mathbf{J})$ is the 4-current. The first step is to expand

$$\frac{1}{|\mathbf{r} - \mathbf{x}|} \approx \frac{1}{r} \left(1 + \frac{\hat{\mathbf{r}} \cdot \mathbf{x}}{r} + \frac{3(\hat{\mathbf{r}} \cdot \mathbf{x})^2 - x^2}{2r^2} + \mathcal{O}(x^3/r^3) \right). \quad (4.186)$$

In electro- or magneto-statics, we would obtain the static scalar potential

$$\phi_{\text{static}}(\mathbf{r}) = \int d^3x \frac{\rho(\mathbf{x})}{|\mathbf{r} - \mathbf{x}|} = \frac{q}{r} + \frac{\mathbf{d} \cdot \hat{\mathbf{r}}}{r^2} + \frac{\overset{\leftrightarrow}{\mathbf{Q}} \cdot \hat{\mathbf{r}} \hat{\mathbf{r}}}{r^3}, \quad (4.187)$$

where the total charge is $q = \int \rho d^3x$, the dipole moment is

$$\mathbf{d} = \int \rho \mathbf{x} d^3x, \quad (4.188)$$

the electric quadrupole moment tensor⁹ (traceless) is

$$\overset{\leftrightarrow}{\mathbf{Q}} = \frac{1}{2} \int \rho \left(3\mathbf{x}\mathbf{x} - x^2 \overset{\leftrightarrow}{\mathbf{I}} \right) d^3x, \text{ or } Q_{ij} = \frac{1}{2} \int \rho (3x_i x_j - x^2 \delta_{ij}) d^3x, \quad (4.189)$$

and $\overset{\leftrightarrow}{\mathbf{I}} = \text{diag}(1, 1, 1)$ is the identity matrix. The static electric dipole fields are

$$\mathbf{E}_{\text{static,dip}} = \nabla \frac{\mathbf{d} \cdot \hat{\mathbf{r}}}{r^2} = \frac{3(\mathbf{d} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} - \mathbf{d}}{r^3}. \quad (4.190)$$

The static vector potential is given by (retaining the lowest-order term)

$$\begin{aligned} \mathbf{A}_{\text{static}}(\mathbf{r}) &\approx \frac{1}{rc} \int d^3x \left(1 + \frac{\hat{\mathbf{r}} \cdot \mathbf{x}}{r} \right) \mathbf{J} \\ &= \underbrace{\frac{1}{rc} \int \mathbf{x} \partial_t \rho d^3x}_{= 0 \text{ in static limit}} + \underbrace{\frac{\boldsymbol{\mu} \times \hat{\mathbf{r}}}{r^2}}_{\text{magnetic dipole}} + \underbrace{\frac{1}{2r^2c} \int \mathbf{x} (\hat{\mathbf{r}} \cdot \mathbf{x}) \partial_t \rho d^3x}_{= 0 \text{ in static limit}}, \end{aligned} \quad (4.191)$$

where the magnetic dipole moment is

$$\boldsymbol{\mu} = \frac{1}{2c} \int (\mathbf{x} \times \mathbf{J}) d^3x. \quad (4.192)$$

The first term in eq. (4.191) is obtained by noticing $0 = \int \nabla \cdot (x_i \mathbf{J}) d^3x = \int (J_i + x_i \nabla \cdot \mathbf{J}) d^3x$ (using Gauss theorem to convert the volume integral into a surface integral at very large

⁹This definition of the electric quadrupole moment follows Griffith's book but differs from Jackson's book by a factor of 2.

distances) and $\partial_t \rho + \nabla \cdot \mathbf{J} = 0$ (charge conservation). Deriving the third term in eq. (4.191) involves the following integral

$$\int [(\hat{\mathbf{r}} \cdot \mathbf{J}) \mathbf{x} + (\hat{\mathbf{r}} \cdot \mathbf{x}) \mathbf{J}] d^3 \mathbf{x} = \int \mathbf{x} (\hat{\mathbf{r}} \cdot \mathbf{x}) \partial_t \rho d^3 \mathbf{x}, \quad (4.193)$$

which can be shown from the vanishing volume integral of $\nabla \cdot (x_i x_j \mathbf{J}) = x_i x_j (\nabla \cdot \mathbf{J}) + x_i J_j + x_j J_i = -x_i x_j \partial_t \rho + x_i J_j + x_j J_i$.

For the static case, $\partial_t \rho = 0$, so this term does not contribute to $\mathbf{A}_{\text{static}}$ and we obtain the static magnetic dipole fields

$$\mathbf{B}_{\text{static,dip}} = \nabla \times \frac{\boldsymbol{\mu} \times \hat{\mathbf{r}}}{r^2} = \frac{3(\boldsymbol{\mu} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} - \boldsymbol{\mu}}{r^3}, \quad (4.194)$$

which has an identical form as the electric field of an electric dipole of same amplitude $|\mathbf{d}| = |\boldsymbol{\mu}|$. As we will see later, the consequence of this identical form is that a rotating magnetic dipole $\boldsymbol{\mu}$ will radiate at the same power as an electric dipole of the same amplitude and rotating at the same angular frequency and inclination angle between the dipole moment and the rotation axis. In the following, we focus on a time-dependent system, where the $\partial_t \rho$ term is important and it contributes to radiative dipole fields.

For our purpose of obtaining the radiative power of the system up to the B-dipole or E-quadrupole order, we actually do not even need the $\mathcal{O}(x^2/r^2)$ term in the expansion of $|\mathbf{r} - \mathbf{x}|^{-1}$. Note that *radiative electric quadrupole* potential come from the combination of two time-dependent expansion terms that are different from the *static electric quadrupole* potential of $\mathbf{d} \cdot \hat{\mathbf{r}}/r^2$.

For the time-dependent case, the retarded time can be expanded as follows

$$t_{\text{ret}} = t - c^{-1}|\mathbf{r} - \mathbf{x}| \approx t - r/c + \frac{\hat{\mathbf{r}} \cdot \mathbf{x}}{c} + \mathcal{O}\left(\frac{x^2}{rc}\right), \quad (4.195)$$

where we have only retained the lowest-order expansion term (sufficient for our purpose). The currents are slowly changing, so $\vec{J}(t_{\text{ret}}, \mathbf{x})$ is very close to $\vec{J}_0 \equiv \vec{J}(t - r/c, \mathbf{x})$, i.e. the retarded time for the entire current system is nearly the same, except for some small deviation of the order $x/c \ll \omega^{-1}$. Thus, the key step is to expand the time-dependent 4-current density as

$$\vec{J}(t_{\text{ret}}, \mathbf{x}) \approx \vec{J}_0 + \frac{\hat{\mathbf{r}} \cdot \mathbf{x}}{c} \partial_t \vec{J}_0 + \mathcal{O}\left(\frac{\omega^2 x^2 \vec{J}_0}{c^2}\right), \quad (4.196)$$

where $\vec{J}_0 = \vec{J}(t - r/c, \mathbf{x})$.

Then, we can write the vector potential as

$$\begin{aligned}\mathbf{A}(t, \mathbf{r}) &= \frac{1}{rc} \int d^3x \left(1 + \frac{\hat{\mathbf{r}} \cdot \mathbf{x}}{r} \right) \left(\mathbf{J} + \frac{\hat{\mathbf{r}} \cdot \mathbf{x}}{c} \partial_t \mathbf{J} \right)_{t-r/c} \\ &= \left[\underbrace{\frac{\dot{\mathbf{d}}}{rc}}_{\text{rad E-dipole}} + \underbrace{\frac{\dot{\boldsymbol{\mu}} \times \hat{\mathbf{r}}}{rc}}_{\text{rad B-dipole}} + \underbrace{\frac{\ddot{\tilde{\mathbf{Q}}} \cdot \hat{\mathbf{r}}}{3rc^2}}_{\text{rad E-quadrupole}} + \underbrace{\frac{\boldsymbol{\mu} \times \hat{\mathbf{r}}}{r^2}}_{\text{static B-dipole}} + \dots \right]_{t-r/c},\end{aligned}\quad (4.197)$$

where we have made use of $\partial_t \approx d/dt$ as particles are assumed to be moving at non-relativistic speeds, and hence

$$\begin{aligned}\int x \partial_t \rho d^3x &\approx \frac{d}{dt} \mathbf{d} = \dot{\mathbf{d}}, \\ \frac{1}{c} \int (\hat{\mathbf{r}} \cdot \mathbf{x}) \partial_t \mathbf{J} d^3x &\approx \frac{d}{dt} \left[\boldsymbol{\mu} \times \hat{\mathbf{r}} + \frac{d}{dt} \frac{1}{2c} \int \mathbf{x} (\hat{\mathbf{r}} \cdot \mathbf{x}) \rho d^3x \right] \approx \dot{\boldsymbol{\mu}} \times \hat{\mathbf{r}} + \frac{\ddot{\tilde{\mathbf{Q}}} \cdot \hat{\mathbf{r}}}{3c},\end{aligned}\quad (4.198)$$

where

$$\tilde{\mathbf{Q}} \equiv \frac{3}{2} \int \rho \mathbf{x} \mathbf{x} d^3x = \overset{\leftrightarrow}{\mathbf{Q}} + \frac{1}{2} \overset{\leftrightarrow}{\mathbf{I}} \int \rho x^2 d^3x \quad (4.199)$$

is similar to the electric quadrupole moment tensor $\overset{\leftrightarrow}{\mathbf{Q}}$ (eq. 4.189) but differs slightly. The radiation electric quadrupole term in the vector potential depends on $\tilde{\mathbf{Q}} \cdot \hat{\mathbf{r}} = \overset{\leftrightarrow}{\mathbf{Q}} \cdot \hat{\mathbf{r}} + (1/2) \hat{\mathbf{r}} \int \rho x^2 d^3x$, and we see that the additional term is along the radial direction $\hat{\mathbf{r}}$. Fortunately, at large distances from the source, the radiation fields \mathbf{B} and \mathbf{E} are both perpendicular to $\hat{\mathbf{r}}$ (since EM waves in vacuum are transverse), this additional term that is proportional to $\hat{\mathbf{r}}$ does not make a difference practically (since only the fields are measurable anyway), so we can simply replace $\tilde{\mathbf{Q}}$ with the electric quadrupole moment $\overset{\leftrightarrow}{\mathbf{Q}}$. The scalar potential can be obtained similarly as the vector potential

$$\begin{aligned}\phi(t, \mathbf{r}) &= \frac{1}{r} \int d^3x \left(1 + \frac{\hat{\mathbf{r}} \cdot \mathbf{x}}{r} \right) \left(\rho + \frac{\hat{\mathbf{r}} \cdot \mathbf{x}}{c} \partial_t \rho + \frac{1}{2} \left(\frac{\hat{\mathbf{r}} \cdot \mathbf{x}}{c} \right)^2 \partial_t^2 \rho \right)_{t-r/c} \\ &= \left[\underbrace{\frac{q}{r}}_{\text{static Coulomb}} + \underbrace{\frac{\hat{\mathbf{r}} \cdot \dot{\mathbf{d}}}{rc}}_{\text{rad E-dipole}} + \underbrace{\frac{\ddot{\tilde{\mathbf{Q}}} \cdot \hat{\mathbf{r}} \hat{\mathbf{r}}}{3rc^2}}_{\text{rad E-quadrupole}} + \underbrace{\frac{\hat{\mathbf{r}} \cdot \dot{\mathbf{d}}}{r^2}}_{\text{static E-dipole}} + \dots \right]_{t-r/c}.\end{aligned}\quad (4.200)$$

To obtain the radiation fields at large distances, since we only need to know $\mathbf{B} = \nabla \times \mathbf{A}$ and then $\mathbf{E} = \mathbf{B} \times \hat{\mathbf{r}}$, without using the scalar potential ϕ . Note that, since \mathbf{A} is obtained by evaluating the RHS in eq. (4.197) at the regarded time at the coordinate origin $t - r/c$,

we obtain $\nabla \times \mathbf{A} \approx -c^{-1}\hat{\mathbf{r}} \times d(\text{RHS})/dt$ at large distances because the derivative should be taken wrt. the retarded time $t - r/c$ which depends on \mathbf{r} . Following this procedure, we finally obtain the *radiation* fields

$$\begin{aligned}\mathbf{B} &= -\frac{\hat{\mathbf{r}} \times \ddot{\mathbf{d}}}{rc^2} + \frac{\hat{\mathbf{r}} \times (\hat{\mathbf{r}} \times \ddot{\boldsymbol{\mu}})}{rc^2} - \frac{\hat{\mathbf{r}} \times (\overset{\leftrightarrow}{\ddot{\mathbf{Q}}} \cdot \hat{\mathbf{r}})}{3rc^3} + \dots, \\ \mathbf{E} &= \underbrace{\frac{\hat{\mathbf{r}} \times (\hat{\mathbf{r}} \times \ddot{\mathbf{d}})}{rc^2}}_{\text{rad E-dipole}} + \underbrace{\frac{\hat{\mathbf{r}} \times \ddot{\boldsymbol{\mu}}}{rc^2}}_{\text{rad B-dipole}} + \underbrace{\frac{\hat{\mathbf{r}} \times [\hat{\mathbf{r}} \times (\overset{\leftrightarrow}{\ddot{\mathbf{Q}}} \cdot \hat{\mathbf{r}})]}{3rc^3}}_{\text{rad E-quadrupole}} + \dots\end{aligned}\quad (4.201)$$

where the terms on the RHS are evaluated at the retarded time $t - r/c$. In the electric quadrupole radiation term above, we have replaced the unfamiliar quantity $\tilde{\mathbf{Q}}$ with the traceless electric quadrupole moment $\overset{\leftrightarrow}{\ddot{\mathbf{Q}}}$.

With the radiation electric/magnetic fields at large radii (i.e. the *radiation zone*), the time-averaged emitting power per solid angle is given by

$$\left\langle \frac{dP}{d\Omega} \right\rangle = \frac{cr^2}{4\pi} \langle E^2 \rangle, \quad (4.202)$$

and the total power can be obtained by integrating over the entire sphere

$$\langle P \rangle = \int_{4\pi} d\Omega \left\langle \frac{dP}{d\Omega} \right\rangle. \quad (4.203)$$

Since the radiation electric quadrupole field depends on the third time derivative of the quadrupole moment, we expect the total emitting power to be proportional to $P_{\text{E-quad}} \propto E^2 \propto \omega^6$, where $\omega \sim t_{\text{var}}^{-1}$ is the typical angular frequency for the temporal variation of the charge distribution in the system. On the other hand, the electric dipole emitting power is proportional to $P_{\text{E-dip}} \propto \omega^4$. For a typical charge distribution, we roughly expect the electric quadrupole emitting power to be a factor of $(L\omega/c)^2 \sim (v/c)^2 \ll 1$ weaker than the electric dipole power, where L is the size of the system and $v \sim \omega L$ is the typical (non-relativistic) speed of the particles. For the same reason, for a typical system of free charges, one often expects the electric dipole moment to be of the order $|\mathbf{d}| \sim qL$ (where q is the total charge of the emitting particles) and the magnetic dipole moment to be of the order $|\boldsymbol{\mu}| \propto qLv/c$, so the power of magnetic dipole emission is weaker than the electric dipole power by a factor of $(v/c)^2$.

In the following, we mainly discuss the electric/magnetic dipole emission.

4.4.2 Electric/Magnetic dipole radiation

In the case that the time variations in \mathbf{d} and $\boldsymbol{\mu}$ are *linearly polarized* and have a Fourier component at frequency ω , these two moments can be described by

$$\mathbf{d}_0 e^{-i\omega t}, \boldsymbol{\mu}_0 e^{-i\omega t}, \quad (4.204)$$

where \mathbf{d}_0 and $\boldsymbol{\mu}_0$ are the real amplitudes (a non-zero phase at $t = 0$ makes no difference to the time-averaged emission properties). Then the radiation electric field is given by

$$\mathbf{E} = \frac{k^2 e^{i(kr - \omega t)}}{r} [-\hat{\mathbf{r}} \times (\hat{\mathbf{r}} \times \mathbf{d}_0) - \hat{\mathbf{r}} \times \boldsymbol{\mu}_0 + \dots], \quad (4.205)$$

where $k = \omega/c$ for waves propagating in vacuum. The time-averaged angular distribution of the emitting power for only electric or magnetic dipole is

$$\left\langle \frac{dP}{d\Omega} \right\rangle = \frac{c}{8\pi} \text{Re}(r^2 E E^*) = \frac{k^4 c}{8\pi} \sin^2 \Theta \times (|d_0|^2 \text{ or } |\mu_0|^2), \quad (4.206)$$

where $\Theta = \text{acos}(\hat{\mathbf{r}} \cdot \hat{\mathbf{d}})$ or $\text{acos}(\hat{\mathbf{r}} \cdot \hat{\boldsymbol{\mu}})$, the time averaging gives a factor of 1/2, and in the second step we have assumed a single dipole moment that is linearly polarized (circular polarization would give a higher radiating power by a factor of 2). For the linearly polarized case, the electric/magnetic dipole emission has a time-averaged total power of

$$\langle P \rangle_{\text{lin}} = \frac{\omega^4}{3c^3} \times (|d_0|^2 \text{ or } |\mu_0|^2), \quad (4.207)$$

where d_0 or μ_0 is the amplitude of the dipole moment oscillation at frequency ω .

For a *linearly polarized* electric dipole with arbitrary time dependence

$$\mathbf{d}(t) = d(t)\hat{\mathbf{x}},$$

the radiative spectrum would in general be broad. Since the radiative E-field at the observer's position is proportional to \ddot{d} , we first carry out the Fourier transform of \ddot{d} by performing integral by parts,

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \ddot{d}(t) e^{i\omega t} dt = -\omega^2 \tilde{d}(\omega), \quad (4.208)$$

where $\tilde{d}_R(\omega) = (2\pi)^{-1} \int_{-\infty}^{\infty} [\text{Re } d(t)] e^{i\omega t} dt$ is the Fourier transform of the real part of the changing dipole $\text{Re } d(t)$. Then, we see that the Fourier transform of the real part of the radiative E-field is given by

$$\tilde{\mathbf{E}}_R(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} [\text{Re } \mathbf{E}(t)] e^{i\omega t} dt = -\omega^2 \tilde{d}_R(\omega) \frac{\hat{\mathbf{r}} \times (\hat{\mathbf{r}} \times \hat{\mathbf{x}})}{rc^2}. \quad (4.209)$$

The specific fluence, energy per area per unit angular frequency for the entire pulse, is given by

$$F_\omega = c|\tilde{E}_R(\omega)|^2 = \frac{\omega^4 |\tilde{d}_R(\omega)|^2}{r^2 c^3} \sin^2 \Theta, \quad (4.210)$$

where $\Theta = \text{acos}(\hat{r} \cdot \hat{x})$ is the angle between the observer's line of sight \hat{r} and the polarization of the dipole moment \hat{x} . Then, the total radiated energy spectrum of the entire pulse along all viewing angles is

$$\frac{dW}{d\omega} = r^2 \int d\Omega F_\omega = \frac{8\pi\omega^4 |\tilde{d}_R(\omega)|^2}{3c^3}. \quad (4.211)$$

The radiated energy spectrum for a time-dependent linearly polarized magnetic dipole $\mu(t) = \mu(t)\hat{x}$ is the same as in the above equation as long as $|\tilde{d}_R(\omega)|$ is replaced by the magnitude of the Fourier transform $\tilde{\mu}_R(\omega) = (2\pi)^{-1} \int_{-\infty}^{\infty} [\text{Re } \mu(t)] e^{i\omega t} dt$.

Although the above discussion apply to a linearly polarized dipole, an arbitrarily polarized dipole can be considered as the superposition of linear components. For instance, circular polarization is the sum of two orthogonal linear polarizations with a phase difference of $\pi/2$. For a system where the tip of the dipole moment vector undergoes circular motion at an angular frequency ω , the time-averaged total emitting power is

$$\langle P \rangle_{\text{cir}} = \frac{2\omega^4}{3c^3} \times (|d_0|^2 \text{ or } |\mu_0|^2). \quad (4.212)$$

This is the emitting power of a rotating electric $\mathbf{d} = d_0(\hat{x} \cos \omega t + \hat{y} \sin \omega t)$ or magnetic dipole $\mu = \mu_0(\hat{x} \cos \omega t + \hat{y} \sin \omega t)$ that remains perpendicular to the rotational axis \hat{z} . The emitting power can be roughly understood in the following picture. For example, let us take a magnetic dipole rotating at an angular frequency ω . The dipolar magnetic field at a distance r is roughly given by $B(r) \sim \mu_0/r^3$. Due to rotation, the magnetic field lines become strongly bent near the *light cylinder radius* $r_{lc} \equiv c/\omega$, because the fields must be moving near the speed of light to co-rotate with dipole moment — this means that co-rotation can not be maintained. Instead, the rotational bending of the magnetic field lines creates an electric field of magnitude $E(r_{lc}) \sim |(\boldsymbol{\omega} \times \mathbf{r}) \times \mathbf{B}| \sim B(r_{lc}) \sim \mu_0/r_{lc}^3$, so the Poynting flux is of the order $S(r_{lc}) \sim E^2(r_{lc})c/(4\pi)$ and the total radiating power is given by $P \sim 4\pi r_{lc}^2 S(r_{lc}) \sim r_{lc}^2 B^2(r_{lc})c \sim \mu_0^2 c/r_{lc}^4$, which reproduces $P \sim \omega^4 \mu_0^2/c^3$. The same argument also applies to a rotating electric dipole.

4.4.3 Rayleigh scattering

Consider a small scatterer (e.g. a small dust grain, a molecule, or an atom) that is made of a simple dielectric material. When placed in a uniform and static E-field \mathbf{E} , the charge distribution inside the scatterer quickly develops an electric dipole moment that is linearly proportional to the external field

$$\mathbf{d} = V\chi_e \mathbf{E}, \quad (4.213)$$

where $V \sim a^3$ is the volume of the scatter, a is the size, and χ_e is the dimensionless electric susceptibility (which is linearly proportional to the density of the material and is related to dielectric constant $\epsilon = 1 + 4\pi\chi_e$). For atoms (or the simplest molecules), it is sometimes more convenient to define a polarizability¹⁰ α_e such that $\mathbf{d} = \alpha_e \mathbf{E}$, where α_e is of the order a few times a_0^3 and $a_0 \equiv \hbar^3/(m_e e^2) = 0.53 \text{ \AA}$ is the Bohr radius.

For an oscillating E-field of an incident linearly polarized EM wave, $\mathbf{E} = \mathbf{E}_0 e^{-i\omega t}$, the static-field limit in eq. (4.213) applies when the characteristic angular frequency ω_0 of the orbital motions of electrons inside the scatterer is much higher than the wave frequency ω . For atomic material, typical orbits of electrons around their nuclei have angular frequency $\omega_0 \sim \alpha c/a_0 \approx 4 \times 10^{16} \text{ rad s}^{-1}$, where $\alpha \approx 1/137$ is the fine structure constant (as the typical orbital speed is $v_0 \sim \alpha c$) and $a_0 = \hbar^2/(m_e e^2) = 0.53 \text{ \AA}$ is the orbital radius according to Bohr's model (see 4.4.5). Thus, the static limit applies as long as (i) the incident wave frequency $\nu \ll 10^{16} \text{ Hz}$, i.e. for the optical/IR frequencies or lower, and (ii) the size of the scatterer a is much less than the wavelength $\lambda = c/\nu$. Note that condition (ii) makes sure that the wave electric field across the scatterer is uniform and coherent.

In the static limit, we obtain an electric dipole moment that oscillates at the same frequency as the incident EM wave

$$\mathbf{d} = V \chi_e \mathbf{E}_0 e^{-i\omega t}. \quad (4.214)$$

Then, the time-averaged scattered power per solid angle is

$$\left\langle \frac{dP}{d\Omega} \right\rangle = \frac{V^2 \chi_e^2}{8\pi c^3} \omega^4 |E_0|^2 \sin^2 \Theta, \quad \Theta = \arccos(\hat{\mathbf{r}} \cdot \hat{\mathbf{d}}), \quad (4.215)$$

which gives the differential cross-section

$$\frac{d\sigma_R}{d\Omega} = \frac{dP/d\Omega}{E_0^2 c / (8\pi)} = \frac{V^2 \chi_e^2 \omega^4}{c^4} \sin^2 \Theta, \quad \text{for } a \ll \lambda. \quad (4.216)$$

The total scattering cross-section is

$$\sigma_R = \int \frac{d\sigma_R}{d\Omega} d\Omega = \frac{8\pi}{3} \frac{V^2 \chi_e^2 \omega^4}{c^4}. \quad (4.217)$$

This is called *Rayleigh scattering*, where the scattering cross-section has the characteristic ω^4 dependence on frequency.

Since the electric susceptibility χ_e is linearly proportional to the density of the *macroscopic* material, we have $\sigma_R \propto V^2 \chi_e^2 \propto (\rho_{gr} V)^2$, where ρ_{gr} is the density of the material and $V \sim a^3$ is the volume of each scatterer. This means that the total scattering cross-section of a dust

¹⁰In the more general case, the polarizability is a tensor with three principle directions — the induced dipole moment \mathbf{d} is only parallel to \mathbf{E} when the E-field is along one of the principle directions. Here, we consider an isotropic scatterer whose polarizability tensor is given by $\alpha_e \times \text{diag}\{1, 1, 1\}$ and α_e is a scalar.



Figure 4.17: Sunset at Malibu, CA. Note how the color of the sky changes with elevation.

grain is proportional to its mass squared, which is valid as long as $a \ll \lambda$. We do note discuss light scattering by dust grains that are larger or comparable to the wavelength — such cases have more complicated differential cross-sections that are generally wavelength-dependent (but the dependence is not as strong as $\sigma \propto \omega^4$).

For a gas that is made of individual polarizable molecules, since their positions are randomly distributed (giving rise to random phase retardations), the amplitude of the scattered waves traveling in directions different from the incident beam are incoherent sum of the waves from each individual molecules. Therefore, the scattering attenuation coefficient along a given beam is $\alpha_R = n\sigma_R$, where n is the number density of air molecules and $\sigma_R \propto \omega^4$ is the scattering cross-section per molecule. The intensity in the beam would then decrease as $I(z) \propto I(0)e^{-\alpha_R z}$, where z is the coordinate along the beam and $I(0)$ is the intensity at the incident boundary $z = 0$.

When we are looking at a patch of the sky not directly towards the Sun, the blue color of the sky is due to air molecules (with $a \sim \text{few \AA} \ll \lambda$) scattering the sunlight away from the original beam into our line of sight (see Fig. 4.17). When the Sun is at a low elevation angle (i.e., close to the horizon), the direct beam of sunlight has to pass through a thick layer of atmosphere (called “a large air mass” by astronomers), so the blue-wavelength light gets more attenuated than red wavelengths and the Sun looks reddened. Routine calibration of the attenuation by the Earth’s atmosphere is carried out at every observatory so as to accurately measure the brightness and color of astronomical objects at different wavelengths (see e.g., [S. Noll et al. 2012](#)). Although the sunlight is unpolarized, the scattered light by air molecules is partially linearly polarized. It can be shown that (see §4.3.2), for an unpolarized point source, the direction of polarization of the scattered

light is perpendicular to the plane that contains the observer, the scatterer, and the source. Readers are referred to [Gal et al. \(2001\)](#) for the polarization patterns of the Sun-lit and Moon-lit skies.

One might want to apply the optical theorem (eq. 4.173) to obtain the cross-section of Rayleigh scattering. The procedure is as follows. First, we calculate the amplitude of the scattered wave in the forward direction (along the z -axis). Taking the polarization of the incident wave \mathbf{E} and hence the induced electric dipole moment \mathbf{d} to be along \hat{x} , we obtain the complex amplitude of the scattered wave at position $z = L$ on the z -axis

$$E_{\text{sca}} = \frac{k^2}{L} d_0 e^{ikL}, \quad (4.218)$$

where $d_0 = V\chi_e E_0$ is the amplitude of the oscillating electric dipole moment and E_0 is the real amplitude of the incident wave. The total wave amplitude is $E_{\text{tot}} = E_0 e^{ikL} + (k^2/L)d_0 e^{ikL}$. Thus, the forward scattering amplitude is given by

$$f_0 = k^2 V \chi_e. \quad (4.219)$$

If χ_e is real, which is the case for an ideal oscillator without damping (or frictional losses), we find that f_0 is real. If we naïvely apply the optical theorem, $\sigma_{\text{tot}} = (4\pi/k)\text{Im}(f_0)$, the result is $\sigma_{\text{tot}} = 0$! This contradicts with our earlier conclusion about the non-zero cross-section for Rayleigh scattering (eq. 4.217). It turns out that our result for the amplitude of the scattered wave E_{sca} is missing a term due to radiative reaction (which is important so as to obey energy conservation). In an earlier treatment of Thomson scattering, we have seen that radiative reaction indeed gives an imaginary part to the forward scattering amplitude (see eq. 4.171). Similarly, when the effects of radiative reaction is included, we obtain

$$f_0 = k^2 V \chi_e + i \frac{2k^5 V^2 \chi_e^2}{3}, \quad (4.220)$$

and this gives the correct cross-section for Rayleigh scattering $\sigma_R = (4\pi/k)\text{Im}(f_0)$.

Since we have already arrived here, it is useful to further discuss the general situation where the electric susceptibility is complex, i.e., $\chi_e = \text{Re}(\chi_e) + i\text{Im}(\chi_e)$. Then, the total (absorption+scattering) cross-section is given by

$$\sigma_{\text{tot}} = \frac{4\pi}{k} \text{Im}(f_0) = 4\pi k V \text{Im}(\chi_e) + \frac{8\pi k^4 V^2}{3} [\text{Re}(\chi_e)]^2, \quad (4.221)$$

where the first term is due to absorption and the second term is due to (Rayleigh) scattering. We see that, in the limit of small scatterers $a \ll \lambda$ and if the electric susceptibility is independent of wavelength, the absorption cross-section scales as $\sigma_{\text{abs}} \propto \omega V$ whereas the scattering cross-section scales as $\sigma_R \propto \omega^4 V^2$. Thus, Rayleigh scattering is only important at the highest frequencies (where ω^4 overwhelms the V^2 term for very small scatterers), whereas the absorption term usually dominates towards longer wavelengths.

4.4.4 Rotating diatomic molecule as an electric dipole

We consider a rigidly rotating diatomic molecule with atomic masses m_1 and m_2 , and the moment of inertia along an axis perpendicular to the line connecting the two nuclei is

$$I = \mu r_{\text{eq}}^2, \quad \mu = \frac{m_1 m_2}{m_1 + m_2}, \quad (4.222)$$

where μ is the reduced mass and r_{eq} is the equilibrium length of the molecular bond (e.g., in the ground vibrational state). The rotational angular momentum around this axis is

$$L = J\hbar, \quad (4.223)$$

where $J = 0, 1, 2, \dots$ is the rotational quantum number. Classically, the angular momentum is given by $L = I\omega$ (ω being the angular frequency) and the rotational energy is $E = I\omega^2/2 = L^2/(2I)$. In the quantum theory, the L^2 operator (which commutes with the Hamiltonian) has eigenvalues of $L^2 = J(J+1)\hbar^2$, so the energy associated with rotational level J is

$$E_J = J(J+1)\hbar^2/(2I). \quad (4.224)$$

Classically, the angular momentum vector \mathbf{L} can have different orientations, which can be described by the projection $L_z = L \cos \theta$, where θ is the angle between \mathbf{L} and a given z -axis. Quantum mechanically, the eigenvalues of the L_z operator are quantized such that $L_z = m_J\hbar$, with $m_J = -J, -(J-1), \dots, J-1, J$. For a molecule in free motion (in the absence of EM fields), there is no preferred direction in the choice of the z -axis, so the rotational energy for a given (J, m_J) state only depends on J , and we say that the J -th energy level has a degeneracy factor of

$$g_J = 2J + 1. \quad (4.225)$$

For a thermal distribution at temperature T , the occupation number η in each distinct quantum state as specified by (J, m_J) is given by the Boltzmann distribution $\eta \propto e^{-E_J/(k_B T)}$. Thus, the number of molecules in the J -th energy level is given by

$$\frac{N_J}{N_0} = (2J+1) e^{-E_J/(k_B T)}, \quad (4.226)$$

where N_0 is the number of molecules in the ground rotational state ($J = 0$).

Now we consider that the molecule undergoes radiative transitions via the emission/absorption of a photon. Since a photon has spin angular momentum \hbar , the change in rotational quantum number must be $\Delta J = \pm 1$. For the $J \rightarrow J-1$ transition, the frequency of the emitted photon is given by

$$\omega = (E_J - E_{J-1}) / \hbar = J\hbar/I, \quad \text{for } J \rightarrow J-1, \quad (4.227)$$

which is equal to the classical rotational frequency L/I for the *upper* energy level. This makes sense in the classical picture, where a rotating dipole emits monochromatic waves at the rotational frequency. This motivates us to estimate the rate of rotational transitions or the Einstein-A coefficient in the classical picture.

Suppose the diatomic molecule has a permanent electric dipole moment d , which is given by the charge distribution inside the molecule. The emitting power of a rotating dipole $\langle P \rangle_{\text{cir}}$ (for circular polarization) is given by eq. (4.212), so one infers the $J \rightarrow J - 1$ transition rate to be

$$A_{J \rightarrow J-1} \simeq \frac{2\omega^4/(3c^3)}{\hbar\omega} d^2 = \frac{2\omega^3}{3\hbar c^3} d^2, \quad \text{for } J \gg 1. \quad (4.228)$$

Our classical result agrees with that from the quantum theory in the limit of $J \gg 1$

$$A_{J \rightarrow J-1} = \frac{4\omega^3}{3\hbar c^3} |d_{J \rightarrow J-1}|^2 = \frac{4\omega^3}{3\hbar c^3} \frac{Jd^2}{2J+1}, \quad \text{for any } J, \quad (4.229)$$

where the factor of $|d_{J \rightarrow J-1}|^2 = Jd^2/(2J+1)$ is the electric dipole matrix element for the $J \rightarrow J - 1$ transition.

For instance, a carbon monoxide (CO) molecule in its ground vibrational state has electric dipole moment of $d = 0.110$ Debye, where $1 \text{ Debye} = 10^{-18} \text{ esu cm}$. The reduced mass of the $^{12}\text{C}^{16}\text{O}$ molecule is $\mu = 12 \times 16 / (12 + 16) \text{ amu} = 1.14 \times 10^{-23} \text{ g}$. The $J = 1 \rightarrow 0$ transition has frequency $\nu_{1 \rightarrow 0} = 115.3 \text{ GHz}$, so we obtain the equilibrium bond length to be $r_{\text{eq}} = 1.13 \text{ \AA}$. The Einstein-A coefficient for this transition is

$$A_{\text{CO}(1 \rightarrow 0)} = 7.2 \times 10^{-8} \text{ s}^{-1} = (2.3 \text{ yr})^{-1}. \quad (4.230)$$

We note that the Einstein-A coefficient is proportional to the ω^3 , so high-frequency transitions occur much faster, provided that the high- J levels are significantly collisionally excited. Let us define a characteristic temperature T_J corresponding to rotational level J by $k_B T_J / 2 \equiv E_J$ (such that the rotational degree of freedom takes an energy of $k_B T_J / 2$). For CO molecules in the ground vibrational state, we obtain

$$T_J \equiv 2E_J/k_B = 11.06 \text{ K} \frac{J(J+1)}{2}. \quad (4.231)$$

Collisional excitation of rotational level J requires a minimum gas temperature $T_{\min} \sim T_J$, so we obtain $T_{\min} \propto J^2$ in the limit $J \gg 1$.

4.4.5 Hydrogen recombination lines as electric dipole radiation

A semi-classical understanding of the hydrogen atom is the Bohr's model. Bohr proposes that the electron is in a non-relativistic circular orbit as confined by the proton's Coulomb

force. The only difference between the electron's orbit and that of a planet around a star is that the angular momentum is quantized

$$L_n = n\hbar, \quad (4.232)$$

where $n = 1, 2, 3, \dots$ is later known to be the principle quantum number. This is because the circumference $2\pi a_n$ (a_n being the radius of the orbit) must be an integer number of the de Broglie wavelength $\hbar/(m_e v_n)$ (where v_n is the orbital speed) so as to make a standing wave. For a given quantum number n , the orbital radius a_n is uniquely determined by the orbital angular momentum $L_n = a_n m_e v_n$ combined with the balance between Coulomb and centrifugal forces $e^2/a_n^2 = m_e v_n^2/a_n$, and we obtain $L_n = \sqrt{m_e e^2 a_n} = n\hbar$, i.e.,

$$a_n = n^2 a_0, \quad a_0 \equiv \frac{\hbar^2}{m_e e^2} = 0.529 \text{ \AA}, \quad (4.233)$$

where a_0 is the Bohr radius. The orbital velocity is

$$v_n = \frac{e^2}{n\hbar} = \frac{\alpha c}{n}, \quad \alpha \equiv \frac{e^2}{\hbar c} \approx \frac{1}{137}, \quad (4.234)$$

where α is the *fine-structure constant*. Since $v_n/c = \alpha/n \ll 1$, we see that the electron's orbit is indeed non-relativistic. The relativistic effects become more important for hydrogen-like atoms with larger nuclear charge numbers $Z \gg 1$, because $v_n/c = \alpha Z/n$.

The force balance in the circular orbit leads to the important result that kinetic energy $T = m_e v_n^2/2$ and the potential energy $V = -e^2/a_n$ satisfy the *virial theorem*, $2T + V = 0$, so the orbital energy is

$$E_n = T + V = \frac{V}{2} = -\frac{e^2}{2a_n} = -\frac{\text{Ry}}{n^2}, \quad (4.235)$$

where Ry is the *Rydberg energy* given by

$$\text{Ry} \equiv \frac{e^2}{2a_0} = \frac{1}{2}\alpha^2 m_e c^2 = 13.6 \text{ eV}. \quad (4.236)$$

Another way of writing the orbital energy is

$$E_n = -\frac{1}{2}m_e v_n^2 = \frac{\alpha^2}{2n} m_e c^2. \quad (4.237)$$

For other hydrogen-like atoms with nuclear charge number Z , we have $E_n = -Z^2 \alpha^2 m_e c^2 / (2n)$, which becomes comparable to $-m_e c^2$ for very heavy nuclei with $Z \gtrsim 100$ (for which relativistic effects must be taken into account).

The $n+1 \rightarrow n$ transition is conventionally denoted as “ $n\alpha$ ” (e.g., the $3 \rightarrow 2$ Balmer line is 2α) corresponds to a photon frequency of

$$\omega_{n\alpha} = \frac{E_{n+1} - E_n}{\hbar} = \frac{\text{Ry}}{\hbar} \frac{2n+1}{n^2(n+1)^2}, \quad (4.238)$$

which reproduces the hydrogen spectral series¹¹ (Lyman, Balmer, Paschen, etc.). The angular frequency of the lower-energy (n) Bohr orbit is given by

$$\Omega_n = \frac{v_n}{a_n} = \frac{\alpha}{n^3} \frac{c}{a_0} = \frac{\text{Ry}}{\hbar} \frac{2}{n^3}. \quad (4.239)$$

We see that $\Omega_n \approx \omega_{n\alpha}$ in the limit $n \gg 1$, meaning that the angular frequency of the orbit is equal to the frequency of the emitted waves, as expected from the emission of a classical dipole oscillating at frequency Ω_n .

Let us take one step further to see if the $n+1 \rightarrow n$ transition rate, as given by the Einstein-A coefficient $A_{n+1 \rightarrow n}$, expected from the classical orbit is in reasonable agreement with that given by the quantum theory. A charge $q = -e$ in a classical, circular orbit with radius a_{n+1} has electric dipole moment $\mathbf{d} = -e\mathbf{a}_{n+1}$ and the emitting power $\langle P \rangle_{\text{cir}} = 2\Omega_{n+1}^4 d^2 / (3c^3)$ is given by eq. (4.212), so one infers the Einstein-A to be

$$A_{n+1 \rightarrow n} \sim \frac{\langle P \rangle_{\text{cir}}}{\hbar\omega_{n\alpha}} = \frac{2\alpha^4}{3} \frac{c}{a_0} \frac{n^2}{(n+1/2)(n+1)^6} \propto \frac{1}{n^5} \text{ for } n \gg 1. \quad (4.240)$$

The factor of α^4 makes the transition rather “slow”, as it occurs in a timescale during which the electron has undergone about $\omega_{n\alpha}/A_{n+1 \rightarrow n} \sim \alpha^{-3}n^2 \sim 10^6 n^2$ orbits. This means that the natural width of the transition, $\Delta\omega = A$, is very narrow — the fractional width is $\Delta\omega/\omega \sim \alpha^3/n^2 \sim 10^{-6}n^{-2}$. However, for hydrogen-like atoms with much larger nuclear charge numbers $Z \gg 1$, the fractional width is much larger (since $A_{n+1 \rightarrow n} \propto Z^4$ and $\omega_{n\alpha} \propto Z^2$) and we obtain

$$\frac{\Delta\omega}{\omega} \sim Z^2 \frac{\alpha^3}{n^2}.$$

For the K α ($n = 1$) transition of Fe with $Z = 26$, we obtain a fractional natural width of the order $\Delta\omega/\omega \sim 10^{-4}$. [Krause & Oliver \(1979\)](#) provided measurements of natural widths of K α lines of elements heavier than Neon.

According to quantum theory for the hydrogen atom, an electron’s orbit is described by a wave function with three quantum numbers (n, ℓ, m), and when averaged over all $n+1 \rightarrow n$

¹¹If one correctly uses the reduced mass $\mu = m_e m_p / (m_e + m_p)$ as in the [reduced 2-body problem](#), the Bohr’s model successfully predicts the wavelengths of hydrogen spectral series to a fractional accuracy of the order $\alpha^2 \sim 10^{-4}$ (the errors due to relativistic effects).

transitions allowed by selection rules with the relevant statistic weights, the asymptotic ($n \gg 1$) Einstein-A coefficient is a factor of 2 less than in eq. (4.240), i.e.

$$A_{n+1 \rightarrow n} \approx \frac{\alpha^4}{3} \frac{c}{a_0} \frac{1}{n^5} \approx 5.3 \times 10^9 \text{ s}^{-1} \frac{1}{n^5}, \quad \text{for } n \gg 1. \quad (4.241)$$

Draine (2011) provided the following fitting formula based on [Wiese et al. \(1966\)](#)

$$A_{n+1 \rightarrow n} \approx 6.13 \times 10^9 \text{ s}^{-1} \frac{1}{(n + 0.7)^5}, \quad (4.242)$$

which is accurate to better than 1% for $n \geq 4$ and only makes an 8% error for $n = 1$.

4.4.6 Hydrogen 21cm line as magnetic dipole radiation

Although the 21cm hyperfine-structure transition ($\nu = 1.420 \text{ GHz}$) of ground state hydrogen atom is purely a quantum mechanical effect due to interactions between the proton's and electron's spins, a classical picture is sometimes helpful in our understanding of this famous spin-flip transition. The two particles can be considered as two tiny current loops each of which has a magnetic dipole moment with dipolar B-fields around them. The magnetic moment associated with spin angular momentum \mathbf{S} is given by

$$\frac{\boldsymbol{\mu}}{\mu_0} = g \frac{\mathbf{S}}{\hbar}, \quad (4.243)$$

where $\mu_0 = e\hbar/(2mc)$ is the magneton¹² (unit for magnetic moment) for particle mass m , g is a “g-factor” of order unity, e is the elementary charge, and \hbar is the reduced Planck constant. An electron has $g = -2.0023 \approx -2$ and $S = \hbar/2$, so its spin magnetic moment

$$\mu_e \approx -\mu_B = -9.27 \times 10^{-21} \text{ erg G}^{-1}, \quad (4.244)$$

where $\mu_B = e\hbar/(2m_e c)$ is called the *Bhor magneton*. A proton has $g = 5.586$ and the same spin angular moment $S = \hbar/2$ as an electron, but, due to its much larger mass m_p , the magnetic moment is much smaller $\mu_p = 1.41 \times 10^{-23} \text{ erg G}^{-1}$. The interactions between these two magnetic moments, $\boldsymbol{\mu}_p$ and $\boldsymbol{\mu}_e$, give rise to the 21cm transition (see [D. Griffiths' notes](#) for details). The classical analog is the interaction between two tiny current loops — a heavier but weaker one (the proton) that does not move or flip, and a lighter but stronger one (the electron) that not only spatially moves around (electron motion in the atom) but also precesses slowly. In this classical picture, the precession of the electron magnetic moment $\boldsymbol{\mu}_e$ in the dipolar B-fields given by the proton magnetic moment leads to changing magnetic dipole moment $\ddot{\boldsymbol{\mu}}_e \neq 0$, and hence there will be magnetic dipole emission from the system.

¹²This definition follows from the fact that a positive charge of $q = e$ undergoing circular motion with an orbital angular momentum \mathbf{L} has magnetic moment of $\boldsymbol{\mu} = \mu_0(\mathbf{L}/\hbar)$, i.e. $g = 1$ for classical orbital motion.

The emitting power by a classical magnetic dipole rotator with angular frequency ω is given by $\langle P \rangle_{\text{cir}} = 2\omega^4\mu_e^2/(3c^3)$ (eq. 4.212), for the case where $\boldsymbol{\mu}_e$ is perpendicular to the rotational axis. A rough estimate of the Einstein-A coefficient would be (see eq. 4.228) $\langle P \rangle_{\text{cir}}/(\hbar\omega)$. The result from quantum theory is a factor of 2 greater than the above estimate,

$$A_{21\text{ cm}} = \frac{4\omega^3}{3\hbar c^3} \mu_B^2 = 2.86 \times 10^{-15} \text{ s}^{-1} \approx (11 \text{ Myr})^{-1}, \quad (4.245)$$

as the magnetic dipole matrix element is $|\mu_e|^2 \approx \mu_B^2$.

4.4.7 Rotating neutron star as a magnetic dipole

The B-fields near a static magnetic dipole moment $\boldsymbol{\mu}$ is given by (see eq. 4.194)

$$\mathbf{B} = \frac{3(\boldsymbol{\mu} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} - \boldsymbol{\mu}}{r^3}. \quad (4.246)$$

At a given radius $r = R$, the dipolar B-field strength at the magnetic equator (where $\hat{\mathbf{r}} \perp \boldsymbol{\mu}$) is

$$B_{\text{eq}} = \mu/R^3 = 10^{12} \text{ G } \mu_{30} R_{10\text{ km}}^3, \quad (4.247)$$

where $R \sim 10 \text{ km}$ is the typical radius of a neutron star. At the magnetic poles (where $\hat{\mathbf{r}} \parallel \boldsymbol{\mu}$), the field is twice stronger $B_{\text{pole}} = 2\mu/R^2$. Noted that the magnetic fields near the surface of a neutron star is most likely much more complicated than a dipole due to complex charge and current distributions in the neutron star interior — there should be higher-order multipoles which may dominate the surface B-field strength (at least in some regions). However, if we go to much larger radii far from the neutron star's surface $r \gg R$, the dipole field is expected to dominate. For this reason, the surface dipole field strength B_{eq} is only a convenient way of describing the magnetic dipole moment μ , whereas the realistic total magnetic field strength near the surface is likely much stronger than B_{eq} .

Suppose the dipole rotates at an angular frequency ω around a spin axis that is misaligned with the magnetic axis by a constant angle χ (called the *magnetic inclination angle*), then the radiative magnetic dipole power in vacuum is

$$L_{\text{B,vacuum}} = \frac{2}{3} \frac{\omega^4 \mu^2}{c^3} \sin^2 \chi, \quad (4.248)$$

where the factor of 2 comes from the fact that a rotating dipole can be considered as two orthogonal, linearly polarized dipoles, and $\sin \chi$ is the projection factor for the variable component of the total dipole moment. This energy loss causes the star to spin down. It should be noted that an aligned rotator ($\chi = 0$) in vacuum does not produce any radiation. In fact, an aligned rotator in vacuum does not produce any electric field in the inertial frame of the center of mass of the dipole (the magnetic field lines stay dipolar to a radius of infinity). However, as we see below, if the dipole is surrounded by a corotating, conducting

plasma (as is the case in stellar/planetary magnetospheres), the plasma develops a non-zero net charge density which produces an electric field. A major consequence of the plasma-induced electric field is that even an aligned rotator still lose energy by driving a plasma-loaded by electromagnetically dominated wind, and hence the system would always lose rotational energy (in contrary to eq. 4.248, which only applies to a dipole in vacuum).

In the following, we consider the simplest case of an *aligned rotator* with $\boldsymbol{\mu}$ parallel to the rotational axis as denoted by the angular frequency vector $\boldsymbol{\omega}$. We use the spherical coordinate (r, θ, ϕ) with polar axis $\hat{\mathbf{z}} \parallel \boldsymbol{\mu} \parallel \boldsymbol{\Omega}$ to describe the system, which is axisymmetric. The misaligned case (with $\chi \neq 0$) will be discussed later.

For a spin angular frequency $\omega = 2\pi/P$ (P being the rotational period), the *light cylinder* radius r_{lc} is where the corotating speed equals to c , i.e.

$$r_{lc} = c/\omega \simeq 4.8 \times 10^9 \text{ cm}(P/\text{s}). \quad (4.249)$$

The light cylinder is important for a plasma-filled magnetosphere, because fluid motion cannot exceed the speed of light. As see below, the magnetic field lines near the light cylinder cannot keep up with the angular frequency of the dipole rotator. Instead, the azimuthal bending of magnetic field lines leads to a toroidal component (in the azimuthal direction along $\hat{\phi}$ in spherical coordinate) which becomes comparable to the poloidal component (inside the plane for a given azimuthal angle ϕ) close to the light cylinder. The toroidal component is physically produced by the current associated with the rotating charge distribution in the magnetosphere.

Let us consider radii much below the light cylinder where rotational effects only weakly perturb the dipolar B-field. The unperturbed dipolar B-field strength at the light cylinder in the equatorial plane is given by

$$B_{lc} = B_{eq}(R/r_{lc})^3, \quad (4.250)$$

where B_{eq} is the B-field strength at the equator of the stellar surface and the r^{-3} scaling comes from the dipolar field geometry. We would like to know which field lines stretch beyond the light cylinder (and hence become *open field lines*) and which field lines stay closed within the light cylinder. For $\boldsymbol{\mu} = \mu \hat{\mathbf{z}}$, the dipolar B-field at position $\mathbf{r} = (r, \theta, \phi)$ in polar coordinates has components

$$B_r = \mathbf{B} \cdot \hat{\mathbf{r}} = \frac{2\mu \cos \theta}{r^3}, \quad B_\theta = \mathbf{B} \cdot \hat{\theta} = \frac{\mu \sin \theta}{r^3}, \quad B_\phi = \mathbf{B} \cdot \hat{\phi} = 0. \quad (4.251)$$

We would like to find the trajectory of a given B-field line in the poloidal plane (for now we are ignoring the toroidal component due to rotation). This is given by the slope of the tangent direction

$$\frac{dr}{r d\theta} = \frac{B_r}{B_\theta} = \frac{2 \cos \theta}{\sin \theta}, \quad (4.252)$$

which can be integrated and the result is $r/\sin^2 \theta = \text{const}$ for a given field line. The field line equation can be written as

$$r = r_{\text{eq}} \sin^2 \theta, \quad (4.253)$$

where r_{eq} is the maximum radius this field line reaches when it intersects the magnetic equatorial plane.

Now, going back to the rotating picture, the field line that intersects the equator at the light cylinder radius $r_{\text{eq}} = r_{\text{lc}}$ (called the *last open field line*) has a magnetic colatitude angle θ_{pc} on the stellar surface, and the angular size of the *polar cap* θ_{pc} is given by

$$\sin^2 \theta_{\text{pc}} = R/r_{\text{lc}}, \text{ or } \theta_{\text{pc}} = \arcsin \sqrt{\omega R/c} = 0.83^\circ (R/10 \text{ km})^{1/2} (P/1 \text{ s})^{-1/2}. \quad (4.254)$$

The polar cap region is important because particles can freely flow along these field lines beyond the light cylinder to infinity — this property makes the plasma near these field lines very active in accelerating particles to high energies and producing observable radiation. The field lines at larger polar angles $\theta > \theta_{\text{pc}}$ on the surface of the star are closed (with both ends anchored on the stellar surface) and charge particles on these field lines cannot escape to infinity.

Another way of obtaining the angular size of the polar cap is to look at the magnetic flux through this region

$$\Phi_B = \int_0^{\theta_{\text{pc}}} B_r(r = R, \theta) 2\pi R^2 \sin \theta d\theta = \frac{2\pi \mu \sin^2 \theta_{\text{pc}}}{R}. \quad (4.255)$$

In the absence of rotation, this magnetic flux (which is proportional to the number of B-field lines) must go through the equatorial plane at radii $r > r_{\text{lc}}$, so we have

$$\Phi_B = \int_{r_{\text{lc}}}^{\infty} B_\theta(r, \theta = \pi/2) 2\pi r dr = 2\pi \mu / r_{\text{lc}}, \quad (4.256)$$

and again we conclude $\sin^2 \theta_{\text{pc}} = R/r_{\text{lc}}$.

The issue of the spin-down power in eq. (4.248) is that it is based on a dipole rotating in vacuum. Realistic neutron star magnetosphere is filled with plasma (in most of the volume) due to charge pulled out from the stellar surface, pair creation, and particles' streaming motion along B-field lines, as first argued by [Goldreich and Julian \(1969\)](#).

In the magnetosphere filled with plasma, the EM force per unit volume (or *force density*) on the plasma is given by

$$\mathbf{f} = \rho \mathbf{E} + \frac{\mathbf{J}_\perp}{c} \times \mathbf{B}, \quad (4.257)$$

where ρ is the charge density, \mathbf{J}_\perp is the current density perpendicular to the local B-field. In realistic neutron star magnetosphere with strong B-fields, the EM force dominates over

the plasma inertia (since magnetic energy density greatly exceeds plasma's kinetic energy density, see later) and other external forces such as gravity¹³ and pressure gradient. The plasma dynamics is determined by the *force-free condition* of $\mathbf{f} \approx 0$.

For plasma corotating with the magnetic dipole at velocity $\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}$ (valid for $r \ll r_{lc}$), the current component perpendicular to the corotating dipolar B-field is given by $\mathbf{J}_\perp = \rho \mathbf{v}$. Therefore, we obtain

$$\mathbf{E} = -(\mathbf{v}/c) \times \mathbf{B} = -\frac{\boldsymbol{\omega} \times \mathbf{r}}{c} \times \mathbf{B}, \text{ for } r \ll r_{lc}, \quad (4.258)$$

which is called the *corotational electric field* — as it simply means that $\mathbf{E}' = 0$ in the comoving frame of the corotating plasma. The charge density corresponding to this E-field is given by

$$\begin{aligned} \rho &= \frac{\nabla \cdot \mathbf{E}}{4\pi} = \frac{1}{4\pi c} [(\boldsymbol{\omega} \times \mathbf{r}) \cdot (\nabla \times \mathbf{B}) - (\nabla \times (\boldsymbol{\omega} \times \mathbf{r})) \cdot \mathbf{B}] \\ &= \frac{(\boldsymbol{\omega} \times \mathbf{r}) \cdot (\nabla \times \mathbf{B})}{4\pi c} - \frac{\boldsymbol{\omega} \cdot \mathbf{B}}{2\pi c}, \end{aligned} \quad (4.259)$$

where we have made use of the following vector identity $\nabla \times (\boldsymbol{\omega} \times \mathbf{r}) = 2\boldsymbol{\omega}$. Since we are only considering radii much below the light cylinder (relativistic numerical treatment is required for the region $r \sim r_{lc}$), one can show that the first term above is of the order $\rho(v/c)^2 \ll \rho$. This is because in a steady state $(\nabla \times \mathbf{B})_\perp = 4\pi \mathbf{J}_\perp / c = 4\pi \rho \mathbf{v} / c$, where \perp means the component perpendicular to the corotating dipolar B-field (or along the corotating velocity). Therefore, we can ignore the first term in the charge density expression and obtain

$$\rho_{GJ} \approx -\frac{\boldsymbol{\omega} \cdot \mathbf{B}}{2\pi c} = \frac{\omega \mu}{2\pi r^3 c} [3(\hat{\boldsymbol{\omega}} \cdot \hat{\mathbf{r}})^2 - 1], \text{ for } r \ll r_{lc}, \quad (4.260)$$

which is called the *Goldreich-Julian charge density* (the first expression is general, and the second expression is for an aligned rotator with dipolar B-fields). The physical meaning of ρ_{GJ} is that, in contrast to the vacuum case, the magnetosphere is filled up with plasma that is sufficiently dense to screen the electric field component parallel to the magnetic field so as to make sure $\mathbf{E} \cdot \mathbf{B} = 0$ everywhere. In this picture, each magnetic field line is an equipotential (as it is assumed to be a perfect conductor in the force-free limit). The toroidal current $\mathbf{J}_\perp = \rho_{GJ}(\boldsymbol{\omega} \times \mathbf{r})$ produces another B-field component which is a small perturbation (of the order v/c) to the initial dipolar B-field and the perturbation becomes important close to the light cylinder (i.e., at $r \sim r_{lc}$, the poloidal and toroidal B-field components are comparable).

When do the charged particles in the magnetosphere come from? Suppose the charge density temporarily deviates from the ρ_{GJ} due to deficit of particle supply in some local

¹³The gravitational acceleration $g = GM/R^2$ only corresponds to the electric force of an extremely weak E-field of the order $m_p g / q \sim 0.5 \text{ esu}$ (m_p being proton mass). Typical E-field strength at a distance r from the neutron star is of the order $E \sim \omega r B \sim 10^8 \text{ esu} \mu_{30} (r/10 \text{ km})^{-2}$.

regions — such a region is called a *magnetospheric gap*. For a local charge density $\rho \ll \rho_{\text{GJ}}$, the electric field component parallel to the local B-field is not efficiently screened and the potential difference along the B-field line is given by $d^2\phi/ds^2 \sim 4\pi\rho_{\text{GJ}}$ (from Gauss's law), where s is the displacement along the local B-field. For an unscreened gap of height h along the dipolar B-field line, the potential drop is given by

$$\Delta\phi_{\text{gap}} \sim 4\pi\rho_{\text{GJ}}h^2 \sim \frac{Bh^2}{r_{\text{lc}}} = 6 \times 10^{12} \text{ V} (h/10^4 \text{ cm})^2 (B/10^{12} \text{ G}) (P/\text{s})^{-1}. \quad (4.261)$$

Even for a very small gap, an electron/positron can be accelerated by the unscreened E-field to a very higher Lorentz factor. Initially, a gap may have a very low charge density $\rho \ll \rho_{\text{GJ}}$, but the seed particles (electrons or positrons) will have very high Lorentz factors such that they produce synchrotron and curvature emission, and the resulting photons will then interact with the B-fields (or other photons) to produce secondary electrons/positrons by Quantum Electrodynamics processes. Such pair creation processes are believed to restore the Goldreich-Julian charge density such that $\rho = \rho_{\text{GJ}}$ and hence the E-field component parallel to the dipolar B-field in most regions in the magnetosphere will be efficiently screened.

The steady state number density of charge particles in the magnetosphere can be considered to be a (highly uncertain) *multiplicity factor* κ higher than ρ_{GJ} , which is the minimum net charge density needed to screen the parallel E-field (and hence it is widely believed $\kappa \geq 1$). The ratio between particles' rest-mass energy density and that of the magnetic fields is then given by

$$\frac{U_{\text{particles}}}{U_{\text{EM}}} = \frac{\kappa(|\rho_{\text{GJ}}|/e)m_e c^2}{B^2/4\pi} \sim \kappa \frac{\omega}{\omega_B} \sim 10^{-18.5} \kappa (B/10^{12} \text{ G})^{-1} (P/\text{s})^{-1}, \quad (4.262)$$

where $\omega_B = eB/m_e c$ is the cyclotron frequency of an electron. We see that the energy density in the neutron star magnetosphere is entirely dominated by EM fields for any reasonable multiplicity factor κ .

For an aligned rotator, we find the θ -component of the E-field just above and just below the stellar surface to be

$$E_\theta(r \approx R) = \mathbf{E} \cdot \hat{\boldsymbol{\theta}} = -\frac{\omega R \sin \theta}{c} B_r, \quad (4.263)$$

which can be understood by noticing the fact that $B_r \hat{\mathbf{r}}$ is continuous across the stellar surface (based on $\nabla \cdot \mathbf{B} = 0$) and that $E_\theta \hat{\boldsymbol{\theta}}$ is also continuous across the stellar surface (based on $\nabla \times \mathbf{E} = -c^{-1} \partial_t \mathbf{B} = 0$ in a steady state). The next step is to see the potential drop from the pole to a given polar angle $\theta < \pi/2$ along the meridian

$$\Delta\phi(\theta) = \int_0^\theta E_\theta R d\theta = -\frac{\mu\omega}{Rc} \sin^2 \theta, \quad (4.264)$$

where the minus sign is for the magnetic northern hemisphere (where B_r is positive). The potential drop across the polar cap from the pole to the last open field line is¹⁴

$$\phi_{\text{pc}} \equiv \Delta\phi(\theta_{\text{pc}}) = \frac{\mu\omega^2}{c^2} = -1.3 \times 10^{13} \text{ V} \mu_{30} (P/1 \text{ s})^{-2}. \quad (4.265)$$

This potential drop is analogous to the case of a Faraday disk (or “unipolar inductor”) — consider a conducting thin disk of radius R rotating at frequency ω in a uniform B-field along the rotational axis, the potential drop between the center and the rim of the disk is given by $\Delta\phi = \int_0^R \frac{\omega Br}{c} dr = \omega BR^2/(2c) = \omega\Phi_B/(2\pi c)$, where $\Phi_B = \pi R^2 B$ is the magnetic flux threading the disk.

What is the meaning of the potential drop in eq. (4.265)? Near the neutron star surface where B-fields are very strong, particles are confined to move only along the B-field lines, so the potential drop across field lines can be maintained. However, at large distances (near and beyond the light cylinder) where the B-fields are much weaker, particles’ Larmor motion can bring them across field lines and in this situation they can be accelerated to an energy of the order $q\phi_{\text{pc}}$ due to the potential drop. We see that an electron moving along the open field lines would gain an energy of 10^{13} eV (for our fiducial parameters in eq. 4.265), corresponding to a very high Lorentz factor of the order 10^7 . These electrons radiate synchrotron and curvature emission and the photons may undergo pair production to create more particles. If each primary electron/positron produces κ secondary ones (here κ is an uncertain multiplicity factor) due to pair production in the magnetosphere, then each particle would acquire an average energy of $\phi_{\text{pc}}e/\kappa$, which is usually much greater than the electron rest-mass energy $m_e c^2$ (meaning that the secondary particles will be highly relativistic). A very conservative upper limit multiplicity factor is given by $\kappa_{\text{max}} = \phi_{\text{pc}}e/(m_e c^2)$, which is of the order 10^7 for $\phi_{\text{pc}} = 10^{13}$ V. Realistically, we expect $\kappa \ll \kappa_{\text{max}}$ because when the secondary particles are only mildly relativistic, they can no longer produce daughter particles.

The E-field near the light cylinder is given by $E_{\text{lc}} \sim B_{\text{lc}}$, because the co-rotation speed approaches the speed of light. Since \mathbf{E} is roughly perpendicular \mathbf{B} , we obtain the Poynting flux $S = |\mathbf{E} \times \mathbf{B}|c/4\pi \sim B_{\text{lc}}^2 c/4\pi$. This multiplied by an area of $4\pi r_{\text{lc}}^2$ (for a roughly isotropic emitter), we can estimate the total energy loss from the spinning neutron star

$$L_{\text{B}} \sim B_{\text{lc}}^2 r_{\text{lc}}^2 c = \frac{\mu^2 \omega^4}{c^3}, \quad (4.266)$$

which means that even an aligned rotator would lose energy and spin down. The difference from the vacuum case is that the plasma-filled magnetosphere launches a plasma wind whose energy density is dominated by EM fields (called *Poynting-dominated wind*). The

¹⁴To convert CGS units to Volt, one simply needs to multiply by a factor of $e/(e\text{V}/\text{erg}) \approx 300$, where e is the electron charge unit and $1\text{eV} = 1.6 \times 10^{-12}$ erg.

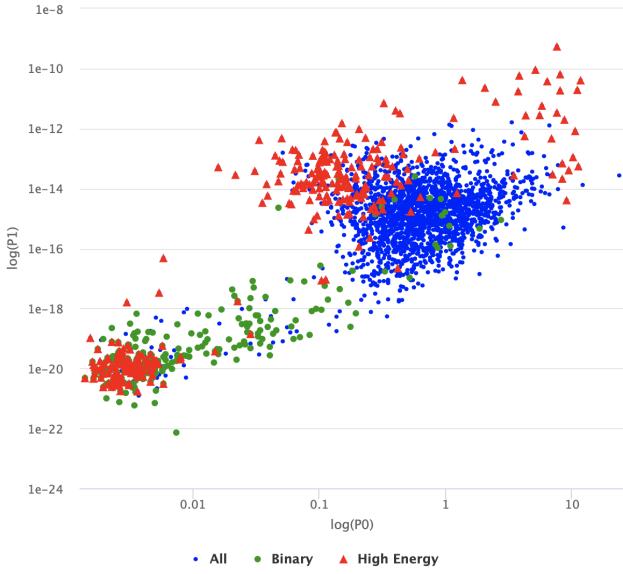


Figure 4.18: Pulsars in the [ANTF](#) catalog (Manchester et al. 2005). Axes: $P_0 = P/\text{s}$ and $P_1 = \dot{P}$. The green dots are binary pulsars, and red triangle ones are detected in the X-ray band or higher energies. Blue dots are radio pulsars.

particle outflowing rate in the wind is roughly given by the number flux $\kappa(|\rho_{\text{GJ}}|/e)c$ (κ being the multiplicity factor) near the stellar surface at the magnetic pole multiplied by the area of the polar cap

$$\dot{N} \sim \kappa(|\rho_{\text{GJ}}|/e)c \times \pi \theta_{\text{pc}}^2 R^2 = \frac{\kappa \mu \omega^2}{ec}, \quad (4.267)$$

which also equals to the number flux of particles crossing the light cylinder. The average energy per particle in the wind is $L_B/\dot{N} \sim \mu \omega^2 e / (\kappa c^2) = \phi_{\text{pc}} e / \kappa$, in agreement with the earlier expectation based on the potential drop across the polar cap. Since particles carry very little mass, the wind is expected to accelerate to a high Lorentz factor as it propagates to large radii $r \gg r_{\text{lc}}$. If the wind power is mainly carried by particle's kinetic motion instead of Poynting flux at sufficiently large distances (although this has not been fully verified by numerical simulations), the terminal wind Lorentz factor would be

$$\Gamma = \frac{L_B}{\dot{N} m_e c^2} \sim \frac{\phi_{\text{pc}} e}{\kappa m_e c^2} = 2.6 \times 10^7 \kappa^{-1} \mu_{30} (P/\text{s})^{-2}. \quad (4.268)$$

Detailed numerical simulations including plasma effects (in the force-free limit) show that the spin-down power for a star with external dipole B-fields at an arbitrary magnetic

misalignment angle χ is given by (Spitkovsky 2006)

$$L_B \approx \frac{\mu^2 \omega^4}{c^3} (1 + \sin^2 \chi) = 5.8 \times 10^{31} (1 + \sin^2 \chi) \text{ erg s}^{-1} \mu_{30}^2 (P/\text{s})^{-4}, \quad (4.269)$$

which means that a perpendicular rotator spins down twice faster than an aligned rotator. [Tchekhovskoy et al. \(2016\)](#) provided a semi-analytic explanation of the $(1 + \sin^2 \chi)$ dependence of the spin-down power.

The moment of inertial of a neutron star is approximately $I \simeq 10^{45} \text{ g cm}^2$ (roughly given by $0.3MR^2 = 8 \times 10^{44} \text{ g cm}^2 M_{1.4M_\odot} R_{10\text{ km}}^2$), so the spin-down power is related to the change in angular frequency

$$I\omega\dot{\omega} = L_B. \quad (4.270)$$

From the observed period changing rate \dot{P} , one can infer the magnetic dipole moment

$$\mu = 0.82 \times 10^{30} \text{ G cm}^3 I_{45}^{1/2} \left(\frac{P\dot{P}}{10^{-15} \text{ sec}} \right)^{1/2} (1 + \sin^2 \chi)^{-1/2}. \quad (4.271)$$

The *spin-down timescale* is defined as

$$\tau_{\text{sd}} = \frac{I\omega^2/2}{L_B} = \frac{\omega}{2|\dot{\omega}|} = \frac{P}{2\dot{P}} = 1.1 \times 10^7 \text{ yr} I_{45} \mu_{30}^{-2} (P/\text{s})^2 (1 + \sin^2 \chi)^{-1}. \quad (4.272)$$

For a constant magnetic dipole moment, then the pulsar spin rate evolution should follow $\dot{\omega} \propto -\omega^n$ and $n = 3$, where n is called the *braking index*. If the initial angular frequency at birth is much shorter than the current one and the magnetic dipole moment stays constant over time, then the age of a pulsar can be estimated by $\tau_{\text{age}} = \omega/2|\dot{\omega}| = \tau_{\text{sd}}$. However, the measured breaking indices, $n = \omega\ddot{\omega}/\dot{\omega}^2 = 2 - P\ddot{P}/\dot{P}^2$, from pulsars are often less than 3, which means that the physical situation is more complicated than the simple picture described here.

4.5 Homework

Prob. 10. Jupiter emits bright radio emission that extends from decameters ($\lambda \sim 10 \text{ m}$, DAM) wavelength up to decimeters ($\lambda \sim 10 \text{ cm}$, DIM), as shown in Fig. 4.19. DAM is produced by cyclotron emission non-relativistic electrons; whereas the DIM comes from synchrotron emission by (a much smaller population of) relativistic electrons. The sharp cut-off of DAM near 30 MHz corresponds to the maximum electron cyclotron frequency at the planet's surface near the magnetic poles. Estimate the surface magnetic field strength at Jupiter's magnetic poles.

Prob. 11. Dust grains in the interstellar medium may be rapidly spinning as a result of various torques. One of them is radiative torque, which is non-zero if one side of a dust grain persistently

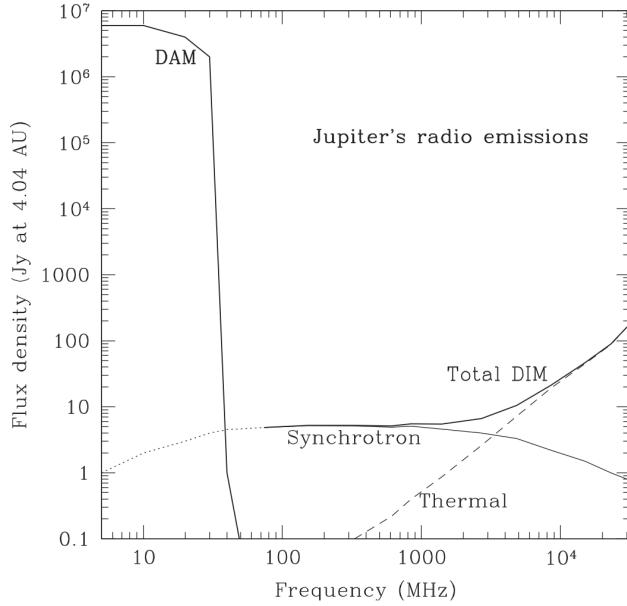


Figure 4.19: Schematic radio spectrum of Jupiter taken from [de Pater \(2004\)](#).

receives more radiation than the other side. Grains can be charged due to UV or X-ray ionization as well as collisions with free electrons. In general, the center of charge distribution is not the same as the center of mass distribution, so grains may have a net electric dipole moment. Let us consider an electric dipole moment of $d = eZa$, for a grain of radius a and effective charge number of Z (which is generally not an integer).

- (i) Estimate the spin angular frequency ω at which the spin kinetic energy is comparable to the motional kinetic energy, i.e., $(1/2)I\omega^2 = (3/2)k_B T$, where $I = (2/5)ma^2$ is the moment of inertia, $m = 4\pi\rho a^3/3$ is the mass, and ρ is the mean density of the grain material.
- (ii) For $a = 10\text{\AA}$ (a very small one!), $\rho = 2\text{ g cm}^{-3}$, $T = 100\text{ K}$, $Z = 0.1$, estimate the frequency and power of the electric dipole radiation from the dust grain for an inclination angle of $\chi = 45^\circ$ (between the dipole moment vector and the spin vector). Note that, for a rotating dipole with an inclination angle of χ , the emitting power is $P = (2/3)\omega^4 d^2 \sin^2 \chi/c^3$.
- (iii) For a cloud of dust grains described in (ii) with mass column density $\Sigma = 10^{-2} M_\odot \text{ pc}^{-2}$ (assuming this mass is all in small dust grains with $a = 10\text{\AA}$), estimate the surface intensity of the electric dipole emission assuming an optically thin cloud. Compared your intensity with that of the cosmic microwave background (CMB), which is a blackbody with temperature $T_{\text{CMB}} = 2.7\text{ K}$.

Prob. 12.

Consider a magnetic white dwarf whose dipolar B-field strength at the magnetic equator is $B_{\text{eq}} = 8 \times 10^8 \text{ G}$. Its mass, radius, moment of inertial, magnetic dipole moment are $M = 1.3 M_\odot$ (a very

massive one), $R = 2 \times 10^8$ cm, $I \approx 0.3MR^2$, and $\mu = B_{\text{eq}}R^3$, respectively. The rotational period is $P = 7$ min (a rapid rotator). These parameters are taken from [Caiazzo et al. \(2021\)](#).

- (i) [***not required**] Assuming an aligned rotator, estimate the potential drop across the polar cap of the white dwarf. If a proton is accelerated by this potential drop, what would be its Lorentz factor?
- (ii) For a magnetic inclination angle of $\chi = 45^\circ$, use the Spitkovsky formula for a plasma-filled magnetosphere (eq. 4.269) to calculate the spin-down luminosity L_B and spin-down time τ_{sd} .
- (iii) Assuming that the spin-down luminosity is in the form of a relativistic wind, estimate the radius of termination shock where the wind is stopped by the interstellar medium (ISM). Take the ISM to be made of hydrogen gas with a number density of $n_H = 0.1 \text{ cm}^{-3}$ and the relative velocity between the white dwarf and the ISM to be $v = 10 \text{ km s}^{-1}$. The radius of the termination shock r_T is roughly given by the balance between the ram pressure and wind kinetic pressure: $p_{\text{ram}} = nm_p v^2 = p_{\text{wind}} = L_B / (4\pi r_T^2 c)$.

References

- Jackson, *Classical Electrodynamics* (John Wiley & Sons, New York, 1975).
- Rybicki & Lightman, *Radiative processes in astrophysics* (A Wiley-Interscience Publication, New York, 1979).
- Landau & Lifshitz, *Mechanics* (Butterworth-Heinemann, 3rd Edition, Oxford, 1976)
- Landau & Lifshitz, *Classical Theory of Fields* (Pergamon, 4th Edition, Oxford, 1975)
- F. F. Chen, *Introduction to Plasma Physics and Controlled Fusion* (Springer International Publishing, 3rd Edition, Switzerland, 2016)
- Goldreich & Julian, 1969, ApJ, 157, 869
- Spitkovsky, 2006, ApJL, 648, 51
- Manchester, Hobbs, Teoh, & Hobbs, 2005, AJ, 129, 1993
- Lorimer, Bailes, McLaughlin, Narkevic, Crawford, 2007, Science, 318, 777
- Thornton, Stappers, Bailes, et al., 2013, Science, 341, 53

Chapter 5

Plasma

In this chapter, we discuss the basic properties of astrophysical plasmas and the physics governing how EM waves propagate through a plasma.

5.1 Introduction

The simplest plasma consists free charged particles. In a sufficiently large volume, there is equal amount of positive and negative charges, so the system is overall charge neutral. Many key ideas in plasma physics are related to the deviation from and recovery to charge neutrality. In this section, we introduce three phenomena.

- (i) At sufficiently small lengthscales, the Coulomb potential of individual particles dominates and this will attract charges of the opposite sign.
- (ii) Deviation from charge neutrality driven by external forces will build up large-scale Coulomb field which then tries to bring the system back to charge neutrality and, in doing so, particles undergo *plasma oscillation* at the plasma's natural frequency.
- (iii) An incoming EM wave's electric field drives particles of different charge signs to move in opposite directions, and this produces oscillating currents which produce secondary waves that will interfere with the incoming EM wave.

5.1.1 Charge neutrality and Debye shielding

Let us consider a spatially uniform plasma with ion and electron number densities satisfying $Zn_{i,0} = n_{e,0}$, where Z is the charge number of the ions¹. Electrons and ions have the same kinetic temperature T . Now suppose we add a new ion into the system and choose a

¹The discussion in this subsection also applies to electron-positron plasma as long as we take $Z = 1$.

coordinate system centered on its position. The excessive positive charge quickly creates a Coulomb potential

$$\phi_c(r) = Ze/r, \quad (5.1)$$

which will cause the concentration of electrons and repulsion of ions near the position of the newly added ion. When equilibrium is achieved, the number densities of electrons and ions are given by the Boltzmann distribution under the Coulomb potential

$$\begin{aligned} n_e(r) &= n_{e,0} \exp\left(\frac{e\phi_c}{k_B T}\right) \approx n_{e,0} \left(1 + \frac{e\phi_c}{k_B T}\right), \\ n_i(r) &= n_{i,0} \exp\left(-\frac{Ze\phi_c}{k_B T}\right) \approx n_{i,0} \left(1 - \frac{Ze\phi_c}{k_B T}\right), \end{aligned} \quad (5.2)$$

where the approximations are based on the fact that the energy perturbation by the Coulomb field of the newly added ion is small for typical radius r , i.e. $Ze^2/r \ll k_B T$ (we will later show that this is well satisfied as long as the plasma's Debye number in eq. 5.13 is much greater than unity). Thus, we obtain the net charge density at distance r from the newly added ion

$$\rho(r) = -en_e(r) + Zen_i(r) + Ze\delta(r) = -e^2 n_{e,0} \frac{1+Z}{k_B T} \frac{Ze}{r} + Ze\delta(r), \quad (5.3)$$

where $\delta(r)$ is a delta-function describing the charge density provided by the newly added ion and it is normalized as $\int_0^\infty \delta(r) 4\pi r^2 dr = 1$. The total charge within radius r is

$$Q(< r) = \int_0^r \rho(r) 4\pi r^2 dr = Ze \left[1 - (r/r_D)^2\right], \quad (5.4)$$

where r_D is the *Debye length*

$$r_D = \sqrt{\frac{2}{1+Z}} \left(\frac{k_B T}{4\pi e^2 n_{e,0}} \right)^{1/2}. \quad (5.5)$$

Since $Q(< r_D) = 0$, we see that the Coulomb field of the newly added ion is completely shielded within the Debye length. The Debye length increases as the plasma gets hotter (because particles with larger kinetic energies are more difficult to be confined by the Coulomb potential) and as the density drops (because there are less particles to shield the Coulomb potential).

As a result of shielding, our approximation for the Coulomb potential in eq. (5.1) is only correct at $r \ll r_D$. To obtain a self-consistent solution for the potential $\phi(r)$, we would need to solve the Poisson equation

$$\nabla^2 \phi(r) = 4\pi\rho(r), \quad (5.6)$$

with the boundary conditions of $\phi(r \rightarrow 0) = Ze/r$ and $\phi(r \rightarrow \infty) = 0$. The charge density is given by the Boltzmann distribution of particles under the potential $\phi(r)$,

$$\rho(r > 0) = -e^2 n_{e,0} \frac{1+Z}{k_B T} \phi(r) = -\frac{\phi(r)}{2\pi r_D^2}, \quad (5.7)$$

where we have excluded the central region (as the solution is already known) and used the definition of the Debye length r_D in eq. (5.5).

We look for solutions in the form of

$$\phi(r) = \frac{Ze}{r} y(r), \quad (5.8)$$

where $y(r)$ is an unknown function with the asymptotic behavior of $y(r \rightarrow 0) = 1$. In terms of our one-dimensional spherical coordinate r , the Laplacian is given by $\nabla^2 = r^{-2} \partial_r(r^2 \partial_r)$. Then, we obtain $\nabla^2 \phi = Zer^{-1} d^2y/dr^2$, so the Poisson equation for $\phi(r)$ becomes a differential equation for $y(r)$ as follows

$$\frac{d^2y}{dr^2} + \frac{2}{r_D^2} y = 0. \quad (5.9)$$

The general solution takes the form of

$$y = a_1 e^{-\sqrt{2}r/r_D} + a_2 e^{\sqrt{2}r/r_D}, \quad (5.10)$$

where a_1 and a_2 are constants. The boundary conditions of $y(r \rightarrow 0) = 1$ and $\phi(r \rightarrow \infty) = 0$ then require that $a_1 = 1$ and $a_2 = 0$.

Therefore, the self-consistent solution to the Poisson equation is

$$\phi(r) = \frac{Ze}{r} e^{-\sqrt{2}r/r_D}, \quad (5.11)$$

which is called the *Coulomb potential with Debye shielding*.

For an electron-proton ($Z = 1$) plasma with number density $n_{i,0} = n_{e,0} = n$ and temperature T , the Debye length is given by

$$r_D(e-p) = \left(\frac{k_B T}{4\pi e^2 n} \right)^{1/2} = 690 \text{ cm} \left(\frac{T}{10^4 \text{ K}} \right)^{1/2} \left(\frac{n}{\text{cm}^{-3}} \right)^{-1/2}. \quad (5.12)$$

The number of electrons inside a sphere of radius r_D is

$$N_D = n \frac{4\pi}{3} r_D^3 = 1.4 \times 10^9 \left(\frac{T}{10^4 \text{ K}} \right)^{3/2} \left(\frac{n}{\text{cm}^{-3}} \right)^{-1/2}, \quad (5.13)$$

which is called the *Debye number* (or the *plasma parameter*). A defining criterion for a plasma is $N_D \gg 1$. Since ionized gas in various astrophysical environments are well in the regime of $N_D \gg 1$, so astrophysical plasmas are common (e.g., ionized phases of the interstellar medium, interior of non-degenerate stars, solar wind, Earth's ionosphere and magnetosphere, accretion flow near black holes, etc).

One can also show that N_D is of the order the ratio between the thermal energy $k_B T$ and the Coulomb energy $|-e^2/r_D|$ for a typical particle inside the “Debye cloud”, so our Taylor expansion in eq. (5.2) is justified when $N_D \gg 1$. This also means that the plasma pressure is dominated by thermal motion of particles. The Coulomb attraction between a given particle and the opposite-charged ones inside the “Debye cloud” acts to *decrease* the total pressure of the plasma, and the ratio between the (negative) Coulomb correction P_{Coul} and the thermal pressure $P = 2n k_B T$ is of the order

$$|P_{\text{Coul}}|/P \sim 1/N_D \propto n^{1/2} T^{-3/2}. \quad (5.14)$$

We see that the Coulomb correction may become appreciable for very dense and cold plasma. In the solar interior, the average temperature is $T \sim \sqrt{GM_\odot m_p/(k_B R_\odot)} \sim 10^7 \text{ K}$ and the average electron number density is $n = 3M_\odot/(4\pi R_\odot^3 m_p) \sim 10^{24} \text{ cm}^{-3}$, so we obtain $|P_{\text{Coul}}|/P \sim 1/N_D \sim 1\%$ — indeed a very small correction.

The above discussion shows that a given particle is surrounded by a “Debye cloud” of $N_D \gg 1$ particles which work together to shield the Coulomb potential. The fractional deviation from charge neutrality inside the “Debye cloud” is of the order $N_D^{-1} \ll 1$. The phenomenon of Debye shielding is a *collective behavior*, which is distinct from 2-body collisions between particles. Such a collective behavior is made possible by the long-range Coulomb forces and involves coherent motion for a large number of particles, which is on top of the random thermal motion of individual particles.

5.1.2 Plasma oscillation (Langmuir wave)

Another phenomenon of collective behavior is plasma oscillation. If we displace a large patch of electrons by a small distance x away from their equilibrium positions while keeping the protons (or other positive charge carriers) stationary, this creates a positively charged thin layer (due to electron under-density) and a negatively charged thin layer (due to electron over-density). This is shown in the left panel of Fig. 5.1. These two thin layers have thickness of x and charge column densities of opposite signs $\pm enx$. A uniform electric field along the direction of the displacement is formed between these two layers, and it is easy to obtain the electric field strength (using Gauss’s theorem)

$$E = 4\pi enx. \quad (5.15)$$

Thus, each displaced electron feels a restoring force

$$m_e \ddot{x} = -eE = -4\pi e^2 nx \quad (5.16)$$

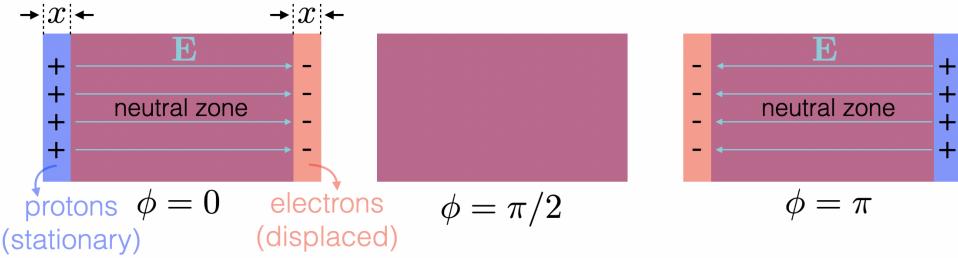


Figure 5.1: Plasma oscillation initiated by collective displacement of a large patch of electrons (shown in red) by x , while the protons (shown in blue) are kept stationary. The three panels show three different oscillatory phases. In the middle panel (for $\phi = \pi/2$), protons and electrons are temporarily overlapping perfectly, resulting in no electric field, but the electrons' inertial cause them to overshoot to as to maintain the oscillation.

This gives rise to a harmonic oscillation at the *electron plasma frequency*

$$\omega_p = \sqrt{\frac{|\ddot{x}|}{x}} = \left(\frac{4\pi e^2 n}{m_e} \right)^{1/2} \Rightarrow \nu_p = \frac{\omega_p}{2\pi} = 8.98 \times 10^3 \text{ Hz} \left(\frac{n}{\text{cm}^{-3}} \right)^{1/2}. \quad (5.17)$$

We see that a plasma can *temporarily* deviates from charge neutrality and will always try to restore charge neutrality on a *plasma timescale* of ω_p^{-1} . In the absence of damping, the displaced electrons will oscillate back and forth near their equilibrium positions — this forms a standing wave oscillation called the *Langmuir wave*.

A plasma can be considered as a harmonic oscillator with natural frequency ω_p . When interacting with an incoming EM wave, a plasma can be considered as a driven harmonic oscillator, and in this picture, the system oscillates at the *driving frequency* = EM wave frequency.

In the above analysis, we have ignored the thermal motion of electrons (taking the “cold plasma” limit) as well as the contribution to the Langmuir oscillation by protons. These complications will be discussed in later sections and they only cause small modifications to the oscillation frequency as compared to ω_p .

One potential source of damping is due to Coulomb collisions between particles. For instance, if electrons have thermal motion at temperature T , the timescale for electron-electron Coulomb collisions is the deflection time (see Ch. 6)

$$t_{\text{def}} \simeq 10^4 \text{ sec} \left(\frac{T}{10^4 \text{ K}} \right)^{3/2} \left(\frac{n}{\text{cm}^{-3}} \right)^{-1}. \quad (5.18)$$

The ratio between the deflection timescale and the plasma timescale is

$$\omega_p t_{\text{def}} \simeq 6 \times 10^8 \left(\frac{T}{10^4 \text{ K}} \right)^{3/2} \left(\frac{n}{\text{cm}^{-3}} \right)^{-1/2}. \quad (5.19)$$

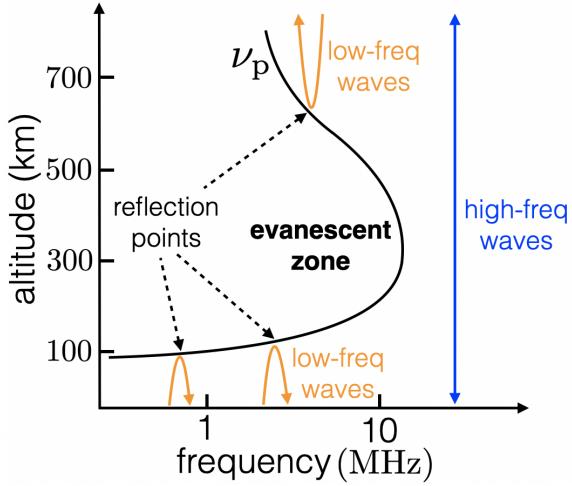


Figure 5.2: A crude representation of the plasma frequency profile of the Earth’s ionosphere. Low-frequencies waves with $\nu < \nu_p$ will be reflected from the ionosphere, whereas high-frequency waves with $\nu \gg 10$ MHz will be able to transmit through the ionosphere.

Most astrophysical plasmas (e.g., ionized phases of the interstellar medium, the solar corona/wind, the intergalactic medium) satisfy the condition of $\omega_{\text{pt,def}} \gg 1$, so the effects of Coulomb collisions can be ignored — such plasmas are called *collisionless*². Note that there are much faster damping mechanisms that do not rely on collisions, and we will later discuss Landau damping as an example.

The Debye length is related to the plasma frequency by

$$r_D \sim v_{\text{th}}/\omega_p, \quad (5.20)$$

where $v_{\text{th}} \sim \sqrt{k_B T/m_e}$ is the electron thermal speed. We see that a typical electron moves a Debye length r_D over one plasma timescale, so we expect the plasma to locally deviate from charge neutrality on spatial lengthscales $r \sim r_D$ — this is consistent with our conclusion in §5.1.1.

Another lengthscale related to the plasma frequency is the *skin depth*

$$\ell_{\text{skin}} = c/\omega_p, \quad (5.21)$$

which is the depth to which low-frequency EM waves ($\omega \ll \omega_p$) can penetrate into a plasma. It takes a plasma timescale ω_p^{-1} for the particles to set up a charge distribution to screen the nearly static (as $\omega \ll \omega_p$) electric field of the incoming EM waves and, during the

²In a radiation-pressure dominated plasma, it is possible that collisions between electrons and photons are important in shaping the electron distribution function.

plasma timescale, the waves can only propagate a distance of the order c/ω_p before being strongly modified. In the absence of damping (= absorption), low-frequency EM waves are totally reflected by a plasma like visible light bouncing off a mirror. Fig. 5.2 demonstrates the interaction between radio waves and the Earth's ionosphere.

5.1.3 Linear plane waves in quasi-homogeneous, quasi-stationary plasma

We take the the curl of Faraday's law of induction to obtain the general wave equation

$$\nabla \times (\nabla \times \mathbf{E}) + \frac{1}{c^2} \partial_t^2 \mathbf{E} = -\frac{4\pi}{c^2} \partial_t \mathbf{J}, \quad (5.22)$$

where the LHS describes wave propagation and the RHS contains the source term in the form of *changing* current density \mathbf{J} . We study self-consistent solutions where the EM fields are coupled to the charged particles by their equation of motion and their emission.

The incoming EM wave is assumed to be sufficiently weak such that the oscillatory motion of charged particles due to the action of the EM wave are non-relativistic. Referring back to [Chapter 4], this assumption corresponds to a very small non-linearity parameter $a_0 = E_0 e / (m_e \omega c) \ll 1$, where E_0 is the real amplitude of the EM wave and $\omega = 2\pi\nu$ is the wave angular frequency. This limits the energy density of the EM wave

$$U_{\text{EM}} = E_0^2 / (8\pi) \ll 3 \times 10^{27} (\nu / 10^{15} \text{Hz})^2 \text{eV/cm}^{-3}, \quad \text{for linear EM wave,} \quad (5.23)$$

where $\nu \sim 10^{15} \text{Hz}$ roughly corresponds to the optical band. This turns out to be a very loose constraint because the radiation energy density is of the order 1eV/cm^{-3} in the interstellar medium and of the order $10^{12} \text{eV/cm}^{-3}$ near the surface of the Sun.

Our assumption of non-relativistic oscillatory motion allows us to ignore non-linear terms that are proportional to the wave amplitude squared (E_0^2) or higher powers. In this limit, as we show below, the current density \mathbf{J} will be linearly proportional to the amplitude of the incoming EM wave. Thus, the wave equation (5.22) is *linear*, which means that any two wave modes \mathbf{E}_1 and \mathbf{E}_2 would independently create their own currents \mathbf{J}_1 and \mathbf{J}_2 , and that the total solution is given by the linear superposition of these two modes.

Besides the requirement of weak EM waves, we also restrict our discussion to plasmas that satisfy the following three conditions:

- (i) The unperturbed plasma is quasi-homogeneous such that the spatial variation of its properties (e.g., density and magnetic field) occur on very a large lengthscale $L \gg \lambda$, where λ is the wavelength of the EM wave in consideration.
- (ii) The unperturbed plasma is in a quasi-stationary state with temporal variability timescale $T \gg \omega^{-1}$, where ω is the EM wave frequency.

- (iii) The plasma is located sufficiently far away from the source of the EM waves at distances $r \gg \lambda$, so we are in the plane-wave limit.

These conditions make our problem one dimensional and the solution takes the following general (plane wave-like) form

$$\mathbf{E}(\mathbf{r}, t) e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)}, \quad (5.24)$$

where $\mathbf{E}(\mathbf{r}, t)$ is the *complex amplitude* which carries information about the polarization as well as slow spatial/temporal variations (on lengthscale L and timescale T mentioned above), and $e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)}$ describes the rapid phase oscillation of the wave. If we ignore any spatial/temporal variations of the unperturbed plasma, then the complex amplitude would be a constant and it will be denoted as \mathbf{E}_0 . A solution in the form of eq. (5.24) that satisfies the wave equation (5.22) is an *eigenmode of the EM wave-plasma system*. The physical solution for given initial and boundary conditions can be constructed as the real part of a linear superposition of all the eigenmodes.

5.2 Cold non-magnetized plasma

In this section, we will ignore any thermal motion and assume the unperturbed plasma to be homogeneous and stationary.

5.2.1 Linearized wave equation

The non-relativistic equation of motion for a particle of rest-mass m and charge q is

$$m\dot{\mathbf{v}} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}/c), \quad (5.25)$$

where $\dot{\mathbf{v}} = d\mathbf{v}/dt = \partial_t \mathbf{v} + (\mathbf{v} \cdot \nabla) \mathbf{v}$ is the total derivative. We ignore the two high-order terms $(\mathbf{v} \cdot \nabla) \mathbf{v}$ and $q\mathbf{v} \times \mathbf{B}/c$ because they are proportional to the wave amplitude squared (E_0^2), so the equation of motion simplifies to

$$\partial_t \mathbf{v}(\mathbf{r}, t) = (q/m)\mathbf{E}(\mathbf{r}, t). \quad (5.26)$$

The particle number density $n(\mathbf{r}, t)$ satisfies the continuity equation

$$\partial_t n + \nabla \cdot (n\mathbf{v}) = 0. \quad (5.27)$$

Now comes the key trick. The number density can be written in the following form

$$n(\mathbf{r}, t) = n_0 + \delta n(\mathbf{r}, t), \quad (5.28)$$

where n_0 is the unperturbed number density and δn denotes the perturbation. Then, the continuity equation becomes

$$\partial_t(\delta n) + n_0 \nabla \cdot \mathbf{v} = -\nabla \cdot (\delta n \mathbf{v}) \approx 0, \quad (5.29)$$

We have ignored the third term on the LHS as it is proportional to the wave amplitude squared (E_0^2).

The above results are for any particle species, so we keep them general. In the absence of thermal motion, the current density is given by

$$\mathbf{J} = \sum_s q_s n_s(\mathbf{r}, t) \mathbf{v}_s(\mathbf{r}, t) = \sum_s q_s [n_{s,0} + \delta n_s(\mathbf{r}, t)] \mathbf{v}_s(\mathbf{r}, t), \quad (5.30)$$

where we have summed over the contributions from each of the particle species as denoted by the subscript s and the number density has been split into the unperturbed and perturbed terms ($n_{s,0}$ and δn_s respectively). The wave equation involves the time derivative of the current density

$$\partial_t \mathbf{J} = \sum_s q_s [\mathbf{v} \partial_t (\delta n_s) + n_{s,0} \partial_t \mathbf{v}_s + \delta n_s \partial_t \mathbf{v}_s] \approx \sum_s q_s n_{s,0} \partial_t \mathbf{v}_s, \quad (5.31)$$

where all the high-order terms ($\propto E_0^2$) have been ignored. Now we make use of eq. (5.26) for $\partial_t \mathbf{v}_s$, and obtain

$$\partial_t \mathbf{J} = \sum_s \frac{n_{s,0} q_s^2}{m_s} \mathbf{E}(\mathbf{r}, t). \quad (5.32)$$

This is then plugged back into the general wave equation (5.22) and we obtain the following *linearized wave equation*

$$\nabla \times (\nabla \times \mathbf{E}) + \frac{1}{c^2} \partial_t^2 \mathbf{E} = -\frac{\omega_p^2}{c^2} \mathbf{E}, \quad (5.33)$$

where we have defined a new plasma frequency that includes all particle species

$$\omega_p = \left(\sum_s \frac{4\pi n_{s,0} q_s^2}{m_s} \right)^{1/2}. \quad (5.34)$$

For a proton-electron plasma with unperturbed electron and proton number density of n_0 , we find our new plasma frequency to be

$$\omega_p = \left(\frac{4\pi n_0 e^2}{m_e} \right)^{1/2} \left(1 + \frac{m_e}{m_p} \right)^{1/2} \approx (1 + 2.7 \times 10^{-4}) \left(\frac{4\pi n_0 e^2}{m_e} \right)^{1/2}, \quad (5.35)$$

which is extremely close to the electron plasma frequency defined in eq. (5.33). Therefore, for practical purposes, it is appropriate to ignore the contribution from protons.

In the next subsection, we solve the linearized wave equation (5.26) to obtain the eigenmodes of the EM wave-plasma system.

5.2.2 Langmuir and EM modes

For a homogeneous and stationary plasma, we consider the following plane-wave ansatz

$$\mathbf{E} = \mathbf{E}_0 e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)}, \quad (5.36)$$

where \mathbf{E}_0 is a constant complex amplitude (independent of \mathbf{r} and t), \mathbf{k} is the wavevector, and ω is the wave frequency. This does not lose generality, as the final solution corresponding to given initial and boundary conditions can always be reconstructed as the superposition of many such plane waves. Often, we would like to understand how an incoming EM wave at a given frequency ω propagates inside a plasma, and to answer this question, we would seek for the relation between the (unknown) wavevector \mathbf{k} and the (known) frequency ω — such a relation of $\mathbf{k}(\omega)$ is called the *dispersion relation*³.

We plug the plane-wave ansatz into the linearized wave equation (5.26) and obtain

$$c^2 \mathbf{k} \times (\mathbf{k} \times \mathbf{E}_0) + (\omega^2 - \omega_p^2) \mathbf{E}_0 = 0. \quad (5.37)$$

Let us orient our z -axis such that $\mathbf{k}(\omega) = k(\omega)\hat{\mathbf{z}}$, then we obtain⁴

$$k^2 c^2 (\hat{\mathbf{z}} \cdot \mathbf{E}_0 - \mathbf{E}_0) + (\omega^2 - \omega_p^2) \mathbf{E}_0 = 0. \quad (5.38)$$

Using the Cartesian components, we obtain the following matrix equation

$$\begin{pmatrix} \omega^2 - \omega_p^2 - k^2 c^2 & 0 & 0 \\ 0 & \omega^2 - \omega_p^2 - k^2 c^2 & 0 \\ 0 & 0 & \omega^2 - \omega_p^2 \end{pmatrix} \begin{pmatrix} E_{0,x} \\ E_{0,y} \\ E_{0,z} \end{pmatrix} = 0. \quad (5.39)$$

This comes down to solving the eigenvalue problem for a matrix (and it is a simple diagonal one). For the above equation to hold for all \mathbf{E}_0 , the determinant of the matrix must vanish, meaning that

$$(\omega^2 - \omega_p^2 - k^2 c^2)^2 (\omega^2 - \omega_p^2) = 0. \quad (5.40)$$

There are two branches of solutions corresponding to two different modes

$$\omega^2 = \omega_p^2 + k^2 c^2, \quad \text{for the transverse EM mode } \mathbf{E}_0 \perp \mathbf{k}, \quad (5.41)$$

and

$$\omega^2 = \omega_p^2, \quad \text{for the longitudinal Langmuir mode } \mathbf{E}_0 \parallel \mathbf{k}. \quad (5.42)$$

³Another way of setting up the problem is to solve for the wave frequency for a given wavevector. This approach would give a dispersion relation in the form of $\omega(\mathbf{k})$, which is equivalent to our result of $\mathbf{k}(\omega)$. Our setup is motivated by the fact that, when an EM wave propagates inside a (weakly) spatially inhomogeneous but temporally stationary plasma, the wave frequency ω is conserved whereas the wavevector \mathbf{k} is modified due to refraction according to Snell's law (like light refraction in the air-water interface). Of course, in a spatially homogeneous but temporally (weakly) variable plasma, the wavevector is conserved while the frequency will be modified.

⁴Here we have used the identity: using the identity $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}$.

From the discussion in §5.1.2, we are familiar with the Langmuir mode, which only has the $E_{0,z}$ component parallel to the wavevector \mathbf{k} .

For the Langmuir mode, the wavenumber does not depend on frequency, which means that the oscillation at all wavelengths have the same frequency $\omega = \omega_p$. This is because, for longer wavelengths λ , particles undergo larger displacements $\Delta z \simeq k^{-1} = \lambda/(2\pi)$ which produces a stronger electric field with a restoring force that is proportional to the displacement $E_0 = 4\pi e n_0 \Delta z$ (eq. 5.15). This solution must break down at extremely long wavelengths such that the non-linearity parameter $a_0 \equiv E_0 e / (m_e \omega c) = \omega_p \Delta z / c \simeq \omega_p / (kc) \gtrsim 1$, because non-linear effects would become important. This limits the wavelength of Langmuir waves to be less than the plasma skin depth: $k^{-1} \lesssim \ell_{\text{skin}} = c/\omega_p$. For a finite-temperature plasma, there is another physical effect called Landau damping that rapidly damps the amplitude of Langmuir waves for wavelengths longer than the Debye length $r_D \sim v_{\text{th}}/\omega_p$ (where v_{th} is the electron thermal speed). This will be discussed later. Realistic Langmuir waves in a finite temperature plasma always have wavelength $\lambda \lesssim r_D$ and they all oscillate near the plasma frequency ω_p .

The other mode is transverse EM mode, which is very similar to the transverse vacuum EM mode where the $E_{0,x}$ and $E_{0,y}$ components propagate independently — in both cases the eigenmodes are linearly polarized. The only difference is that the dispersion relation in a plasma involves the plasma frequency ω_p in the following way

$$kc = \sqrt{\omega^2 - \omega_p^2}, \quad (5.43)$$

where, without losing generality, we have ignored the other case with $kc = -\sqrt{\omega^2 - \omega_p^2}$ as it simply corresponds to waves propagating along the $-z$ direction. The *refractive index* of the plasma is given by

$$\tilde{n}(\omega) \equiv \frac{kc}{\omega} = \sqrt{1 - \omega_p^2/\omega^2}. \quad (5.44)$$

The key property of the above dispersion relation is that when $\omega < \omega_p$, the wavenumber becomes purely imaginary $k = \pm i|k|$, and the corresponding plane-wave solution becomes

$$\mathbf{E}(z, t) = \mathbf{E}_0 e^{\mp |k|z} e^{-i\omega t}, \quad \text{at low frequencies } \omega < \omega_p. \quad (5.45)$$

One of the solution has an exponentially growing amplitude $\mathbf{E}_0 e^{+|k|z}$, which is inconsistent with energy conservation. The other one has exponentially decreasing amplitude $\mathbf{E}_0 e^{-|k|z}$, and this corresponds to an evanescent wave.

We conclude that ω_p is the *cut-off frequency* — no EM waves with $\omega < \omega_p$ can propagate inside the plasma. This is because the plasma can react sufficiently fast (on a timescale ω_p^{-1}) so as to shield the (slower changing) electric field of the EM wave. The consequence is that a low-frequency EM wave striking on the surface of a dense plasma cloud will be reflected (in

the absence of damping). A plasma cloud with a boundary is necessarily inhomogeneous, and we expect the EM wave to be refracted due to spatially varying refractive index (see §5.4 later).

High-frequency EM waves with $\omega > \omega_p$ can propagate in a plasma which has a refractive index $\tilde{n} < 1$. The general solution for EM waves inside a plasma is given by the linear superposition of monochromatic plane-wave components, in the form of an inverse Fourier transformation

$$\mathbf{E}(z, t) = \int_{\omega_p}^{\infty} \tilde{\mathbf{E}}(\omega) e^{i[k(\omega)z - \omega t]} d\omega, \quad (5.46)$$

where $\tilde{\mathbf{E}}(\omega)d\omega$ is the complex amplitude of the monochromatic component near frequency ω , the function $\tilde{\mathbf{E}}(\omega)$ has two Cartesian components along \hat{x} and \hat{y} given by the initial and boundary conditions, and the wavenumber $k(\omega)$ depends on the wave frequency via eq. (5.43). All the information about the EM wave propagation is contained in the E-field solution (5.46) and the essential characteristic will be discussed in the next subsection.

5.2.3 Phase speed, group speed, and wave dispersion

For simplicity, let us consider the scalar case of the electric field solution (eq. 5.46) that is applicable to a given Cartesian component or linear polarization ($E = E_x$ or E_y). We will consider a general dispersion relation where the wavenumber $k(\omega)$ is an arbitrary non-linear scalar function of the wave frequency ω . For a wavepacket whose spectral power is concentrated in a narrow range of frequencies near ω_0 (a Gaussian wavepacket is an example). In the limit $|\omega - \omega_0| \ll \omega_0$, we Taylor expand the dispersion relation to the quadratic order

$$\begin{aligned} k(\omega) &\approx k(\omega_0) + \left. \frac{dk}{d\omega} \right|_{\omega_0} (\omega - \omega_0) + \frac{1}{2} \left. \frac{d^2k}{d\omega^2} \right|_{\omega_0} (\omega - \omega_0)^2 \\ &\approx \frac{\omega_0}{v_{ph}} + \frac{\omega - \omega_0}{v_g} + \frac{D}{2}(\omega - \omega_0)^2, \end{aligned} \quad (5.47)$$

where in the second row we have defined the following constants (all evaluated at ω_0)

$$v_{ph} \equiv \frac{\omega}{k(\omega)}, \quad v_g \equiv \left[\frac{dk(\omega)}{d\omega} \right]^{-1} = \frac{d\omega}{dk}, \quad D \equiv \frac{d^2k(\omega)}{d\omega^2}. \quad (5.48)$$

It will later on be shown that v_{ph} = phase speed, v_g = group speed, and D = dispersion constant (related to the rate of wave dispersion). For each frequency component, the phase of oscillation changes along the propagation according to the phase factor $e^{i(kz - \omega t)}$. The *phase speed* v_{ph} describes the motion of the equal-phase front $\phi = kz - \omega t = \text{const.}$

Our plan is to plug the approximate dispersion relation into the wave solution (eq. 5.46).

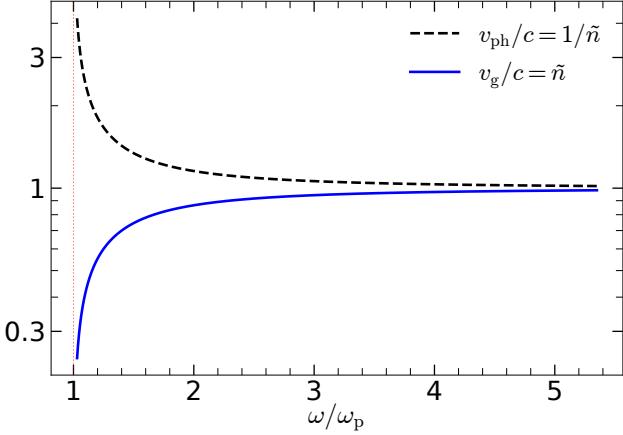


Figure 5.3: Phase and group speeds for EM waves in a non-magnetized plasma.

First, we consider only up to the linear order of the Taylor expansion, and the electric field solution is given by

$$E(z, t) \approx e^{ik_0 z(1-v_{\text{ph}}/v_g)} \int_{-\infty}^{+\infty} \tilde{E}(\omega) e^{-i\omega(t-z/v_g)} d\omega, \quad (5.49)$$

where $k_0 \equiv k(\omega_0)$. Let us define an effective ‘‘light-cone’’ coordinate

$$\xi = t - z/v_g, \quad (5.50)$$

Then, the integral is simply an inverse Fourier transform of the frequency spectrum $\tilde{E}(\omega)$ and it gives the shape of the waveform in terms of our ‘‘light-cone’’ coordinate ξ , i.e.,

$$E(\xi) = \int_{-\infty}^{+\infty} \tilde{E}(\omega) e^{-i\omega\xi} d\omega. \quad (5.51)$$

Thus, the electric field solution has the same shape as the initial waveform but at a shifted position along the ‘‘light-cone’’, i.e.,

$$E(z, t) = E(\xi) e^{ik_0 z(1-v_{\text{ph}}/v_g)}. \quad (5.52)$$

The shape of the wavepacket’s envelope, $|E(z, t)| = |E(\xi)|$ (the phase factor is unimportant here), is preserved in the ‘‘light-cone’’ coordinate $\xi = t - z/v_g$, so the entire pattern propagates at velocity $v_g = d\omega/dk$ which is the called *group speed*. This is the speed at which the energy content as well as the information in the wavepacket is carried forward.

Then, we consider the approximate dispersion relation up the quadratic term, and the electric field solution is

$$E(z, t) \approx e^{i(k_0 z - \omega_0 t)} \int_{-\infty}^{+\infty} \tilde{E}_0(\omega) e^{-i\xi(\omega - \omega_0)} e^{iDz(\omega - \omega_0)^2/2} d\omega, \quad (5.53)$$

where $k_0 = k(\omega_0)$ and $\xi = t - z/v_g$. To make further progress, let us take an example of a Gaussian wavepacket whose frequency spectrum is given by the normal distribution function

$$\tilde{E}(\omega) = \frac{1}{\sqrt{2\pi}\sigma_\omega} \exp\left[-\frac{(\omega - \omega_0)^2}{2\sigma_\omega^2}\right], \quad (5.54)$$

where σ_ω is the frequency dispersion. The wavepacket is initialized at $z = 0$ and it propagates toward the $+z$ direction. In this case, one can show that eq. (5.53) is in the form of a complex Gaussian integral and the result is given by the theorem in eq. (1.26). After some tedious work, we arrive at the following waveform solution

$$E(z, t) \approx \frac{1}{\sigma_\omega \sigma_t(z)} \exp\left(-\frac{\xi^2}{2\sigma_t^2(z)}\right) e^{i\phi(z,t)}, \quad (5.55)$$

where $\phi(z, t)$ is a complicated phase factor (not important for our purpose here) and σ_t describes the temporal width of the wavepacket

$$\sigma_t(z) = \frac{\sqrt{1 + D^2 \sigma_\omega^4 z^2}}{\sigma_\omega}. \quad (5.56)$$

We find that the center of the wavepacket (where $\xi = 0$) moves at the group speed $z = v_g t$ and that the temporal width of the wavepacket expands during the propagation — this is the effect of *wave dispersion*. At early time, when the wavepacket's center has only moved a small distance $z = v_g t \ll 1/(\sigma_\omega^2 |D|)$, the amount of dispersion is not significant; but after propagating a sufficiently long distance $z \gg 1/(\sigma_\omega^2 |D|)$, the temporal width of the wavepacket grows linearly with distance to the origin

$$\sigma_t \approx |D| \sigma_\omega z, \quad \text{for } z \gg 1/(\sigma_\omega^2 |D|), \quad (5.57)$$

and the peak amplitude drops as $(\sigma_\omega \sigma_t)^{-1} \approx (|D| \sigma_\omega^2 z)^{-1} \propto z^{-1}$ so the total energy contained in the wavepackets stays conserved.

For the special case of vacuum EM waves $k(\omega) = \omega/c$, we see that both the phase and group speeds are equal to c , i.e.,

$$v_{\text{ph}} = v_g = c, \quad \text{in vacuum}, \quad (5.58)$$

and since $D = d^2k/d\omega^2 = 0$, there is no dispersion. Thus, for a linear dispersion relation, the shape of a wavepacket is exactly preserved along the propagation.

For the case of an EM wave propagating in a non-magnetized plasma, the dispersion relation is given by eq. (5.43), so we obtain the phase speed, group speed, and dispersion constant

$$v_{\text{ph}} = \frac{\omega}{k} = \frac{c}{\sqrt{1 - \omega_p^2/\omega^2}}, \quad v_g = \frac{d\omega}{dk} = \sqrt{1 - \omega_p^2/\omega^2} c, \quad D = \frac{d^2k}{d\omega^2} = -\frac{\omega_p^2/c}{(\omega^2 - \omega_p^2)^{3/2}}. \quad (5.59)$$

We find that the phase speed in a non-magnetized plasma is always faster⁵ than c but the group speed is always less than c (in fact $v_{\text{ph}}v_g = c^2$). These two characteristic speeds are shown in Fig. 5.3. The group velocity is an increasing function of wave frequency $dv_g/d\omega > 0$ and it approaches c as $\omega \rightarrow \infty$, which means that the propagation of the highest frequency EM waves (e.g., X-rays) are essentially not affected by the plasma.

Another important feature of a plasma is the dispersion of EM waves. This is essentially because different frequency components propagate at different group speed $v_g(\omega)$.

Let us consider two nearby frequencies ω and $\omega + \Delta\omega$. The difference in the group speeds of the wave components at these two frequencies is $\Delta v_g = (dv_g/d\omega)\Delta\omega$. Suppose the plasma extends from 0 (the left boundary) to z (the right boundary), we would like to know the difference in the *arrival times* t_a after the waves propagate through the plasma

$$\Delta t_a = t_a(\omega + \Delta\omega) - t_a(\omega) = z \left[\frac{1}{v_g(\omega + \Delta\omega)} - \frac{1}{v_g(\omega)} \right] = -\frac{dv_g}{d\omega} \frac{1}{v_g^2(\omega)} z \Delta\omega = Dz \Delta\omega, \quad (5.60)$$

where D is the same as the dispersion constant defined earlier (eq. 5.47)

$$D = -\frac{dv_g}{d\omega} \frac{1}{v_g^2(\omega)} = \frac{d^2k}{d\omega^2} = -\frac{\omega_p^2/c}{(\omega^2 - \omega_p^2)^{3/2}}. \quad (5.61)$$

Note that eq. (5.60) is in agreement with the temporal dispersion of a Gaussian wavepacket in eq. (5.57) under the following interpretation: $\Delta\omega \rightarrow \sigma_\omega$ and $\Delta t \rightarrow \sigma_t$. The reason Δt and D are negative is because higher frequency waves have faster group speeds and hence will arrive at the right boundary of the plasma earlier.

In the limit of $\omega \gg \omega_p$ (well satisfied for radio waves propagating in e.g., the interstellar medium), we find

$$\frac{t_a(\omega + \Delta\omega) - t_a(\omega)}{\Delta\omega} = \frac{dt_a}{d\omega} = -\frac{\omega_p^2 z/c}{\omega^3}. \quad (5.62)$$

By integrating $dt_a/d\omega$, we obtain the difference in the arrival time for the wave component at frequency ω as compared to that at infinite frequency,

$$t_a(\omega) - t_a(\infty) = - \int_{\omega}^{\infty} \frac{dt_a}{d\omega} d\omega = \frac{\omega_p^2 z/c}{2\omega^2} = \frac{2\pi e^2}{m_e c} \frac{DM}{\omega^2} = 4.15 \text{ ms} \frac{DM}{\text{pc cm}^{-3}} (\nu/\text{GHz})^{-2}, \quad (5.63)$$

where we have used the electron plasma frequency in eq. (5.17) for electron number density n and defined the *dispersion measure* (DM) as the column density of free electrons along the propagation

$$DM = \int_0^z n dz. \quad (5.64)$$

⁵This does not violate causality as information travels at the group speed. An interesting implication of $v_{\text{ph}} > c$ is that cosmic ray particles with an arbitrarily large Lorentz factor does *not* produce Cherenkov radiation inside a plasma as the physical speed is less than the phase speed of EM waves in the medium.

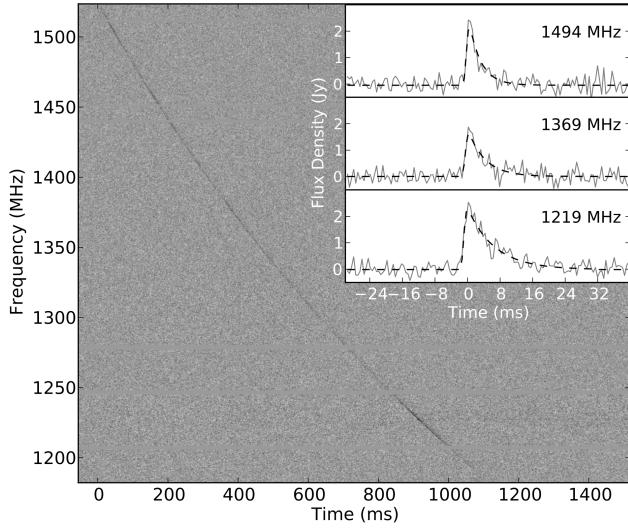


Figure 5.4: Temporal-spectral data for one of the earliest discovered fast radio bursts, FRB110220 (Thornton et al. 2013). In each frequency channel, an observer records the flux density as a function of time (see the inset). The high-frequency Fourier components of the radio pulse arrive earlier than the low-frequency components, and the arrival time has the characteristic frequency dependence of ν^{-2} as predicted. Moreover, one also sees that the duration of the pulse is longer at lower frequencies, and this is the result of multi-path propagation — lower frequency waves are refracted more by the density fluctuations in the intervening plasma and this leads to a larger spread in path-lengths (see §5.4.3).

This is in agreement with eq. (4.166) which was derived based on the superposition of the incident wave and the electron-scattered wave. The dispersion measure is conventionally expressed in units of parsec per cubic centimeter. This time delay due to plasma dispersion is measurable if the initial width of the pulse is sufficiently narrow. The dispersion delay of a fast radio burst is shown in Fig. 5.4.

5.3 Cold magnetized plasma

Now we consider that the unperturbed plasma has a large-scale magnetic field \mathbf{B}_p . The electric and magnetic fields of the wave are denoted as \mathbf{E} and \mathbf{B} . Again in this section, we ignore any thermal motion and assume the unperturbed plasma to be homogeneous and stationary. We orient the z -axis of our coordinate system to be along the direction of the plasma magnetic field, i.e.,

$$\mathbf{B}_p = B_p \hat{\mathbf{z}}. \quad (5.65)$$

A non-relativistic particle would gyrate around the plasma B-field at the *cyclotron frequency*

that is independent of the particle's velocity,

$$\omega_B = \frac{qB_p}{mc}. \quad (5.66)$$

As the electric field of an incoming EM wave pushes the particles away from their initial (equilibrium) positions, this has two effects: (i) the plasma deviates from charge neutrality and then a Coulomb electric field drives the plasma oscillation on a timescale of ω_p^{-1} ; (2) particles gain speed and then start to gyrate around the plasma B-field on a timescale of ω_B^{-1} . Now our system has three different frequencies: ω_p , ω_c , and the wave frequency ω . The interplay among them leads to interesting propagation effects.

Note that the plasma B-field can be provided by persistent currents inside or outside of the plasma region of interest. A consequence of the collisionless nature of most astrophysical plasmas is that they are nearly perfect conductors (due to very low Ohmic resistivity). This means that large-scale currents that are driven by plasma motion can persist and hence the plasma stays magnetized.

5.3.1 General dispersion relation — an eigenvalue problem

The non-relativistic equation of motion for a particle of rest-mass m and charge q is

$$m\dot{\mathbf{v}} = q[\mathbf{E} + \mathbf{v} \times (\mathbf{B} + \mathbf{B}_p)/c]. \quad (5.67)$$

We ignore the two high-order terms $(\mathbf{v} \cdot \nabla)\mathbf{v}$ (so $d\mathbf{v}/dt \approx \partial_t \mathbf{v}$) and $q\mathbf{v} \times \mathbf{B}/c$ because they are proportional to the wave amplitude squared (E_0^2), so we obtain the linearized equation of motion

$$\partial_t \mathbf{v}(\mathbf{r}, t) = (q/m)[\mathbf{E}(\mathbf{r}, t) + \mathbf{v}(\mathbf{r}, t) \times \mathbf{B}_p/c]. \quad (5.68)$$

For a plane-wave solution in the form of $\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)}$ (where \mathbf{E}_0 being a constant complex amplitude), we see that the velocity solution must also have the form of $\mathbf{v} = \mathbf{v}_0 e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)}$ where \mathbf{v}_0 is a constant complex amplitude. This means that $\partial_t \mathbf{v} = -i\omega \mathbf{v}$, so the above partial differential equation becomes an algebraic equation that only involves the complex amplitudes (as the phase factor $e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)}$ cancels out on both sides)

$$\frac{-i\omega m}{q} \mathbf{v}_0 = \mathbf{E}_0 + \mathbf{v}_0 \times \mathbf{B}_p/c. \quad (5.69)$$

A useful way of thinking about the complex amplitudes is that they are the Fourier transforms of the position- and time-dependent oscillatory waveform.

The above algebraic equation can be solved using the following trick. We cross-multiply both sides of eq. (5.69) by $\times \mathbf{B}_p$ and then eliminate $\mathbf{v}_0 \times \mathbf{B}_p$. The result is

$$-\frac{\omega^2 m^2 c^2}{q^2} \mathbf{v}_0 + \frac{i\omega m c^2}{q} \mathbf{E}_0 = c \mathbf{E}_0 \times \mathbf{B}_p - (\mathbf{v}_0 \cdot \mathbf{B}_p) \mathbf{B}_p - B_p^2 \mathbf{v}_0. \quad (5.70)$$

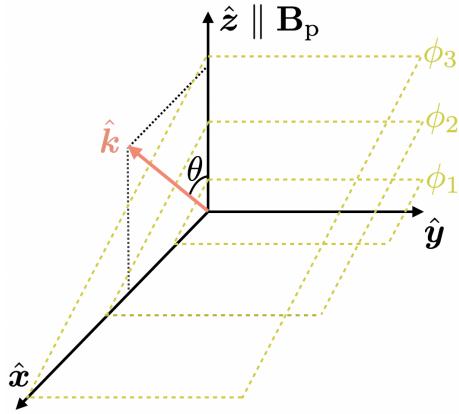


Figure 5.5: Geometry for plane-wave solutions in a magnetized plasma. The plasma B-field \mathbf{B}_p is along \hat{z} , and the wavevector \hat{k} is inside the x - z plane at an angle θ from the \hat{z} axis. The planar equal-phase fronts (marked by ϕ_1 , ϕ_2 and ϕ_3) at a given time t are shown by the yellow dashed lines.

Let us consider the z -component of the above equation (parallel to the plasma B-field), and since the RHS of the above equation is perpendicular to $\mathbf{B}_p \propto \hat{z}$, we obtain

$$v_{0,\parallel} = \frac{iq}{m\omega} E_{0,\parallel}, \quad (5.71)$$

where $v_{0,\parallel} = \mathbf{v}_0 \cdot \hat{z}$ and $E_{0,\parallel} = \mathbf{E}_0 \cdot \hat{z}$. Let us denote the perpendicular components of the velocity amplitude and electric field amplitude as $\mathbf{v}_{0,\perp}$ and $\mathbf{E}_{0,\perp}$, respectively. The perpendicular component of the velocity amplitude can be then solved from eq. (5.70)

$$\begin{aligned} \mathbf{v}_{0,\perp} &= \frac{q/m}{\omega^2 - q^2 B_p^2 / (m^2 c^2)} \left(i\omega \mathbf{E}_{0,\perp} - \frac{q}{mc} \mathbf{E}_0 \times \mathbf{B}_p \right) \\ &= \frac{q/m}{\omega^2 - \omega_B^2} (i\omega \mathbf{E}_{0,\perp} - \omega_B \mathbf{E}_0 \times \hat{z}), \end{aligned} \quad (5.72)$$

where ω_B is the electron cyclotron frequency (eq. 5.66).

The current density associated with the particle's oscillatory motion is given by $\mathbf{J} \approx n_0 q \mathbf{v}$, where n_0 is the unperturbed number density and we have ignored a high-order term due to the perturbation to the number density. We see that, for our linear plane-wave solution, the current density can also be written in the form of $\mathbf{J} = \mathbf{J}_0 e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)}$, and the complex amplitude of the current density is given by

$$\mathbf{J}_0 = n_0 q \mathbf{v}_0. \quad (5.73)$$

The above consideration is only for particles of a single species.

The total current is given by the contributions from the oscillatory motions of all particle species $\mathbf{J}_0 = \sum_s n_{0,s} q_s \mathbf{v}_{0,s}$. The parallel and perpendicular (wrt. to $\hat{\mathbf{z}}$ or the plasma B-field) components of the complex amplitude of the current density are given by

$$J_{0,\parallel} = \sum_s \omega_{p,s}^2 \frac{iE_{0,\parallel}}{4\pi\omega}, \quad \mathbf{J}_{0,\perp} = \frac{1}{4\pi} \left[\sum_s \frac{\omega_{p,s}^2}{\omega^2 - \omega_{B,s}^2} i\omega \mathbf{E}_{0,\perp} + \sum_s \frac{\omega_{p,s}^2 \omega_{B,s}}{\omega^2 - \omega_{B,s}^2} \mathbf{E}_{0,\perp} \times \hat{\mathbf{z}} \right], \quad (5.74)$$

where we have defined the cyclotron frequency $\omega_{B,s} = q_s B_p / (m_s c)$ and plasma frequency $\omega_{p,s} = \sqrt{4\pi q_s^2 n_{0,s}^2 / m_s}$ for each species.

For an electron-proton plasma ($s = e, p$), we see that $J_{0,\parallel}$ is dominated by electron plasma oscillation, but $\mathbf{J}_{0,\perp}$ is not necessarily dominated by electrons. Due to the mass scalings $\omega_{p,s}^2 \propto m_s^{-1}$ and $\omega_{B,s} \propto m_s^{-1}$, protons actually dominate the $i\omega \mathbf{E}_{0,\perp}$ term at wave frequencies below a fraction of the electron cyclotron frequency $\omega \ll \sqrt{m_e/m_p} \omega_{B,e}$ (because of protons' larger mass). We see that even for the simplest cold, magnetized electron-proton plasma, the system is controlled by a large number of free parameters (the characteristic frequencies $\omega, \omega_{p,s}, \omega_{B,s}$ and direction of propagation θ), so it is difficult to map out the entire parameter space. Here, we only illustrate the general method.

Equation (5.74) shows that the complex amplitude of the oscillatory current density driven by an EM wave can be written as the *generalized Ohm's law*

$$\mathbf{J}_0 = \overset{\leftrightarrow}{\sigma} \cdot \mathbf{E}_0, \quad (5.75)$$

where $\overset{\leftrightarrow}{\sigma}$ is called the *complex conductivity tensor*.

For the non-magnetized case, since the plasma is isotropic, the complex conductivity tensor is diagonal $\overset{\leftrightarrow}{\sigma} = \overset{\leftrightarrow}{\sigma} I$, where $\overset{\leftrightarrow}{I} = \text{diag}(1, 1, 1)$ is the identity matrix, so we have the usual Ohm's law $\mathbf{J}_0 = \sigma \mathbf{E}_0$ with a scalar conductivity

$$\sigma = \frac{i\omega_p^2}{4\pi\omega}, \quad \text{for non-magnetized plasma.} \quad (5.76)$$

The conductivity being imaginary simply means that the oscillatory current density has a phase difference of $\pi/2$ wrt. the oscillatory wave electric field. This is conceptually the same as in the case of a pendulum where the oscillation in velocity has a $\pi/2$ phase difference wrt. the oscillation in position. For the magnetized case, the conductivity tensor $\overset{\leftrightarrow}{\sigma}$ is much more complicated due to cyclotron motion of the charged particles.

Going back to the general wave equation (5.22), we use the version that only involves the complex amplitudes of the EM wave and the current density

$$\mathbf{k} \times (\mathbf{k} \times \mathbf{E}_0) + \frac{\omega^2}{c^2} \mathbf{E}_0 = -\frac{4\pi i\omega}{c^2} \mathbf{J}_0 = -\frac{4\pi i\omega}{c^2} \overset{\leftrightarrow}{\sigma} \cdot \mathbf{E}_0. \quad (5.77)$$

We then plug the complex amplitude of the current density into the above equation and arrive at the wave equation in tensor form

$$\mathbf{k} \times (\mathbf{k} \times \mathbf{E}_0) + \frac{\omega^2}{c^2} \overset{\leftrightarrow}{K} \cdot \mathbf{E}_0 = 0, \quad (5.78)$$

where $\overset{\leftrightarrow}{K}$ is called the *dielectric tensor* and is given by

$$\overset{\leftrightarrow}{K} = \overset{\leftrightarrow}{I} + \frac{4\pi i}{\omega} \overset{\leftrightarrow}{\sigma} = \begin{pmatrix} 1 - \epsilon_1 & i\epsilon_2 & 0 \\ -i\epsilon_2 & 1 - \epsilon_1 & 0 \\ 0 & 0 & 1 - \epsilon_3 \end{pmatrix}, \quad (5.79)$$

where $\overset{\leftrightarrow}{I} = \text{diag}(1, 1, 1)$ and we have used the following notations

$$\epsilon_1 = \sum_s \frac{\omega_{p,s}^2}{\omega^2 - \omega_{B,s}^2}, \quad \epsilon_2 = \sum_s \frac{\omega_{B,s} \omega_{p,s}^2 / \omega}{\omega^2 - \omega_{B,s}^2}, \quad \epsilon_3 = \frac{\sum_s \omega_{p,s}^2}{\omega^2}, \quad (5.80)$$

On the other hand, we need to specify the direction of propagation for the EM wave. Without losing generality, let us place the wavevector \mathbf{k} in the x - z plane at an angle θ from $\hat{\mathbf{z}} (\parallel \mathbf{B}_p)$ (see Fig. 5.5), so its unit vector has Cartesian coordinates

$$\hat{\mathbf{k}} = (\sin \theta, 0, \cos \theta). \quad (5.81)$$

Note that we only need to consider the angles $0 \leq \theta \leq \pi$, because the cases with $\pi/2 < \theta \leq \pi$ can be analyzed in the same way as the ones with $\pi - \theta$ if we simply flip the sign of the wavevector k (or flip our $\hat{\mathbf{z}}$ -axis). Thus, eq. (5.78) can be written in the following form

$$\overset{\leftrightarrow}{M} \cdot \mathbf{E}_0 = 0, \quad \overset{\leftrightarrow}{M} = \left[\tilde{n}^2 \left(\frac{\mathbf{k}\mathbf{k}}{k^2} - \overset{\leftrightarrow}{I} \right) + \overset{\leftrightarrow}{K} \right], \quad (5.82)$$

where $\tilde{n} \equiv kc/\omega$ is the refractive index and

$$\frac{\mathbf{k}\mathbf{k}}{k^2} - \overset{\leftrightarrow}{I} = \begin{pmatrix} -\cos^2 \theta & 0 & \sin \theta \cos \theta \\ 0 & -1 & 0 \\ \sin \theta \cos \theta & 0 & -\sin^2 \theta \end{pmatrix}, \quad (5.83)$$

is a geometrical factor where $\cos \theta = \hat{\mathbf{k}} \cdot \hat{\mathbf{z}}$. For eq. (5.82) to have non-trivial solution (with $\mathbf{E}_0 \neq 0$), the determinant $\det(\overset{\leftrightarrow}{M})$ must vanish. The eigenvectors of the matrix $\overset{\leftrightarrow}{M}$ correspond to the polarizations of the eigenmodes, and the eigenvalues are the corresponding dispersion relation for each of the eigenmodes. The *general dispersion relation* is a scalar equation

$$\det \overset{\leftrightarrow}{M} = \det \left[\tilde{n}^2 \left(\frac{\mathbf{k}\mathbf{k}}{k^2} - \overset{\leftrightarrow}{I} \right) + \overset{\leftrightarrow}{K} \right] = 0. \quad (5.84)$$

In the following, we will consider an electron-ion plasma in two regimes: one at very weak plasma magnetization and high wave frequencies with $\omega \gg \omega_{B,e}$, and the other one where the plasma is infinitely magnetized $\omega_{B,e/p} \rightarrow \infty$. The first case applies to electromagnetic waves propagating in the interstellar medium, and the second case corresponds to waves in the magnetosphere of neutron stars and accreting black holes. For this two regimes, it is sufficient to assume that the protons are stationary, meaning that the oscillatory currents are dominated by electrons⁶.

In the following sections, we will use the notations $\omega_p = \sqrt{4\pi q^2 n_{e,0}/m_e}$ for the electron plasma frequency and $\omega_B = eB_p/(m_e c)$ for the electron cyclotron frequency. Then, the components of the complex amplitude of the current density are given by

$$J_{0,\parallel} = \frac{i\omega_p^2}{4\pi\omega} E_{0,\parallel}, \quad \mathbf{J}_{0,\perp} = \frac{\omega_p^2(i\omega\mathbf{E}_{0,\perp} + \omega_B\mathbf{E}_0 \times \hat{\mathbf{z}})}{4\pi(\omega^2 - \omega_B^2)}, \quad \text{for electrons only.} \quad (5.85)$$

and the conductivity and dielectric tensors are given by

$$\overset{\leftrightarrow}{\sigma} = \frac{i\omega_p^2\omega}{4\pi(\omega^2 - \omega_B^2)} \begin{pmatrix} 1 & -i\omega_B/\omega & 0 \\ i\omega_B/\omega & 1 & 0 \\ 0 & 0 & 1 - (\omega_B/\omega)^2 \end{pmatrix}, \quad (5.86)$$

$$\overset{\leftrightarrow}{K} = \overset{\leftrightarrow}{I} + \frac{4\pi i}{\omega} \overset{\leftrightarrow}{\sigma} = \begin{pmatrix} 1 - \epsilon_1 & i\epsilon_2 & 0 \\ -i\epsilon_2 & 1 - \epsilon_1 & 0 \\ 0 & 0 & 1 - \epsilon_3 \end{pmatrix}, \quad (5.87)$$

where

$$\epsilon_1 = \frac{\omega_p^2}{\omega^2 - \omega_B^2}, \quad \epsilon_2 = \frac{\omega_B\omega_p^2/\omega}{\omega^2 - \omega_B^2}, \quad \epsilon_3 = \frac{\omega_p^2}{\omega^2}, \quad (5.88)$$

are the simplified versions (for electrons only) of the variables in eq. (5.80). It should be noted that, if the readers are interested in plasmas where species other than electrons (e.g., positrons or protons) carry significant currents, then one needs to use the general case of eq. (5.80) in the dielectric tensor — this is beyond our scope here.

5.3.2 Propagation parallel to B-field

Let us first consider wavemodes that propagate along the B-field, $\mathbf{k} \parallel \mathbf{B}_p$ or $\theta = 0$. In this case, the linearized wave equation is given by

$$\overset{\leftrightarrow}{M} \cdot \mathbf{E}_0 = 0, \quad (5.89)$$

⁶This assumption breaks down for the current $\mathbf{J}_{0,\perp}$ perpendicular to the plasma B-field at very low frequencies $\omega \ll \sqrt{m_e/m_p} \omega_{B,e}$. We will explain the consequences of relaxing this assumption in §5.3.5 where low-frequency Alfvén and fast magnetosonic modes are discussed.

where

$$\hat{\vec{M}} = \left[\tilde{n}^2 \left(\frac{\mathbf{k}\mathbf{k}}{k^2} - \hat{\vec{I}} \right) + \hat{\vec{K}} \right] = \begin{pmatrix} 1 - \tilde{n}^2 - \epsilon_1 & i\epsilon_2 & 0 \\ -i\epsilon_2 & 1 - \tilde{n}^2 - \epsilon_1 & 0 \\ 0 & 0 & 1 - \epsilon_3 \end{pmatrix}, \quad (5.90)$$

and $\tilde{n} = k(\omega)c/\omega$ is the refractive index. Thus, the dispersion relation is given by

$$\det \left(\hat{\vec{M}} \right) = 0 \Rightarrow (1 - \epsilon_3) [(1 - \tilde{n}^2 - \epsilon_1)^2 - \epsilon_2^2] = 0. \quad (5.91)$$

We therefore find the following two branches of solutions with *dispersion relations*

$$\begin{cases} \epsilon_3 = 1 \Rightarrow \omega = \omega_p, k = \text{any}, \\ \tilde{n}_\pm^2 = k_\pm^2 c^2 / \omega^2 = 1 - \epsilon_1 \pm \epsilon_2 = 1 - \frac{\omega_p^2}{\omega^2 - \omega_B^2} \pm \frac{\omega_B \omega_p^2 / \omega}{\omega^2 - \omega_B^2}, \end{cases} \quad (5.92)$$

which corresponds to three eigenmodes of a magnetized plasma: the longitudinal Langmuir mode (the branch with $\omega = \omega_p$), and the two transverse electromagnetic modes (in the other branch, one for \tilde{n}_+ and one for \tilde{n}_-). The polarization properties of each of these modes can be found by plugging their dispersion relations back into the linearized wave equation (5.89).

For the Langmuir mode with $\omega = \omega_p$, the linearized wave equation becomes

$$\begin{pmatrix} 1 - \tilde{n}^2 - \epsilon_1 & i\epsilon_2 & 0 \\ -i\epsilon_2 & 1 - \tilde{n}^2 - \epsilon_1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} E_{0,x} \\ E_{0,y} \\ E_{0,z} \end{pmatrix} = 0. \quad (5.93)$$

We then find that $E_{0,x} = E_{0,y} = 0$ because of the non-zero determinant of the 2×2 submatrix involving the x and y components, $1 - \tilde{n}^2 - \epsilon_1 \pm \epsilon_2 \neq 0$. Thus, for the Langmuir mode, the only non-zero amplitude component is $E_{0,z}$, which means that this is longitudinal oscillation. In fact, for such a longitudinal oscillation where electrons only move along the B-field lines, the system is identical to a non-magnetized plasma because the Lorentz force vanishes $\mathbf{v} \times \mathbf{B}_p = 0$.

Let us then look at the electromagnetic modes for which $\tilde{n}_\pm^2 = 1 - \epsilon_1 \pm \epsilon_2$. For these two modes, the linearized wave equation is

$$\begin{pmatrix} \mp\epsilon_2 & i\epsilon_2 & 0 \\ -i\epsilon_2 & \mp\epsilon_2 & 0 \\ 0 & 0 & 1 - \epsilon_3 \end{pmatrix} \cdot \begin{pmatrix} E_{0,x} \\ E_{0,y} \\ E_{0,z} \end{pmatrix} = 0. \quad (5.94)$$

We find that $E_{0,z} = 0$ (because $1 - \epsilon_3 \neq 0$) and that the complex amplitude of the two transverse components are related to each other by

$$\epsilon_2(E_{0,x} \mp iE_{0,y}) = 0. \quad (5.95)$$

If $\epsilon_2 \neq 0$, we obtain two circularly polarized modes⁷ with $E_{0,x} = \pm iE_{0,y}$ for the two modes with refractive indices \tilde{n}_\pm , respectively.

5.3.3 Propagation along an arbitrary direction

In the more general case of an arbitrary propagation direction ($0 < \theta \leq \pi/2$), the full expression for the dispersion relation is very lengthy and difficult to understand. To simplify the expression somewhat, we use a more convenient variable

$$Y \equiv \tilde{n}^2 - 1 + \omega_p^2/\omega^2, \quad (5.96)$$

where $\tilde{n}(\omega) = k(\omega)c/\omega$ is the refractive index. Then, the dispersion relation for *electrons only* is given by

$$\begin{aligned} & \left(Y + \frac{\omega_B^2 \omega_p^2}{\omega^2(\omega^2 - \omega_B^2)} \right) \left[Y + \left(\cos^2 \theta - \frac{Y \sin^2 \theta}{1 - \omega_p^2/\omega^2} \right) \frac{\omega_B^2 \omega_p^2}{\omega^2(\omega^2 - \omega_B^2)} \right] \\ &= \frac{\omega_B^2 \omega_p^4}{\omega^2(\omega^2 - \omega_B^2)^2} \left(\cos^2 \theta - \frac{Y \sin^2 \theta}{1 - \omega_p^2/\omega^2} \right), \end{aligned} \quad (5.97)$$

where no approximation has been made regarding the propagation direction θ and the hierarchy of the three frequencies ω , ω_p , and ω_B , except that $\omega \neq \omega_p$ and $\omega \neq \omega_B$ (to avoid resonances). The dispersion relation is generally a quadratic equation for Y , so it has two solutions which correspond to two eigenmodes. The general polarization properties of these two eigenmodes are rather complicated, so we will only discuss two limiting cases of a weakly magnetized plasma (§5.3.4) and an infinitely magnetized plasma (§5.3.5).

The propagating solutions have $\tilde{n}^2 > 0$, and each such propagating eigenmode permit a positive and negative \tilde{n} (or wavenumber k) that simply correspond to the same physical mode propagating in opposite directions. Note that only in the special case of $\mathbf{k} \parallel \mathbf{B}_p$ ($\theta = 0$), the non-propagating longitudinal Langmuir mode with $\omega = \omega_p$ is allowed.

5.3.4 Weakly magnetized plasma: Faraday rotation and conversion

In this subsection, we consider the case of a weakly magnetized electron-proton plasma where the EM wave frequency is much above the cyclotron resonance $\omega \gg \omega_B$. Regarding the plasma frequency, we will also restrict ourselves to the limit of $\omega \gg \omega_p$, because in the opposite limit of $\omega \ll \omega_p$ the EM wave is below the cutoff frequency and hence becomes evanescent. Thus, the following discussion applies if

$$\omega \gg \max(\omega_p, \omega_B). \quad (5.98)$$

⁷Later in §5.3.5, we will discuss the case with infinite magnetization $B \rightarrow \infty$ and in that case $\epsilon_2 \rightarrow 0$, so we do not recover the two circularly polarized modes discussed here.

A physical example would be radio waves propagating through the interstellar medium.

We will discuss three different cases: (1) parallel ($\theta = 0$), (2) quasi-parallel ($\theta \approx \pi/2$), and (3) quasi-perpendicular ($\theta \approx \pi/2$). The meanings of cases (2) and (3) will become clear soon. We will show that, in cases (1) and (2), the propagating electromagnetic eigenmodes are primarily circularly polarized transverse waves. In these two cases, the most important consequence of the plasma B-field is that the left and right circularly polarized (LCP and RCP) modes propagate at different phase (and group) speeds, and this leads to the famous effect of *Faraday rotation*. In case (3), we will see that the electromagnetic eigenmodes are primarily linearly polarized transverse waves (called the O and X modes), and the plasma B-field causes these two modes to propagate at different phase/group speeds, and this leads to the effect of *Faraday conversion* — an elliptically polarized wave containing a mixture of O and X modes may convert between elliptical and linear polarizations back and forth.

Parallel case ($\theta = 0$) and Faraday rotation

We focus on the branch of propagating electromagnetic waves whose dispersion relation is given by (cf. eq. 5.92)

$$\begin{aligned}\tilde{n}_\pm^2 &= \left(\frac{k_\pm c}{\omega}\right)^2 = 1 - \frac{\omega_p^2}{\omega^2 - \omega_B^2} \pm \frac{\omega_B \omega_p^2 / \omega}{\omega^2 - \omega_B^2} \\ &\approx 1 - \frac{\omega_p^2}{\omega^2} \pm \frac{\omega_B \omega_p^2}{\omega^3},\end{aligned}\tag{5.99}$$

where the approximation applies for a weakly magnetized plasma with $\omega \gg \omega_B$. In the following, we take the solution with positive k_\pm without losing generality (the negative solutions simply correspond to flipping the z -axis) and write the dispersion relation as

$$\tilde{n}_\pm = \frac{k_\pm c}{\omega} \approx 1 - \frac{\omega_p^2}{2\omega^2} \pm \frac{\omega_B \omega_p^2}{2\omega^3},\tag{5.100}$$

where we have additionally assumed $\omega \gg \omega_p$ and ignored a higher-order term $\mathcal{O}(\omega_p^4/\omega^4)$.

If we plug the dispersion relation of $\tilde{n}_\pm^2 = 1 - \epsilon_1 \pm \epsilon_2$ back into the wave equation (5.89), this leads to $E_{0,x} = \pm i E_{0,y}$, which means that the two electromagnetic modes are circularly polarized. This motivates us to define the base vectors for the right circularly polarization (RCP) and left circular polarization (LCP) modes as

$$\hat{\mathbf{R}} = (\hat{x} - i\hat{y})/\sqrt{2}, \quad \hat{\mathbf{L}} = (\hat{x} + i\hat{y})/\sqrt{2},\tag{5.101}$$

so the complex amplitudes of the two modes corresponding to \tilde{n}_\pm can be written as

$$\mathbf{E}_{0,+} = E_0 \hat{\mathbf{R}}, \quad \mathbf{E}_{0,-} = E_0 \hat{\mathbf{L}}.\tag{5.102}$$

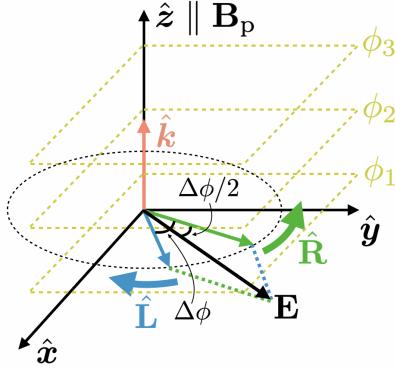


Figure 5.6: High-frequency, $\omega \gg \max(\omega_p, \omega_B)$, plane-wave solution in the parallel case ($\theta = 0$), where both plasma B-field \mathbf{B}_p and the wavevector \mathbf{k} are along the z -axis. The RCP mode has electric vector rotating counter-clockwise inside the x - y plane, whereas the electric vector of the LCP mode rotates the other way. Propagation leads to a phase difference of $\Delta\phi$ between the two modes, which then causes Faraday rotation of the linear polarization angle by an angle $\Delta\phi/2$. The eigenmodes are generally circularly polarized as long as $\theta = \arccos(\hat{\mathbf{k}} \cdot \hat{\mathbf{B}}_p)$ is not very close to $\pi/2$, and the Faraday rotation angle $\Delta\phi/2$ still applies as long as we consider the component of the plasma B-field projected onto the wavevector.

The RCP (\tilde{n}_+) and LCP (\tilde{n}_-) modes propagate at different phase and group speeds, and the consequences will be discussed in the following.

The phase speed difference leads to the effect of *Faraday rotation*. Consider that a monochromatic wave is initially (at $z = 0$) linearly polarized, meaning that the RCP and LCP modes have the same amplitude and phase, so the initial complex amplitude⁸ can be written as $\mathbf{E}_0(z = 0) = (\hat{\mathbf{R}} + \hat{\mathbf{L}})/\sqrt{2} = \hat{\mathbf{x}}$ (using the circular basis in eq. 5.101 and normalized to unity physical amplitude). After propagating for a distance z along the plasma B-field ($\mathbf{B}_p \parallel \hat{z}$), the complex amplitude becomes

$$\mathbf{E}_0(z) = \frac{1}{\sqrt{2}} (\hat{\mathbf{R}} e^{ik_+ z} + \hat{\mathbf{L}} e^{ik_- z}) = \frac{1}{\sqrt{2}} e^{ik_- z} (\hat{\mathbf{R}} e^{i\Delta\phi(z)} + \hat{\mathbf{L}}), \quad (5.103)$$

where the phase difference between the two modes is given by

$$\Delta\phi(z) = (k_+ - k_-)z = \frac{\omega z \omega_B \omega_p^2}{c} \frac{\omega^2}{\omega^3}. \quad (5.104)$$

The superposition of two RCP and LCP modes of the same amplitude but with an arbitrary phase difference is always linearly polarized, and this can be seen by re-writing the complex

⁸In our notation here, the full waveform is $\mathbf{E}_0 e^{-i\omega t}$ and the physical electric field is $\text{Re}(\mathbf{E}_0 e^{-i\omega t})$.

amplitude by pulling a phase factor of $e^{i\Delta\phi/2}$ out of the both modes

$$\begin{aligned}\mathbf{E}_0(z) &= \frac{1}{\sqrt{2}} e^{i(k_z z + \Delta\phi/2)} \left(\hat{\mathbf{R}} e^{i\Delta\phi/2} + \hat{\mathbf{L}} e^{-i\Delta\phi/2} \right) \\ &= \frac{1}{2} e^{i(k_z z + \Delta\phi/2)} \left((\hat{\mathbf{x}} - i\hat{\mathbf{y}}) e^{i\Delta\phi/2} + (\hat{\mathbf{x}} + i\hat{\mathbf{y}}) e^{-i\Delta\phi/2} \right) \\ &= e^{i(k_z z + \Delta\phi/2)} [\hat{\mathbf{x}} \cos(\Delta\phi/2) + \hat{\mathbf{y}} \sin(\Delta\phi/2)].\end{aligned}\quad (5.105)$$

This shows that, by propagating for a distance of z , the Jones vector in linear basis of $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ has undergone the following change

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} \cos \frac{\Delta\phi}{2} \\ \sin \frac{\Delta\phi}{2} \end{pmatrix} e^{i(k_z z + \Delta\phi/2)}. \quad (5.106)$$

We see that the effect of Faraday rotation has shifted the polarization angle χ by

$$\Delta\chi = \frac{\Delta\phi(z)}{2} = \frac{\omega_B \omega_p^2 \ell}{2\omega^2 c}. \quad (5.107)$$

We will later show that the effect of Faraday rotation for the simple case of wave propagation parallel ($\theta = \arccos(\hat{\mathbf{k}} \cdot \mathbf{B}_p) = 0$) to the plasma B-field can be generalized for any quasi-parallel case as long as θ is not extremely close to $\pi/2$. In the more general case of a finite inclination angle between $\hat{\mathbf{k}}$ and the plasma B-field, we find the polarization angle shift to be given by

$$\Delta\chi = \text{RM} \lambda^2 \quad (5.108)$$

where $\lambda = c/\nu$ is the wavelength and the *rotation measure* (RM) is defined as

$$\text{RM} = \frac{e^3}{2\pi m_e^2 c^4} n z B_{p,\parallel} = 0.812 \text{ rad m}^{-2} \frac{\int n B_{p,\parallel} dz}{\text{cm}^{-3} \mu\text{G pc}}, \quad (5.109)$$

Note that we have denoted the projected plasma B-field along the line of sight as $\mathbf{B}_p \cdot \hat{\mathbf{k}} = B_p \cos \theta = B_{p,\parallel}$. The effect of Faraday rotation is demonstrated in Fig. 5.7.

If the wavepacket is initially linearly polarized with the same polarization angle at all frequencies (i.e. the intrinsic χ_0 is independent of frequency), then after propagating through a magnetized plasma the observed polarization angle profile will be

$$\chi(\lambda) = \chi_0 + \text{RM} \lambda^2. \quad (5.110)$$

Observationally, one can measure $\chi(\lambda)$ from the Stokes parameters, so this method can be used to infer the RM of the intervening plasma as well as the intrinsic polarization angle of the source.

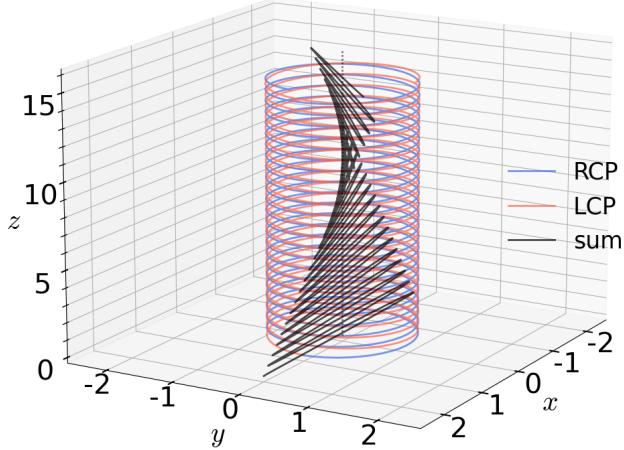


Figure 5.7: The effect of Faraday rotation as a result of different phase speeds for the RCP and LCP modes. In this example, there is a fractional difference of 1% between the wavenumbers for the two modes, and the two modes have the same amplitude and frequency. The red, blue, and black curves trace the tips of the electric vectors for the RCP component, LCP component, and the sum of the two (linear polarized), respectively.

Another consequence of the different refractive indices for the RCP and LCP modes is that they have different group speeds (a general finite inclination angle $\theta \neq \pi/2$ is included)

$$\frac{v_{g,\pm}}{c} = \left(\frac{dk_\pm}{d\omega} \right)^{-1} = 1 - \frac{\omega_p^2}{2\omega^2} \pm \frac{\omega_B \omega_p^2 \cos \theta}{\omega^3}. \quad (5.111)$$

After propagating a large distance $z = \int dz$, the Fourier components at different angular frequencies ω will arrive at different times

$$\begin{aligned} t_{a,\pm}(\omega) - t_a(\infty) &= \frac{\int dz}{c} \frac{\omega_p^2}{2\omega^2} \mp \frac{\int dz}{c} \frac{\omega_B \omega_p^2 \cos \theta}{\omega^3} \\ &= 4.15 \text{ ms} \frac{\text{DM}}{\text{pc cm}^{-3}} \left(\frac{\text{GHz}}{\nu} \right)^2 \mp 2.86 \times 10^{-11} \text{ s} \frac{\text{RM}}{\text{rad m}^{-2}} \left(\frac{\text{GHz}}{\nu} \right)^3. \end{aligned} \quad (5.112)$$

The first term involving the dispersion measure ($\text{DM} = \int n_e dz$) typically dominates the frequency-dependent arrival time, but the second term might have some observable consequences in an extremely magnetized plasma. For instance, for $\text{RM} = 10^5 \text{ rad m}^{-2}$, the arrival time difference between the RCP and LCP modes is about $0.6 \mu\text{s} (\nu/\text{GHz})^{-3}$. If this is longer than the duration of the wavepacket, then a linearly polarized wavepacket will split into two temporally separated RCP and LCP components. However, the duration of

the wavepacket increases as a result of plasma dispersion (the DM term), so the two modes are generally not expected to detach from each other as long as $\omega_B \ll \omega$.

General propagating angles ($\theta \neq 0$)

Going back to the dispersion relation (eq. 5.97) for electromagnetic waves propagating along a general angle θ , for a weakly magnetized, dilute plasma with $\omega \gg \max(\omega_p, \omega_B)$, we find $Y = \tilde{n}^2 - 1 + \omega_p^2/\omega^2 = \ll \omega_p^2/\omega^2$ to be only a small perturbation term. Therefore, it is appropriate to drop the $Y \sin^2 \theta / (1 - \omega_p^2/\omega^2)$ term (which shows up twice) and other higher-order small terms, and then the dispersion relation becomes

$$Y^2 + Y(1 + \cos^2 \theta) \frac{\omega_B \epsilon}{\omega} - \epsilon^2 \cos^2 \theta = 0. \quad (5.113)$$

where we have used the following short-hand notation

$$\epsilon \equiv \omega_B \omega_p^2 / \omega^3 \ll 1. \quad (5.114)$$

This quadratic equation has two solutions (corresponding to two eigenmodes),

$$Y_{\pm} = \frac{1 + \cos^2 \theta}{2} \frac{\omega_B \epsilon}{\omega} \left(-1 \pm \sqrt{1 + \frac{4 \cos^2 \theta}{(1 + \cos^2 \theta)^2} \frac{\omega^2}{\omega_B^2}} \right), \quad \text{for } \omega \gg \max(\omega_p, \omega_B). \quad (5.115)$$

Once the dispersion relation for a given mode is solved, one can plug it back into the (approximated) wave equation to obtain the polarization. In the limit $\omega \gg \max(\omega_p, \omega_B)$, the dielectric tensor is roughly given by

$$\overset{\leftrightarrow}{K} \approx (1 - \omega_p^2/\omega^2) \overset{\leftrightarrow}{I} + \begin{pmatrix} 0 & i\epsilon & 0 \\ -i\epsilon & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (5.116)$$

and we obtain an approximate wave equation

$$\left[\tilde{n}^2 \begin{pmatrix} -\cos^2 \theta & 0 & \sin \theta \cos \theta \\ 0 & -1 & 0 \\ \sin \theta \cos \theta & 0 & -\sin^2 \theta \end{pmatrix} + \left(1 - \frac{\omega_p^2}{\omega^2} \right) \overset{\leftrightarrow}{I} + \begin{pmatrix} 0 & i\epsilon & 0 \\ -i\epsilon & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right] \cdot \begin{pmatrix} E_{0,x} \\ E_{0,y} \\ E_{0,z} \end{pmatrix} = 0. \quad (5.117)$$

In the following, we discuss two limiting regimes: the quasi-parallel case ($\theta \not\approx \pi/2$) and the quasi-perpendicular case ($\theta \approx \pi/2$).

Quasi-parallel case

We first study the regime of $\cos \theta \gg \omega_B/\omega$. We call this regime *quasi-parallel* but the name is a bit misleading because we actually do not require the wavevector to be very nearly parallel to the plasma B-field. At high wave frequencies $\omega \gg \omega_B$, the following discussion applies as long as $\theta \not\approx \pi/2$. Thus, this case is much more commonly encountered than the quasi-perpendicular case to be discussed later. Radio waves propagating in the interstellar medium should generally be in this regime.

Our Y parameter in eq. (5.115) reduces to⁹

$$Y_{\pm} \approx \pm \epsilon \cos \theta, \quad \text{for } \cos \theta \gg \omega_B/\omega, \quad (5.118)$$

Then, the dispersion relation becomes

$$\tilde{n}_{\pm}^2 = \left(\frac{k_{\pm}c}{\omega} \right)^2 = 1 - \frac{\omega_p^2}{\omega^2} \pm \epsilon \cos \theta, \quad \epsilon = \frac{\omega_B \omega_p^2}{\omega^3}, \quad (5.119)$$

or

$$\tilde{n}_{\pm} = \frac{k_{\pm}c}{\omega} \approx 1 - \frac{\omega_p^2}{2\omega^2} \pm \frac{\epsilon \cos \theta}{2}. \quad (5.120)$$

These two modes are very nearly circularly polarized: the \tilde{n}_+ mode corresponds to right circular polarization (RCP) and the \tilde{n}_- mode is for left circular polarization (LCP). By inserting the dispersion relation (5.119) back into the approximate wave equation (5.117), one obtains the following ratios

$$E_{0,x}/E_{0,y} = \pm i \cos \theta, \quad E_{0,z}/E_{0,x} = -\tan \theta (1 \pm \epsilon / \cos \theta). \quad (5.121)$$

If we define another unit vector $\hat{x}' = (\cos \theta, 0, -\sin \theta) = \hat{y} \times \hat{k}$ and then the base vectors for the RCP and LCP modes can be written as

$$\hat{\mathbf{R}} = (\hat{x}' - i\hat{y}) / \sqrt{2}, \quad \hat{\mathbf{L}} = (\hat{x}' + i\hat{y}) / \sqrt{2}. \quad (5.122)$$

When we ignore the very small term proportional to ϵ (for a non-magnetized plasma), the $+$ and $-$ modes are perfectly circularly polarized

$$\mathbf{E}_{0,+} \propto \hat{\mathbf{R}} \text{ and } \mathbf{E}_{0,-} \propto \hat{\mathbf{L}}. \quad (5.123)$$

If we include the ϵ term (for non-zero magnetization), then there is a very small electric field component along the \hat{k} direction

$$|\mathbf{E}_0 \cdot \hat{k}| / |\mathbf{E}_0| = \epsilon \sin \theta, \quad (5.124)$$

⁹The more accurate result is $Y_{\pm} \approx \pm \epsilon \cos \theta - \omega_B \epsilon (1 + \cos^2 \theta) / (2\omega)$, and we are dropping the second term which is smaller than the first term by a factor of $\omega_B/\omega \ll 1$ for most propagation directions (and the second term is unimportant for Faraday rotation).

which only vanishes when $\theta = 0$ or $\mathbf{k} \parallel \mathbf{B}_p$). This shows that + and - modes are nearly but not exactly circularly polarized, but practically they can be considered as RCP and LCP modes.

Like what has been discussed earlier in the parallel case, the RCP (\tilde{n}_+) and LCP (\tilde{n}_-) modes propagate at different phase and group speeds. The phase speed difference leads to the effect of Faraday rotation. If we consider a monochromatic wave that is initially linearly polarized (the RCP and LCP modes having the same amplitude and phase), after propagating for a distance ℓ along the direction of the wavevector $\hat{\mathbf{k}}$, the two modes will acquire a phase difference of

$$\Delta\phi(\ell) = (k_+ - k_-)\ell = \frac{\omega\ell}{c}\epsilon \cos\theta. \quad (5.125)$$

It is possible to show that this phase difference leads to Faraday rotation of the polarization angle χ by

$$\Delta\chi = \frac{\Delta\phi(\ell)}{2} = \frac{\omega_B\omega_p^2\ell\cos\theta}{2\omega^2c} = \text{RM } \lambda^2, \quad (5.126)$$

where $\lambda = 2\pi c/\omega$ is the wavelength and

$$\text{RM} = \frac{e^3}{2\pi m_e^2 c^4} n \ell B_{p,\parallel} \quad (5.127)$$

is the rotation measure.

Quasi-perpendicular case

When the wavevector \mathbf{k} is very nearly perpendicular to the plasma B-field \mathbf{B}_p such that $\cos\theta \ll \omega_B/\omega$, there are two linearly polarized modes: one has its electric vector parallel to \mathbf{B}_p and is hence called the O-mode ("O" for ordinary); the other one has its electric vector perpendicular to the \mathbf{k} - \mathbf{B}_p plane and is hence called the X-mode ("X" for extraordinary).

Taking the limit of $\cos\theta \approx 0$ in eq. (5.115), we obtain

$$Y_O = 0, \text{ and } Y_X = -\omega_B^2\omega_p^2/\omega^4, \quad \text{for } \cos\theta \ll \omega_B/\omega, \quad (5.128)$$

and the corresponding dispersion relations are

$$\begin{aligned} \tilde{n}_O^2 &= 1 - \omega_p^2/\omega^2, \text{ for O-mode with } \mathbf{E}_0 \parallel \mathbf{B}_p, \\ \tilde{n}_X^2 &= 1 - \omega_p^2/\omega^2 - \omega_B^2\omega_p^2/\omega^4, \text{ for X-mode with } \mathbf{E}_0 \perp (\mathbf{k} \& \mathbf{B}_p). \end{aligned} \quad (5.129)$$

For the O-mode, the plasma magnetic field does not play any role as particles' oscillatory motion is parallel to \mathbf{B}_p , so the refractive index is identical to the case in a non-magnetized

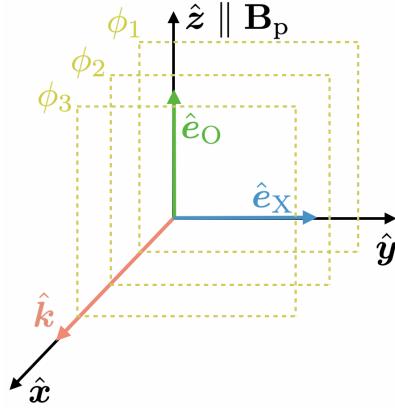


Figure 5.8: High-frequency, $\omega \gg \max(\omega_p, \omega_B)$, plane-wave solution in the *quasi-perpendicular regime* ($\theta \approx \pi/2$). The plasma B-field \mathbf{B}_p is along \hat{z} , and the wavevector \mathbf{k} is parallel to the x -axis. The X-mode has electric vector parallel to \hat{y} , whereas the O-mode has electric vector parallel to \hat{z} .

plasma. For the X-mode, the plasma magnetic field plays a role to reduce the refractive index as compared to that of the O-mode. One can plug the dispersion relations for these two modes back into the wave equation to calculate their polarizations. The O-mode is nearly linearly polarized with electric field components

$$E_{0,x} = E_0 \cot \theta \approx 0, \quad E_{0,y} = 0, \quad E_{0,z} \approx E_0, \quad \text{for O-mode.} \quad (5.130)$$

The X-mode is also nearly linearly polarized with electric field components¹⁰

$$E_{0,x} = i\epsilon E_{0,y} \approx 0, \quad E_{0,y} \approx E_0, \quad E_{0,z} = -\frac{\sin \theta \cos \theta}{\epsilon \omega_B / \omega + \cos^2 \theta} E_{0,x} \approx 0, \quad \text{for X-mode.} \quad (5.131)$$

The difference in phase speeds, $v_{ph} = c/\tilde{n}$, between these two modes can lead to an interesting phenomenon called *Faraday conversion*. Suppose a wavepacket is initially linear polarized with both O-mode and X-mode components. After propagating for a distance ℓ within the plasma (while remaining in the quasi-perpendicular regime), these two modes will acquire a phase difference

$$\Delta\phi = (k_O - k_X)\ell = \frac{\omega\ell}{c} (\tilde{n}_O - \tilde{n}_X) = \frac{\omega_B^2 \omega_p^2 \ell}{\omega^3 c} \propto \frac{n B_p^2 \ell}{\nu^3}, \quad (5.132)$$

where n is the plasma electron density. We see that, along the propagation, the wave polarization can convert from linear to elliptical and back and forth. The criterion for

¹⁰To obtain the polarization of the X-mode in the high-frequency limit $\omega \gg \max(\omega_p, \omega_B)$, one has to use the full dielectric tensor in eq. (5.79) (the approximation in eq. 5.116 is not sufficient) together with the refractive index with higher-order terms $\tilde{n}_X^2 = 1 - \omega_p^2/(\omega^2 - \omega_B^2) - \omega_B^2 \omega_p^4/\omega^6$.

significant Faraday conversion is $\Delta\phi \gtrsim 1$ rad. It should be noted that, if $\omega_B \ll \omega$ (for radio waves propagating in the interstellar medium), Faraday conversion requires a fine-tuned geometry because the wavevector must be extremely close to being perpendicular to the plasma B-field, $\cos\theta \ll \omega_B/\omega \ll 1$.

5.3.5 Infinite magnetization: Alfvén, fast magnetosonic, O, and X modes

Next we consider that the plasma is extremely magnetized such that the cyclotron frequencies for all particle species are practically infinite for the EM waves we are interested in. We take the limit of $\omega_B \rightarrow \infty$, so the only possible oscillating currents are parallel to the plasma B-field — charged particles only move along the B-field lines like beads on a wire. In this limit, the dielectric tensor (eq. 5.79) reduces to

$$\overset{\leftrightarrow}{K} \approx \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 - \omega_p^2/\omega^2 \end{pmatrix}, \quad \text{for } \omega_B \rightarrow \infty, \quad (5.133)$$

and the dispersion relation (eq. 5.97) reduces to

$$(\tilde{n}^2 - 1) [\tilde{n}^2 (\omega_p^2 \cos^2 \theta - \omega^2) + \omega^2 - \omega_p^2] = 0. \quad (5.134)$$

The above expression applies to any plasma frequencies as long as $\omega_p \ll \omega_B$. We find the following two branches of solutions

$$\begin{aligned} \tilde{n}_{O/A}^2 &= \frac{\omega^2 - \omega_p^2}{\omega^2 - \omega_p^2 \cos^2 \theta}, \quad \text{for O-mode or Alfvén mode,} \\ \tilde{n}_{X/fms}^2 &= 1, \quad \text{for X-mode or fast magnetosonic mode.} \end{aligned} \quad (5.135)$$

The function $\tilde{n}_{O/A}(\omega)$ is shown in Fig. 5.9.

The polarization of these modes can be found by plugging the dispersion relation into the wave equation under the approximate dielectric tensor (5.133). The eigenvector corresponding to the $\tilde{n}_{O/A}$ dispersion relation is described by

$$E_{0,z}/E_{0,x} = -\frac{\tan \theta}{1 - \omega_p^2/\omega^2}, \quad E_{0,y} = 0. \quad (5.136)$$

This shows a linearly polarized EM wave with electric vector inside the \mathbf{k} - \mathbf{B}_p plane. The electric vector is generally not perpendicular to either the wavevector $\hat{\mathbf{k}} = (\sin \theta, 0, \cos \theta)$ or the plasma B-field $\mathbf{B}_p \parallel \hat{\mathbf{z}}$. In the high-frequency limit of $\omega \gg \omega_p$ (but below the cyclotron frequency), we obtain

$$\tilde{n}_O^2 \approx 1 - \frac{\omega_p^2}{\omega^2} \sin^2 \theta \approx 1, \quad E_{0,z}/E_{0,x} \approx -\tan \theta, \quad E_{0,y} = 0. \quad (5.137)$$

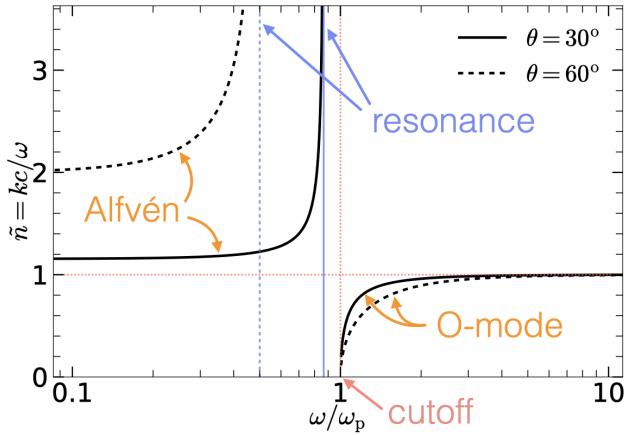


Figure 5.9: The refractive indices of the Alfvén mode (low-frequency branch) and the O-mode (high-frequency branch) in the limit of an infinite plasma magnetic field. The black solid and dashed lines are for two different propagation directions with $\theta = 30^\circ$ and $\theta = 60^\circ$, respectively. In between the resonance frequency ($\tilde{n} \rightarrow \infty$) and cutoff frequency ($\tilde{n} = 0$) of the two branches, there is an evanescent region where $\cos \theta < \omega/\omega_p < 1$.

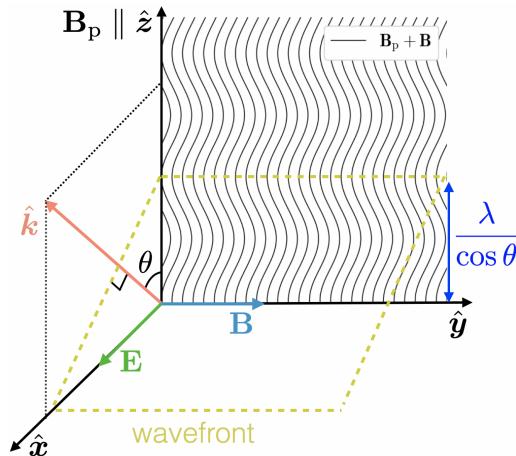


Figure 5.10: Alfvén wave with an inclined wavevector \hat{k} wrt. the plasma B-field \mathbf{B}_p . Only the magnetic field lines inside the y - z plane are shown. Here, \hat{k} is inside the x - z plane, $\mathbf{B}_p \parallel \hat{z}$, $\mathbf{E} \parallel \hat{x}$, and $\mathbf{B} \parallel \hat{y}$. Note that the B-field lines do not get compressed as $\mathbf{B} \cdot \mathbf{B}_p = 0$.

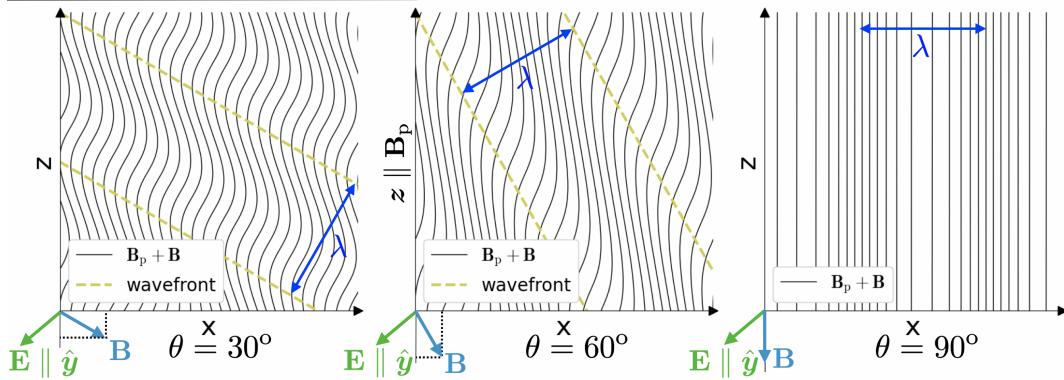


Figure 5.11: Fast magnetosonic wave propagating at different inclination angles θ between $\hat{\mathbf{k}}$ and \mathbf{B}_p . Here, both $\hat{\mathbf{k}}$ and \mathbf{B} are inside the x - z plane, $\mathbf{B}_p \parallel \hat{\mathbf{z}}$, and $\mathbf{E} \parallel \hat{\mathbf{y}}$. Note the compression of B-field lines as $\mathbf{B} \cdot \mathbf{B}_p \neq 0$.

This is called the *O-mode*, as it has electric vector inside the \mathbf{k} - \mathbf{B}_p plane and perpendicular to the wavevector $\hat{\mathbf{k}}$. On the other hand, in the low-frequency limit of $\omega \ll \omega_p$, we obtain

$$\tilde{n}_A^2 \approx \frac{1}{\cos^2 \theta} \left(1 - \frac{\omega^2 \tan^2 \theta}{\omega_p^2} \right) \approx \frac{1}{\cos^2 \theta}, \quad E_{0,x} = E_0, \quad E_{0,y} = E_{0,z} = 0. \quad (5.138)$$

This is called the *Alfvén mode*. As illustrated in Fig. 5.10, the Alfvén mode has electric vector inside the \mathbf{k} - \mathbf{B}_p plane and perpendicular to the plasma B-field, and the wave magnetic field¹¹ $\mathbf{B}_0 = (c/\omega)\mathbf{k} \times \mathbf{E}_0$ is perpendicular to the \mathbf{k} - \mathbf{B}_p plane. The dispersion relation of the Alfvén mode can be written as

$$\omega = c\mathbf{k}_A \cdot \hat{\mathbf{B}}_p, \quad (5.139)$$

where $\hat{\mathbf{B}}_p$ is a unit vector along the unperturbed plasma B-field. Thus, the group speed of the Alfvén mode is given by

$$\mathbf{v}_{g,A} = \frac{d\omega}{d\mathbf{k}_A} = c\hat{\mathbf{B}}_p. \quad (5.140)$$

We see that the energy of the wavepacket propagates along the unperturbed plasma B-field instead of the wavevector.

The eigenvector correspond to the $\tilde{n}_{X/fms}$ dispersion relation is described by

$$E_{0,x} = E_{0,z} = 0, \quad E_{0,y} = E_0. \quad (5.141)$$

¹¹Remind ourselves that $-\partial_t \mathbf{B} = c\nabla \times \mathbf{E}$ and, for a plane wave where both $\mathbf{B}(\mathbf{r}, t)$ and $\mathbf{E}(\mathbf{r}, t)$ depend on spacetime coordinates as $e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)}$, we can do the following replacements $\partial_t \rightarrow -i\omega$ and $\nabla \rightarrow i\mathbf{k}$.

This is called the *X-mode* at high frequencies $\omega > \omega_p$ and the *fast magnetosonic (fms) mode*¹² at low frequencies $\omega < \omega_p$. The propagation of the X-mode and the fms-mode are unaffected by the plasma in the limit of infinite magnetization, because the electric vector is perpendicular to the plasma B-field but charged particles are not allowed to move perpendicular to the B-field. This is why the dispersion relation is the same as that in vacuum. The wave magnetic field $\mathbf{B}_0 = (c/\omega)\mathbf{k} \times \mathbf{E}_0$ is inside the \mathbf{k} - \mathbf{B}_p plane.

The fms and Alfvén modes can be considered as magnetic perturbations to the plasma B-field in the picture of magnetohydrodynamics. A major difference between the fms mode and the Alfvén mode is that the former involves compression (as $\mathbf{B} \cdot \mathbf{B}_p \neq 0$) of the B-field lines whereas the latter does not (as $\mathbf{B} \cdot \mathbf{B}_p = 0$). Thus, Alfvén waves are transverse incompressible magnetic oscillations (like shear waves), whereas fms waves are compressible (like pressure waves).

Finally, we comment on the case of finite plasma magnetization. The dimensionless *magnetization parameter* of a plasma is defined as

$$\sigma_B = \frac{B_p^2}{4\pi\rho c^2}, \quad (5.142)$$

where $\rho = nm_p$ is the mass density of the plasma. At low wave frequencies $\omega \ll \omega_{B,p}$, we cannot ignore protons' contribution to the current component $\mathbf{J}_{0,\perp}$ (perpendicular to the plasma B-field) and, as we show below, the proton inertia does play an important role in setting the speed of wave propagation (e.g., the Alfvén speed).

Going back to the full expression for the dielectric tensor in eq. (5.79), we find that ϵ_1 term is dominated by protons when $\omega \ll \omega_{B,p}$, because

$$\epsilon_1 = \sum_{s=e,p} \frac{\omega_{p,s}^2}{\omega^2 - \omega_{B,s}^2} \approx - \sum_{s=e,p} \frac{\omega_{p,s}^2}{\omega_{B,s}^2} \approx -\frac{1}{\sigma_B}. \quad (5.143)$$

Linearly polarized Alfvén and fms modes only exist when $\epsilon_2 \ll 1$ (i.e., $\omega \gg \omega_{p,e}^2/\omega_{B,e}$), because otherwise the off-diagonal terms $\pm i\epsilon_2$ will lead to significant circular polarization (and we would get the so-called whistler mode). Ignoring ϵ_2 terms and considering finite magnetization, we obtain the approximate dielectric tensor

$$\overset{\leftrightarrow}{K} \approx \begin{pmatrix} 1 + \sigma_B^{-1} & 0 & 0 \\ 0 & 1 + \sigma_B^{-1} & 0 \\ 0 & 0 & 1 - \omega_p^2/\omega^2 \end{pmatrix}, \quad \omega_p = \omega_{p,e}. \quad (5.144)$$

From $\det(\overset{\leftrightarrow}{M}) = 0$, we further obtain

$$(\tilde{n}^2 - 1 - \sigma_B^{-1}) [\tilde{n}^2 (\cos^2 \theta - (1 - \sigma_B^{-1} \sin^2 \theta) \omega^2 / \omega_p^2) + (1 + \sigma_B^{-1})(\omega^2 / \omega_p^2 - 1)] = 0. \quad (5.145)$$

¹²For a finite temperature plasma (with non-zero sound speed), there is also a slow magnetosonic mode, which is not considered here.

Thus, the X/fms modes (with electric vector perpendicular to the \mathbf{k} - \mathbf{B}_p plane) have dispersion relation

$$\tilde{n}_{X/\text{fms}}^2 = 1 + \sigma_B^{-1}, \quad (5.146)$$

which means that the phase and group speeds for these two modes are given by

$$\frac{\omega}{k_{X/\text{fms}}} = \frac{d\omega}{dk_{X/\text{fms}}} = c \sqrt{\frac{\sigma_B}{1 + \sigma_B}}. \quad (5.147)$$

On the other hand, the O/Alfvén modes (with electric vector inside the \mathbf{k} - \mathbf{B}_p plane) have dispersion relation

$$\tilde{n}_{O/A}^2 = \frac{(\omega_p^2 - \omega^2)(1 + \sigma_B^{-1})}{\omega_p^2 \cos^2 \theta - \omega^2(1 - \sigma_B^{-1} \sin^2 \theta)}. \quad (5.148)$$

In the low-frequency limit, we have the Alfvén mode dispersion relation

$$\tilde{n}_A^2(\omega \ll \omega_p) \approx \frac{1 + \sigma_B^{-1}}{\cos^2 \theta} \Rightarrow \omega \approx \sqrt{\frac{\sigma_B}{1 + \sigma_B}} c \mathbf{k}_A \cdot \hat{\mathbf{B}}_p. \quad (5.149)$$

Then, the group velocity of the Alfvén mode in the low-frequency limit is given by

$$\mathbf{v}_{g,A}(\omega \ll \omega_p) = \frac{d\omega}{d\mathbf{k}_A} \approx v_A \hat{\mathbf{B}}_p, \quad (5.150)$$

where we have defined the *Alfvén speed*¹³ for a plasma with finite magnetization

$$v_A \equiv \sqrt{\frac{\sigma_B}{1 + \sigma_B}} c. \quad (5.151)$$

This means that an Alfvén wavepacket propagates along the plasma B-field lines at the Alfvén speed. Note that the X-mode and fms-mode have both phase and group speeds equal to the the Alfvén speed.

5.4 Weakly inhomogeneous plasma

In this section, we discuss the propagation of EM waves in weakly inhomogeneous plasmas first in the simple picture of ray optics and then in the full wave optics picture.

¹³The Lorentz factor corresponding to the Alfvén speed is given by $\gamma_A = (1 - \beta_A^2)^{-1/2} = \sqrt{1 + \sigma_B}$ (where $\beta_A \equiv v_A/c$), and the corresponding 4-speed is given by $\gamma_A \beta_A = \sqrt{\sigma_B}$.

5.4.1 Geometrical optics: Eikonal approximation

Wave propagation in a weakly inhomogeneous and slowly varying plasma can be described by geometrical optics under the *eikonal*¹⁴ approximation, which has been developed long before Maxwell's theory of light as electromagnetic waves. Geometrical optics describes light propagation in a picture of "rays" but does not account for wave effects such as diffraction and interference. It applies under the following three conditions: (i) the wavelength λ is much smaller than the lengthscale L for the spatial variations of plasma properties, (ii) the wave oscillation timescale ω^{-1} is much shorter than the timescale T over which the plasma evolves, and (iii) the size of the plasma in the transverse direction (perpendicular to the direction of wave propagation) is much larger than the Fresnel length $r_F = \sqrt{\lambda r}$ where r is the distance to the source or the observer. The last condition simply means that the aperture must contain a large number of Fresnel zones (see §5.4.3). Another limitation of geometric optics is that the temporal/spatial spreading of a short-duration (or wide-band) wavepacket due to plasma dispersion is ignored. Apart from the limitations above, we discuss the widely used eikonal approximation, which is general for any wave modes in any medium (not necessarily a plasma).

The general, complex solution of the wave equation can be written in the form of

$$\mathbf{E}(\mathbf{r}, t) e^{i\phi(\mathbf{r}, t)}, \quad (5.152)$$

and then let us define

$$\mathbf{k} \equiv \nabla\phi, \quad \omega \equiv -\partial_t\phi, \quad (5.153)$$

which is a natural generalization of a plane wave in a homogeneous plasma. Since the medium is only weakly inhomogeneous, the wave is nearly planar: the wave vector $\mathbf{k}(\mathbf{r}, t)$ is along the normal direction of the wavefront, the frequency $\omega(\mathbf{r}, t)$ describes the sinusoidal oscillation at a fixed point in space (the fields at different locations may in general oscillate at different frequencies), and the complex amplitude $\mathbf{E}(\mathbf{r}, t)$ varies with space and time only very slowly.

The eikonal approximation describes that an eigenmode of a locally homogeneous and stationary plasma obeys its dispersion relation in a small region of spacetime

$$\omega = W(\mathbf{k}, \mathbf{r}, t). \quad (5.154)$$

For instance, an electromagnetic wave mode propagating in an isotropic non-magnetized plasma has $W(\mathbf{k}, \mathbf{r}, t) = \sqrt{k^2 c^2 + \omega_p^2(\mathbf{r}, t)}$, where ω_p is the local plasma frequency which may vary weakly with position and time. The Alfvén mode propagating in the magnetosphere of a neutron star, where $\omega_B \gg \omega_p \gg \omega$, has $W(\mathbf{k}, \mathbf{r}, t) = ck \cdot \hat{\mathbf{B}}_p(\mathbf{r}, t)$ and $\hat{\mathbf{B}}_p$

¹⁴Eikonal originates from the Greek word *εικων* (eikon in English alphabet) which means icon or image.

describes the orientation of the local plasma magnetic field which may vary weakly with position and time. The invariance of the dispersion relation $\omega - W = 0$ for a given eigenmode along a given ray means that

$$\frac{d\omega}{dt} - \frac{dW}{dt} = \frac{d\omega}{dt} - \frac{\partial W}{\partial t} - \frac{\partial W}{\partial \mathbf{k}} \cdot \frac{d\mathbf{k}}{dt} - \frac{\partial W}{\partial \mathbf{r}} \cdot \frac{d\mathbf{r}}{dt} = 0, \quad (5.155)$$

where we have used the *Lagrangian derivative* d/dt meaning the change of a certain quantity with time *along the path* and partial derivative wrt. a vector \mathbf{x} ($= \mathbf{k}$ or \mathbf{r}) is equivalent to the gradient $\nabla_{\mathbf{x}} = \sum_{i=1}^3 \hat{\mathbf{e}}_i \partial_{x_i}$. Suppose a wavepacket starts from initial conditions (\mathbf{k}_0, ω_0) at spacetime coordinate (\mathbf{r}_0, t_0) , the path described by the following differential equations will automatically satisfy the invariance of the dispersion (eq. 5.155),

$$\frac{d\omega}{dt} = \frac{\partial W}{\partial t}, \quad \frac{d\mathbf{k}}{dt} = -\frac{\partial W}{\partial \mathbf{r}}, \quad \frac{d\mathbf{r}}{dt} = \frac{\partial W}{\partial \mathbf{k}}. \quad (5.156)$$

This is called the *ray-tracing equation*, which can be formally derived from the variational method (as done by Hamilton in the context of classical mechanics). For a given plasma setup and initial conditions for a wave mode, the path the wave takes is deterministic and unique, so the one described by eq. (5.156) must be the only physical solution.

In fact, eqs. (5.156) are indeed the Hamilton equations for a classical particle, with W playing the role of the Hamiltonian \mathcal{H} and \mathbf{k} acting like the canonical momentum \mathbf{P} . This can be physically understood if one considers that waves in the geometrical regime (ignoring wave effects) are made of discrete photons, each of which has energy $\hbar\omega$ and momentum $\hbar\mathbf{k}$. Then, the worldline for each photon, which has Hamiltonian $\mathcal{H} = \hbar W(\mathbf{k}, \mathbf{r}, t)$, can be described by the canonical Hamilton equations

$$\frac{d\mathcal{H}}{dt} = \frac{\partial \mathcal{H}}{\partial t}, \quad \frac{d\mathbf{P}}{dt} = -\frac{\partial \mathcal{H}}{\partial \mathbf{r}}, \quad \frac{d\mathbf{r}}{dt} = \frac{\partial \mathcal{H}}{\partial \mathbf{P}}. \quad (5.157)$$

Going back to eqs. (5.156), we see that the first expression tells us how the wave frequency changes with time $d\omega/dt$ along the path. If the unperturbed plasma properties have no time dependence, then W has no time dependence and hence the wave frequency ω is conserved along the propagation. For instance, the properties of the ionized gases in the interstellar medium are variable on timescales much longer than ω^{-1} for radio waves, so one can practically ignore the time dependence of the plasma when considering the propagation of radio waves in the ray optics regime (however, this conclusion does not hold in the wave optics regime, see §5.4.3). The second expression in eqs. (5.156) tells us how the wavevector or the local gradient of oscillatory wave phase changes with time $d\mathbf{k}/dt$ along the path. The spatial variation of the wavevector causes refraction, an example of which is light propagating from air into water. The third expression gives the *group velocity* of the wavepacket

$$\mathbf{v}_g = d\mathbf{r}/dt, \quad (5.158)$$

and the trajectory of the wavepacket is given by the time integral of the group velocity $\mathbf{r}(t) = \mathbf{r}_0 + \int_{t_0}^t \mathbf{v}_g dt$.

Knowing the trajectory of the wavepacket, it is straightforward to calculate how the wave amplitude is transported, because the photon number density n_{ph} satisfies the continuity equation (ignoring absorption, emission, and scattering here)

$$\partial_t n_{\text{ph}} + \nabla \cdot (n_{\text{ph}} \mathbf{v}_g) = 0, \quad (5.159)$$

where $n_{\text{ph}} \mathbf{v}_g$ is the photon number flux. Since the phase-averaged energy density of the wave is given by $U \propto |E_0|^2$, where E_0 is the wave amplitude, the number density is related to the wave amplitude by

$$n_{\text{ph}}(\mathbf{r}, t) = \frac{U(\mathbf{r}, t)}{\hbar \omega(\mathbf{r}, t)} \propto \frac{|E_0(\mathbf{r}, t)|^2}{\omega(\mathbf{r}, t)}. \quad (5.160)$$

Let us illustrate the ray-tracing equation with a concrete example. Consider a plasma with one-dimensional spatial variations along the z direction (which is not necessarily along the plasma B-field in this example). The plasma is stationary so the wave frequency stays constant along the propagation $\omega = \text{const}$. For simplicity, we also assume the plasma to be isotropic so the dispersion relation does not depend on the direction of the wavevector, and we write

$$\omega = W(\mathbf{k}, \mathbf{r}, t) = \frac{kc}{\tilde{n}(z)}, \quad (5.161)$$

where \tilde{n} is the refractive index. The time evolution of the wavevector is given by

$$\frac{d\mathbf{k}}{dt} = -\frac{\partial W}{\partial \mathbf{r}} = \frac{kc}{\tilde{n}^2} \frac{d\tilde{n}}{dz} \hat{\mathbf{z}} \Rightarrow \frac{dk_z}{dt} = \frac{kc}{\tilde{n}^2} \frac{d\tilde{n}}{dz}, \quad \frac{dk_x}{dt} = 0. \quad (5.162)$$

where we have taken the wavevector to be inside the x - z plane (without losing generality), and we see that the wavenumber k_x perpendicular to the direction of the plasma variation stays constant along the propagation. The group velocity is given by

$$\mathbf{v}_g = \frac{d\mathbf{r}}{dt} = \frac{\partial W}{\partial \mathbf{k}} = \frac{c}{\tilde{n}} \hat{\mathbf{k}}, \quad (5.163)$$

which shows that the wavepacket propagates along the direction of the wavevector (as expected for an isotropic plasma).

Let us denote the angle between the wavevector and the gradient of plasma variation as θ (meaning that $\cos \theta = \hat{\mathbf{k}} \cdot \hat{\mathbf{z}}$), then we write $k_x = k \sin \theta$, and the invariance of k_x and the wave frequency ω along the ray means that

$$\frac{d(k \sin \theta)}{dt} = \frac{\omega}{c} \frac{d(\tilde{n} \sin \theta)}{dt} = 0. \quad (5.164)$$

We see that $\tilde{n} \sin \theta$ is conserved along the ray — this is the *Snell's law*.

5.4.2 *Adiabatic walking (under construction)

The polarization of a given wave mode is transported such that the electric vector maintains the eigenmode characteristics along the path. For instance, the polarization of the Alfvén mode is such that the electric vector is within the $\mathbf{k} \cdot \hat{\mathbf{B}}_p$ plane and perpendicular to $\hat{\mathbf{B}}_p$, so it will stay so along the ray. On the other hand, since an X-mode wave is polarized such that the electric vector is perpendicular to the $\mathbf{k} \cdot \hat{\mathbf{B}}_p$ plane, if the plasma has a spatially varying orientation of the B-field direction $\hat{\mathbf{B}}_p$, then the electric vector of the X-mode wave will rotate together with the changing $\mathbf{k} \cdot \hat{\mathbf{B}}_p$ plane along the ray.

For simplicity, we consider an X-mode wave propagating in a one-dimensional plasma with slowly varying B-field orientation. Since the dispersion relation $\omega = ck$ has no explicit dependence on position, one immediately knows that $d\mathbf{k}/dt = 0$ from the ray-tracing equation and hence the wavevector \mathbf{k} (including its direction) stays unchanged along the ray. The wavepacket propagates at a group speed of $v_g = dr/dt = c\hat{\mathbf{k}}$. Let us assume the unperturbed state of the plasma to be time-independent, so the wave frequency ω also stays constant.

(to be continued)

5.4.3 *Wave diffraction by a phase screen (under construction)

5.5 *Warm plasma (under construction)

5.6 Homework

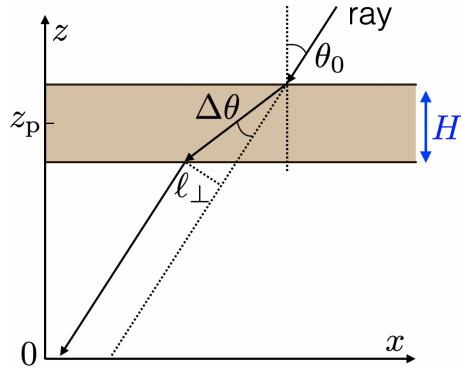


Figure 5.12: Geometry for radio waves propagating through the ionosphere.

Prob. 13. The free electron density profile of the Earth's ionosphere changes significantly with time due to diurnal, seasonal, and solar coronal variability. The ionosphere is only partially ionized with its composition dominated by neutral molecules (at height $z \lesssim 200$ km above the

ground) and atoms ($200 \lesssim z \lesssim 2000$ km); it only becomes nearly fully ionized at $z \gtrsim 2000$ km with its composition dominated by H^+ (this region is called the plasmasphere). Let us simplify the ionosphere into a uniform planar layer as shown in Fig. 5.12. The center of the layer is located at a height $z_p \sim 300$ km, the thickness of the layer is $H \sim 100$ km, and the electron density is taken to be uniform $n \sim 10^6 \text{ cm}^{-3}$.

- (i) Estimate the dispersion measure of our simplified one-zone ionosphere for a ray at an inclination angle $\theta_0 = 45^\circ$ from the vertical direction. What is the time delay Δt caused by the plasma dispersion at a frequency $\nu = 1 \text{ GHz}$ as compared to a signal traveling at c ? What is the phase delay $\Delta\phi = 2\pi\nu\Delta t$? Is the phase delay important for radio interferometric observations?
- (ii) For the same ray in (i), suppose that the Earth's magnetic field has a component parallel to the ray $B_{p,\parallel} \sim 0.3 \text{ G}$. Estimate the rotation measure for the ray as well as the Faraday rotation angle (expressed in radians) at frequency $\nu = 1 \text{ GHz}$.
- (iii) Our one-zone ionosphere has a constant refractive index $\tilde{n} = (1 - \omega_p^2/\omega^2)^{1/2}$, where ω_p is the plasma frequency corresponding to electron density n . A given ray from a distant radio source is refracted twice at the upper and lower boundaries of our idealized ionosphere layer (in the realistic ionosphere, the refraction is more gradual but the net result is the same). Suppose that the incident ray makes an angle θ_0 from the z axis (meaning that the altitude angle is $\pi/2 - \theta_0$). The Snell's law tells us $d(\tilde{n} \sin \theta) = d\tilde{n} \sin \theta - \tilde{n} \cos \theta d\theta = 0$, so we obtain the deflection angle at each refraction interface

$$\Delta\theta = \frac{\tan \theta}{\tilde{n}} (\tilde{n} - 1) \approx \tan \theta_0 (\tilde{n} - 1) \approx \frac{\omega_p^2}{2\omega^2} \tan \theta_0, \quad (5.165)$$

where the approximations are under the assumption of $\omega \gg \omega_p$ (for very weak refraction). Due to refraction, there is a transverse shift ℓ_\perp between the propagation directions before and after passing through the ionosphere. In the limit of $\Delta\theta \ll 1 \text{ rad}$, one can geometrically show that the angular shift of the apparent source position viewed from a ground-based observer is given by

$$\delta_{\text{app}} \simeq (H/z_p) \Delta\theta \simeq \Delta\theta/3. \quad (5.166)$$

The peak electron density around local noon is of the order $n_p \sim 10^6 \text{ cm}^{-3}$, whereas it drops to around 10^5 cm^{-3} near local midnight (due to reduced solar UV flux). Estimate the angular shift δ_{app} (expressed in arcminutes) at noon and at midnight for $\theta_0 = 45^\circ$ and $\nu = 100 \text{ MHz}$.

Prob. 14. The ionized intergalactic medium (IGM) has mean electron number density $n \sim 2 \times 10^{-7} \text{ cm}^{-3}$. The electron column density or DM up to redshift z is roughly $(850z) \text{ pc cm}^{-3}$, which corresponds to a dispersion time delay of about $(1.8z) \text{ sec}$ at $\nu = 1.4 \text{ GHz}$. The large DMs of the first few fast radio bursts were used to argue for their extragalactic origin before the discovery of their host galaxies (e.g., [Lorimer et al. 2007](#)). However, since the inter-particle separation in the IGM is about 10^2 cm which is much longer than the wavelength of the EM waves considered $\lambda \simeq 20 \text{ cm}$ (for $\nu = 1.4 \text{ GHz}$), do you think the dispersion delay in eq. (5.63) still applies? [hint: think about the picture where the dispersion delay is caused by electron scattering in eq. (4.166)]

Prob. 15. * not required This exercise is related to low-frequency Alfvén waves propagating through the magnetosphere of a strongly magnetized neutron star. Consider a magnetic flux tube

with two boundary surfaces S_1 and S_2 on the ends. Show that the wave energy flowing in through S_1 must come out from S_2 with out loss, i.e., $\int_{S_1} U \mathbf{v}_g \cdot d\mathbf{S} = \int_{S_2} U \mathbf{v}_g \cdot d\mathbf{S}$.

(i) If the flux tube is a part of a dipolar field with strength $B_p \propto r^{-3}$ (r being the radius to the center of the star), show that the amplitude of the Alfvén wave goes as $B \propto r^{-3/2}$.

(ii) If (i) is true, the fractional amplitude of the wave increases rapidly with radius as $B/B_p \propto r^{3/2}$, so the wave may become *nonlinear* at a critical radius r_c where $B/B_p \sim 1$. Estimate r_c for the following parameters: neutron star magnetic field strength at the surface $B_p(r = R) = 10^{14}$ G, wave magnetic field strength at the surface $B(r = R) = 10^{10}$ G, and neutron star radius $R = 10$ km.

References

- F. F. Chen, *Introduction to Plasma Physics and Controlled Fusion* (Springer International Publishing, 3rd Edition, Switzerland, 2016)
- Jackson, *Classical Electrodynamics* (John Wiley & Sons, New York, 1975).
- Landau & Lifshitz, *Mechanics* (Butterworth-Heinemann, 3rd Edition, Oxford, 1976)
- Rybicki & Lightman, *Radiative processes in astrophysics* (A Wiley-Interscience Publication, New York, 1979).
- Thorne & Blandford, *Modern Classical Physics* (Princeton University Press, New Jersey, 2017).

Chapter 6

Free-free Process

This chapter discusses the interaction between two charged particles that are unbound from each other and the radiation associated with this interaction.

6.1 Coulomb collision

In this section, we study the dynamics of a hyperbolic encounter between two non-relativistic charged particles approaching each other from infinity. Both particles' trajectories are deflected due to Coulomb interaction. It will be shown later (see §6.2) that the radiative losses are negligible for non-relativistic encounters, so we will ignore radiation for now.

6.1.1 Impulse approximation and Coulomb relaxation

Let us first consider the limit of weak encounters where each particle's trajectory is only slightly deflected. In this limit, the deflection angle θ for a given projectile can easily be obtained under the *impulse approximation* (equivalent to the *Born approximation* in quantum mechanics) where we assume that the projectile moves at a constant velocity along a straight line. For simplicity, we will also assume the motion of the target particle to be negligible — this assumption can be relaxed by considering the reduced 2-body problem (see §6.1.3). As we will see later, processes like Coulomb relaxation and free-free emission/absorption can be understood in the limit small deflection angles $\theta \ll 1$ rad.

As shown in Fig. 6.1, let us consider that the projectile's path is approximately a straight line and that the target stays nearly at rest. Along the path of the projectile, the Coulomb force on the particle in the direction perpendicular to the velocity vector is given by

$$F_{\perp}(x) = \frac{q_1 q_2 b}{(x^2 + b^2)^{3/2}}, \quad (6.1)$$

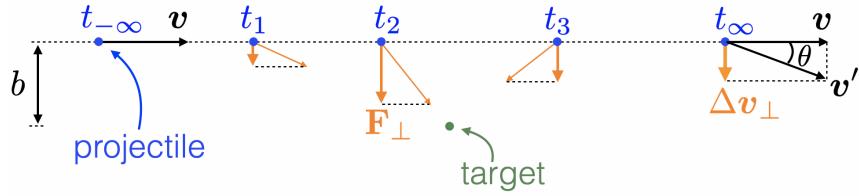


Figure 6.1: Geometry for the impulse approximation where we ignore the deflection and assume the projectile moves at a constant velocity along a straight line. This approximation is valid to the first order in the deflection angle θ .

where q_1 and q_2 are the charges of the projectile and target (respectively), and x is the distance to the point of minimum separation along the path. The projectile moves from $x = -\infty$ to $+\infty$, and its velocity is $v = dx/dt$. The time-integrated momentum change in the perpendicular direction is given by

$$\Delta p_{\perp} = \int F_{\perp} dt = \int \frac{F_{\perp}}{v} dx = \frac{q_1 q_2 b}{v} \int_{-\infty}^{\infty} \frac{dx}{(x^2 + b^2)^{3/2}} = \frac{2 q_1 q_2}{b v}, \quad (6.2)$$

where we have used $\int_0^{\infty} dx/(x^2 + 1)^{3/2} = \int_0^{\pi/2} \cos \phi d\phi = 1$ by change of variable $x = \tan \phi$. For an initial momentum of $p = m_1 v$, the momentum kick in the perpendicular direction gives a *deflection angle* of

$$\theta \approx \Delta p_{\perp}/p = 2 q_1 q_2 / (m_1 v^2 b). \quad (6.3)$$

The deflection angle reaches $\theta = 1 \text{ rad} = 57$ degrees at a critical impact parameter

$$b_{57} \approx 2 |q_1 q_2| / (m_1 v^2), \quad (6.4)$$

i.e., when the Coulomb potential energy at the closest approach $|q_1 q_2|/b$ is close to the kinetic energy of the projectile $(1/2)m_1 v^2$. A more careful calculation shows that the critical impact parameter for 1 rad deflection $b_{57} = 1.83 |q_1 q_2| / (m_1 v^2)$ (see eq. 6.39), so our expression for the deflection angle (eq. 6.3) under the impulse approximation is fairly accurate. At larger deflection angles, we expect the impulse approximation to break down significantly. The following discussion applies for $\theta \ll 1 \text{ rad}$ or $b \gg b_{57}$.

It can be geometrically shown that the momentum change in the direction parallel to the incident velocity is given by

$$\Delta p_{\parallel} \approx -\Delta p_{\perp}^2 / (2p) \approx -p \theta^2 / 2, \quad (6.5)$$

and hence $|\Delta p_{\parallel}| \ll \Delta p_{\perp}$ for small deflection angles. This can be understood because the Coulomb force in the parallel direction F_{\parallel} is positive for half of the time and negative for the other half, and this cancellation causes Δp_{\parallel} to be a higher-order term. Note that Δp_{\parallel} is negative, because the total magnitude of the momentum stays unchanged in the

center-of-mass frame before and after the hyperbolic encounter. We conclude that, for weak encounters, the dominant change in the projectile's velocity/momentum is in the direction *perpendicular* to the initial velocity.

Let us then consider a fast projectile (q_1, m_1) interacting with a sea of slow targets (q_2, m_2). The number density of targets is n_2 . This setup is reasonable for the following situations: (i) an electron ($1 \rightarrow e$) interacting with heavier ions ($2 \rightarrow i$), or (ii) a fast cosmic ray particle interacting with a relatively cold plasma where the particles' thermal motions are much slower. In these situations, we would like to know the timescales for the deflection and energy loss of the projectile.

The rate of collisions with impact parameters in the range b to $b + db$ (for a differential cross-sectional area of $d\sigma = 2\pi b db$) is given by

$$dR = n_2 v d\sigma, \quad (6.6)$$

which is in units of sec^{-1} and v is the velocity of the fast projectile. In each of these collisions, the projectile receives a momentum kick of Δp_\perp in the *plane* perpendicular to the initial velocity. The direction of Δp_\perp in the perpendicular plane is random, and this means that over time, the cumulative change in the momentum vector Δp undergoes a 2D random walk¹ in the plane perpendicular to the initial velocity. This is the effect of *Coulomb relaxation*, as shown in Fig. 6.2 (note that rare, large-angle deflections are not captured in the model here). The consequence of random walk is that the mean-squared perpendicular momentum grows with time

$$\begin{aligned} \frac{d}{dt}(\Delta p_\perp)^2 &= \int (\Delta p_\perp)^2 dR = \int (\Delta p_\perp)^2 n_2 v d\sigma \\ &= \frac{8\pi n_2 (q_1 q_2)^2}{v} \int_{b_{\min}}^{b_{\max}} \frac{db}{b}. \end{aligned} \quad (6.7)$$

where one notices that the integral over the impact parameter diverges logarithmically.

The divergence in the limit of low impact parameters (or large deflection angles) is entirely due to our impulse approximation — if one uses the full expression for Δv_\perp (see eq. 6.40 later) without assuming $\theta \ll 1 \text{ rad}$, the divergence near $b \rightarrow 0$ goes away (as we will see when calculating the large angle cross-section in eq. 6.43) and the result is close to that obtained by using a minimum impact parameter $b_{\min} \sim b_{57}$ in eq. (6.7). Quantum effects are important for impact parameters less than the positional uncertainty of the projectile $\hbar/(m_1 v)$ based on the Uncertainty Principle, which will be discussed later.

The divergence at $b \rightarrow \infty$ is more problematic! For this reason, we provide a physical restriction of b_{\max} beyond which the Coulomb potential of the target, $\phi = q_2/r$, is shielded

¹Random walk for N uncorrelated steps gives a cumulative mean-squared displacement of $\langle \mathbf{r}_N^2 \rangle = \langle (\sum_i \Delta \mathbf{r}_i)^2 \rangle = N \langle \Delta \mathbf{r}^2 \rangle$, where $\langle \Delta \mathbf{r}^2 \rangle$ is the mean-squared displacement for each step.

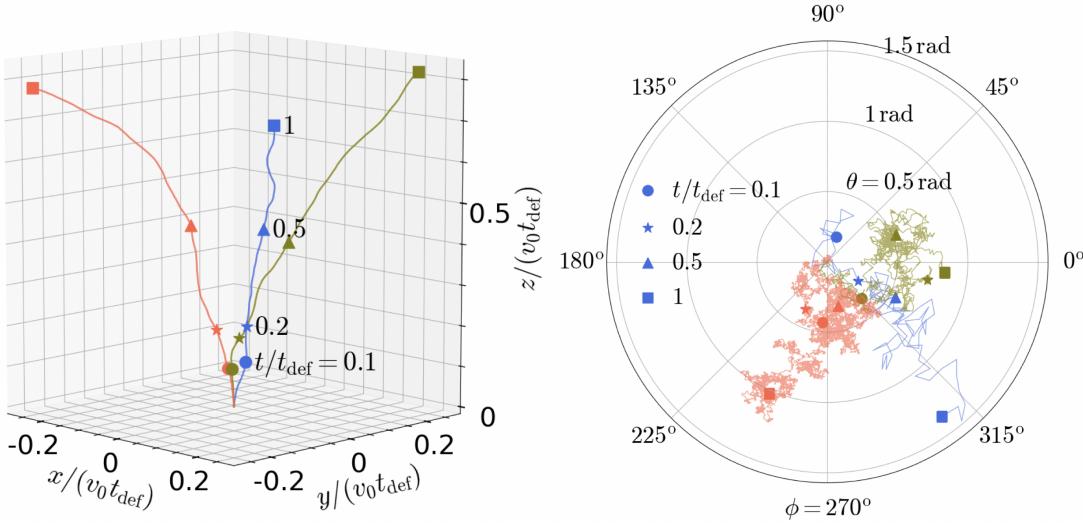


Figure 6.2: Simulations of a fast particle with initial velocity $v_0 \hat{z}$ undergoing Coulomb deflections by a sea of static perturbers on a deflection timescale $t_{\text{sim}} = t_{\text{def}}$ (eq. 6.9). The three cases differ in the number of Monte Carlo steps $N = 10^2$ (blue), 10^3 (green), 10^4 (red), and the statistical results are independent of N (as long as $\gg 1$). In each step, the velocity perturbation is taken to be $\Delta \mathbf{v} = \Delta v_{\perp} \hat{\mathbf{e}}_{\perp} + \Delta v_{\parallel} \hat{\mathbf{v}}$, where $\Delta v_{\perp} = v \sqrt{t_{\text{sim}} / (N t_{\text{def}})}$, $\Delta v_{\parallel} = -\Delta v_{\perp}^2 / (2v)$, $\hat{\mathbf{v}}$ is along the current velocity \mathbf{v} , and $\hat{\mathbf{e}}_{\perp}$ is randomly chosen in the plane perpendicular to $\hat{\mathbf{v}}$. The left panel shows the particle's trajectory. The right panel shows the evolution of the velocity's direction as described by polar angles θ and ϕ (while the magnitude stays constant $v \approx v_0$).

by other particles of opposite charge. In a plasma, the maximum impact parameter is determined by the effect of *Debye shielding* (see Ch. 5), which leads to a shielded potential for the target particle in the form of $\phi = q_2/re^{-\sqrt{2}r/r_D}$ and r_D is the Debye radius (eq. 5.12). If a fast projectile interacts with nuclei in atomic material, the maximum impact parameter would be of the order the atomic size.

For now, we write the integral over the impact parameter in the following symbolic form

$$\int_{b_{\min}}^{b_{\max}} \frac{db}{b} = \ln(b_{\max}/b_{\min}) \equiv \ln \Lambda, \quad (6.8)$$

which is called the *Coulomb logarithm*. It will be shown later (§6.1.2) that b_{\max} is many orders of magnitude greater than b_{\min} and that a typical value for the Coulomb logarithm is $\ln \Lambda \sim 20$, so this factor cannot be ignored.

When the cumulative mean-squared perpendicular momentum change reaches $(\Delta p_{\perp})^2 \sim p^2 = (m_1 v)^2$, then the direction of the projectile's velocity would be *cumulatively* deflected by a large angle ($\sim 90^\circ$), and this leads to the *deflection timescale*

$$t_{\text{def}} = \frac{(m_1 v)^2}{(d/dt)(\Delta p_{\perp})^2} = \frac{m_1^2 v^3}{8\pi n_2 (q_1 q_2)^2 \ln \Lambda}, \quad (6.9)$$

where v is the velocity of the projectile and n_2 is the number density of the targets. An important result from the scaling of $t_{\text{def}} \propto m_1^2 v^3$ is that faster and heavier projectile particles are much harder to be deflected. For the electron-ion collision case ($1 \rightarrow e$ and $2 \rightarrow i$), the deflection time is given by

$$t_{\text{def}}(e-i) = \frac{m_e^2 v_e^3}{8\pi n_i Z^2 e^4 \ln \Lambda}, \quad (6.10)$$

where Z is the ion charge number. For electron-electron collisions, if the projectile electron moves much faster than the thermal motion of the target electrons, then we have

$$t_{\text{def}}(e-e) = t_{\text{def}}(e-i)|_{Z \rightarrow 1, n_i \rightarrow n_e} = \frac{m_e^2 v_e^3}{8\pi n_e e^4 \ln \Lambda}. \quad (6.11)$$

When the typical thermal speed of target electrons is comparable to the speed of the projectile electron, one must take into account the motions of both particles during the collision. The resulting $t_{\text{def}}(e-e)$ differ from eq. (6.11) by a factor of order unity.

Realistic astrophysical plasmas are magnetized, and the consequence is that the deflection of particles' trajectories is usually dominated by Larmor motion instead of Coulomb collisions. The realistic deflection timescale for a non-relativistic electron is given by the inverse of the cyclotron frequency

$$\omega_B^{-1} = \frac{m_e c}{e B} = 5.7 \times 10^{-2} \text{s}^{-1} (B/\mu\text{G})^{-1}. \quad (6.12)$$

However, Coulomb collisions are important for the energy exchange between particles.

Let us still consider the simplified case that the target particles have negligible motion initially (before the encounter with the projectile). Along with each momentum transfer of Δp_{\perp} , the *target* gains energy by an amount $(\Delta p_{\perp})^2/(2m_2)$. This energy must come from the projectile, and this means that the energy-loss rate is given by

$$\frac{dE_1}{dt} = -(2m_2)^{-1} \frac{d}{dt} (\Delta p_{\perp})^2 = -\frac{4\pi n_2 (q_1 q_2)^2}{m_2 v} \ln \Lambda \propto \frac{n_2}{v}. \quad (6.13)$$

The *energy-loss timescale* or *thermalization timescale* for the projectile is

$$t_{\text{th}} = \frac{(1/2)m_1 v^2}{\dot{E}_1} = \frac{m_1 m_2 v^2}{(d/dt)(\Delta p_{\perp})^2} = \frac{m_2}{m_1} t_{\text{def}}. \quad (6.14)$$

The reason that the energy loss timescale is also the thermalization timescale is as follows. Based on the $t_{\text{th}} \propto v^2$ scaling, as the projectile slows down, it will lose energy on a shorter and shorter timescale. Thus, after the first energy-loss timescale, the projectile will be brought into thermal equilibrium with the slower target particles. We also see that, if the two interacting species have the same mass $m_1 = m_2$, then the energy loss timescale is roughly equal to the deflection timescale — meaning that particles of the same species will reach thermal equilibrium due to energy exchanges on a deflection timescale.

Here are some concrete examples. A fast electron with an initial velocity v_e loses a significant fraction of its kinetic energy to target (much slower) electrons on a timescale of

$$t_{\text{th}}(\text{e-e}) \simeq t_{\text{def}}(\text{e-e}) \simeq \frac{m_e^2 v_e^3}{8\pi n_e e^4 \ln \Lambda} \simeq 9.5 \times 10^3 \text{ sec} \left(\frac{T_e}{10^4 \text{ K}} \right)^{3/2} \frac{\text{cm}^{-3}}{n_e} \frac{20}{\ln \Lambda}, \quad (6.15)$$

where in the second expression we have expressed the electron velocity in terms of an electron kinetic temperature defined as $(3/2)k_B T_e = (1/2)m_e v_e^2$. The thermalization timescale for ion-ion collisions is given by

$$t_{\text{th}}(\text{i-i}) \simeq t_{\text{def}}(\text{i-i}) \simeq 4.1 \times 10^5 \text{ sec} \left(\frac{T_i}{10^4 \text{ K}} \right)^{3/2} \left(\frac{m_i}{m_p} \right)^{1/2} \frac{\text{cm}^{-3}}{Z^4 n_i} \frac{20}{\ln \Lambda}, \quad (6.16)$$

where we have used the ion kinetic temperature defined as $(3/2)k_B T_i = (1/2)m_i v_i^2$. On the other hand, electron-ion collisions have a rather long thermalization timescale

$$t_{\text{th}}(\text{e-i}) = \frac{m_i}{m_e} t_{\text{def}}(\text{e-i}) \simeq 1.7 \times 10^7 \text{ sec} \left(\frac{T_e}{10^4 \text{ K}} \right)^{3/2} \frac{m_i}{m_p} \frac{\text{cm}^{-3}}{Z^2 n_i} \frac{20}{\ln \Lambda}. \quad (6.17)$$

For the special case of a proton-electron plasma (fully ionized hydrogen gas) where the ion temperature is equal to the electron temperature, then the different thermalization

timescales are given by

$$t_{\text{th}}(\text{e-p}) : t_{\text{th}}(\text{p-p}) : t_{\text{th}}(\text{e-e}) = \frac{m_p}{m_e} : \sqrt{\frac{m_p}{m_e}} : 1, \quad \text{for } Z = 1, n_p = n_e, T_p = T_e. \quad (6.18)$$

This shows that electrons tend to quickly reach a Maxwellian distribution at temperature T_e by e-e Coulomb scatterings. After that, ions will reach a Maxwellian distribution at temperature T_i by i-i Coulomb scatterings. At last, the much slower energy exchange between electrons and ions will eventually lead to an equilibrium between the two temperatures $T_i = T_e$.

In many physical situations of a hot, dilute plasma (e.g., black hole accretion disks at very low accretion rates, intracluster medium in galaxy clusters, supernova remnants), the electron-ion thermalization timescale may be longer than the dynamical timescale of the system, and we may be dealing with a *two-temperature plasma* with $T_i \neq T_e$. An interesting example is the case of a supernova remnant. Right after a strong shock passes through a hydrogen plasma, protons are heated to a high temperature of $k_B T_p = (3/16) m_p v_s^2$ (where v_s is the shock speed) whereas electrons have a much lower temperature by a factor of m_e/m_p — almost all the thermal energy in the shock-heated plasma is initially in the protons. Then, it takes a very long time, $t_{\text{th}}(\text{e-p}) \simeq 1.7 \times 10^4 \text{ yr} (T_e/10^7 \text{ K})^{3/2} (n_p/\text{cm}^{-3})^{-1}$, for electron-proton temperature equilibrium² to be established via Coulomb collisions.

6.1.2 Coulomb logarithm

The simple picture of two-body Coulomb collision breaks down when $b \rightarrow 0$ or $b \rightarrow \infty$. This subsection discusses the physical limits for the impact parameter, b_{\min} and b_{\max} , that should be used when computing the Coulomb logarithm (eq. 6.8).

The impulse approximation (by assuming the trajectory to be a straight line) breaks down for impact parameters less than the critical one b_{57} for a deflection angle of $1 \text{ rad} = 57$ degrees,

$$b_{57} \approx \frac{2|q_1 q_2|}{m_1 v^2} = 1.1 \times 10^{-7} \text{ cm} Z \left(\frac{T_e}{10^4 \text{ K}} \right)^{-1}, \quad (6.19)$$

where in the 2nd expression we have expressed the kinetic energy in terms of a kinetic temperature $(3/2)k_B T_e = (1/2)m_1 v^2$ for an electron projectile and taken $|q_1 q_2| = Ze^2$ for a target charge number of Z . For $b \lesssim b_{57}$, the deflection angle is capped near $\mathcal{O}(1) \text{ rad}$, and the amount of momentum exchange between the two particles is capped near $m_1 v$. The very small number of encounters at $b \ll b_{57}$ do not contribute appreciably to the Coulomb

²How do we know that the electrons are indeed much colder than the protons as the theory suggests? Inelastic collisions of electrons with ions can leave the ion in an excited state or a different ionization state, so the electron temperature can be inferred by studying the atomic lines (see e.g., [Ghavamian et al. 2013](#)).

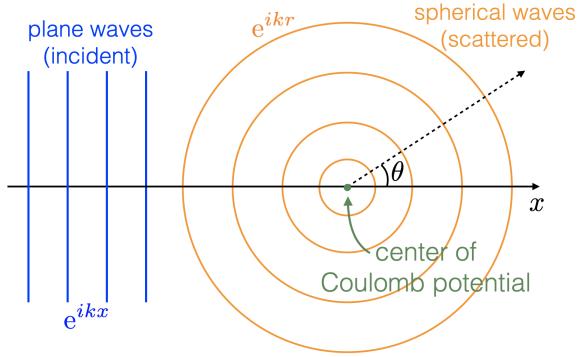


Figure 6.3: Quantum mechanic picture: the incident plane waves (shown by the blue wavefronts) are scattered into spherical waves (orange) in the form of $A(\theta)e^{ikr}/r$, where the amplitude of the scattered waves depends on the scattering angle θ according to the Rutherford differential cross-section $|A(\theta)|^2 \propto d\sigma/d\Omega$ (eq. 6.42).

relaxation. For this reason, it is appropriate to take the minimum impact parameter in the Coulomb logarithm to be

$$b_{\min}^{(1)} \sim b_{57}. \quad (6.20)$$

On the other hand, the projectile's momentum $p = m_1 v$ corresponds to positional uncertainty of $\Delta x \gtrsim \hbar/p$ according to the Uncertainty Principle. This motivates us to consider another minimum impact parameter

$$b_{\min}^{(2)} \sim \frac{2\hbar}{m_1 v} = 3.4 \times 10^{-8} \text{ cm} \left(\frac{T_e}{10^4 \text{ K}} \right)^{-1/2}, \quad (6.21)$$

where the second expression is for $v = \sqrt{3k_B T_e / m_e}$ and $m_1 = m_e$. Note that we include a factor of 2 by taking $b_{\min}^{(2)} \sim 2\hbar/p$ so as to be more consistent with the quantum mechanical result³. The projectile's wave function (or the probability distribution of its position) is “spread out” in a region of size $\mathcal{O}(\hbar/p)$. For $b \lesssim b_{\min}^{(2)}$, the wave nature of the projectile particle becomes important. A schematic picture for the quantum mechanical Coulomb scattering is shown in Fig. 6.3. In this picture, the wave packet of the projectile is *diffracted* by the target — just like electromagnetic waves being diffracted by an obstacle of size comparable to the wavelength.

In practice, which one of the two, $b_{\min}^{(1)}$ or $b_{\min}^{(2)}$, should we take? The answer is to take the

³One may consider $b_{\min}^{(2)}$ to be roughly the full-width half-maximum of the positional uncertainty. We are not too worried about the factor of 2 here, because the Coulomb logarithm only depends weakly on b_{\min} in a logarithmic manner.

more conservative one

$$b_{\min} \sim \max(b_{\min}^{(1)}, b_{\min}^{(2)}) = \frac{2|q_1 q_2|}{m_e v^2} \max\left(1, \frac{v}{v_{\text{crit}}}\right), \quad (6.22)$$

where the critical projectile velocity v_{crit} at which $b_{\min}^{(1)} = b_{\min}^{(2)}$ is given by

$$v_{\text{crit}} = \frac{|q_1 q_2|}{\hbar} = Z\alpha c, \quad (6.23)$$

and in the second expression we have used $|q_1 q_2| = Ze^2$ for charge number Z and the fine-structure constant $\alpha = e^2/(\hbar c) \approx 1/137$. For an electron projectile, the critical velocity corresponds to a critical kinetic energy

$$E_{\text{crit}} = \frac{1}{2} m_e v_{\text{crit}}^2 = Z^2 \text{Ry}, \quad (6.24)$$

where $\text{Ry} = \alpha^2 m_e c^2 / 2 = 13.6 \text{ eV}$ is the Rydberg energy. The corresponding kinetic temperature is

$$T_{e,\text{crit}} = 2E_{\text{crit}}/(3k_B) = 1.0 \times 10^5 \text{ K}. \quad (6.25)$$

We see that, if the electron kinetic temperature is sufficiently high, we should take the quantum limit of

$$b_{\min}(T_e \gtrsim 10^5 \text{ K}) \sim b_{\min}^{(2)} \sim 2\hbar/(m_e v);$$

whereas for lower electron temperatures, the minimum impact parameter should be taken as that for 1 rad deflection

$$b_{\min}(T_e \lesssim 10^5 \text{ K}) \sim b_{\min}^{(1)} \sim b_{57}.$$

It should be emphasized that, although the deflection (eq. 6.7) and energy loss (eq. 6.13) of the projectile are dominated by weak encounters with large impact parameters, close encounters cannot be overlooked because they are responsible for free-free emission of the highest energy photons near frequency $\omega_{\max} \sim v/b_{\min}$ (which may be up to $(1/2)m_e v^2/\hbar$) as well as collisional excitation/ionization of the inner shell electrons of the target ions.

Let us then discuss the three possible choices for the maximum impact parameter b_{\max} .

The first one is the Debye length r_D — the Coulomb potential of an ion in a plasma is screened by electrons at radii $r \gtrsim r_D$. The shielded Debye Coulomb potential around an ion of charge Ze is given by $\Phi(r) = (Ze/r)e^{-\sqrt{2}r/r_D}$ (see Ch. 5). For the case of an electron-proton plasma with electron number density n and temperature T , the Debye length is given by

$$r_D(\text{e-p}) = \left(\frac{k_B T}{4\pi e^2 n} \right)^{1/2} = 690 \text{ cm} \left(\frac{T}{10^4 \text{ K}} \right)^{1/2} \left(\frac{n}{\text{cm}^{-3}} \right)^{-1/2}. \quad (6.26)$$

This motivates us to take the maximum impact parameter

$$b_{\max}^{(1)} \sim r_D. \quad (6.27)$$

This is indeed the correct choice for the purpose of obtaining the deflection timescale (eq. 6.10) or thermalization timescale (eq. 6.14) in an ionized gas. For $b_{\min} \sim 10^{-7}$ cm and $b_{\max} \sim 10^3$ cm, we obtain $\ln \Lambda \simeq 23$.

The second choice is for a fast charged particle moving through atoms (e.g., neutral hydrogen gas or solid material). In this case, the target electrons have bound orbits around nuclei. For an orbit with angular frequency Ω_{orb} , then the maximum impact parameter would be

$$b_{\max}^{(2)} \sim v/\Omega_{\text{orb}}. \quad (6.28)$$

This is because at $b > b_{\max}^{(2)}$, the Coulomb interaction between our projectile and target would occur on a timescale longer than the atomic orbital timescale and is hence suppressed. In Bohr's model for atomic orbits, the ionization energy I is of the order $I \sim \hbar\Omega_{\text{orb}}$ (see eq. 4.239), so we obtain $b_{\max} \sim \hbar v/I$. In reality, the target electrons may be in different orbits with different ionization energies, and this motivates us to take $b_{\max} \sim \hbar v/\bar{I}$, where \bar{I} is of the order the *average* binding energy of bound electrons and its value is related to the composition of the material. On the other hand, if the minimum impact parameter is taken to be related to the quantum mechanical limit $b_{\min} \sim \hbar/(m_1 v)$, then we obtain $\Lambda = b_{\max}/b_{\min} \sim m_1 v^2/\bar{I}$. Conventionally, the Coulomb logarithm for atomic targets is written as follows (so as to be consistent with the Bethe formula, see §6.5.3)

$$\ln \Lambda = \ln \left(\frac{2m_1 v^2}{\bar{I}} \right), \quad (\text{for atomic targets}) \quad (6.29)$$

where \bar{I} is a constant obtained by fitting to the experimental data and typically $\bar{I} \simeq 10Z$ eV for atomic targets of charge number Z .

Finally, if an observer is interested in free-free emission at angular frequency ω (which is much greater than the plasma frequency ω_p), then she would only care about Coulomb collisions with impact parameters less than

$$b_{\max}^{(3)} \sim \frac{v}{\omega} = 1.1 \times 10^{-2} \text{ cm} \left(\frac{T_e}{10^4 \text{ K}} \right)^{1/2} \left(\frac{\nu}{\text{GHz}} \right)^{-1}, \quad (6.30)$$

where the 2nd expression is for $v = \sqrt{3k_B T_e/m_e}$ (v being velocity of the projectile electron) and $\omega = 2\pi\nu$. This is another choice for the upper bound of the impact parameter in the Coulomb logarithm and this will be used when we study the free-free emission spectrum in §6.2. We note that Coulomb scatterings with impact parameters near the Debye length r_D would only produce EM emission at frequencies $\omega \lesssim v/r_D \sim \omega_p$ (where we have taken $v \sim v_{\text{th}} = \sqrt{k_B T/m_e}$, since the acceleration of the projectile occurs on a timescale of r_D/v). Such low-frequency EM waves cannot propagate in the plasma (!), as they only penetrate to a plasma skin depth $\lambda_{\text{skin}} = c/\omega_p$.

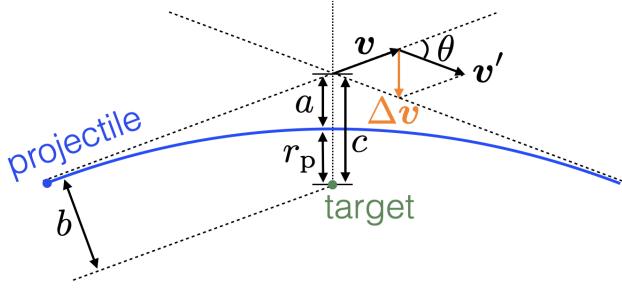


Figure 6.4: Coulomb collision as a reduced 2-body problem — the target is considered stationary in this consideration. The two particles have opposite charges in this case. The deflection angle of the projectile is $\theta = \arccos(\mathbf{v} \cdot \mathbf{v}')$. One can geometrically show that $c^2 = a^2 + b^2$, $\cot(\theta/2) = b/a$, and $r_p = c - a$.

6.1.3 *Deflection angle in the reduced 2-body problem

Let us consider a non-relativistic Coulomb collision between two charges q_1 and q_2 , and their masses are m_1 and m_2 . For given initial velocities, our goal is to calculate the deflection angles for both particles. Since the Coulomb potential only depends on the relative position between the two particles, this is the classical reduced 2-body problem. Let us define the reduced mass, relative position vector, and relative velocity vector as

$$\mu = \frac{m_1 m_2}{m_1 + m_2}, \quad \mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2, \quad \mathbf{v} = \dot{\mathbf{x}}, \quad (6.31)$$

where $\mathbf{x}_i (i = 1, 2)$ are the positions of the two particles in the center-of-momentum frame. In the center-of-momentum frame, the total kinetic energy is given by

$$T = \frac{1}{2} m_1 \dot{\mathbf{x}}_1^2 + \frac{1}{2} m_2 \dot{\mathbf{x}}_2^2 = \frac{1}{2} \mu \dot{\mathbf{x}}^2, \quad (6.32)$$

and the total angular momentum is

$$\mathbf{L} = m_1 \mathbf{x}_1 \times \dot{\mathbf{x}}_1 + m_2 \mathbf{x}_2 \times \dot{\mathbf{x}}_2 = \mu \mathbf{x} \times \dot{\mathbf{x}}. \quad (6.33)$$

The potential energy is $V = q_1 q_2 / |\mathbf{x}|$, so the Lagrangian is

$$\mathcal{L} = T - V = \frac{1}{2} \mu \dot{\mathbf{x}}^2 - \frac{q_1 q_2}{|\mathbf{x}|}. \quad (6.34)$$

The Hamiltonian is given by the total energy $\mathcal{H} = T + V$, which is conserved. This shows that the system is entirely described by a single particle of mass μ moving in a static potential $V(|\mathbf{x}|)$.

Initially, the two particles are far apart, $|\mathbf{x}| \rightarrow \infty$. They have relative speed v and impact parameter b . The trajectory of the reduced single particle is described by a hyperbola (as

in the Kepler's problem with positive energy). The closest separation (pericenter radius) r_p and the velocity v_p at pericenter satisfy energy and angular momentum conservations

$$v_0 b = r_p v_p, \quad \frac{1}{2} \mu v^2 = \frac{1}{2} \mu v_p^2 + \frac{q_1 q_2}{r_p}, \quad (6.35)$$

and the solution is

$$r_p = b_{90} + \sqrt{b_{90}^2 + b^2}, \quad (6.36)$$

where we have defined

$$b_{90} \equiv \frac{q_1 q_2}{\mu v^2}, \quad (6.37)$$

which will later be shown to be the impact parameter for 90° deflection angle (and in that sense, we use $b_{90} = |q_1 q_2|/\mu v^2$). For a hyperbola described by $x^2/a^2 - y^2/b^2 = 1$ (here b is the impact parameter), the distance from the focal point to the center ($x = y = 0$) is $c = \sqrt{a^2 + b^2}$. For the case where the two particles carry opposite charges $q_1 q_2 < 0$, the pericenter radius is $r_p = c - a$, so we obtain $c + a = b^2/r_p$ and hence

$$a = \frac{1}{2} \left(\frac{b^2}{r_p} - r_p \right). \quad (6.38)$$

Our goal is to obtain the deflection angle θ — the angle between the initial velocity \mathbf{v} and the final velocity \mathbf{v}' after the scattering. The velocity's magnitude stays unchanged before and after the collision $|\mathbf{v}'| = |\mathbf{v}|$, and the deflection is described by the change in velocity vector $\Delta \mathbf{v} = \mathbf{v}' - \mathbf{v}$. For $q_1 q_2 < 0$, we obtain $\tan(\theta/2) = a/b = -b_{90}/b$. The case of $q_1 q_2 > 0$ can be similarly worked out by using $r_p = c + a$ and the result is $\tan(\theta/2) = b_{90}/b$.

To conclude, the *deflection angle* θ is given by (using the critical impact parameter b_{90} for 90° deflection angle)

$$\tan(\theta/2) = \frac{|q_1 q_2|}{\mu v^2 b} = \frac{b_{90}}{b}, \quad b_{90} \equiv \frac{|q_1 q_2|}{\mu v^2}, \quad (6.39)$$

which applies to any impact parameter b provided that classical description of the Coulomb collision is valid. The vectorial change in velocity $\Delta \mathbf{v} = \mathbf{v}' - \mathbf{v}$ has two components parallel and perpendicular to the incident velocity vector

$$\begin{aligned} \Delta v_{\parallel} &= \Delta \mathbf{v} \cdot \hat{\mathbf{v}} = -\Delta v \sin(\theta/2) = -2v \sin^2(\theta/2), \\ \Delta v_{\perp} &= \Delta v \cos(\theta/2) = v \sin \theta, \\ |\Delta \mathbf{v}| &= 2v \sin(\theta/2), \end{aligned} \quad (6.40)$$

which are exact for all deflection angles. We notice that, for small deflection angles, the velocity change $\Delta \mathbf{v}$ is mainly in the direction perpendicular to the initial velocity \mathbf{v} .

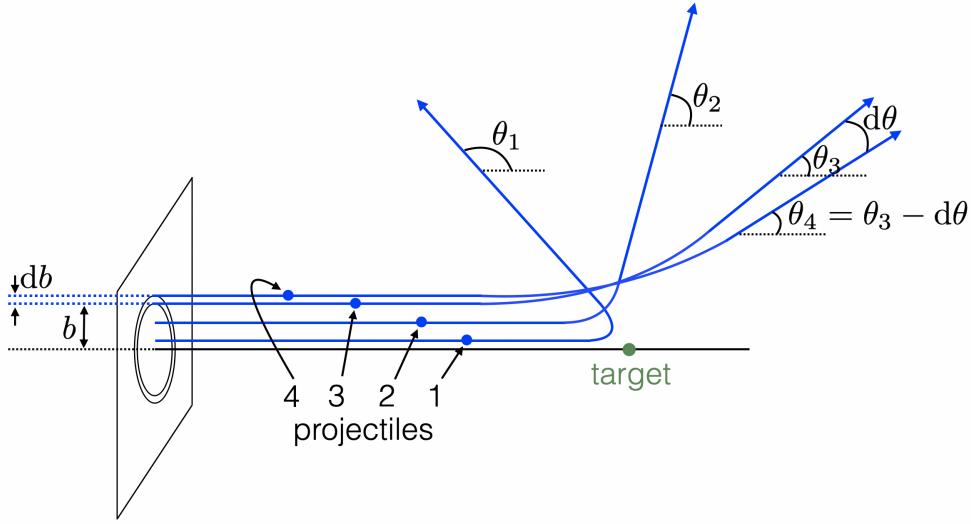


Figure 6.5: Trajectories depend on the impact parameter b . The projectiles in the area $d\sigma = 2\pi bdb$ are scattered by the target (of the same charge sign) into a solid angle $d\Omega = 2\pi \sin \theta d\theta$.

6.1.4 *Rutherford's formula for the differential cross-section

As shown in Fig. 6.5, for a large number of projectiles with impact parameters in the range from b to $b + db$ (this corresponds to an area $d\sigma = 2\pi bdb$ in the direction \perp to the incident velocity), what is the solid angle $d\Omega$ they are scattered into? We differentiate eq. (6.39) and obtain

$$\frac{d\theta}{db} = -\frac{|q_1 q_2|}{\mu v^2 b^2} (\cos \theta + 1) \Rightarrow d\Omega = 2\pi \sin \theta \left| \frac{d\theta}{db} \right| db = 2\pi \sin \theta (\cos \theta + 1) \frac{|q_1 q_2|}{\mu v^2 b^2} db, \quad (6.41)$$

where the negative sign simply means that the deflection angle decreases for larger impact parameters. Therefore, we obtain the following ratio (using $b_{90} = |q_1 q_2| / \mu v^2$)

$$\frac{d\sigma}{d\Omega} = \frac{2\pi bdb}{d\Omega} = \frac{b_{90}^2}{4 \sin^4(\theta/2)}, \quad (6.42)$$

which is called *Rutherford's differential cross-section* — the cross-section per solid angle that the projectile is scattered into.

When attempting to calculate the *total* cross-section by integrating over all the deflection angles $\sigma = \int (d\sigma/d\Omega) d\Omega \sim \int_0^\pi \theta^{-3} d\theta$, one immediately finds that the integral diverges near $\theta \rightarrow 0$. This shows that Coulomb scatterings are dominated by a very large number of small-deflection encounters — we will discuss this point in §6.1.1.

It is sometimes useful to define the *large deflection-angle cross-section* for $\theta > 90^\circ$, which is given by

$$\sigma_{>90} = \int_{\theta=\pi/2}^{\pi} \frac{d\sigma}{d\Omega} d\Omega = \pi b_{90}^2. \quad (6.43)$$

This means that each target has a cross-sectional area of πb_{90}^2 within which the projectile will be deflected by more than 90° in one scattering. Given this result, one might naively guess that the projectile's velocity gets significantly deflected after a timescale the order $1/(n\sigma_{90}v) \sim \mu^2 v^3 / (\pi n q_1^2 q_2^2)$ (where n is the number density of targets), but this turns out to be an overestimation of the deflection timescale (by a factor of $8 \ln \Lambda \sim \mathcal{O}(10^2)$, see eq. 6.10), because the cumulative deflection over time is in fact dominated by a large number of weak encounters each of which has a small deflection angle.

6.2 Free-free emission

In the treatment of Coulomb collisions in §6.1, we have ignored radiative losses, which is an excellent approximation for non-relativistic or mildly relativistic particles. In this section, our goal is to calculate the properties of the EM emission associated with a Coulomb collision between an electron and an ion — this is called *free-free emission* because the initial and final states of the electron are free instead of bound.

Instead of the general 2-body problem, we will restrict us to the simpler case where the projectile is an electron and the target is an ion — the essential assumption is that the projectile moves at a high velocity such that the target essentially stays at rest during the collision process. It should also be noted that electron-electron or proton-proton collisions do not produce changing electric dipole moment, so there will be essentially no free-free emission at the dipole order (although deflection and energy exchange can still occur).

6.2.1 Single-velocity case

Our starting point is the radiation field from an accelerating charge (our projectile electron), but instead of viewing it in the frame where the target ion is at rest, we will switch to the comoving frame of the initial velocity of the electron. In this comoving frame (see Fig. 6.6), we place the electron at the coordinate origin, and the ion moves at a constant velocity $\mathbf{v} = v\hat{\mathbf{x}}$ with equation of motion described by $(x = vt, y = -b, z = 0)$ for $t \in (-\infty, \infty)$. The calculation is similar to that in the Thomson scattering problem except that here it is the Coulomb field of the ion that accelerates the electron which then radiates.

In the non-relativistic limit $v \ll c$, the ion's Coulomb field at the electron's position is

$$\mathbf{E}(t) = \frac{Ze}{[(vt)^2 + b^2]^{3/2}} (b\hat{\mathbf{y}} - vt\hat{\mathbf{x}}), \quad (6.44)$$

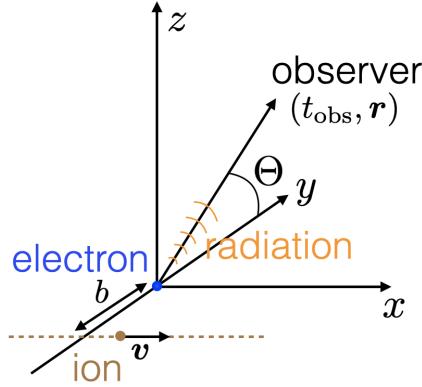


Figure 6.6: Free free emission from electron-ion Coulomb collision, as viewed in the rest frame of the unperturbed electron. The Coulomb field from the moving ion accelerates the electron which then radiates EM waves.

The relativistic version of the above Coulomb field will be used in §6.5 to calculate the free-free emission for relativistic encounters, but here we keep the discussion non-relativistic. This E-field causes the electron to undergo acceleration and the characteristic timescale for this interaction is

$$\tau = \frac{b}{v}. \quad (6.45)$$

Based on this timescale, we expect the free-free emission to be significant only at frequencies $\omega \lesssim \tau^{-1} = v/b$. As we shall see below, the free-free spectrum indeed cuts off exponentially at $\omega \gg \tau^{-1}$. For sufficiently large impact parameters $b \gg b_{90} = Ze^2/(m_e v^2)$ (where b_{90} is the impact parameter for a 90° deflection), the electron's position can be considered to be near the origin all the time. The electron's acceleration is given by

$$\mathbf{a} = \dot{\mathbf{v}}_e = -\frac{\mathbf{E}_e}{m_e}, \quad (6.46)$$

which has two components along \hat{x} and \hat{y} . Based on the acceleration, it is straightforward to calculate the radiation field at any observer's time and position. Before doing that, let us calculate the total radiation energy emitted by the entire Coulomb collision process, which is given by the time integral of the Larmor power (see Ch. 4)

$$W = \frac{2e^2}{3c^3} \int_{-\infty}^{\infty} \dot{v}^2(t) dt = \frac{2Z^2 e^6}{3m_e^2 c^3} \int_{-\infty}^{\infty} \frac{dt}{[(vt)^2 + b^2]^2} = \frac{\pi Z^2 e^6}{3m_e^2 c^3} \frac{1}{vb^3}, \quad (6.47)$$

where we have used $\int_0^\infty dx/(x^2+1)^2 = \pi/4$ by change of variable $x = \tan \phi$. This radiative energy loss should be compared with that due to momentum kick on the ion $(\Delta p_\perp)^2/2m_i$ (eq. 6.2), and the ratio between the two energy loss terms is

$$\frac{W}{(\Delta p_\perp)^2/2m_i} = \frac{\alpha m_i}{12m_e} \left(\frac{v}{c}\right)^2, \quad (6.48)$$

where $\alpha = e^2/\hbar c \approx 1/137$ is the fine structure constant. For a proton target $m_i = m_p$, we have $\alpha m_p/(12m_e) = 1.12$. Thus, for non-relativistic electrons $v \ll c$, the radiative loss from the electron is subdominant compared to the (negative) work done by the ion's Coulomb force. However, in a proton-electron plasma, the energy loss from the fast electron is dominated by electron-electron collisions (since $t_{\text{th}}(\text{e-p})/t_{\text{th}}(\text{e-e}) = m_p/m_e$), and therefore, the relevant energy loss ratio should in fact be

$$\frac{W}{(\Delta p_\perp)^2/2m_e} = \frac{\alpha}{12} \left(\frac{v}{c} \right)^2 \ll 1. \quad (6.49)$$

We conclude that, in the non-relativistic limit, radiative loss plays a subdominant role in decelerating the fast electron compared to the energy loss due to Coulomb scattering. The case for relativistic electrons will be discussed in §6.5. It should also be noted that Coulomb scattering conserves energy in the entire plasma system but the energy taken away by free-free emission is permanently lost from an optically thin system. This leads to free-free cooling of the system (the timescale of which will be discussed later).

Let us then consider the emission spectrum, which can be obtained by applying the Parseval's Theorem for Fourier transformation $\int_{-\infty}^{\infty} |\mathbf{a}|^2 dt = 4\pi \int_0^{\infty} |\tilde{\mathbf{a}}|^2 d\omega$ to eq. (6.47), and the result is

$$\frac{dW}{d\omega} = \frac{8\pi e^2}{3c^3} |\tilde{\mathbf{a}}(\omega)|^2. \quad (6.50)$$

The question comes down to the Fourier transformation of the acceleration vector,

$$\tilde{\mathbf{a}}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{a}(t) e^{i\omega t} dt. \quad (6.51)$$

The first calculation of $\tilde{\mathbf{a}}(\omega)$ in this context was performed by [Kramers \(1923\)](#), who, after obtaining the emission spectrum, then used the Einstein relations to obtain the free-free absorption coefficient. We will follow the same procedure.

When calculating $\tilde{\mathbf{a}}(\omega)$, we first consider the low-frequency limit $\omega \ll \tau^{-1} = v/b$ in which $e^{i\omega t} \approx 1$, and hence

$$\tilde{\mathbf{a}}(\omega \ll v/b) \approx \frac{1}{2\pi} \int \dot{\mathbf{v}}_e(t) dt = \frac{\Delta \mathbf{v}_e}{2\pi}, \quad (6.52)$$

which shows that we only need to know the net change in velocity (or the deflection) over the entire scattering process. Based on eq. (6.2), we know that, for weak encounters with small deflection angles $\theta \ll 1$, the change in velocity $\Delta \mathbf{v}_e$ is dominated by the perpendicular component and is given by

$$\Delta \mathbf{v}_e \approx -\frac{2Ze^2}{m_e v b} \hat{\mathbf{y}}, \quad \text{for weak deflections.} \quad (6.53)$$

This means that the spectrum of the total radiated energy is

$$\frac{dW}{d\omega}(\omega \ll v/b) = \frac{8Z^2e^6}{3\pi m_e^2 c^3} \frac{1}{v^2 b^2}. \quad (6.54)$$

We see that the low-frequency emission spectrum of a single Coulomb scattering is in fact frequency independent! This can be understood from the fact that the Fourier transform of a delta function (since the acceleration occurs on a timescale $\tau \ll \omega^{-1}$) gives a flat spectrum. One can also show that it is linearly polarized in the plane that contains $\hat{\mathbf{r}}$ (the line of sight) and $\hat{\mathbf{y}}$ (the acceleration perpendicular to the trajectory).

So far we have only considered the low-frequency limit $\omega \ll v/b$. In order to obtain the correct result for $\omega \gtrsim v/b$, one must directly compute the Fourier transform of the acceleration vector in eq. (6.51), which is given by

$$\tilde{\mathbf{a}}(\omega) = \frac{Ze^2}{2\pi m_e v b} \int_{-\infty}^{\infty} \frac{\xi \hat{\mathbf{x}} - \hat{\mathbf{y}}}{(\xi^2 + 1)^{3/2}} e^{i\omega\tau\xi} d\xi = \frac{Ze^2}{\pi m_e v b} \int_0^{\infty} \frac{\xi \sin(\omega\tau\xi) \hat{\mathbf{x}} - \cos(\omega\tau\xi) \hat{\mathbf{y}}}{(\xi^2 + 1)^{3/2}} d\xi, \quad (6.55)$$

where $\tau = b/v$, $\xi = t/\tau$, and we have made use of the symmetries of the integrand. We denote $z = \omega\tau$, and two integrals involved in the above expression can be written in terms of modified Bessel functions of integer orders (Abramowitz & Stegun 1964)

$$zK_0(z) = \int_0^{\infty} \frac{\xi \sin(z\xi)}{(\xi^2 + 1)^{3/2}} d\xi, \quad zK_1(z) = \int_0^{\infty} \frac{\cos(z\xi)}{(\xi^2 + 1)^{3/2}} d\xi, \quad (6.56)$$

which have the following asymptotic behaviors

$$\begin{aligned} zK_0(z) &\approx -z \ln z, \quad zK_1(z) \approx 1, \quad \text{for } z \ll 1, \\ zK_0(z) &\approx \sqrt{\frac{\pi z}{2}} e^{-z} \left(1 - \frac{1}{8z}\right), \quad zK_1(z) \approx \sqrt{\frac{\pi z}{2}} e^{-z} \left(1 + \frac{3}{8z}\right), \quad \text{for } z \gg 1, \end{aligned} \quad (6.57)$$

We see that the energy spectrum at high frequencies rapidly cuts off as $dW/d\omega \propto |\tilde{\mathbf{a}}|^2 \propto \omega e^{-2\omega\tau}$ and that near the maximum frequency $\omega \sim v/b$, the emission has mixed polarizations due to contributions from the acceleration components along both $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$.

The angle-integrated total emission spectrum is given by the sum of the two independent contributions from accelerations along $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$, i.e.

$$\frac{dW}{d\omega} = \frac{8Z^2e^6}{3\pi m_e^2 c^3} \frac{1}{v^2 b^2} \left[(zK_0(z))^2 + (zK_1(z))^2 \right], \quad z = \omega\tau, \quad (6.58)$$

which applies to frequencies $\hbar\omega \ll (1/2)m_e v^2$ such that photon discreteness does not play a role. The argument in the modified Bessel functions is given by $\omega\tau = \omega b/v$. The values of the two Bessel function terms are shown in Fig. 6.7. At low frequencies $\omega\tau \ll 1$, free-free

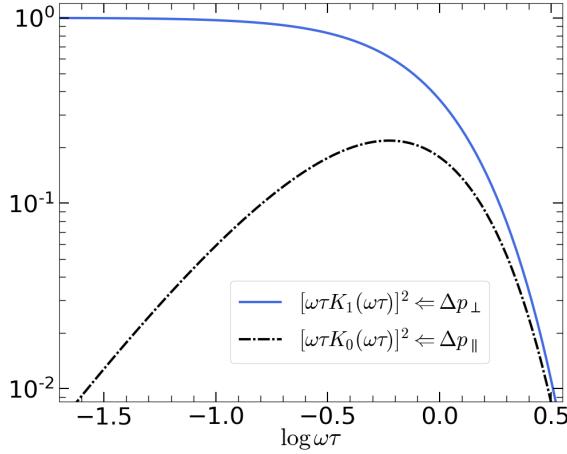


Figure 6.7: The relative contributions to the free-free spectrum from accelerations in the direction perpendicular (the $(\omega\tau K_1(\omega\tau))^2$ term, blue dash-dotted line) and parallel (the $(\omega\tau K_0(\omega\tau))^2$ term, black solid line) to the initial velocity.

emission is dominated by the acceleration perpendicular to the initial velocity, as shown in eq. (6.54).

Then, we move on to consider a wide range of impact parameters. For a plasma with ion number density n_i , an electron with velocity v would undergo Coulomb collisions at a rate $dR = n_i v 2\pi b db$ for impact parameters in the range $(b, b + db)$. Integrating over the full range of impact parameters, we obtain the emitted power per angular frequency

$$P_\omega(v) = \int db 2\pi b n_i v \frac{dW}{d\omega} = \frac{16Z^2 e^6 n_i}{3m_e^2 c^3} \frac{v}{\omega} \ln \Lambda(\omega, v), \quad \text{for } \hbar\omega < m_e v^2 / 2, \quad (6.59)$$

where the Coulomb logarithm is roughly given by

$$\ln \Lambda(\omega, v) = \int_{b_{\min}}^{\infty} \frac{db}{b} \left[(\omega\tau K_0(\omega\tau))^2 + (\omega\tau K_1(\omega\tau))^2 \right] \simeq \ln \frac{b_{\max}}{b_{\min}}. \quad (6.60)$$

In the low-frequency limit $\hbar\omega \ll \hbar\omega_{\max} = m_e v^2 / 2$, the maximum and minimum impact parameters (for given ω and v) are given by

$$b_{\max}(v, \omega) \sim v/\omega, \quad b_{\min}(v) \sim \max \left(\frac{2Ze^2}{m_e v^2}, \frac{2\hbar}{m_e v} \right), \quad (6.61)$$

where we have taken the more conservative one of the two minimum impact parameters $b_{\min}^{(1)}$ (eq. 6.20) and $b_{\min}^{(2)}$ (eq. 6.21). Thus, the Coulomb logarithm is roughly given by

$$\ln \Lambda(\omega \ll \omega_{\max}, v) \simeq \ln \left[\frac{m_e v^3}{2Ze^2 \omega} \min \left(1, \frac{v_{\text{crit}}}{v} \right) \right], \quad v_{\text{crit}} = Z\alpha c, \quad (6.62)$$

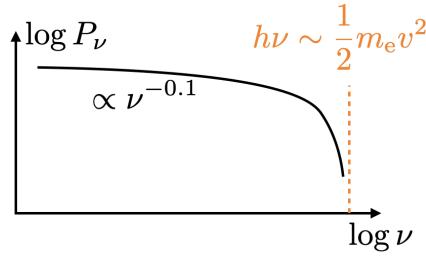


Figure 6.8: Power per unit frequency for a single electron with velocity v . The spectrum is nearly flat ($P_\nu \propto \nu^{-0.1}$) in the low-frequency limit and cuts off near the maximum frequency as restricted by photon discreteness (no photons can be emitted above $h\nu_{\max} = m_e v^2/2$).

where the critical projectile velocity v_{crit} is defined where the two choices for the minimum impact parameter is equal, $b_{\min}^{(2)}/b_{\min}^{(1)} = v/v_{\text{crit}}$.

On the high-frequency end, we know that the maximum possible emitting frequency is given by the quantum limit $\hbar\omega_{\max} = m_e v^2/2$, so the emitting power $P_\omega(v)$ must drop to zero at $\omega > \omega_{\max}$. Unfortunately, the detailed spectral shape for the cut off at $\omega \lesssim \omega_{\max}$ cannot be accurately determined using our classical approach here, except that the emitting power should drop exponentially. A sketch of the emitting spectrum is shown in Fig. 6.8.

It is important to understand the spectral shape at low frequencies $\hbar\omega \ll \hbar\omega_{\max} = m_e v^2/2$. In this limit, the emitting spectrum from a single-velocity electron has a weak dependence on frequency ω with a power-law slope of

$$\frac{d \ln P_\omega}{d \ln \omega} = \frac{1}{\ln \Lambda} \frac{d \ln \Lambda}{d \ln \omega} = -\frac{1}{\ln \Lambda} \simeq -0.1, \quad (6.63)$$

where we have used $\ln \Lambda \sim 10$ at low frequencies. Thus, the low-frequency spectrum for free-free emission is roughly a power-law

$$P_\omega \propto \omega^{-0.1}, \text{ for } \omega \ll \omega_{\max}, \quad (6.64)$$

and this feature also applies to the volume emissivity ($j_\nu \propto \nu^{-0.1}$ at $\nu \ll \nu_{\max}$) when we integrate over the electron velocity distribution. It should be noted that the power per unit frequency P_ν is related to P_ω by

$$P_\nu = \frac{dP}{d\nu} = 2\pi \frac{dP}{d\omega} = 2\pi P_\omega. \quad (6.65)$$

Finally, we calculate the total free-free emitting power of the electron by integrating over the emitting spectrum. Most of the radiation energy is emitted just below the maximum frequency $\hbar\omega_{\max} = m_e v^2/2$, so we can estimate the total power by

$$P_{\text{tot}}(v) = \int d\omega P_\omega \simeq \omega_{\max} P_{\omega \sim \omega_{\max}} \simeq \frac{8Z^2 e^6 n_i v}{3m_e \hbar c^3} = v \sigma_T \frac{\alpha Z^2 n_i m_e c^2}{\pi}, \quad (6.66)$$

where we have used $\ln \Lambda \sim 1$ for $\omega \lesssim \omega_{\max}$. Note that the emitting power in the above expression is written in the form of “velocity \times cross-section \times energy density”— free-free emission may be pictured as an electron with velocity v scattering an effective “Coulomb radiation field” of energy density $\alpha Z^2 n_i m_e c^2 / \pi$ at the Thomson cross-section $\sigma_T = (8\pi/3)r_e^2 = (8\pi/3)e^4/m_e^2 c^4$. In Chapter 7 where we discuss Compton scattering, it will be shown that the inverse-Compton emission power is of the order $c\sigma_T(\gamma^2 U_{\text{rad}})$, where $\gamma^2 U_{\text{rad}}$ is the energy density of the radiation field in the comoving frame of a relativistic electron (and U_{rad} is the energy density in the lab frame). We conclude that a faster electron radiates more free-free emission ($P_{\text{tot}} \propto v$), and it will be shown later that $P_{\text{tot}} \propto \gamma$ for relativistic electrons, where γ is the electron Lorentz factor.

6.2.2 Thermal velocity distribution and cooling timescale

So far we have focused on a single electron with a given velocity. The next step is to consider the velocity distribution function of electrons $f(v)$. For an isotropic velocity distribution, one can easily integrate the single-electron emissivity (power per solid angle per frequency) $P_\nu(v)/4\pi$ over the electron velocity distribution to obtain the volume emissivity

$$j_\nu = n_e \int_{v_{\min}}^{\infty} dv f(v) \frac{P_\nu(v)}{4\pi}, \quad v_{\min} = \sqrt{2h\nu/m_e}, \quad (\text{for isotropic } v\text{-distribution}) \quad (6.67)$$

where we have normalized the velocity distribution such that $\int dv f(v) = 1$ and the minimum velocity v_{\min} is set by the requirement of $h\nu < (1/2)m_e v^2$ since *photons are discrete*—electrons with velocities $v < v_{\min}$ cannot create photons of energy $h\nu$. After obtaining j_ν , one can then calculate the absorption coefficient for any velocity distribution function using the Kirchhoff’s law (which has been discussed in Ch. 2).

The electron velocity distribution “relaxes” to the Maxwellian distribution on the thermalization timescale that is set by Coulomb interactions with other electrons. This thermalization timescale is $t_{\text{th}}(\text{e-e}) \sim 10^4 \text{ s} (T_e/10^4 \text{ K})^{3/2} (n_e/\text{cm}^{-3})^{-1}$ (eq. 6.15) if we express the electron velocity in the temperature form. If the dynamical timescale of the system of interest is much longer than $t_{\text{th}}(\text{e-e})$, then it is appropriate to take the following isotropic, Maxwellian velocity distribution

$$f(v)dv = \frac{4}{\sqrt{\pi}\sigma_v} \frac{v^2}{\sigma_v^2} e^{-v^2/\sigma_v^2} dv, \quad (6.68)$$

which is normalized by imposing $\int_0^\infty f(v)dv = 1$ and the velocity dispersion σ_v is the related to the kinetic temperature T_e by

$$\sigma_v = \sqrt{2k_B T_e/m_e}. \quad (6.69)$$

The root mean square speed is $\sqrt{\langle v^2 \rangle} = \sqrt{3k_B T_e/m_e}$, the mean kinetic energy is $(1/2)m \langle v^2 \rangle = (3/2)k_B T_e$, and the mean speed is $\langle v \rangle = \sqrt{8k_B T_e/\pi m_e}$.

For a thermal velocity distribution at temperature T_e , the free-free emissivity (erg/cm³/s/Hz/sr) is given by the integral in eq. (6.67). Using the single-velocity spectral power in eq. (6.59) together with the Coulomb logarithm in eq. (6.62), we obtain the following result

$$j_\nu = \frac{32}{3\sqrt{\pi}} \frac{Z^2 e^6}{m_e^2 c^3} n_i n_e \times Q(\nu, T), \quad (6.70)$$

where Q is an integral given by

$$\begin{aligned} Q(\nu, T) &= \int_{v_{\min}}^{\infty} dv \frac{v}{\sigma_v^3} e^{-v^2/\sigma_v^2} \ln \Lambda(v) = \frac{1}{\sigma_v} \int_{x_{\min}}^{\infty} dx x e^{-x^2} \ln \Lambda(x) \\ &= \frac{1}{2\sigma_v} \int_{y_{\min}}^{\infty} dy e^{-y} \ln \Lambda(y), \end{aligned} \quad (6.71)$$

and $x \equiv v/\sigma_v$, $y \equiv x^2 = (v/\sigma_v)^2$, and

$$x_{\min}(\nu, T_e) \equiv \frac{v_{\min}}{\sigma_v} = \sqrt{\frac{h\nu}{k_B T_e}}, \quad y_{\min}(\nu, T_e) \equiv \left(\frac{v_{\min}}{\sigma_v} \right)^2 = \frac{h\nu}{k_B T_e}. \quad (6.72)$$

In the following, we discuss the low-frequency ($h\nu \ll k_B T_e$) and high-frequency ($h\nu \gtrsim k_B T$) limits separately.

In the low-frequency limit $h\nu \ll k_B T_e$, the Coulomb logarithm is given by

$$\ln \Lambda(y) \simeq \ln \left[\frac{1}{Z} \frac{(k_B T_e)^{3/2}}{h\nu \sqrt{Ry}} y^{3/2} \min \left(1, \frac{Z}{y^{1/2}} \sqrt{\frac{Ry}{k_B T_e}} \right) \right], \quad (6.73)$$

where $Ry = \alpha^2 m_e c^2 / 2 = 13.6 \text{ eV}$ is the Rydberg energy and we have used $v_{\text{crit}}/v = Z y^{-1/2} \sqrt{Ry/k_B T_e}$. We then carry out the Q -integral in the low-frequency limit such that $y_{\min} = h\nu/k_B T_e \ll 1$. Since most contribution to the Q -integral comes from $y \sim 1$, we can extend the lower limit of the integral from y_{\min} to 0 and take $y \sim 1$ in the Coulomb logarithm term $\ln \Lambda(y) \simeq \ln \Lambda(y \sim 1)$ without introducing much error. Making use of $\int_0^\infty e^{-y} dy = 1$, we obtain

$$Q(\nu \ll k_B T_e/h, T_e) \approx \frac{1}{2\sigma_v} \ln \Lambda(y \sim 1) = \frac{\pi}{2\sqrt{3}\sigma_v} g_{\text{ff}}(\nu, T_e), \quad (6.74)$$

where g_{ff} is called the *Gaunt factor*

$$g_{\text{ff}}(\nu, T_e) = \frac{\sqrt{3}}{\pi} \ln \Lambda(y \sim 1) \simeq \frac{\sqrt{3}}{\pi} \ln \left[\frac{1}{Z} \frac{(k_B T_e)^{3/2}}{h\nu \sqrt{Ry}} \min \left(1, Z \sqrt{\frac{Ry}{k_B T_e}} \right) \right], \quad \text{for } h\nu \ll k_B T_e, \quad (6.75)$$

and the definition of the Gaunt factor contains a factor of $\sqrt{3}/\pi$ in front of the Coulomb logarithm so as to be consistent with the expectation of $g_{\text{ff}} \approx 1$ in the quantum mechanical limit of $h\nu \gtrsim k_B T_e$. We conclude that, in the low-frequency limit $h\nu \ll k_B T_e$, the Gaunt factor has two regimes

$$g_{\text{ff}}(\nu \ll k_B T_e/h, T_e) \simeq \begin{cases} \frac{\sqrt{3}}{\pi} \ln \left[\frac{(k_B T_e)^{3/2}}{h\nu \sqrt{\text{Ry}}} \right] = 6.0 \left[1 - \ln \left(\frac{Z\nu_9}{T_{e,4}^{3/2}} \right) \right], & \text{if } T_e < 1.6 \times 10^5 \text{ K}, \\ \frac{\sqrt{3}}{\pi} \ln \left(\frac{k_B T_e}{h\nu} \right) = 9.3 \left[1 - \ln \left(\frac{\nu_9}{T_{e,6}} \right) \right], & \text{if } T_e > 1.6 \times 10^5 \text{ K}, \end{cases} \quad (6.76)$$

where we have used $\nu_9 = \nu/\text{GHz}$, $T_{e,4} = T_e/10^4 \text{ K}$, and $T_{e,6} = T_e/10^6 \text{ K}$.

Let us then consider the high-frequency limit $h\nu \gtrsim k_B T_e$ where our classical approach breaks down because the photon energy is comparable to the electron energy. When considering the Q-integral in the high-frequency limit $h\nu \gtrsim k_B T_e$, we expect the Gaunt factor to be close to unity, $g_{\text{ff}}(h\nu \gtrsim k_B T_e) = (\sqrt{3}/\pi) \ln \Lambda \simeq 1$, so it is reasonable to take $\ln \Lambda \simeq \pi/\sqrt{3}$ and the Q-integral can be estimated as follows

$$Q(\nu \gtrsim k_B T_e/h, T_e) \simeq \frac{\pi}{2\sqrt{3}\sigma_v} \int_{y_{\min}}^{\infty} e^{-y} dy = \frac{\pi}{2\sqrt{3}\sigma_v} e^{-h\nu/(k_B T_e)}, \quad (6.77)$$

where we have used $y_{\min} = h\nu/(k_B T_e)$.

Finally, we put the two limits together and write the free-free volume emissivity as

$$j_{\nu} = \underbrace{\frac{16\sqrt{\pi}e^6}{3\sqrt{3}m_e^2c^3}}_{\approx 3.00 \times 10^{-33} \text{ cgs}} \underbrace{\left(\frac{m_e}{2k_B T_e} \right)^{1/2}}_{= 1/\sigma_v} \underbrace{Z^2 n_i n_e g_{\text{ff}}(\nu, T_e)}_{\propto \rho^2} e^{-h\nu/k_B T_e}, \quad (6.78)$$

where the frequency-dependent *Gaunt factor for free-free emission*. The Gaunt factor has been calculated using detailed quantum mechanical techniques ([Karzas & Latter 1961](#), [Hummer 1988](#)). The following expression extends the analytic fit provided in Fig. 10.1 in Draine's book from the low-temperature regime ($T_e \lesssim 5 \times 10^4 \text{ K}$) to all non-relativistic temperatures

$$g_{\text{ff}}(\nu, T_e, Z) \approx \ln \left[e + \exp \left(6.0 - \frac{\sqrt{3}}{\pi} \ln \left(\nu_9 T_{e,4}^{-1} \max(0.25, Z T_{e,4}^{-1/2}) \right) \right) \right], \quad (6.79)$$

where $e = 2.71828$ is the base for natural logarithm.

If there are multiple ion species, the free-free emissivity is given by the sum of all

$$\begin{aligned} j_{\nu} &= \frac{16\sqrt{\pi}e^6}{3\sqrt{3}m_e^2c^3} \left(\frac{m_e}{2k_B T_e} \right)^{1/2} n_e \left(\sum_i g_{\text{ff},i} Z_i^2 n_i \right) e^{-h\nu/k_B T_e} \\ &= 5.44 \times 10^{-41} \frac{\text{erg}}{\text{cm}^3 \text{s Hz sr}} \left(\frac{T_e}{10^4 \text{ K}} \right)^{-1/2} \frac{n_e}{\text{cm}^{-3}} \frac{\sum_i g_{\text{ff},i} Z_i^2 n_i}{\text{cm}^{-3}} e^{-h\nu/k_B T_e}, \end{aligned} \quad (6.80)$$

where the sum \sum_i takes into account the contributions from multiple ion species (assumed to be hydrogen-like with charge number Z_i) and $g_{\text{ff},i}(\nu, T_e)$ is the Gaunt factor for a given ion species. It should be noted that the Gaunt factor may be much greater than unity at radio frequencies, so it is not negligible.

For a fully ionized gas with hydrogen and helium mass fractions X and $1 - X$ (solar composition corresponding to $X \approx 0.7$), we have the following approximations $\sum_i g_{\text{ff},i} Z_i^2 n_i \approx g_{\text{ff}}(\nu, T_e, Z = 1) \rho / m_p$ and $n_e \approx (1 + X) \rho / (2m_p)$, where ρ is the mass density and m_p is the proton mass. Thus, we can simplify the emissivity into the following form

$$j_\nu \approx 9.73 \times 10^6 \frac{\text{erg}}{\text{cm}^3 \text{s Hz sr}} (1 + X) g_{\text{ff}}(\nu, T_e) \left(\frac{T_e}{10^4 \text{K}} \right)^{-1/2} \left(\frac{\rho}{\text{g/cm}^3} \right)^2 e^{-h\nu/k_B T_e}, \quad (6.81)$$

for a mixture of H and He (for a cosmic composition) and it is appropriate to take the Gaunt factor g_{ff} for hydrogen ions ($Z = 1$) as they dominate the free-free emission.

We see that free-free emission from a thermal plasma has a shallow power-law spectrum of $j_\nu \propto \nu^{-0.1}$ (due to the Gaunt factor) in the radio band, then transitions to a flat spectrum when $g_{\text{ff}} \approx 1$ (but still $h\nu \ll k_B T_e$), and then exponentially cuts off above $h\nu \sim k_B T_e$.

The frequency-integrated free-free energy loss rate per unit volume is given by

$$\begin{aligned} 4\pi j_{\text{ff}} &= 4\pi \int_0^\infty j_\nu d\nu = \frac{32\pi^{3/2} e^6}{3\sqrt{3} h m_e c^3} \left(\frac{2k_B T_e}{m_e} \right)^{1/2} n_e \sum_i \langle g_{\text{ff},i} \rangle_T Z_i^2 n_i \\ &= \frac{2}{\sqrt{3}\pi} n_e \sigma_T \sigma_v \left(\alpha m_e c^2 \sum_i \langle g_{\text{ff},i} \rangle_T Z_i^2 n_i \right) \\ &\approx 1.42 \times 10^{-25} \frac{\text{erg}}{\text{cm}^3 \text{s}} \left(\frac{T_e}{10^4 \text{K}} \right)^{1/2} \frac{n_e}{\text{cm}^{-3}} \frac{\sum_i \langle g_{\text{ff},i} \rangle Z_i^2 n_i}{\text{cm}^{-3}}, \end{aligned} \quad (6.82)$$

where the *frequency-integrated Gaunt factor* is given by the following numerical fit provided by §10.3 Draine's book (with fractional errors $< 2\%$ for $10^{4.2} < T < 10^{8.2}$ K)

$$\langle g_{\text{ff}} \rangle \approx 1 + \frac{0.44}{1 + 0.058 \left[\ln \left(\frac{T_e/\text{K}}{10^{5.4} Z^2} \right) \right]^2}. \quad (6.83)$$

Since most energy from free-free emission is near frequency $h\nu \sim k_B T_e$ where the gaunt factor $g_{\text{ff}} \approx 1$, we see that the frequency-integrated Gaunt factor $\langle g_{\text{ff}} \rangle_T$ is close to unity.

The free-free cooling timescale t_{ff} is defined as the ratio between the electron thermal energy density $(3/2)n_e k_B T_e$ and the free-free cooling rate,

$$t_{\text{ff}} = \frac{(3/2)n_e k_B T_e}{4\pi j_{\text{ff}}} = 4.6 \times 10^5 \text{yr} \left(\frac{T_e}{10^4 \text{K}} \right)^{1/2} \frac{\text{cm}^{-3}}{\sum_i \langle g_{\text{ff},i} \rangle_T Z_i^2 n_i}. \quad (6.84)$$

For a fully ionized gas with mass density ρ dominated by hydrogen and helium, it is reasonable to use the following approximation $\sum_i Z_i^2 n_i \approx \rho/m_p$. At sufficiently high temperatures $T_e \gtrsim 10^7$ K where free-free cooling dominates over collisionally excited line emission lines and free-bound transitions, the Gaunt factor is very close to unity ($1 \lesssim \langle g_{\text{ff}} \rangle_T \lesssim 1.2$), so it is appropriate to take $\langle g_{\text{ff}} \rangle_T \simeq 1$ and write

$$t_{\text{ff}} \simeq 0.24 \text{ yr} \left(\frac{T_e}{10^7 \text{ K}} \right)^{1/2} \left(\frac{\rho}{10^{-16} \text{ g cm}^{-3}} \right)^{-1}. \quad (6.85)$$

It is interesting to note that the free-free cooling time $t_{\text{ff}} \propto T_e^{1/2}$ decreases as the temperature drops — this means that the plasma cools faster and faster over time.

It is also interesting to compare the free-free cooling timescale for a thermal plasma to the electron-ion thermalization timescale in eq. (6.17),

$$\frac{t_{\text{ff}}}{t_{\text{th(e-i)}}} \simeq 5 \times 10^5 (T_e/10^4 \text{ K})^{-1}, \quad (6.86)$$

which is for fully ionized plasma with solar abundance. Typical ionized gas in the interstellar medium have temperatures $10^4 \lesssim T_e \lesssim 10^6$ K, so the free-free cooling time is much longer than the thermalization time $t_{\text{ff}} \gg t_{\text{th(e-i)}}$, and ions and free electrons share the same kinetic temperature unless other cooling mechanisms or recombination operate on timescales $\ll t_{\text{th(e-i)}}$. However, for a plasma with relativistic electrons $T_e \gtrsim 5 \times 10^9$ K, the free-free cooling timescale becomes shorter than the electron-ion thermalization timescale, so electrons cannot reach energy equipartition with ions via Coulomb collisions. Other *collisionless* plasma processes (involving electric and magnetic fields) will be controlling the electron velocity distribution.

Finally, we mention that, at very high temperatures of $10^8 \lesssim T \lesssim \text{few} \times 10^9$ K (e.g., for the intracluster medium in the galaxy clusters or the gas very close to a black hole), relativistic correction becomes important. The frequency-dependent emissivity is roughly flat ($j_\nu \propto \nu^0$) up to the maximum photon energy comparable to the electron's non-relativistic kinetic energy, $h\nu_{\text{max,nrel}} \sim k_B T/(m_e c^2)$. Most of the frequency-integrated energy $j_{\text{ff}} = \int j_\nu d\nu$ is emitted near $k_B T/(m_e c^2)$. One of the correction terms is the relativistic kinetic energy

$$(\gamma - 1)m_e c^2 \approx \frac{\beta^2 m_e c^2}{2}(1 + 3\beta^4/4), \text{ for } \beta \ll 1. \quad (6.87)$$

Since $\beta^2 \sim k_B T/(m_e c^2)$, we see that the relativistic version of the maximum photon energy is $h\nu_{\text{max,rel}} = h\nu_{\text{max,nrel}}(1 + \xi)$, where $\xi \sim k_B T/(m_e c^2)$ is the relativistic correction term. To obtain a more accurate expression, one must consider the relativistic corrections to the momentum distribution function $f(p)$ as well as the emissivity j_ν , and the final result for

the total emissivity with relativistic correction is given by [Gould \(1980\)](#),

$$j_{\text{ff,rel}} = j_{\text{ff,nrel}} \left(1 + \frac{19k_{\text{B}}T}{24m_{\text{e}}c^2} \right), \text{ for } T \lesssim \text{few} \times 10^9 \text{ K}, \quad (6.88)$$

where $j_{\text{ff,nrel}}$ is the non-relativistic result given by eq. (6.82).

6.2.3 Emission measure and volume-filling factor

For an optically thin plasma, the intensity from free-free emission is proportional to the spatial integral of the emissivity along our line of sight. Since the ion number density is proportional to electron number density (for charge neutrality), it is convenient to define the *emission measure*

$$\text{EM} \equiv \int n_{\text{e}}^2 d\ell. \quad (6.89)$$

Thus, the *optically thin* free-free intensity can be written as (for $\tau_{\nu} \ll 1$)

$$I_{\nu}^{\text{thin}} = \int j_{\nu} d\ell = 5.44 \times 10^{-15} \frac{\text{erg}}{\text{cm}^2 \text{s Hz sr}} \left(\frac{T_{\text{e}}}{10^4 \text{K}} \right)^{-1/2} \frac{\text{EM}}{10^{26} \text{cm}^{-5}} \frac{\sum_i g_{\text{ff},i} Z_i^2 n_i}{n_{\text{e}}} e^{-\frac{h\nu}{k_{\text{B}}T_{\text{e}}}}. \quad (6.90)$$

Some textbooks (e.g., Draine 2011) define $\text{EM} = \int n_{\text{p}} n_{\text{e}} d\ell$ (n_{p} being the number density of protons), which can be measured by the strength of hydrogen recombination lines. Measuring the optically thin free-free intensity gives $\text{EM} \times \sum_i g_{\text{ff},i} Z_i^2 n_i / n_{\text{e}}$ if we use the EM defined in eq. (6.89), and one must estimate the ionization fractions of different species to further deduce EM. For nearly fully ionized gas of cosmic composition (hydrogen to helium number ratio of about 9 : 1), the difference between the two definitions is not large. For the sake of self-consistency, one should be careful when applying different methods to infer the EM.

Based on dimensional analysis, one may relate the EM to the dispersion measure $\text{DM} = \int n_{\text{e}} d\ell$ (column density of free electrons) and write

$$\frac{\text{EM}}{\text{DM}} = \frac{\langle n_{\text{e}}^2 \rangle}{\langle n_{\text{e}} \rangle} \geq \langle n_{\text{e}} \rangle, \quad (6.91)$$

where $\langle \dots \rangle$ denotes the average along a given path and the following inequality is used

$$\langle n_{\text{e}}^2 \rangle - \langle n_{\text{e}} \rangle^2 = \langle (n_{\text{e}} - \langle n_{\text{e}} \rangle)^2 \rangle \geq 0. \quad (6.92)$$

Because EM is heavily weighted by the densest regions along the path whereas DM simply counts the column density of free electrons (which is not necessarily dominated by the densest regions), the interstellar medium of the Milky Way often has $\text{EM}/\text{DM} \gg \langle n_{\text{e}} \rangle$.

To demonstrate the possibility of $\langle n_e^2 \rangle \gg \langle n_e \rangle$, let us consider a spherical distribution cloud made of small, spherical “droplets”. Each droplet has radius of r and uniform electron density of n_e , whereas the space in between droplets has negligible electron density. Each droplet has $4\pi r^3 n_e / 3$ electrons. The entire cloud has radius R and total number of free electrons N_e . Thus, the average electron number density throughout the entire cloud is $\langle n_e \rangle = 3N_e/(4\pi R^3)$ and we can define the *volume-filling factor* for the droplets as

$$f_V \equiv \frac{\langle n_e \rangle}{n_e} = \frac{3N_e}{4\pi R^3 n_e}. \quad (6.93)$$

Given that each droplet has a cross-section of $\sim r^2$ and the number density of droplets inside the cloud is $\sim f_V/r^3$, the mean free path is of the order r/f_V . Thus, for a given line of sight that has a path length $\ell \sim R$ through the cloud, the number of droplets it will intersect is roughly given by $\ell f_V/r$. Since each intercepted droplet contributes an emission measure of the order $n_e^2 r$, we obtain the total emission measure for this line of sight

$$\text{EM} \sim n_e^2 r (\ell f_V/r) = n_e^2 \ell f_V = \langle n_e \rangle^2 \ell / f_V, \quad (6.94)$$

where we have used $n_e = \langle n_e \rangle / f_V$. If we define $\widetilde{\text{EM}} \equiv \langle n_e \rangle^2 \ell$ as the emission measure for a hypothetical cloud with uniform density, then the realistic emission measure for a clumpy electron distribution is given by

$$\text{EM} = \widetilde{\text{EM}} / f_V, \quad (6.95)$$

which may be much larger than the case of a uniform cloud if the volume filling factor is small ($f_V \ll 1$). For this reason, one needs to be careful when trying to infer the gas number density from the emission measure alone.

6.3 Free-free absorption

Once we know the per-particle emissivity as well as the velocity distribution function, it is straightforward to calculate the absorption coefficient using Einstein relations.

6.3.1 Absorption opacity for a thermal plasma

For a (thermal) Maxwell-Boltzmann velocity distribution, the *net* absorption coefficient of a given plasma is given by the Kirchhoff’s law

$$\alpha_\nu = \frac{j_\nu}{B_\nu(T_e)} = \frac{4\sqrt{\pi}e^6}{3\sqrt{3}m_e^2hc} \left(\frac{2m_e}{k_B T_e} \right)^{1/2} \frac{1 - e^{-h\nu/k_B T_e}}{\nu^3} n_e \sum_i g_{\text{ff},i}(\nu, T_e) Z_i^2 n_i, \quad (6.96)$$

where the $1 - e^{-h\nu/k_B T_e}$ factor accounts for contributions from both true absorption (“1”) and stimulated emission (“ $e^{-h\nu/k_B T_e}$ ”). Practically, we only care about free-free absorption

at low frequencies in the Rayleigh-Jeans (RJ) limit $h\nu \ll k_{\text{B}}T_{\text{e}}$. This is because, at higher frequencies $h\nu \gtrsim k_{\text{B}}T_{\text{e}}$, it is often the case that the plasma is either very optically thin or that other sources of opacities (e.g., bound-free, electron scattering) are more important than free-free absorption.

Thus, we write the free-free optical depth in the RJ limit in terms of the emission measure,

$$\tau_{\nu} = \int \alpha_{\nu} d\ell = 1.77 \nu_{\text{GHz}}^{-2} \left(\frac{T_{\text{e}}}{10^4 \text{ K}} \right)^{-3/2} \frac{\text{EM}}{10^{26} \text{ cm}^{-5}} \frac{\sum_i g_{\text{ff},i} Z_i^2 n_i}{n_{\text{e}}} \quad (\text{for } h\nu \ll k_{\text{B}}T_{\text{e}}). \quad (6.97)$$

Note that, the Gaunt factor $g_{\text{ff},i}$ for each species i depends on frequency, temperature, and ion charge, and it is usually not negligible at radio frequencies.

Using the approximation for g_{ff} in eq. (6.79), one can calculate the free-free optical depth τ_{ν} for any gas composition. In the simplest case of a fully ionized gas with a single ion species of charge number Z and for $T_{\text{e}} \lesssim 1.6 \times 10^5 \text{ K}$, the optical depth is well described by the following approximation

$$\tau_{\nu} \approx 1.0 Z^{0.9} \nu_{\text{GHz}}^{-2.1} \frac{\text{EM}}{10^{25} \text{ cm}^{-5}} \left(\frac{T_{\text{e}}}{10^4 \text{ K}} \right)^{-1.35} \quad (6.98)$$

The free-free self-absorption frequency ν_{a} is defined where $\tau_{\nu_{\text{a}}} = 1$, and it is approximately given by

$$\nu_{\text{a}} \approx 1.0 \text{ GHz } Z^{0.4} \left(\frac{\text{EM}}{10^{25} \text{ cm}^{-5}} \right)^{0.47} \left(\frac{T_{\text{e}}}{10^4 \text{ K}} \right)^{-0.6}. \quad (6.99)$$

For solar composition, the main contributors are from H and He, and the contribution from heavier ions are only at the percent level. In the case of singly ionized He (for $T_{\text{e}} \sim 10^4 \text{ K}$), the free-free optical depth is higher than the H-only case by a factor of $\simeq 1 + 0.1 = 1.1$ for number ratio $N_{\text{He}}/N_{\text{H}} \simeq 0.1$; whereas if He is fully ionized (for $T_{\text{e}} \gtrsim 10^5 \text{ K}$), then we obtain an enhancement factor of $\simeq 1 + 0.1 \times 2^2 = 1.4$ as compared to the H-only case.

The free-free opacity for fully ionized gas of cosmic composition (primarily H and He) is given by

$$\kappa_{\text{ff},\nu}(\text{RJ}) = \frac{\alpha_{\nu}}{\rho} \approx 1.06 \times 10^{-2} \text{ cm}^2 \text{ g}^{-1} g_{\text{ff}}(\nu, T_{\text{e}}, Z = 1) \frac{n_{\text{e}}}{\text{cm}^{-3}} \nu_{\text{GHz}}^{-2} \left(\frac{T_{\text{e}}}{10^4 \text{ K}} \right)^{-3/2}, \quad (6.100)$$

where we have approximated $\sum_i g_{\text{ff},i} Z_i^2 n_i \approx g_{\text{ff}}(\nu, T_{\text{e}}, Z = 1) \rho / m_{\text{p}}$ using the Gaunt factor for hydrogen ($Z = 1$) and taken $[1 - e^{-h\nu/k_{\text{B}}T_{\text{e}}}] (k_{\text{B}}T_{\text{e}}/h\nu) \approx 1$ in the RJ limit.

6.3.2 Rosseland- and Planck-mean opacities

In this section, we estimate the Planck-mean (κ_P) and Rosseland-mean (κ_R) opacities for free-free absorption. Fortunately, for the integrals involved in the Planck-mean and Rosseland-mean opacities, since most contribution comes from the highest frequencies $h\nu \sim k_B T_e$ where the $g_{ff} \simeq 1$, we simply ignore the Gaunt factor in the following estimates.

The Planck-mean opacity is given by

$$\begin{aligned}\kappa_P &= \frac{\pi}{\sigma_{SB} T^4} \int_0^\infty \kappa_\nu B_\nu(T) d\nu \simeq \text{const} \cdot \frac{n_e}{T^{4.5}} \int_0^\infty \frac{1 - e^{-h\nu/k_B T}}{h\nu^3} B_\nu d\nu \\ &\simeq \frac{4\sqrt{\pi}e^6}{3\sqrt{3}m_e^2c} \left(\frac{2m_e}{k_B T}\right)^{1/2} \frac{n_e}{m_p} \frac{\pi}{\sigma_{SB} T^4} \frac{2k_B T}{hc^2} = 3.8 \times 10^{-14} \frac{\text{cm}^2}{\text{g cm}^{-3}} \left(\frac{T}{10^4 \text{K}}\right)^{-3.5}.\end{aligned}\quad (6.101)$$

The usage of κ_P is primarily for computing the rate of momentum transport from radiation to matter, under the assumption that the radiation spectrum has a blackbody shape at the same temperature as the electrons in the plasma (both denoted as T here).

On the other hand, the Rosseland-mean opacity is used for computing the rate of radiative energy transport in a radiation-matter mixture under LTE in a very optically thick medium (e.g., in stellar interior). The Rosseland-mean opacity is given by

$$\begin{aligned}\kappa_R^{-1} &= \frac{\pi}{acT^3} \int_0^\infty \kappa_\nu^{-1} \partial_T B_\nu(T) d\nu \simeq \text{const} \cdot \frac{1}{n_e T^{2.5}} \int_0^\infty \frac{h\nu^3}{1 - e^{-h\nu/k_B T}} \partial_T B_\nu d\nu \\ &\simeq \left[\frac{4\sqrt{\pi}e^6}{3\sqrt{3}m_e^2c} \left(\frac{2m_e}{k_B T}\right)^{1/2} \frac{n_e}{m_p} \right]^{-1} \frac{\pi}{acT^3} \frac{2k_B^7 T^6}{h^5 c^2} \int_0^\infty \frac{x^7 e^{2x} dx}{(e^x - 1)^3} \\ &\Rightarrow \kappa_R \simeq 1.2 \times 10^{-15} \frac{\text{cm}^2}{\text{g cm}^{-3}} \left(\frac{T}{10^4 \text{K}}\right)^{-3.5},\end{aligned}\quad (6.102)$$

where $a = 4\sigma_{SB}/c$ is the radiation density constant and we have obtained $\int_0^\infty x^7 e^{2x} (e^x - 1)^{-3} dx = 5.10 \times 10^3$ from numerical integration. We find that the Rosseland-mean opacity is much smaller than the Planck-mean opacity. This is because the integral in κ_R is stronger weighted towards higher frequencies where the κ_ν is lower. Both κ_P and κ_R have the same density and temperature scalings $\kappa \propto \rho T^{-3.5}$.

6.3.3 Free-free spectrum without electron scattering

In this subsection, we calculate the broadband free-free spectrum from a uniform thermal slab with temperature T_e and emission measure EM. We will first consider the simpler case where the optical depth for Thomson scattering is negligible $\tau_s \ll 1$.

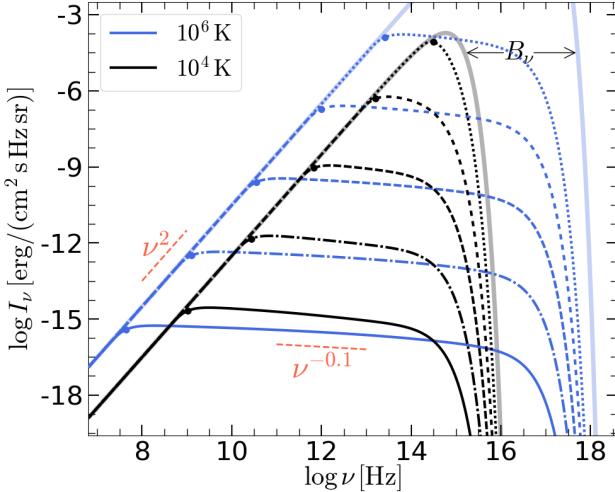


Figure 6.9: Specific intensity I_ν for the free-free emission from a uniform slab of thermal proton-electron plasma at different temperatures $T_e = 10^4$ and 10^6 K. Different dashes are for emission measures $\text{EM} = 10^{25}, 10^{28}, 10^{31}, 10^{34}, 10^{37} \text{ cm}^{-5}$ (from bottom to up). The filled circle on each line denotes where $\tau_{\nu_a} = 1$. We assume $\tau_s \ll 1$ for Thomson scattering. The Planck functions at the two temperatures are shown by the thick solid lines.

Ignoring scattering, the radiative transfer equation along a ray that is normal to the slab is $dI_\nu/ds = j_\nu - \alpha_\nu I_\nu$. Using Kirchhoff's law $j_\nu = \alpha_\nu B_\nu(T_e)$, we write

$$\frac{dI_\nu}{ds} = \alpha_\nu [B_\nu(T_e) - I_\nu], \quad (6.103)$$

which can easily be integrated as B_ν does not depend on coordinate s along the ray (for a uniform slab). The result is

$$I_\nu = B_\nu(T_e)(1 - e^{-\tau_\nu}), \quad (6.104)$$

where $\tau_\nu = \int \alpha_\nu ds$ is the optical depth along the ray (see eq. 6.97). The free-free self-absorption frequency ν_a is defined by $\tau_{\nu_a} = 1$. We find that, at low frequencies $\nu \ll \nu_a$, the intensity approaches the Planck function $B_\nu(T_e)$, and that at high frequencies $\nu \gg \nu_a$, the intensity is given by $I_\nu \approx \tau_\nu B_\nu = \int j_\nu ds \propto \text{EM}$ as the slab is optically thin. The broadband free-free spectra are shown in Fig. 6.9 for a number of cases with different emission measures and electron temperatures.

6.3.4 *Free-free spectrum with electron scattering

Let us then consider the case where the slab has very large Thomson scattering optical depth $\tau_s \gg 1$. Since the opacity due to free-free absorption rapidly drops with frequency, we expect the opacity due to Thomson scattering to be more important than free-free

absorption at sufficiently high frequencies (e.g., the optical band). The interplay between free-free absorption and Thomson scattering at different frequencies makes the emerging spectrum more complicated than the Thomson thin case considered above.

Using eq. (6.100) for the free-free absorption opacity $\kappa_{\text{ff},\nu}$ in the RJ limit and the Thomson opacity $\kappa_s = n_e \sigma_T / \rho$ (frequency independent), we obtain the ratio between the two

$$\frac{\kappa_{\text{ff},\nu}}{\kappa_s} \simeq 12.6 g_{\text{ff}} \rho_{-20} \nu_{\text{GHz}}^{-2} (T_e/10^4 \text{ K})^{-3/2}, \quad (6.105)$$

where $g_{\text{ff}}(\nu, T_e, Z = 1)$ is the Gaunt factor for hydrogen and we have used $\rho_{-20} = \rho/10^{-20} \text{ g cm}^{-3}$ as a reference point for the gas density. Then, we obtain the critical frequency ν_b at which the two opacities are equal to each other

$$\nu_b \simeq 7.9 \text{ GHz } [g_{\text{ff}}(\nu_b)/5]^{1/2} \rho_{-20}^{1/2} (T_e/10^4 \text{ K})^{-3/4}, \text{ for } \kappa_{\text{ff},\nu} = \kappa_s. \quad (6.106)$$

Using the critical frequency ν_b , we write the free-free optical depth $\tau_{\text{ff}}(\nu)$ in terms of the Thomson optical depth τ_s ,

$$\tau_{\text{ff}}(\nu) \simeq \tau_s (\nu/\nu_b)^{-2}, \quad (6.107)$$

where we have ignored the very weak frequency-dependence of the Gaunt factor.

The interplay between absorption and scattering is captured by the effective optical depth

$$\tau_{\text{eff}}(\nu) \simeq \sqrt{\tau_{\text{ff}}(\nu) + \tau_s}, \quad (6.108)$$

where we have ignored a factor of $\sqrt{3}$ (see eq. 2.271) so as for our formulism to be consistent from optically thick to optically thin limits. Even for $\tau_s \gg 1$, we expect the slab to be *effectively* optically thin $\tau_{\text{eff}} \ll 1$ at sufficiently high frequencies where $\tau_{\text{ff}}(\nu) \ll \tau_s$. This motivates us to define another critical frequency ν_t at which $\tau_{\text{eff}} = 1$. The subscript in ν_t is for “translucent” (following [Felten & Rees 1972](#)), which means that light can pass through, but the detailed shape of the source is lost due to scatterings.

We are interested in the limit of $\tau_s \gg 1$, so we know $\nu_t \gg \nu_b$, so we take the limit $\tau_{\text{eff}} \simeq \sqrt{\tau_{\text{ff}} \tau_s} \simeq \tau_s (\nu/\nu_b)^{-1}$ and obtain

$$\nu_t \simeq \tau_s \nu_b \gg \nu_b, \text{ for } \tau_{\text{eff}}(\nu_t) = 1 \text{ (and } \tau_s \gg 1\text{).} \quad (6.109)$$

In the following, we discuss the broadband spectrum based on ν_b and ν_t .

- (i) At $\nu \ll \nu_b$, we have $\tau_{\text{eff}} \simeq \tau_{\text{ff}} \gg \tau_s \gg 1$, so the plasma is absorption dominated and optically thick. In this case, we expect the emerging intensity to be equal to that of a blackbody $I_\nu \approx B_\nu(T_e) \propto \nu^2$ (in the RJ limit).
- (ii) At $\nu \gg \nu_t$, all the free-free emission generated inside the slab will escape, so we obtain $I_\nu = \int j_\nu ds = I_\nu(\text{thin})$ as in the optical thin case.

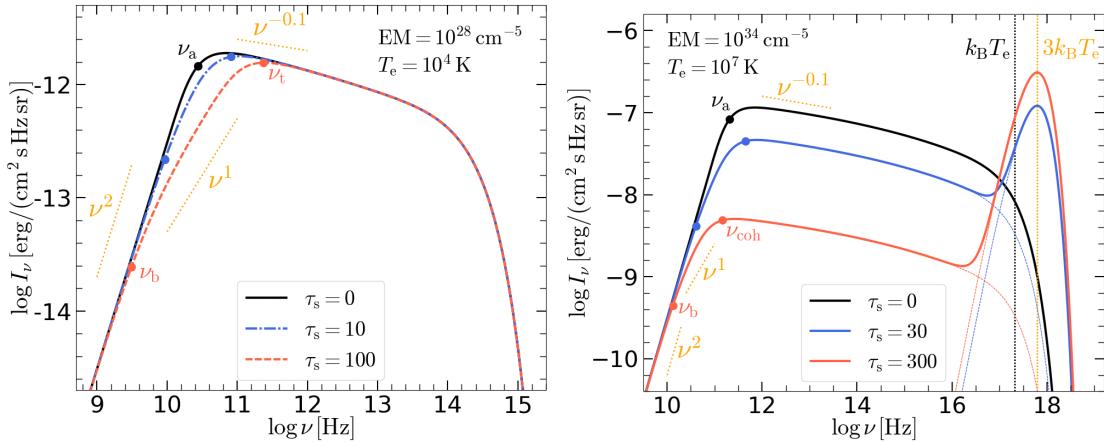


Figure 6.10: Broadband spectrum for free-free emission from a uniform slab of thermal proton-electron plasma. In each panel, we fix the emission measure and electron temperature, and different Thomson optical depths τ_s are considered. The effects of Comptonization are ignored for the Compton parameter $y < 1$ (all cases, left panel); whereas if $y > 1$ ($\tau_s = 30, 300$ cases, right panel), low-energy photons are Compton upscattered and form the Wien component with peak energy at $3k_B T_e$.

- (iii) In between these two critical frequencies $\nu_b \ll \nu \ll \nu_t$, the slab is scattering dominated (as $\tau_s \gg \tau_{ff}$) and effectively optically thick. In this regime, the observed emission only comes from the surface layer with effective optical thickness $\Delta\tau_{eff} \sim 1$, which corresponds to a geometrical thickness of

$$\Delta\ell \simeq \frac{\ell}{\tau_s} \frac{\nu}{\nu_b}, \quad (6.110)$$

where ℓ is the total thickness of the slab. The observed intensity can be estimated from the emission from the surface layer of thickness $\Delta\ell$, so we obtain

$$I_\nu \simeq j_\nu \Delta\ell = j_\nu \ell \frac{1}{\tau_s} \frac{\nu}{\nu_b}. \quad (6.111)$$

Thus, we obtain a power-law spectrum $I_\nu \propto \nu$ in between the two critical frequencies. This is a demonstration that $\tau_{eff}(\nu) \gg 1$ does not necessarily lead to a thermal (blackbody) *emerging spectrum* near frequency ν — it only means that the radiation field near frequency ν in the deep interior layers where $\tau_{eff}(\nu) \gg 1$ is thermalized (but the emerging spectrum only comes from the surface layers with $\tau_{eff}(\nu) \lesssim 1$).

Motivated by the discussion in §2.6 about the radiative transfer in a slab including both scattering and absorption, we find the following approximation for the broad-band spec-

trum (slightly different from eq. 2.322 for consistency with asymptotic limits)

$$I_\nu \simeq 2bB_\nu \frac{\mathcal{E}(1+b) + \mathcal{E}^{-1}(1-b) - 2}{\mathcal{E}(1+b)^2 - \mathcal{E}^{-1}(1-b)^2} \rightarrow \begin{cases} \tau_{\text{ff}} B_\nu = I_\nu^{\text{thin}}, & \text{for } \tau_{\text{eff}} \ll 1, \\ 2bB_\nu/(1+b), & \text{for } \tau_{\text{eff}} \gg 1, \end{cases} \quad (6.112)$$

where $\mathcal{E} \equiv \exp\left[\sqrt{\tau_{\text{ff}}(\tau_{\text{ff}} + \tau_s)}\right] \approx \exp(\tau_{\text{eff}})$ and $b \equiv \sqrt{\tau_{\text{ff}}/(\tau_{\text{ff}} + \tau_s)} \simeq \tau_{\text{ff}}/\tau_{\text{eff}}$. The above result is shown in the left panel of Fig. 6.10. The Thomson thick cases may be relevant for dense gas in a very intense ionizing radiation field (e.g., dense circumstellar medium of a supernova explosion).

Note that the above discussion ignores modification of photon energies by repetitive Compton scatterings (i.e., Comptonization, see §7.4.1). For this reason, the above results apply for systems with Compton- y parameter $y = \tau_s^2 4k_B T_e / (m_e c^2) \ll 1$, where τ_s^2 is the typical number of scatterings a photon undergoes before escaping and $4k_B T_e / (m_e c^2)$ is the fractional energy gain in each Compton scattering.

Even when $y \gg 1$, eq. (6.112) still applies at sufficiently low frequencies where the free-free absorption mean-free path $\ell_{\text{ff}}(\nu)$ is less than $\ell_s m_e c^2 / (4k_B T_e)$. This is because, over a timescale $\ell_{\text{ff}}(\nu)/c$, a photon undergoes $\ell_{\text{ff}}(\nu)/\ell_s$ scatterings and its energy does not get strongly modified by Compton scatterings as long as $[\ell_{\text{ff}}(\nu)/\ell_s] 4k_B T_e / (m_e c^2) < 1$. This motivates us to define another critical frequency ν_{coh} below which Compton scattering can be considered as *coherent* (i.e., the photon frequency is modified by no more than a factor of order unity), and we obtain

$$\nu_{\text{coh}} = \nu_b \left(\frac{m_e c^2}{4k_B T_e} \right)^{1/2} = 3.0 \times 10^{12} \text{ Hz} \left(\frac{g_{\text{ff}}(\nu_b)}{5} \right)^{1/2} \rho_{-20}^{1/2} \left(\frac{T_e}{10^4 \text{ K}} \right)^{-5/4}, \quad (6.113)$$

where we have made use of eq. (6.106) and $\ell_{\text{ff}}(\nu)/\ell_s = \kappa_s/\kappa_{\text{ff},\nu} = (\nu/\nu_b)^2$. Thus, even when $y \gg 1$, our ignorance of Comptonization is justified at $\nu < \nu_{\text{coh}}$. Note that, since $\nu_t \simeq \tau_s \nu_b$, the ratio between the critical frequencies for translucence ν_t and coherence ν_{coh} is given by

$$\frac{\nu_t}{\nu_{\text{coh}}} \simeq \tau_s \left(\frac{4k_B T_e}{m_e c^2} \right)^{1/2} = y^{1/2}. \quad (6.114)$$

In the following, we consider the limit of $y \gg 1$ and discuss how the emerging spectrum at $\nu > \nu_{\text{coh}}$ is modified by Comptonization. The entire slab can be divided into two layers: (1) the surface layer with Thomson thickness $\Delta\tau_s \sim \sqrt{m_e c^2 / (4k_B T_e)}$ such that the corresponding “Compton- y thickness” is $\Delta y = \Delta\tau_s^2 4k_B T_e / (m_e c^2) \sim 1$; and (2) a deeper layer with $y \gg 1$.

The surface layer has its own translucent frequency $\nu'_t \sim \nu_{\text{coh}} \simeq \nu_b \sqrt{m_e c^2 / (4k_B T_e)}$, and the contribution to the emerging spectrum from the surface layer is not affected by Comptonization and is given by eq. (6.112) but only for a Thomson thickness of $\Delta\tau_s \sim \sqrt{m_e c^2 / (4k_B T_e)}$. The fractional thickness of the surface layer is $\Delta\tau_s / \tau_s \sim y^{-1/2}$.

As for the emission from the deeper layer, it is possible to show (see §7.4.5) that, if $y \gg 1$, Compton scatterings of all the emitted photons at $\nu > \nu_{\text{coh}}$ will reach saturation, and that the equilibrium spectrum is described by the Bose-Einstein distribution. In the limit of $3k_B T \gg h\nu_{\text{coh}}$ (meaning that seed photons get significantly up-scattered during Comptonization), the Bose-Einstein distribution reduces to the Wien spectrum (see Fig. 7.12 in §7.4.5)

$$I_\nu^{\text{Wien}} = C \nu^3 e^{-h\nu/k_B T_e}, \quad (6.115)$$

where $C = \text{const}$ is to be determined. The normalization constant C of the Wien spectrum is determined by the fact that Comptonization conserves photon number. We calculate the total *emitted* photon number at $\nu > \nu_{\text{coh}}$ from the intensity of the unmodified (optically thin) spectrum from free-free emission I_ν^{thin} (eq. 6.90),

$$\int_0^\infty \frac{I_\nu^{\text{Wien}}}{\nu} d\nu = 2C \frac{(k_B T_e)^3}{h^3} = \int_{\nu_{\text{coh}}}^\infty \frac{I_\nu^{\text{thin}}}{\nu} d\nu \approx \frac{1}{2} \ln \left(\frac{k_B T_e}{h\nu_{\text{coh}}} \right) I_{\nu=\nu_{\text{coh}}}^{\text{thin}}, \quad (6.116)$$

where we have taken the Gaunt factor as $g_{\text{ff}}(\nu, T_e) \approx (\sqrt{3}/\pi) \ln(k_B T_e/h\nu)$ for $T_e > 1.6 \times 10^5 \text{ K}$ and considered the low-frequency limit $h\nu_{\text{coh}} \ll k_B T_e$ (these two conditions are usually satisfied when Comptonization of free-free emission is important). In the above equation, we have used $\int x^{-1} \ln(x) dx = (1/2) \ln^2(x) + \text{const}$ and $\int_0^\infty x^2 e^{-x} dx = \Gamma(3) = 2$ where Γ stands for the Gamma-function. Thus, we find the normalization constant C and hence the Wien spectrum for the photons in Compton equilibrium is given by

$$I_\nu^{\text{Wien}} \approx \frac{1}{4} \ln \left(\frac{k_B T_e}{h\nu_{\text{coh}}} \right) I_{\nu=\nu_{\text{coh}}}^{\text{thin}} \left(\frac{h\nu}{k_B T_e} \right)^3 e^{-h\nu/(k_B T_e)}, \quad \text{for } y \gg 1, \quad (6.117)$$

which describes the emerging spectrum contributed by the deep layer.

The above result (eq. 6.117) is only for applicable for extremely large $y \gg 1$ (or $y \gg y_{\text{coh}}$, see below), whereas for smaller Compton- y parameters such that $e^y < k_B T_e/(h\nu_{\text{coh}})$, the photons near energy $h\nu_{\text{coh}}$ cannot be upscattered to energies $\sim k_B T_e$. For this reason, we define a critical Compton- y parameter

$$y_{\text{coh}} \equiv \ln \left(\frac{k_B T_e}{h\nu_{\text{coh}}} \right). \quad (6.118)$$

In the intermediate regime of $1 \ll y \ll y_{\text{coh}}$, we can simply replace ν_{coh} in eq. (6.117) with the minimum frequency ν_{min} for Compton upscattering into the Wien component

$$h\nu_{\text{min}} \approx \max [k_B T_e e^{-y}, h\nu_{\text{coh}}]. \quad (6.119)$$

The peak frequency of the Wien spectrum is located at $x_{\text{pk}} = h\nu_{\text{pk}}/k_B T_e = 3$ where $x_{\text{pk}}^3 e^{-x_{\text{pk}}} \approx 4/3$. The peak frequency of the emission from the surface layer of thickness

$\Delta y \sim 1$ or $\Delta\tau_s/\tau_s \sim y^{-1/2}$ is located at ν_{coh} , above which the surface layer is translucent. Thus, the ratio between peak intensity of the Wien spectrum and peak intensity of the surface layer emission is

$$\begin{aligned} \frac{I_{\nu=\nu_{pk}}^{\text{Wien}}}{y^{-1/2} I_{\nu=\nu_{coh}}^{\text{thin}}} &\approx \frac{4}{3} \frac{y^{1/2}}{4} \frac{I_{\nu=\nu_{min}}^{\text{thin}}}{I_{\nu=\nu_{coh}}^{\text{thin}}} \ln \left(\frac{k_B T_e}{h\nu_{min}} \right) = \frac{y^{1/2}}{3} \frac{\ln^2 [k_B T_e / (h\nu_{min})]}{\ln [k_B T_e / (h\nu_{coh})]} \\ &= \frac{y^{1/2} y_{coh}}{3} \min [(y/y_{coh})^2, 1]. \end{aligned} \quad (6.120)$$

Since above ratio is usually much greater than unity for $y \gg 1$, the majority of the photons emitted above frequency ν_{coh} are upscattered into the Wien component and only a small fraction of the photons remain at low frequencies. Thus, Comptonization raises the average photon energy to $\langle \epsilon \rangle_W = 3k_B T_e$, which is much greater than the average photon energy for optically thin free-free emission⁴ $\langle \epsilon \rangle_{ff} \ll k_B T_e$.

The sum of the contributions from the surface layer with $\Delta y \sim 1$ and the deep layer at $y \gg 1$ are shown in the right panel of Fig. 6.10. Readers who are interested in more details are referred to [Felten & Rees \(1972\)](#), but one should be aware that their Fig. 3 did not include the Gaunt factor (which should be included for more accurate calculations as most photons are emitted at low frequencies).

6.3.5 *Absorption of strong EM waves

When the EM wave amplitude is very large such that the typical oscillation speed of electrons $v_0 = E_0 q / (m_e \omega)$ (assuming non-relativistic) exceeds that of the thermal motion $\sigma_v = \sqrt{2k_B T_e / m_e}$, the plasma is effectively “hotter”. The free-free absorption coefficient, which scales as the electron temperature as $\alpha_\nu \propto T_e^{-3/2}$, is suppressed by a factor of the order $(v_0/\sigma_v)^{-3}$. The absorption coefficient for strong EM waves should be modified into

$$\alpha_\nu \simeq \alpha_\nu(E_0 \approx 0) \left(1 + \frac{(E_0 q / \omega)^2}{2m_e k_B T_e} \right)^{-3/2}, \quad (6.121)$$

where $\alpha_\nu(E_0 \approx 0)$ is the absorption coefficient for low-amplitude EM waves.

6.4 Radio/Infrared continuum spectrum of an ionized wind

The free-free opacity increases dramatically towards low frequencies $\kappa_{ff,\nu} \propto \nu^{-2}$, and this means that for a hot star launching an ionized wind, the photospheric radii in the radio/infrared bands may be much farther out than that in the optical/UV bands. This leads to “excessive” radio/infrared emission with a shallower spectral slope $F_\nu \propto \nu^{\approx 0.6}$

⁴If we only include photons down to ν_{coh} , then $\langle \epsilon \rangle_{ff} / k_B T_e \approx 2 / \ln^2(k_B T_e / h\nu_{coh}) = 2/y_{coh}^2 \ll 1$.

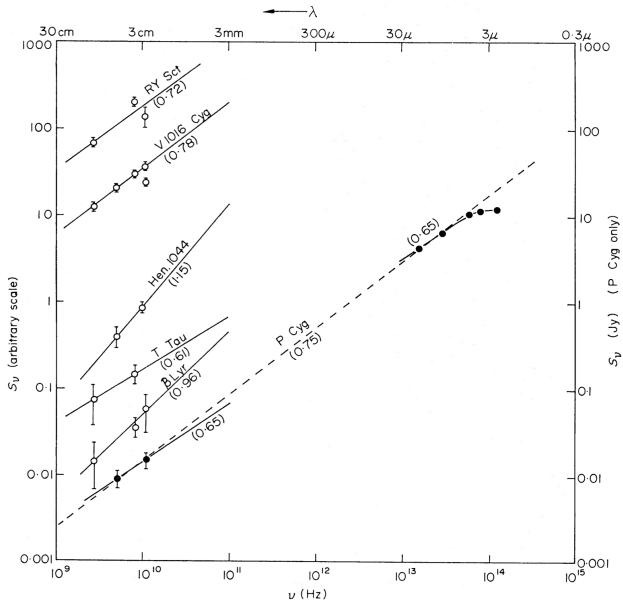


Figure 6.11: Radio and infrared spectra for a number of stars with ionized wind (Wright & Barlow 1975). The black filled circles are for P Cygni (distance $d \approx 1.6$ kpc), whose actual flux densities are shown on the right axis.

than the Rayleigh-Jeans law $F_\nu \propto \nu^2$ for a fixed photospheric radius. This phenomenon is shown in Fig. 6.11. The underlying model is first constructed by Wright & Barlow (1975). Here, we explain the key points.

Consider a spherical wind with mass loss rate of \dot{M} and constant velocity v_w , and the ratio between the two is

$$\frac{\dot{M}}{v_w} = 6.3 \times 10^{12} \text{ g cm}^{-1} \frac{\dot{M}}{10^{-5} M_\odot \text{ yr}^{-1}} \left(\frac{v_w}{10^3 \text{ km s}^{-1}} \right)^{-1}. \quad (6.122)$$

The density profile of the wind is given by

$$\rho(r) = \frac{\dot{M}}{4\pi r^2 v_w} = 8.0 \times 10^{-13} \text{ g cm}^{-3} \frac{\dot{M}/v_w}{10^{13} \text{ g cm}^{-1}} \left(\frac{r}{10^{12} \text{ cm}} \right)^{-2}, \quad (6.123)$$

which applies to radii $r \gg R_*$ (R_* being the stellar radius) because in the wind acceleration region near the stellar surface the wind velocity may be radius dependent. The gas has hydrogen and helium mass fractions X and $1 - X$, and it is assumed to be fully ionized so we obtain the electron number density

$$n_e = \frac{(1+X)\rho}{2m_p} \simeq 4.0 \times 10^{11} \text{ cm}^{-3} \frac{1+X}{1.7} \frac{\dot{M}/v_w}{10^{13} \text{ g cm}^{-1}} \left(\frac{r}{10^{12} \text{ cm}} \right)^{-2}, \quad (6.124)$$

the free-free absorption opacity at frequency $\nu = 10^{14}\nu_{14}$ Hz (cf. eq. 6.100)

$$\kappa_{\text{ff},\nu} \simeq 0.106 \text{ cm}^2 \text{ g}^{-1} \nu_{14}^{-2} g_{\text{ff}}(\nu, T_e) \frac{n_e}{10^{11} \text{ cm}^{-3}} \left(\frac{T_e}{10^4 \text{ K}} \right)^{-3/2}, \quad (6.125)$$

and the electron scattering opacity

$$\kappa_s = 0.2(1 + X) \text{ cm}^2 \text{ g}^{-1}. \quad (6.126)$$

The observed luminosity at a given frequency is set by the emission near the thermalization radius $r_{\text{th},\nu}$ where the effective optical depth equals to unity

$$\tau_{\text{eff}}(r_{\text{th},\nu}) = \sqrt{3\tau_{\text{ff},\nu}(\tau_s + \tau_{\text{ff},\nu})} \simeq 1, \quad (6.127)$$

where

$$\tau_{\text{ff},\nu}(r) = \rho \kappa_{\text{ff},\nu} r / 3, \quad \tau_s(r) = \rho \kappa_s r$$

are the absorption and scattering optical depths from radius r to infinity. At low frequencies such that $\kappa_{\text{ff},\nu} \gg \kappa_s$, the thermalization radius is given by $\tau_{\text{ff},\nu}(r_{\text{th},\nu}) \simeq 1/\sqrt{3}$, and we obtain

$$r_{\text{th,low } \nu} \simeq 5.8 \times 10^{11} \text{ cm} \nu_{14}^{-2/3} g_{\text{ff}}^{1/3} \left(\frac{T_e}{10^4 \text{ K}} \right)^{-1/2} \left(\frac{\dot{M}/v_w}{10^{13} \text{ g cm}^{-1}} \right)^{2/3} \left(\frac{1+X}{1.7} \right)^{1/3}. \quad (6.128)$$

On the other hand, at high frequencies where electron scattering is important such that $\kappa_{\text{ff},\nu} \ll \kappa_s$, the thermalization radius is given by $\tau_s \tau_{\text{ff},\nu} = 1/3$, and we obtain

$$r_{\text{th,high } \nu} \simeq 5.5 \times 10^{11} \text{ cm} \nu_{14}^{-1/2} g_{\text{ff}}^{1/4} \left(\frac{T_e}{10^4 \text{ K}} \right)^{-3/8} \left(\frac{\dot{M}/v_w}{10^{13} \text{ g cm}^{-1}} \right)^{3/4} \left(\frac{1+X}{1.7} \right)^{1/2}. \quad (6.129)$$

The critical frequency ν_c that separates these two regimes is

$$\nu_c \simeq 1.4 \times 10^{14} \text{ Hz} g_{\text{ff}}^{1/2} \left(\frac{\dot{M}/v_w}{10^{13} \text{ g cm}^{-1}} \right)^{-1/2} \left(\frac{T_e}{10^4 \text{ K}} \right)^{-3/4} \left(\frac{1+X}{1.7} \right)^{-1}. \quad (6.130)$$

Then, the thermalization radius can be written as

$$r_{\text{th},\nu} \simeq 4.6 \times 10^{11} \text{ cm} \frac{\dot{M}/v_w}{10^{13} \text{ g cm}^{-1}} \frac{1+X}{1.7} \max \left[(\nu/\nu_c)^{-2/3}, (\nu/\nu_c)^{-1/2} \right]. \quad (6.131)$$

At high frequencies, the thermalization radius from the above equation may be below the photospheric radius R_* of the star, and in that case free-free emission from the wind is unimportant. At sufficiently low frequencies, the thermalization radius can be very far

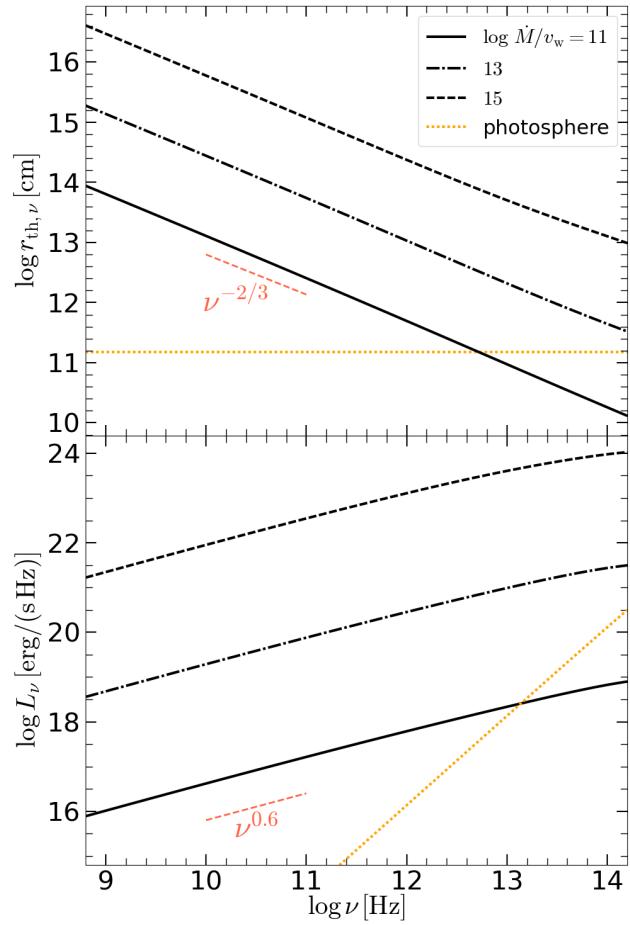


Figure 6.12: Free-free emission from ionized wind with different mass loss rates $\dot{M}/v_w [\text{g cm}^{-1}]$. The blackbody emission from a massive star with luminosity $10^{38} \text{ erg s}^{-1}$ and photospheric temperature $5 \times 10^4 \text{ K}$ is shown by an orange dotted line for comparison.

from the stellar photosphere and this leads to the deviation of the stellar spectrum from the Rayleigh-Jeans law (which only applies for the case that the thermalization radius is independent of frequency).

For stars with high mass loss rates we expect $r_{\text{th},\nu} \gg R_*$ at infrared and radio frequencies. Photons emitted above radius $r_{\text{th},\nu}$ will arrive at the observer without further absorption (but they may be scattered). The observed specific luminosity is roughly given by

$$L_\nu = \frac{dL}{d\nu} \simeq \int_{r_{\text{th},\nu}}^{\infty} dr 4\pi r^2 \times 4\pi j_\nu(r) \simeq 4\pi r_{\text{th},\nu}^3 \times 4\pi j_\nu(r_{\text{th},\nu}), \quad (6.132)$$

where the emissivity (in units of $\text{erg cm}^{-3} \text{s}^{-1} \text{Hz}^{-1} \text{sr}^{-1}$) is given by

$$j_\nu(r) = \rho \kappa_{\text{ff},\nu} B_\nu(T_e) \simeq 1.0 \times 10^{-17} g_{\text{ff}} \left(\frac{\dot{M}/v_w}{10^{13} \text{g cm}^{-1}} \right)^2 \left(\frac{r}{10^{12} \text{cm}} \right)^{-4} \left(\frac{T_e}{10^4 \text{K}} \right)^{-1/2}. \quad (6.133)$$

Using the thermalization radius in eq. (6.131), we obtain the specific luminosity

$$\begin{aligned} L_\nu &\simeq 2.7 \times 10^{21} \frac{\text{erg}}{\text{s Hz}} \left(\frac{\dot{M}/v_w}{10^{13} \text{g cm}^{-1}} \right)^{4/3} \left(\frac{1+X}{1.7} \right)^{-1/3} \nu_{14}^{2/3} g_{\text{ff}}^{2/3} \\ &\times \min \left[1, 1.4 \left(\frac{\dot{M}/v_w}{10^{13} \text{g cm}^{-1}} \right)^{-1/2} \left(\frac{T_e}{10^4 \text{K}} \right)^{-3/4} \left(\frac{1+X}{1.7} \right)^{-1} \nu_{14}^{-1} g_{\text{ff}}^{1/2} \right]^{1/6}. \end{aligned} \quad (6.134)$$

At low frequencies where electron scattering is unimportant, the spectrum is a power-law $L_\nu \propto (\nu g_{\text{ff}})^{2/3} \propto \nu^{\approx 0.6}$; whereas at high frequencies, when photon trapping by electron scattering is important, the spectrum becomes $L_\nu \propto \nu^{1/2} g_{\text{ff}}^{3/4} \propto \nu^{\approx 0.4}$.

The free-free spectrum for different wind mass loss rates are shown in Fig. 6.12, where we assume that the gas is fully ionized with hydrogen mass fraction $X = 0.7$ and electron temperature $T_e = 3 \times 10^4 \text{K}$ (the results depend weakly on these two parameters). We see that it is possible to infer the wind mass loss rate from a hot star by measuring the fluxes in the radio/infrared bands. However, this is complicated by the fact that the density distribution in the wind may be clumpy and since the emissivity scales as $j_\nu \propto \rho^2$, denser clumps produces more free-free emission — this means that the mass loss rate inferred based on a smooth wind density profile is an over-estimate. Another potential complication is that some circum-stellar dust (heated by the star light) will emit in the mid-infrared band, which confuses with the free-free emission.

6.5 *Relativistic bremsstrahlung and Coulomb losses

In this section, we consider the free-free emission from a relativistic electron undergoing Coulomb interactions with an ion as well as the energy loss due to Coulomb interactions

with other electrons. The energy loss rate of a cosmic ray ion moving through matter will also be discussed.

6.5.1 *Emitting power and spectrum

Again, instead of viewing it in the frame where the target ion is at rest, we will go to the comoving frame of the initial velocity of the projectile electron⁵. This is motivated by the Lorentz invariance of the total emission power, which is easier to obtain in the projectile's comoving frame. To be consistent with the notations in the Special Relativity Chapter 3, we would have to denote all quantities in the comoving frame with primes ('), but that would make the notations difficult to read. Therefore, we temporarily omit the primes for now, and we will add the primes in the very end when carrying out the Lorentz transformation back to the lab frame.

- **Projectile electron's comoving frame**

In the comoving frame, we place the electron at the coordinate origin, and the ion moves at a constant velocity $\mathbf{v} = v\hat{\mathbf{x}}$ with equation of motion described by $(x = vt, y = -b, z = 0)$ for $t \in (-\infty, \infty)$. The Coulomb field at the electron's position is given by (see Chapter 3)

$$\mathbf{E}(t) = \frac{\gamma Ze}{[(\gamma vt)^2 + b^2]^{3/2}} (b\hat{\mathbf{y}} - vt\hat{\mathbf{x}}), \quad (6.135)$$

where γ is the relative Lorentz factor between the electron and ion. We see that $|E_y|$ reaches the maximum value of $E_{y,\max} = \gamma Ze/b^2$ at $t = 0$, whereas the field strength $|E_x|$ in the $\hat{\mathbf{x}}$ direction only reaches the maximum value of $2Ze/(3\sqrt{3}b^2)$ at $t = \pm b/(\sqrt{2}\gamma v)$.

For the case of an ultra-relativistic electron with $\gamma \gg 1$, the acceleration by the Coulomb field is dominated by E_y . In fact, in the electron's comoving frame, the ion's Coulomb field can be considered as virtual photons which are Compton scattered by the electron. In Rybicki & Lightman's book, relativistic Bremsstrahlung is described as a Thomson scattering problem (in the low-frequency limit). The reader is referred to [Blumenthal & Gould \(1970\)](#) for more detailed discussion of relativistic bremsstrahlung and the relevant literature.

Here, to keep the discussion analogous to the non-relativistic version, we adopt the classical description. Appropriate quantum mechanical corrections will be added based on our physical intuition. The electron's acceleration is mainly in the $\hat{\mathbf{y}}$ direction

$$\mathbf{a}(t) = \dot{\mathbf{v}}_e \approx -\frac{eE_y}{m_e}\hat{\mathbf{y}} = -\frac{\gamma Ze^2b/m_e}{[(\gamma vt)^2 + b^2]^{3/2}}\hat{\mathbf{y}}. \quad (6.136)$$

⁵In the case of a relativistic cosmic ray particle interacting with a non-relativistic electron in e.g., the interstellar medium, the electron's comoving frame can be considered as the lab frame.

The total radiated energy is given by the time integral of the Larmor power

$$W(b) = \frac{2e^2}{3c^3} \int a^2(t) dt = \frac{\pi\gamma Z^2 e^6}{4m_e^2 c^3} \frac{1}{vb^3}, \quad (6.137)$$

where we have used $\int_0^\infty dx/(x^2 + 1)^3 = 3\pi/16$ by change of variable $x = \tan\phi$.

The radiated energy W is a factor of $3\gamma/4 \gg 1$ greater than that for the non-relativistic case at the same impact parameter (cf. eq. 6.47), for the following reason. The maximum acceleration $|a_{\max}| \sim \gamma Ze^2/(b^2 m_e)$ only lasts for an interaction timescale

$$\tau = b/(\gamma v), \quad (6.138)$$

so we know that most of the radiation energy from free-free emission will be near frequency $\omega_{\max} \sim \tau^{-1} \sim \gamma v/b$ in the comoving frame of the electron.

The Fourier transformation of the time-dependent acceleration is given by

$$\tilde{\mathbf{a}}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{a}(t) e^{i\omega t} dt = -\frac{Ze^2}{\pi m_e v b} \omega \tau K_1(\omega\tau) \hat{\mathbf{y}}, \quad (6.139)$$

where we have made use of the modified Bessel function of order 1 (see eq. 6.56) and it has asymptotic behaviors of $zK_1(z) \approx 1$ for $z \ll 1$ and $zK_1(z) \approx \sqrt{\pi z/2}(1 + 3/(8z))e^{-z}$ for $z \gg 1$. The total radiative spectrum is

$$\frac{dW}{d\omega} = \frac{8\pi e^2}{3c^3} |\tilde{\mathbf{a}}(\omega)|^2 = \frac{8Z^2 e^6}{3\pi m_e^2 c^3} \frac{1}{v^2 b^2} [\omega \tau K_1(\omega\tau)]^2, \quad (6.140)$$

which is identical to the non-relativistic case at low frequencies $\omega \ll \tau^{-1} = \gamma v/b$.

Next, we consider that the electron undergoes Coulomb scatterings in a sea of target ions at a broad range of impact parameters b . In the rest frame of the projectile electron, all ions approach the electron at velocity $v \approx c$ and their number density is γn_i (where n_i is the lab-frame number density of ions). Thus, the specific free-free power at frequency ω is given by (taking $v \approx c$)

$$P_\omega = \int db 2\pi b \gamma n_i v \frac{dW}{d\omega} = \frac{16Z^2 e^6}{3m_e^2 c^3} \frac{\gamma n_i}{c} \ln \Lambda, \quad (\text{for } \hbar\omega < m_e c^2), \quad (6.141)$$

where the Coulomb logarithm is given by

$$\ln \Lambda(\omega) = \int_{b_{\min}}^{\infty} \frac{db}{b} [\omega \tau K_1(\omega\tau)]^2 \simeq \int_{b_{\min}}^{b_{\max}} \frac{db}{b} \simeq \ln \left[\gamma \left(\frac{m_e c^2}{\hbar\omega} - 1 \right) \right], \quad (6.142)$$

where we have enforced Λ to go to zero when $\hbar\omega \geq m_e c^2$ (such photons would have an average energy exceeding $\gamma m_e c^2$ in the lab frame), and the minimum and maximum impact parameters are taken to be

$$b_{\min} \sim \hbar/(m_e c), \quad b_{\max} \sim \gamma c/\omega. \quad (6.143)$$

The above choice of b_{\min} might be very confusing to the reader.

At first thought, one might argue that the minimum impact parameter for free-free emission should be $\tilde{b}_{\min} \sim \gamma\hbar/(m_e c)$ (note that this is *not* \hbar/p). The argument is as follows. For $b < \tilde{b} \sim \gamma\hbar/(m_e c)$, the Coulomb interaction timescale would be $\tau \approx b/(\gamma c) < \hbar/(m_e c^2)$, which means that the electron would oscillate at angular frequency $\tau^{-1} \gtrsim m_e c^2/\hbar$ and hence the average energy of the emitted photons would be $\gtrsim m_e c^2$ in the projectile's comoving frame or $\gtrsim \gamma m_e c^2$ in the lab frame. The emission of such photons must be suppressed by the Klein-Nishina effects, because it is as if a virtual photon of energy greater than $m_e c^2$ trying to scatter off the projectile electron in its rest frame. This argument is correct and indeed the contribution from $b \ll \tilde{b}_{\min}$ to the *total* free-free emission power is negligible. However, only the emission of the *most energetic photons* near frequency of τ^{-1} (which dominate the emission power in the classical picture) is quantum-mechanically suppressed due to Klein-Nishina effects (see Chapter 7), whereas the emission at much lower frequencies $\omega \ll \tau^{-1}$ is still allowed.

On the other hand, for an impact parameter $b < b_{\min} \sim \hbar/(m_e c)$ (our choice above), the momentum uncertainty of the projectile electron in its comoving frame would be $\Delta p \gtrsim m_e c$. If the electron were to radiate a photon, the photon momentum would be of the order Δp (as allowed by momentum conservation), so the photon's energy would be $\Delta pc \gtrsim m_e c^2$ in the comoving frame and $\gtrsim \gamma m_e c^2$ in the lab frame. This would violate energy conservation — free-free emission must be suppressed due to the discrete nature of the emitted photons. Thus, we conclude that, for $b < b_{\min} \sim \hbar/(m_e c)$, our classical picture for free-free emission would break down at *all frequencies*!

It should be noted that it is still possible for the projectile electron to get much closer to the target than $\hbar/(m_e c)$. If the target particle is infinitely heavy, diffraction of the projectile electron's wave packet only becomes important for impact parameters $b \lesssim \hbar/p \approx \hbar/(\gamma m_e c)$. Since the impact parameter is defined in the direction *perpendicular* to the projectile's initial momentum, it is the same between the lab frame and the projectile's comoving frame⁶. In those encounters with impact parameters in between $\hbar/(\gamma m_e c)$ and $\hbar/(m_e c)$, energy/momentum transfer between the projectile and the target is still mediated by Coulomb scattering in the classical picture (see §6.5.3), but the free-free emission of photons must be quantum mechanically suppressed.

Let us then turn to our choice of b_{\max} . For even larger impact parameters $b \gg b_{\max} \sim \gamma c/\omega$, the Coulomb interactions would occur on a timescale $b/(\gamma c) \gg \omega^{-1}$ and hence the emission at the desired frequency ω is suppressed. Our choice of $b_{\max} \sim \gamma c/\omega$ is appropriate for the case of a fully ionized gas where the ion's charge is not screened by bound electrons (and Debye screening can be ignored at sufficiently high emission frequencies).

⁶My apology to the reader for going back and forth between frames, but our goal here is to provide a simple picture that illustrates the confusing physics of relativistic bremsstrahlung.

However, if the medium consists of atomic material, then the projectile only “sees” a large fraction of the nuclear charge Ze when the impact parameter is smaller than the typical orbital radius of bound electrons. A crude representation for the screening of nuclear charge is the *Thomas-Fermi model*, where a nucleus is surrounded by a cloud of Fermi electron gas in the ground state. Inside an atom with characteristic size a , the electron number density is of the order $n_e \sim Z/a^3$. The electrons are confined inside a volume of a^3 , so the Uncertainty Principle gives a typical electron momentum $p \sim \hbar n_e^{1/3} \sim \hbar Z^{1/3}/a$. On the other hand, the balance between Coulomb and centrifugal forces gives $Ze^2/a \sim p^2/m_e$ (for non-relativistic electron orbits), or $p \sim \sqrt{Ze^2 m_e / a}$. Combining these two expressions for the electron momentum, we solve for the characteristic size of an atom

$$a \simeq Z^{-1/3} \frac{\hbar^2}{m_e e^2} = Z^{-1/3} a_0, \quad (6.144)$$

where $a_0 \equiv \hbar^2/(e^2 m_e) = 0.529 \text{ \AA}$ is the Bohr radius. We see that atoms with larger nuclear charge numbers are smaller. The above discussion motivates us to take the following maximum impact parameter

$$b_{\max} \sim Z^{-1/3} a_0 \quad (\text{for atomic targets}). \quad (6.145)$$

For atomic targets, the Coulomb logarithm for free-free emission is roughly given by

$$\ln \Lambda \simeq \ln \left[\frac{\min(\gamma c/\omega, Z^{-1/3} a_0)}{\hbar/(m_e c)} \right] \simeq \min \left(\frac{\gamma m_e c^2}{\hbar \omega}, 137 Z^{-1/3} \right), \quad \text{for atomic targets,} \quad (6.146)$$

where ω is the emission frequency in the projectile electron’s comoving frame and we have used the fine-structure constant $\alpha = e^2/(\hbar c) \approx 1/137$.

Finally, we estimate the (Lorentz-invariant) total emitting power in the electron’s comoving frame with two methods: (i) multiplying the total emitted energy $W(b)$ by the rate of Coulomb encounters at a wide range of impact parameters (only those with $b \gtrsim b_{\min} = \gamma \hbar/(m_e c)$ contribute significantly to the total power), or (ii) integrate the specific power P_ω over frequency (up to $\omega_{\max} = m_e c^2/\hbar$ in the electron’s comoving frame). The results are comparable, so we only provide the result from method (ii) here

$$P_{\text{tot}}(\gamma) = \int_0^{\omega_{\max}} d\omega P_\omega \simeq v \sigma_T \frac{3\alpha Z^2 (\gamma + 1) n_i m_e c^2}{2\pi} \ln(\gamma + 1), \quad (\text{for fully ionized ion targets}), \quad (6.147)$$

where we have made slight adjustments of $c \rightarrow v$ and $\gamma \rightarrow \gamma + 1$ so that the above result is compatible with the non-relativistic case.

For atomic targets, [Bethe and Heitler \(1934\)](#) provide the following result based on relativistic quantum mechanical calculations

$$P_{\text{tot}}(\gamma) = c \sigma_T \frac{3\alpha Z^2 \gamma n_i m_e c^2}{2\pi} \left(\frac{1}{8} + \ln \frac{183}{Z^{1/3}} \right), \quad (6.148)$$

which only applies to ultra-relativistic projectile electrons ($\gamma \gg 1$) interacting with heavy atoms ($Z \gg 1$), as the Thomas-Fermi model breaks down for light elements.

Practically, one can combine the expressions in eqs. (6.147) and (6.148) to obtain the total emitting power that applies to an arbitrarily relativistic (including non-relativistic) projectile electron interacting with atomic materials

$$P_{\text{tot}}(\gamma) = v\sigma_T \frac{3\alpha Z^2(\gamma+1)n_i m_e c^2}{2\pi} \min \left[\ln(\gamma+1), \frac{1}{8} + \ln \frac{183}{Z^{1/3}} \right], \quad (\text{for atomic targets}). \quad (6.149)$$

All the above discussions are in the comoving frame of the projectile electron — remember that the physical situation we have considering is a relativistic electron interacting with a sea of ions/atoms that are nearly static in the lab frame. Luckily, since the total emitting power is Lorentz invariant (see Ch 3), we know that eq. (6.147) also gives the emitting power in the lab frame. Radiative losses causes the Lorentz factor to change at a rate

$$\dot{\gamma}_{\text{ff}} \simeq -\frac{3\alpha Z^2 \ln \gamma}{2\pi} \gamma n_i \sigma_T c, \quad \text{for } \gamma \gg 1. \quad (6.150)$$

Thus, the free-free cooling timescale of the electron in the lab frame is

$$t_{\text{ff}} = \frac{\gamma}{|\dot{\gamma}_{\text{ff}}|} \simeq 4.5 \times 10^7 \text{ yr} \frac{\text{cm}^{-3}}{Z^2 n_i} \frac{10}{\ln \gamma}, \quad \text{for } \gamma \gg 1. \quad (6.151)$$

For an ultra-relativistic Maxwell-Jüttner distribution $f_e(\gamma)d\gamma \propto \gamma^2 e^{-\gamma m_e c^2/(k_B T_e)} d\gamma$ at any temperature $T_e \gg m_e c^2/k_B$, the average free-free power per electron depends on the mean electron Lorentz factor $\langle \gamma \rangle = 3k_B T / (m_e c^2)$. We replace the $v(\gamma+1)$ factor in eq. (6.147) by $n_e c \langle \gamma \rangle$ and obtain the total free-free energy loss rate per unit volume. Then, we combine the result with that in the non-relativistic limit (eq. 6.82) to obtain the following volumetric cooling rate due to electron-ion bremsstrahlung (for arbitrary temperatures)

$$4\pi j_{\text{ff}} \simeq 1.4 \times 10^{-27} \frac{\text{erg}}{\text{cm}^3 \text{s}} n_e T_e^{1/2} \sum_i (\langle g_{\text{ff},i} \rangle_T Z_i^2 n_i) \left[1 + \frac{1.3(T_e/10^{10} \text{ K})}{1 + 0.6(T_e/10^{10} \text{ K})^{1/2}} \right], \quad (6.152)$$

where cgs units are adopted ($n_e \rightarrow n_e/\text{cm}^{-3}$, $T_e \rightarrow T_e/\text{K}$, and the “relativistic correction” factor comes from the following argument. For mildly relativistic temperatures ($T_e \lesssim \text{few} \times 10^9 \text{ K}$), the relativistic correction factor is given by $[1 + 19k_B T_e / (24m_e c^2)]$ (see [Gould 1980](#), and a similar expression is also quoted in Rybicki & Lightman’s book), whereas in the ultra-relativistic limit ($T_e \gg 10^{10} \text{ K}$), the temperature scaling is given by $j_{\text{ff}} \propto T_e$.

We also note that, at relativistic temperatures, electron-electron bremsstrahlung from electric quadrupolar emission (due to vanishing electric dipole moment) also contributes significantly to the total free-free cooling rate of the plasma. [Gould \(1980\)](#) provides the following result that is applicable at mildly relativistic temperatures

$$4\pi j_{\text{ff,ee}}(T_e \lesssim \text{few} \times 10^9 \text{ K}) = 5.6 \times 10^{-37} \frac{\text{erg}}{\text{cm}^3 \text{s}} (n_e/\text{cm}^{-3})^2 (T_e/\text{K})^{3/2}. \quad (6.153)$$

- **Back to the lab frame**

In the following, we denote all quantities in the projectile electron's comoving frame with primes ('), and the unprimed quantities are in the lab frame.

We would like to Lorentz transform the specific power $P'_{\omega'}$ in eq. (6.141) (note the primed notation) from the electron comoving frame to the lab frame. Since $\omega = \gamma\omega'$ (averaged over the photons emitted in all directions) and the total emitting power is Lorentz invariant $P = P'$, the lab-frame specific power is simply given by

$$P_{\omega}(\text{lab frame}) = \frac{dP}{d\omega} = \frac{P'_{\omega'}}{\gamma} = \frac{16Z^2e^6}{3m_e^2c^4} n_i \ln \Lambda. \quad (6.154)$$

and using the lab-frame frequency ω , the Coulomb logarithm should be written as

$$\ln \Lambda(\omega, \text{lab frame}) \simeq \ln \left[\gamma \left(\frac{\gamma m_e c^2}{\hbar \omega} - 1 \right) \right]. \quad (6.155)$$

6.5.2 *Coulomb energy-loss rate

The rate of radiative losses should be compared with that of Coulomb losses to the target electrons. In the following, we stay agnostic to the nature of the projectile particle (it could be an electron, ion, or other charged particle), but the target particle is restricted to be an electron, because most Coulomb energy loss goes to target electrons due to their much lower mass than ions. Below, we ignore the thermal motion of the target electrons if they are unbound and their orbital motion if they are bound to atoms. For free electrons, the typical thermal speed is of the order $v_{\text{th}}/c \sim \sqrt{k_B T_e/m_e} \sim 10^{-3}(T_e/10^4 \text{ K})^{1/2}$, and for bound electrons, the typical orbital speed is of the order $v_{\text{orb}}/c \sim \alpha Z^{2/3} \sim 10^{-2} Z^{2/3}$ (α being the fine-structure constant) in the crude Thomas-Fermi model. Our discussion is only valid for projectile velocity

$$\beta \gg \max \left[10^{-3}(T_e/10^4 \text{ K})^{1/2}, 10^{-2} Z^{2/3} \right]. \quad (6.156)$$

In the lab frame where the targets are initially at rest, each Coulomb scattering gives a perpendicular momentum kick of $\Delta p_{\perp} = 2e^2/(bv)$ to the target electron, which leads to an energy loss of $\Delta p_{\perp}^2/2m_e$ (that is taken away by the target electron). Based on the scattering rate $dR = v n_e 2\pi b db$, we obtain $d(\Delta p_{\perp}^2)/dt = 8\pi n_e e^4 \ln \Lambda/v$ and hence the energy-loss rate for the projectile is given by

$$-\frac{dE}{dt} = \frac{1}{2m_e} \frac{d(\Delta p_{\perp}^2)}{dt} = \frac{4\pi n_e e^4 \ln \Lambda}{m_e \beta c}, \quad (6.157)$$

which has an identical form as the non-relativistic version (eq. 6.13), except that the Coulomb logarithm $\ln \Lambda$ depends on the Lorentz factor (see below). The above equation

applies to both relativistic and non-relativistic projectiles of any type, as long as eq. (6.156) is satisfied. When the projectile is a fast electron, its Lorentz factor evolves as

$$\dot{\gamma}_{\text{Coul}} = \frac{dE/dt}{m_e c^2} = -\frac{3 \ln \Lambda}{2\beta} n_e \sigma_T c. \quad (6.158)$$

Thus, the energy-loss timescale for a projectile electron due to Coulomb scatterings off target electrons is given by

$$t_{\text{Coul}} = \frac{\gamma}{|\dot{\gamma}_{\text{Coul}}|} = 5.3 \times 10^4 \text{ yr} \gamma \beta \frac{\text{cm}^{-3}}{n_e} \frac{20}{\ln \Lambda}, \quad (6.159)$$

Based on $n_e \sim Zn_i$ and the scaling of $t_{\text{Coul}} \propto \gamma$, we see that for sufficiently high electron Lorentz factors ($\gamma \gg 10^3/Z$), free-free cooling due to interactions with ions becomes more important than the energy loss due to Coulomb scattering off other electrons. On the other hand, cooling due to synchrotron emission in a magnetized plasma or inverse-Compton emission in an intense radiation field may dominate over free-free emission.

6.5.3 *Bethe formula and mass stopping power

In this subsection, we discuss the Coulomb logarithm $\ln \Lambda$ in energy loss rate eq. (6.157) based on our physical intuition, and our result will be compared with the Bethe formula. Our discussion applies to both non-relativistic and relativistic projectiles of any type, and the target particle is taken to be an electron. We will focus on atomic medium where the target electrons are initially bound. For a fully ionized plasma medium where the target electrons are unbound, the only difference is that the maximum impact parameter b_{\max} should be taken to be the Debye length r_D (eq. 5.12).

For atomic targets, the maximum impact parameter should be taken as

$$b_{\max} \sim \gamma v / \Omega_{\text{orb}}, \quad (6.160)$$

where Ω_{orb} is the typical orbital angular frequency of electrons. For $b > b_{\max}$, the Coulomb interaction timescale $\tau \sim b/\gamma v$ would be longer than Ω_{orb}^{-1} , and hence the scattering is suppressed. Since $\hbar \Omega_{\text{orb}}$ is of the order mean ionization energy \bar{I} , we write

$$b_{\max} \sim \gamma \beta \hbar c / \bar{I}. \quad (6.161)$$

On the other hand, the minimum impact parameter should be taken as

$$b_{\min} \sim \frac{\hbar}{\gamma_{\text{cm}} \beta_{\text{cm}} m_e c}, \quad (6.162)$$

where γ_{cm} and β_{cm} are the Lorentz factor and velocity of the *target electron* in the center-of-momentum (cm) frame⁷ of the Coulomb interaction, and b_{\min} is close to the corresponding

⁷For a 2-body collision, the “proper” frame to analyze the problem should be the center-of-momentum frame where the two particles collide head-on and the net momentum vanishes.

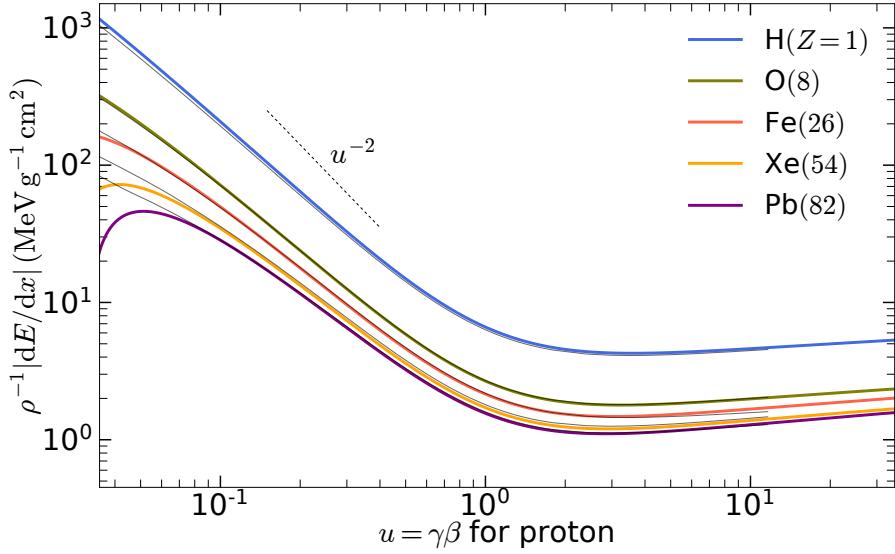


Figure 6.13: Mass stopping power for a fast proton interacting with atomic material of different compositions. The thin black lines shows the results from the [STAR program](#). The discrepancy at the low-velocity end is due to the inaccuracy of Bethe formula when the projectile's velocity is comparable to the atomic orbital speed of target electrons.

de Broglie wavelength. At $b < b_{\min}$, the scattering is described by the diffraction of the target electron's wave packet (see Fig. 6.3).

Let us denote the rest mass of the projectile particle as M , and the velocity of the center-of-mass frame wrt. the lab frame is given by (see eq. 3.86)

$$\beta_{\text{cm}} = \frac{\gamma M}{\gamma M + m_e} \beta, \quad (6.163)$$

The maximum energy gain by the target electron in units of $m_e c^2$ is given by (see eq. 3.89)

$$w_{\max} = 2\gamma_{\text{cm}}^2 \beta_{\text{cm}}^2 = \frac{2\gamma^2 \beta^2}{1 + 2\gamma m_e/M + (m_e/M)^2}. \quad (6.164)$$

After some algebra, one can show

$$\gamma_{\text{cm}} \beta_{\text{cm}} = \sqrt{w_{\max}/2}. \quad (6.165)$$

Thus, the minimum impact parameter is of the order (discarding the factor of $\sqrt{2}$)

$$b_{\min} \sim \frac{\hbar/m_e c}{\sqrt{w_{\max}}}. \quad (6.166)$$

Our intuition has led us to the following approximate result for the argument of the Coulomb logarithm

$$\Lambda = \frac{b_{\max}}{b_{\min}} \propto \frac{\gamma \beta w_{\max}^{1/2}}{\bar{I}}. \quad (6.167)$$

For a very heavy projectile with $M \gg 2\gamma m_e$, we obtain $w_{\max} \approx 2\gamma^2 \beta^2$, so the argument of the Coulomb logarithm scales as $\Lambda \propto \gamma^2$ in the ultra-relativistic limit. On the other hand, if the projectile is not so massive $M \ll 2\gamma m_e$, then we obtain $w_{\max} \simeq \gamma \beta^2 M/m_e$ and hence $\Lambda \propto \gamma^{3/2}$ in the ultra-relativistic limit.

From relativistic quantum mechanical calculations, the accurate result for the Coulomb energy loss rate in atomic medium is given by *Bethe formula* (see [Livingston & Bethe 1937](#))

$$-\frac{dE}{dt} = \frac{4\pi n_e z^2 e^4}{m_e \beta c} \left[\ln \frac{\gamma \beta (2w_{\max})^{1/2}}{\bar{I}/(m_e c^2)} - \beta^2 \right], \quad (6.168)$$

where z is the charge number of the *projectile particle*, n_e is the *total* number density of electrons in the medium, \bar{I} is a constant that depends on the composition of the medium. The mean ionization energy is roughly given by

$$\bar{I} \simeq 12Z \text{ eV}, \quad (6.169)$$

where Z is the atomic number of the target medium. In Bethe formula above, we have ignored the effects of polarization of the medium by the projectile's electric fields and our approximation is appropriate for a dilute gas medium. Our semi-classical arguments are able to reproduce the logarithmic term⁸, and the β^2 term is always subdominant as compared to the logarithmic term. The reader is referred to the *Passage of Particles Through Matter* Chapter in [this review](#) for a more detailed discussion.

It is convenient to define a mass stopping power (in units of $\text{MeV g}^{-1} \text{ cm}^2$)

$$\rho^{-1} \frac{dE}{dx} = \frac{Z}{n_e A m_p} \frac{1}{\beta} \frac{dE}{dt}, \quad (6.170)$$

where Z and A are the atomic number and mass number of the atomic medium, respectively. Fig. 6.13 shows the mass stopping power for a fast proton projectile.

Finally, if the projectile is a fast electron, it is important to account for identical particle effects. For scattering in the center-of-momentum frame, the result for a deflection angle θ is indistinguishable from that for a deflection angle $\pi - \theta$. Thus, the maximum energy gain for the target particle occurs when both particles get deflected by 90° in the center-of-momentum frame (instead of 180° which would correspond to zero energy exchange for

⁸The discrepancy by a constant of proportionality is unimportant as it can always be absorbed into the definition of \bar{I} .

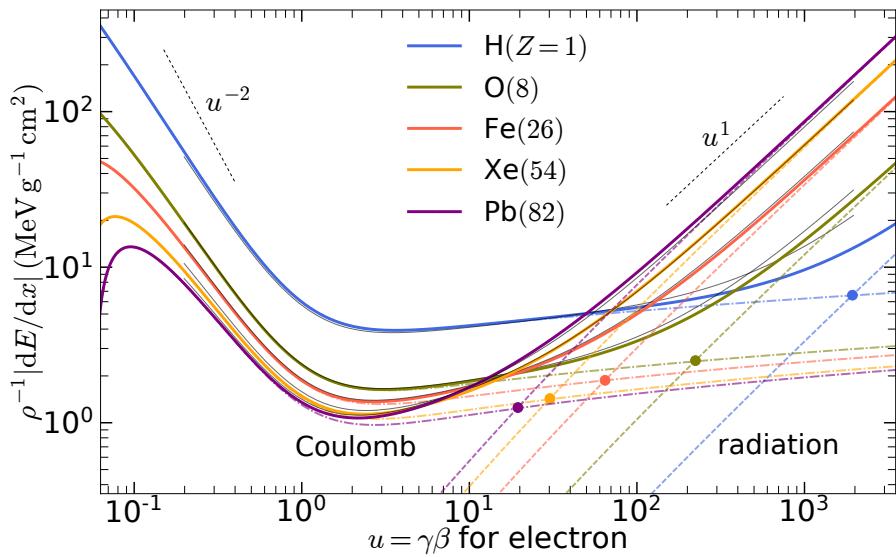


Figure 6.14: Mass stopping power for a fast electron in atomic medium of different compositions, with contributions from Coulomb interactions with bound electrons (dashed lines) and free-free emission (dotted lines). The dots mark where the two contributions are equal. The thin black lines shows the results from the [STAR program](#) on the NIST database (not including medium-polarization effect). The reason for the disagreement with the NIST results in the atomic hydrogen case on the high- γ end is not clear.

identical particles). It is straightforward to show from Lorentz transformation that the energy gain by the target electron for a 90° deflection angle is given by

$$\tilde{w}_{\max} = \gamma_{\text{cm}}^2 \beta_{\text{cm}}^2 = (\gamma - 1)/2, \quad (\text{for electron-electron scattering}) \quad (6.171)$$

where we have made use of eq. (6.163) for β_{cm} and then taken $M = m_e$. We see that the maximum energy gain by the target electron is half of the kinetic energy of the projectile electron. One can then plug the above w_{\max} into Bethe formula to obtain the Coulomb energy-loss rate and the result is not too inaccurate.

A relativistic quantum mechanical treatment of electron-electron scattering (including spin effects) gives the following Coulomb energy-loss rate (see [Uehling 1954](#))

$$-\frac{dE}{dt} = \frac{4\pi n_e e^4}{m_e \beta c} \left[\ln \frac{\gamma \beta \sqrt{(\gamma - 1)/2}}{\bar{I}/(m_e c^2)} + \frac{1 - (2\gamma - 1) \ln 2 + (\gamma - 1)^2/8}{2\gamma^2} \right]. \quad (6.172)$$

The second term in the square bracket is always subdominant as compared to the logarithmic term. In the ultra-relativistic limit, the Coulomb energy-loss rate increases with the projectile electron's Lorentz factor as $\ln \gamma^{3/2}$. Fig. 6.14 shows the mass stopping power for a fast electron projectile interacting with various atomic materials, with contributions from ionization losses and free-free emission.

6.6 Homework

Prob. 16. Fig. 6.15 shows the radio spectrum of the Orion nebula. Assume that the emitting gas is a proton-electron plasma.

- (i) Briefly explain the shape of the spectrum on the low-frequency and high-frequency ends.
- (ii) Estimate (by eye) the free-free self-absorption frequency ν_a where $\tau_{\nu_a} = 1$.
- (iii) Assuming a temperature of $T_e = 10^4$ K, estimate the emission measure of the nebula.

Prob. 17. Fig. 6.16 shows the X-ray spectrum of the Coma Cluster. Assume that the emitting gas is a proton-electron plasma and estimate the following:

- (i) The emission measure of the cluster.
- (ii) The free-free self-absorption frequency ν_a where $\tau_{\nu_a} = 1$.

Prob. 18. The solar wind has mass loss rate of $\dot{M}_\odot \simeq 2 \times 10^{-14} M_\odot \text{ yr}^{-1}$ and typical velocity of $v_w \simeq 500 \text{ km s}^{-1}$, which is comparable to the surface escape speed. We are interested in the region within 0.1 AU from the Sun where the solar wind has a nearly isothermal profile of $T_e \sim 10^6$ K (but the heating mechanisms by various plasma wave modes are not well understood in detail). At very

low radio frequencies, the solar wind becomes optically thick due to free-free absorption and hence the low-frequency emission from near the Sun will be absorbed by the solar wind. Estimate the radius of the “photosphere” of the Sun seen at 1 MHz. What if the Sun has a mass-loss rate of $100 M_{\odot}$?

Prob. 19. Typical Milky Way-like galaxies has total H α line luminosity of the order $10^{41} \text{ erg s}^{-1}$, which comes from the ionization of the interstellar medium by young massive stars and corresponds to a star-formation rate of the order $1 M_{\odot} \text{ yr}^{-1}$ (Kennicutt & Evans 2012). For a total area of the order 10^2 kpc^2 , we find an average surface density of H α luminosity $\Sigma_{\text{H}\alpha} \sim 10^{39} \text{ erg s}^{-1} \text{ kpc}^{-2}$. From atomic physics, we know that ionized gas at typical temperature $T_e \simeq 10^4 \text{ K}$ (appropriate for HII regions) undergoes recombination and emits H α photons at a rate of $\alpha_{\text{H}\alpha} n_{\text{H}}^2$, where n_{H} is the hydrogen number density and $\alpha_{\text{H}\alpha} \simeq 10^{-13} \text{ cm}^3 \text{ s}^{-1}$. For a typical line of sight perpendicular to the galaxy’s star-forming disk, provide order-of-magnitude estimates for the following quantities:

- (i) The emission measure.
- (ii) The free-free self-absorption frequency ν_a where $\tau_{\nu_a} = 1$.
- (iii) The surface density of specific luminosity (in units of $\text{erg kpc}^{-2} \text{ s}^{-1} \text{ Hz}^{-1}$) of free-free emission at 1 GHz.

References

- Abramowitz & Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Dover, New York, 1964)
- Bethe & Heitler, 1934, Proceedings of the Royal Society of London, 146, 83
- Blumenthal & Gould, 1970, Reviews of Modern Physics, 42, 237
- F. F. Chen, *Introduction to Plasma Physics and Controlled Fusion* (Springer International Publishing, 3rd Edition, Switzerland, 2016)
- B. Draine, *Physics of the Interstellar and Intergalactic Medium* (Princeton University Press, New Jersey, 2011)
- Henriksen & Mushotzky, 1986, ApJ, 302, 287
- D. G. Hummer, 1988, ApJ, 327, 477
- Karzas & Latter, 1961, ApJS, 6, 167
- Landau & Lifshitz, *Mechanics* (Butterworth-Heinemann, 3rd Edition, Oxford, 1976)
- Livingston & Bethe, 1937, Reviews of Modern Physics, 9, 245
- Rybicki & Lightman, *Radiative processes in astrophysics* (A Wiley-Interscience Publication, New York, 1979).

- Terzian, & Parrish, 1970, *Astrophysical Letters*, 5, 261
Uehling, E. A., 1954, *Annual Review of Nuclear and Particle Science*, 4, 315
Wright & Barlow, 1975, *MNRAS*, 170, 41

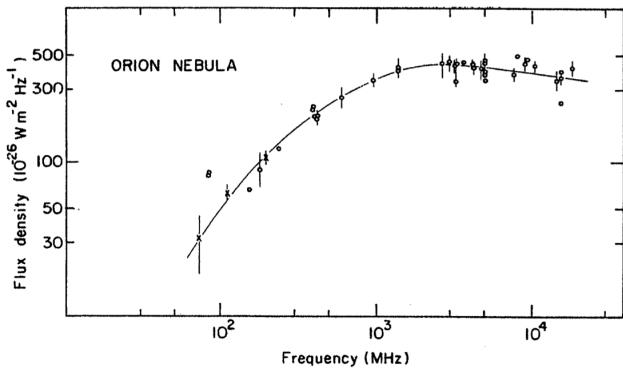


Figure 6.15: Radio spectrum of the Orion nebula (Terzian & Parrish, 1970). The distance to the Orion nebula is $d = 410 \text{ pc}$.

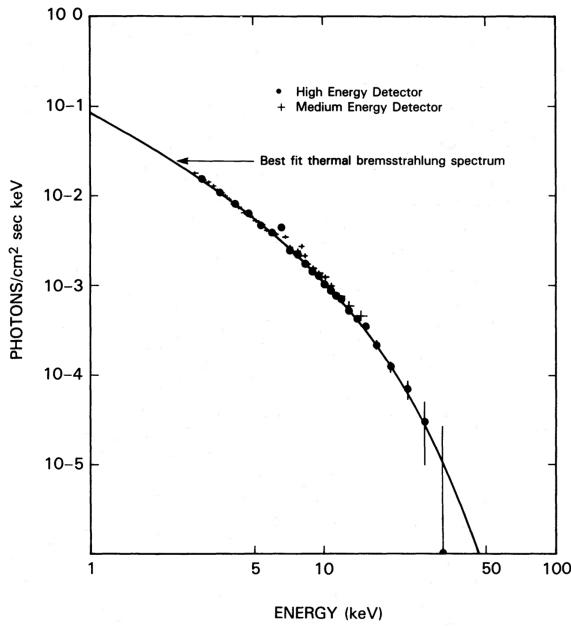


Figure 6.16: The X-ray continuum spectrum of the Coma galaxy cluster is well fit by thermal free-free emission for electron temperature $k_B T_e = 7.6 \text{ keV}$ (Henriksen & Mushotzky 1986). The angular diameter of the X-ray emission region is about 1 degree and this corresponds to a solid angle of $\Omega \simeq 2 \times 10^{-4} \text{ sr}$. The distance to Coma Cluster is $d = 100 \text{ Mpc}$.

Chapter 7

Compton Scattering

This chapter discusses the scattering of a photon by a charge particle.

7.1 Kinematics of the Compton process

Scattering of low-frequency EM waves by a free electron is accurately described by the classical picture of Thomson scattering: a charge that is driven to oscillate by the incoming monochromatic EM wave radiates secondary waves in all directions. If we ignore radiative damping (which is an excellent approximation for low-amplitude EM waves), then the electron oscillates at exactly the same frequency as the incoming wave as in the case of a periodically driven oscillator and hence it radiates at the wave frequency. This can be pictured as photons bouncing off the electron without energy loss (or frequency change) — similar to the elastic scattering of a bouncy ball against a wall. However, from momentum conservation, we know that elastic scattering cannot be the whole story, because for any non-zero angle between the momenta of the incoming and outgoing photons, there must be momentum exchange with the electron. This means that the electron, which has a finite mass (not an infinitely heavy “wall”), must receive a recoil kick. In the frame where the electron is initially at rest, the outgoing photon must have a lower energy than the incoming one.

Let us first work out the energy shift in the scattering process, which will then give the magnitude of the recoil kick received by the electron. Suppose the electron is initially at rest — it is convenient to study Compton scattering in the rest frame of the electron. The scattering process is shown in Fig. 7.1. Let us write the initial 4-momenta of the photon and electron

$$\text{photon: } \hbar(\omega/c, \mathbf{k}), \text{ electron: } m_e c(1, \mathbf{0}), \quad (7.1)$$

where the magnitude of the photon wavevector is $k = |\mathbf{k}| = \omega/c$ in vacuum. After the

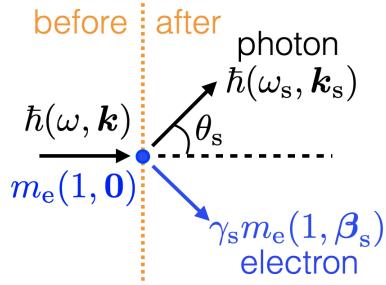


Figure 7.1: Compton scattering process illustrated by the relevant 4-momenta ($c = 1$).

scattering, the 4-momenta of the photon and electron are

$$\text{photon: } \hbar(\omega_s/c, \mathbf{k}_s), \text{ electron: } \gamma_s m_e c(1, \beta_s), \quad (7.2)$$

The total 4-momentum is conserved: $\vec{p}(\text{initial}) = \vec{p}(\text{final})$. We obtain two following two equations

$$\hbar\omega_s + \gamma_s m_e c^2 = \hbar\omega + m_e c^2, \quad \hbar\mathbf{k}_s + \gamma_s \beta_s m_e c = \hbar\mathbf{k}. \quad (7.3)$$

The first equation gives $\gamma_s = 1 + \hbar(\omega - \omega_s)/(m_e c^2)$ and the second one leads to $\gamma_s \beta_s = \hbar(\mathbf{k} - \mathbf{k}_s)/(m_e c)$. Combining these two to eliminate γ_s , we obtain

$$\omega_s = \frac{\omega}{1 + x(1 - \cos \theta_s)}, \quad x \equiv \frac{\hbar\omega}{m_e c^2}, \quad \cos \theta_s \equiv \hat{\mathbf{k}} \cdot \hat{\mathbf{k}}_s, \quad (7.4)$$

where x is the ratio between incident photon's energy and the electron's rest mass energy, and θ_s is the angle between the incoming and outgoing wavevectors. The frequency shift due to Compton scattering is shown on the right panel of Fig. 7.2.

The photon always loses energy by a fractional amount

$$\frac{\Delta\omega}{\omega} = \frac{\omega_s - \omega}{\omega} = -\frac{x(1 - \cos \theta_s)}{1 + x(1 - \cos \theta_s)}. \quad (7.5)$$

Based on the frequency change, we obtain the wavelength change

$$\lambda - \lambda_s = \lambda_C(1 - \cos \theta_s), \quad \lambda_C \equiv h/(m_e c) = 0.024 \text{ \AA}, \quad (7.6)$$

where λ_C is the *Compton wavelength of an electron* (which equals to the de Broglie wavelength for a particle of momentum $m_e c$). One sees that in the classical limit $\hbar \rightarrow 0$ (and hence $x \rightarrow 0$), there is no electron recoil at all and we simply have an elastic collision. However, in a typical quantum mechanical scattering, the electron in fact takes away an order-unity fraction of the incident photon's *momentum*, meaning that

$$\gamma_s \beta_s \sim \mathcal{O}(x), \quad \forall x \in (0, \infty). \quad (7.7)$$

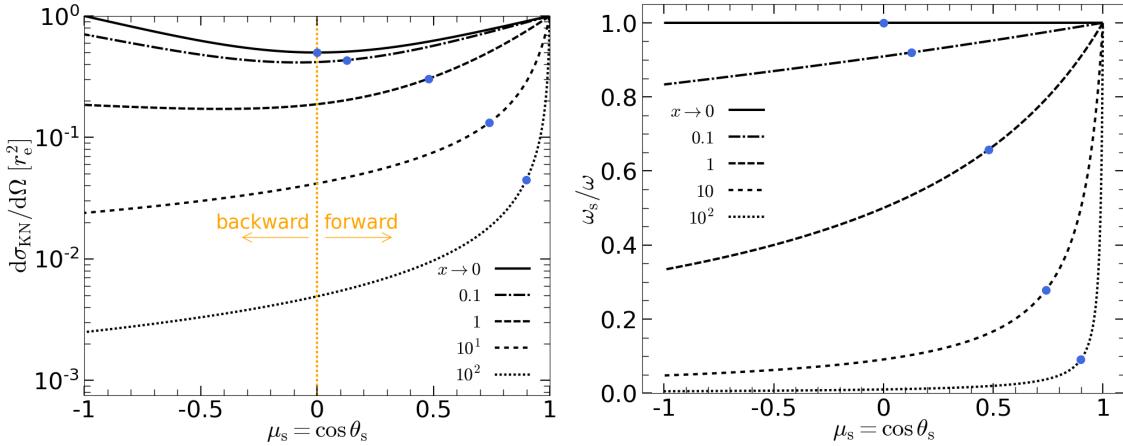


Figure 7.2: *Left panel:* Klein-Nishina differential cross-sections, in units of $r_e^2 = e^4/m_e^2 c^4$, for different photon energies ($x = \hbar\omega/m_e c^2$) and scattering angles. The blue dots show the median scattering angle $\theta_{s,\text{med}}$ (50% photons are scattered to $\theta_s < \theta_{s,\text{med}}$). *Right panel:* Cumulative distribution function (CDF) for the electron's energy gain E_e . The horizontal axis is normalized by the maximum energy gain $E_{e,\text{max}} = 2x^2/(1+2x)$ (the Compton edge).

The dimensionless *energy gain* by the electron is given by

$$\frac{E_e}{m_e c^2} = \gamma_s - 1 = \frac{x^2(1 - \cos \theta_s)}{1 + x(1 - \cos \theta_s)}, \quad (7.8)$$

which has a maximum value called the *Compton edge*,

$$\frac{E_{e,\text{max}}}{m_e c^2} = \frac{2x^2}{1 + 2x}. \quad (7.9)$$

To obtain the probability distributions of the energies of the scattered photon and the recoiling electron, we will need the differential scattering cross-section, which describes the probability distribution of the scattering angle θ_s .

7.2 Klein-Nishina differential cross-section

Based on quantum electrodynamics calculations, the probability distribution of the scattering angle θ_s , for *unpolarized* incident radiation, is described by the *Klein-Nishina differential cross-section*

$$\frac{d\sigma_{\text{KN}}}{d\Omega}(\theta_s) = \frac{r_e^2}{2} \frac{\omega_s^2}{\omega^2} \left(\frac{\omega_s}{\omega} + \frac{\omega}{\omega_s} - 1 + \cos^2 \theta_s \right), \quad (7.10)$$

where $r_e = e^2/m_e c^2$ is the classical electron radius, and the frequency ratio between incoming and scattered photons, ω/ω_s , is given by eq. (7.4). This functional form is shown on the left panel of Fig. 7.2.

The total cross-section is given by (Rybicki & Lightman's eq. 7.5)

$$\sigma_{\text{KN}} = \int \frac{d\sigma_{\text{KN}}}{d\Omega} d\Omega = \frac{3\sigma_T}{4} \left\{ \frac{1+x}{x^3} \left[\frac{2x(1+x)}{1+2x} - \ln(1+2x) \right] + \frac{\ln(1+2x)}{2x} - \frac{1+3x}{(1+2x)^2} \right\}, \quad (7.11)$$

where $\sigma_T = 8\pi r_e^2/3$ is the Thomson cross-section. This is shown on the left panel of Fig. 7.3. The total cross-section has the following asymptotic limits

$$\frac{\sigma_{\text{KN}}}{\sigma_T} \approx \begin{cases} 1 - 2x + \frac{26}{5}x^2, & \text{for } x \ll 1, \\ \frac{3}{8}x^{-1}(\ln 2x + 1/2), & \text{for } x \gg 1. \end{cases} \quad (7.12)$$

A simple analytic approximation for the total KN cross-section is

$$\frac{\sigma_{\text{approx}}}{\sigma_T} \approx \frac{1 + 1.2x}{1 + 1.6x} \frac{1 + x/2}{a + x} \frac{\ln(1 + ax)}{ax}, \quad a = 2e^{1/2} \approx 3.30, \quad \forall x \in (0, \infty), \quad (7.13)$$

which has a fractional error $< 1.7\%$ for any x .

In the $x \gg 1$ limit, Klein-Nishina suppression of the scattering cross-section $\sigma_{\text{KN}} \propto x^{-1}$ can be understood in the following way. The photon-electron system becomes nearly symmetric in the *center of momentum (CoM) frame*, which is moving at a Lorentz factor $\Gamma \approx x^{1/2}$ with respect to the electron's rest frame. In the CoM frame, the two particles have nearly the same energy $x^{1/2}m_e c^2$ and hence relativistic mass of $x^{1/2}m_e$. Since the interaction cross-section scales as the inverse square of the particle masses (for instance the Thomson cross-section $\sigma_T \propto r_e^2 \propto m_e^{-2}$), we obtain $\sigma_{\text{KN}} \propto x^{-1}m_e^{-2}$. We note that there is also a rather weak logarithmic dependence of $\propto \ln(2x)$ which makes decrease in the cross-section slightly shallower than $\sigma_{\text{KN}} \propto x^{-1}$, as can be seen in Fig. 7.3.

In the low-frequency limit $x \ll 1$, $d\sigma_{\text{KN}}/d\Omega$ is very close to the Thomson differential cross-section¹

$$\frac{d\sigma_{\text{KN}}}{d\Omega} \approx \frac{d\sigma_T}{d\Omega} [1 - 2x(1 - \cos \theta_s)], \quad \text{for } x \ll 1, \quad \frac{d\sigma_T}{d\Omega} = \frac{r_e^2}{2}(1 + \cos^2 \theta_s), \quad (7.14)$$

¹Note that the Thomson differential cross-section for *linearly polarized* waves is given by $(d\sigma_T/d\Omega)_{\text{lin}} = r_e^2 \sin^2 \Theta$, where Θ is the angle between the line of sight (or the \mathbf{k}_s vector) and the polarization vector (or the E-field) of the incoming wave. An unpolarized beam of radiation can be decomposed into two orthogonal sets of linearly polarized waves that do not have any phase correlation. We can choose one linear polarization vector to be \perp to the $\mathbf{k}\cdot\mathbf{k}_s$ plane and hence $\Theta_1 = \pi/2$, and then the other linear polarization vector must be in the $\mathbf{k}\cdot\mathbf{k}_s$ plane while \perp to \mathbf{k} (the incoming wavevector) — meaning that the angle between the E-field and \mathbf{k}_s is $\Theta_2 = \pi/2 - \theta_s$. Summing up these two components, we obtain the differential cross-section for *unpolarized waves* $(d\sigma_T/d\Omega)_{\text{unp}} = (1/2)r_e^2(\sin^2 \Theta_1 + \sin^2 \Theta_2) = (1/2)r_e^2(1 + \cos^2 \theta_s)$.

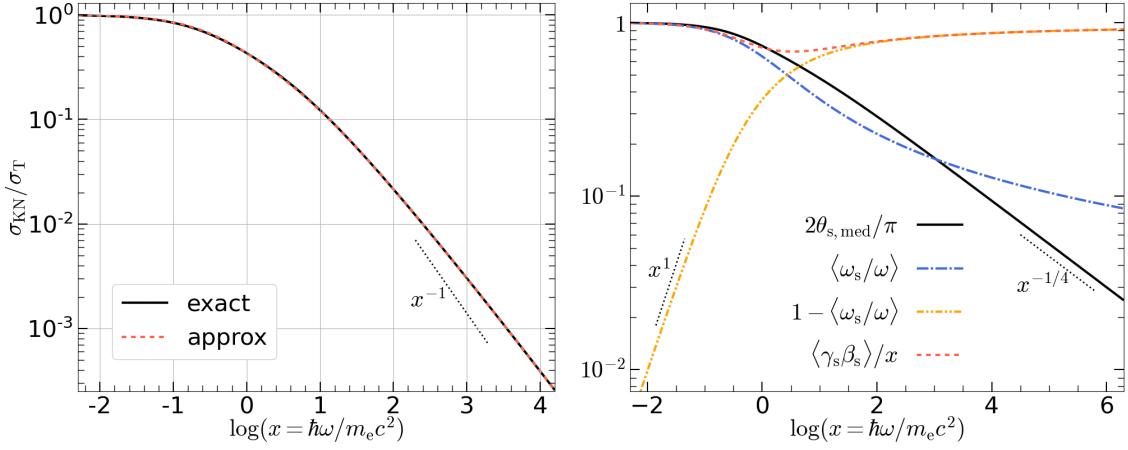


Figure 7.3: *Left panel:* Exact and approximate Klein-Nishina cross-section from eqs. (7.11) and (7.13), respectively. *Right panel:* The median scattering angle $\theta_{s,\text{med}}$, mean energy of the scattered photons $\langle \omega_s/\omega \rangle$, mean recoil momentum of the electron $\langle \gamma_s \beta_s \rangle$ divided by x , as functions of the incoming photon's dimensionless energy x .

where the $2x(1 - \cos \theta_s)$ term only brings in a slight preference for forward scattering ($\theta_s < \pi/2$) as opposed to backward scattering ($\theta_s > \pi/2$).

The right panel of Fig. 7.3 shows various moments of the differential KN cross-section. For instance, in the high-energy limit $x \gg 1$, the differential cross-section is strongly forward-scattering dominated. Let us define the median² scattering angle $\theta_{s,\text{med}}$ by

$$\begin{aligned} 0.5\sigma_{\text{KN}} &= \int_0^{\theta_{s,\text{med}}} \frac{d\sigma_{\text{KN}}}{d\Omega} d\Omega \approx \pi r_e^2 \int_0^{\theta_{s,\text{med}}} \left[\left(\frac{\omega_s}{\omega} \right)^3 + \frac{\omega_s}{\omega} \right] \theta_s d\theta_s \\ &\approx \pi r_e^2 \int_0^{\theta_{s,\text{med}}} \left[\frac{1}{(1 + x\theta_s^2/2)^3} + \frac{1}{1 + x\theta_s^2/2} \right] \theta_s d\theta_s = \frac{\pi r_e^2}{x} \left(\frac{1}{2} + \ln \frac{x\theta_{s,\text{med}}^2}{2} \right), \end{aligned} \quad (7.15)$$

where we have assumed $\theta_{s,\text{med}} \ll 1$ and $x\theta_{s,\text{med}}^2 \gg 1$ (to be justified later). Comparing the above result with the high-energy asymptotic limit of $0.5\sigma_{\text{KN}} \approx \pi r_e^2 (\ln 2x + 1/2) / (2x)$, we obtain

$$\theta_{s,\text{med}} \approx 2^{3/4} e^{-1/8} x^{-1/4} \approx 1.484 x^{-1/4}, \text{ for } x \gg 1, \quad (7.16)$$

which scales rather weakly with incoming photon energy as $x^{-1/4}$. For $x = 10^4$ (or $\hbar\omega \simeq 5 \text{ GeV}$), half of the scattered photons are beamed within an angle of $\approx 8.5^\circ$ from the initial wavevector.

²Note that $\theta_{s,\text{med}}$ is different from the mean scattering angle $\langle \theta_s \rangle$, which is not discussed here.

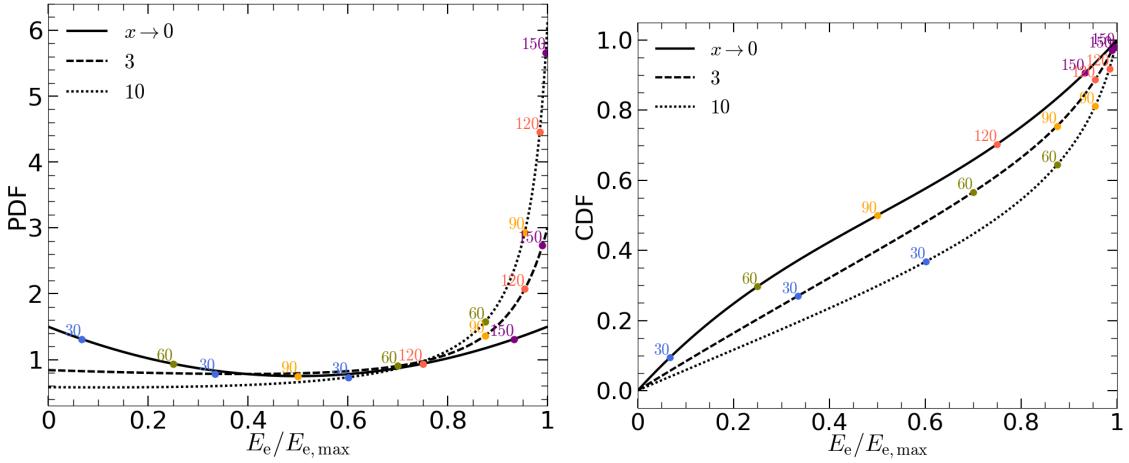


Figure 7.4: Distribution of the recoiling electron’s energy E_e (normalized by the Compton edge $E_{e,\max}$) for different incoming photon energy. The two panels show the probability density function (PDF, left) and the cumulative distribution function (CDF, right). In each case, we mark the scattering angles of 30° , 60° , 90° , 120° , and 150° in colored circles.

Another potentially useful quantity is the mean energy of the scattered photons $\langle \hbar\omega_s \rangle$. In the limit $x \ll 1$, we obtain the following mean ratio in the Thomson regime

$$\langle \omega_s / \omega \rangle \approx 1 - x, \text{ for } x \ll 1. \quad (7.17)$$

The energy loss goes to the electron, which has a mean energy gain of $\langle E_e \rangle = x^2 m_e c^2$ — this is the *Compton heating effect*.

For $x \gg 1$, the photon loses the majority of its energy to the electron, which acquires a momentum kick of roughly $\Delta p_e \sim xm_e c$ and becomes highly relativistic. The mean energy of the scattered photons has the following asymptotic limit

$$\langle \omega_s / \omega \rangle = \frac{\int (\omega_s / \omega) (d\sigma_{\text{KN}} / d\Omega) d\Omega}{\int (d\sigma_{\text{KN}} / d\Omega) d\Omega} \approx \frac{4/3}{\ln(2x) + 1/2}, \text{ for } x \gg 1. \quad (7.18)$$

Combining the asymptotic limits on both ends, we find the following approximation

$$\langle \omega_s / \omega \rangle_{\text{approx}} \approx \frac{8}{5/(1 + 4x/5) + 6 \ln(1 + 2x) + 3}, \quad \forall x \in (0, \infty), \quad (7.19)$$

which has a fractional error $< 1.5\%$ for any x . The corresponding mean energy gain by the electron is given by

$$\langle E_e \rangle = (1 - \langle \omega_s / \omega \rangle) x m_e c^2, \quad \forall x \in (0, \infty). \quad (7.20)$$

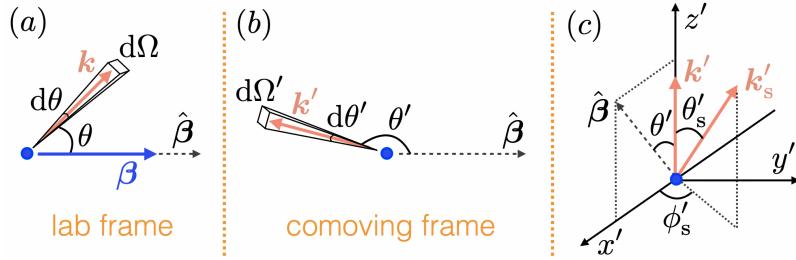


Figure 7.5: Geometry for inverse-Compton scattering. Panel (a): A photon with wavevector \mathbf{k} in the lab frame. Panel (b): The same photon with wavevector \mathbf{k}' in the electron's comoving frame. Panel (c): The photon is now scattered to wavevector \mathbf{k}'_s which is at an angle θ'_s from the incident wavevector \mathbf{k}' , and the azimuthal angle of the scattered wavevector is denoted as ϕ'_s (and $\phi'_s = 0$ or π means that \mathbf{k}'_s is in the $\hat{\beta} \cdot \mathbf{k}'$ plane).

The probability distribution of the recoiling electron's energy E_e is shown in Fig. 7.4, where we see that the probability density has a sharp peak at the Compton edge $E_{e,\max}$ even for $x \gg 1$, despite the fact that the probability of backward scattering ($\theta_s \sim 180^\circ$) is strongly suppressed.

7.3 Single inverse-Compton (IC) scattering

7.3.1 Two Lorentz transformations

Next, we consider a moving electron in the lab frame. Let us first Lorentz transform all lab-frame quantities to the comoving frame of the electron's initial velocity, and then after the Compton scattering, we transform things back to lab frame. Our goal is to study how the radiation field is modified by electron scattering — this is called the *inverse-Compton (IC) process*. We will focus on a single IC scattering by an electron with given velocity β in this section, and the results will then be applied to a group of electrons with a given velocity distribution function in §7.3.4. Repeated scatterings in a gas with modest to high scattering optical depths will be discussed later in §7.4.

Hereafter, comoving-frame quantities are denoted by a prime ('') whereas unprimed quantities are measured in the lab frame. The geometry of the system is shown in Fig. 7.5.

In this problem there are two Doppler factors, which correspond to the momentum directions of the incoming and scattered photons, $\hat{\mathbf{k}}$ and $\hat{\mathbf{k}}_s$, respectively. These are

$$\mathcal{D} = \frac{1}{\gamma(1 - \beta \cdot \hat{\mathbf{k}})} = \gamma(1 + \beta \cdot \hat{\mathbf{k}}'), \quad \mathcal{D}_s = \frac{1}{\gamma(1 - \beta \cdot \hat{\mathbf{k}}_s)} = \gamma(1 + \beta \cdot \hat{\mathbf{k}}'_s). \quad (7.21)$$

Then, the photon energies are related to each other by

$$\epsilon' = \epsilon/\mathcal{D}, \quad \epsilon_s = \epsilon'_s \mathcal{D}_s, \quad (7.22)$$

and

$$\epsilon'_s = \epsilon' \left[1 + \frac{\epsilon'}{m_e c^2} (1 - \cos \theta'_s) \right]^{-1}, \quad (7.23)$$

where the only unknown is the scattering angle θ'_s (between $\hat{\mathbf{k}}'$ and $\hat{\mathbf{k}}'_s$) in the comoving frame. The momentum direction of the scattered photon in the comoving frame $\hat{\mathbf{k}}'_s$ is randomly drawn from the differential cross-section. For each $\hat{\mathbf{k}}'_s$, which includes the value of the scattering angle $\theta'_s = \arccos(\hat{\mathbf{k}}' \cdot \hat{\mathbf{k}}'_s)$, we obtain a one-to-one mapping between the initial and final photon energies in the lab frame

$$\frac{\epsilon_s}{\epsilon} = \frac{\mathcal{D}_s}{\mathcal{D}} \left[1 + \frac{\epsilon/\mathcal{D}}{m_e c^2} (1 - \cos \theta'_s) \right]^{-1}. \quad (7.24)$$

All things are exact up to now.

In the following, we will mainly focus on the Thomson-scattering limit with $\gamma\epsilon/m_e c^2 \ll 1$, which simplifies the calculations significantly and allows us to obtain important and intuitive results. In this limit, the final photon energy is given by

$$\frac{\epsilon_s}{\epsilon} \approx \begin{cases} (\mathcal{D}_s/\mathcal{D}) \left[1 - \frac{\epsilon/\mathcal{D}}{m_e c^2} (1 - \cos \theta'_s) \right], & \text{(retaining linear term } \mathcal{O}\left(\frac{\gamma\epsilon}{m_e c^2}\right)\text{)} \\ \mathcal{D}_s/\mathcal{D}, & \text{(ignoring recoiling)} \end{cases} \quad (7.25)$$

When recoil effects are ignored, the maximum and minimum energies of the scattered photon are given by

$$\frac{1}{\Lambda} < \frac{\epsilon_s}{\epsilon} = \frac{1 - \boldsymbol{\beta} \cdot \hat{\mathbf{k}}}{1 - \boldsymbol{\beta} \cdot \hat{\mathbf{k}}_s} < \Lambda, \quad \Lambda \equiv \frac{1 + \beta}{1 - \beta}, \quad (7.26)$$

where Λ is the maximum Compton boost factor (and in the relativistic limit $\Lambda \approx 4\gamma^2$).

Order-of-magnitude estimate in the relativistic limit

For highly relativistic electron motion and in the Thomson scattering limit (ignoring electron recoiling), we obtain $\epsilon_s/\epsilon \approx \mathcal{D}_s/\mathcal{D}$. The maximum and minimum energies of the scattered photon are $4\gamma^2\epsilon$ and $\epsilon/(4\gamma^2)$, respectively. If the initial photon momentum is at a large angle from the electron's velocity, i.e. $\boldsymbol{\beta} \cdot \hat{\mathbf{k}}$ is not extremely close to 1, then we obtain $\mathcal{D} \sim 1/\gamma$. This means that the incoming photon in the comoving frame has energy $\sim \gamma\epsilon$. In the comoving frame, the photon is scattered at a large angle $\theta'_s \sim 1$ rad, so $\boldsymbol{\beta} \cdot \hat{\mathbf{k}}'_s$ is not very close to 1, and hence $\mathcal{D}_s \sim \gamma$. Therefore, we conclude that for typical relativistic IC scatterings, the photon energies obey the following hierarchy

$$\epsilon : \epsilon' (\approx \epsilon'_s) : \epsilon_s \sim 1 : \gamma : \gamma^2, \quad (7.27)$$

which means that low-energy photons can be very efficiently converted to high-energy ones — each scattering gives a net boost of the order γ^2 . Of course, the energy gain comes from

the kinetic energy of the electron and this restricts the lab-frame energy of the scattered photon to be less than $\epsilon + \gamma m_e c^2$.

IC scatterings by relativistic electrons are very common in many astrophysical sources, e.g., relativistic jets from black holes, pulsar wind nebulae, neutron star magnetosphere, and many other systems that have shocks which can accelerate electrons to high energies. The seed photons may be from the emission by scattering plasma itself (in this case, IC is called “self-Compton”) or of external origin (in this case, IC is called “external-Compton”). When dealing with relativistic electrons, one should always keep in mind the ambient radiation field (e.g., Cosmic Microwave Background, star light).

Order-of-magnitude estimate in the non-relativistic limit

On the other hand, for non-relativistic electrons ($\gamma \approx 1$), the mapping between the initial and final photons energies is given by

$$\frac{\epsilon_s}{\epsilon} \approx 1 + \boldsymbol{\beta} \cdot (\hat{\mathbf{k}}_s - \hat{\mathbf{k}}) - \frac{\epsilon}{m_e c^2} (1 - \cos \theta_s), \quad (\text{for non-relativistic motion}) \quad (7.28)$$

where we have ignored high-order terms $\mathcal{O}(\beta^2)$ or $\mathcal{O}(\beta \epsilon / m_e c^2)$. This shows that the photon may gain or lose energy depending on the scattering direction and initial photon energy. For an isotropic electron velocity distribution, the first-order Doppler shift term $\boldsymbol{\beta} \cdot (\hat{\mathbf{k}}_s - \hat{\mathbf{k}})$ does not contribute to net energy gain or loss — net energy gain or loss due to electron motion comes from higher order terms $\mathcal{O}(\beta^2)$, which will be discussed later in this subsection.

A dividing line between hard (“high energy”) and soft (“low energy”) photons comes from the relative importance of Doppler shifts (the $\boldsymbol{\beta} \cdot (\hat{\mathbf{k}}_s - \hat{\mathbf{k}})$ term) and Compton recoil (the $-\epsilon / m_e c^2$ term). For this reason, we define a critical photon energy

$$\epsilon_c \sim \beta m_e c^2 \sim \sqrt{k_B T m_e c^2} \sim 1 \text{ keV} (T/10^4 \text{ K})^{1/2}, \quad (7.29)$$

where T is the kinetic temperature for a Maxwellian distribution.

For soft photons $\epsilon \ll T_4^{1/2}$ keV, Doppler shifts are more important than recoil effects, and the photon can gain or lose energy at roughly half-half probabilities — this leads to *line broadening by electron scattering* as well as intensity fluctuations of the Cosmic Microwave Background (the Sunyaev-Zel’dovich effect). For instance, a single scattering of Balmer line H α photons ($\epsilon = 1.89$ eV) by an ionized gas of temperature T leads to wavelength broadening of the order $|\Delta\lambda|/\lambda = |\epsilon_s/\epsilon - 1| \simeq \sigma_v/c = 1.8 \times 10^{-3} T_4^{1/2}$, where we used the velocity dispersion $\sigma_v = \sqrt{2k_B T / m_e} = 550 \text{ km s}^{-1} T_4^{1/2}$ for a Maxwellian distribution. Multiple scatterings cause the photon wavelength to undergo a 1D random walk with step-size of the order $\pm \Delta\lambda$.

On the other hand, hard photons with $\epsilon \gg T_4^{1/2}$ keV always get down-scattered, with an average fractional energy loss per-scattering of (ignoring the electron motion)

$$\frac{\langle \Delta\epsilon \rangle}{\epsilon} \approx -\frac{\epsilon}{m_e c^2} \sigma_{\text{KN}}^{-1} \int d\Omega \frac{d\sigma_{\text{KN}}}{d\Omega} (1 - \cos \theta_s) = -\frac{\epsilon}{m_e c^2} (1 - \langle \cos \theta_s \rangle) \approx -\frac{\epsilon}{m_e c^2}, \quad (7.30)$$

where $d\Omega = d\phi_s d\cos \theta_s$ is the differential solid angle for scattered photons and $\langle \cos \theta_s \rangle = \mathcal{O}(\epsilon/m_e c^2)$ (a high-order term) describes the average over all scattered angles.

For instance, an X-ray photon with $\epsilon = 5$ keV on average loses 1% of its energy in each scattering, and if it undergoes 10^2 repeated scatterings (for a slab of scattering optical depth of the order $\tau_s \sim 10$), nearly all its initial energy will be lost; the radiation energy is used to heat up the electrons. In fact, this downward scattering trend always exists for arbitrary photon energies. For a system of given radiation field and electron distribution function, one must consider the competition between downward scatterings due to electron recoils (leading to electron heating) and upward scatterings due to electron motion (causing electrons to cool). This will be discussed in more detail later in §7.3.2.

Mean energy of scattered photons in the Thomson limit

Consider a photon with given wavevector \mathbf{k} scattering off an electron with given velocity vector β in the Thomson scattering limit (ignoring recoil effects, otherwise the math gets messy). What is the mean energy of the scattered photon $\langle \epsilon_s \rangle$? This can be calculated by averaging over all possible scattering angles $\hat{\mathbf{k}}'_s$,

$$\frac{\langle \epsilon_s \rangle}{\epsilon} = \frac{\langle \mathcal{D}_s \rangle}{\mathcal{D}} = \frac{1}{\mathcal{D}_{\sigma_T}} \int d\Omega' \frac{d\sigma_T}{d\Omega'} \mathcal{D}_s(\hat{\mathbf{k}}'_s). \text{ (ignoring recoil)} \quad (7.31)$$

Later on, we will show that $\langle \beta \cdot \hat{\mathbf{k}}'_s \rangle = 0$ and hence

$$\langle \mathcal{D}_s \rangle = \gamma \left(1 + \langle \beta \cdot \hat{\mathbf{k}}'_s \rangle \right) = \gamma, \quad (7.32)$$

which holds for any electron velocity β and photon wavevector \mathbf{k} as long as we are in the Thomson scattering limit. This is essentially because Thomson scattering is forward-backward symmetric in the electron's comoving frame. Then, the mean energy of the scattered photon is given by (for a given collision as specified by β and \mathbf{k})

$$\frac{\langle \epsilon_s \rangle}{\epsilon} \approx \frac{\langle \mathcal{D}_s \rangle}{\mathcal{D}} = \gamma^2 (1 - \beta \cdot \hat{\mathbf{k}}) = \gamma^2 (1 - \beta \cos \theta). \quad (7.33)$$

Based on eq. (7.33), one might naively guess that for an isotropic radiation field *or* electron velocity distribution, the vanishing average of $\langle \beta \cdot \hat{\mathbf{k}} \rangle = 0$ would lead to $\langle \epsilon_s \rangle = \gamma^2 \epsilon$, and hence the average fractional energy gain is $\langle \Delta\epsilon \rangle / \epsilon = \gamma^2 - 1 = \gamma^2 \beta^2$. This is not too far

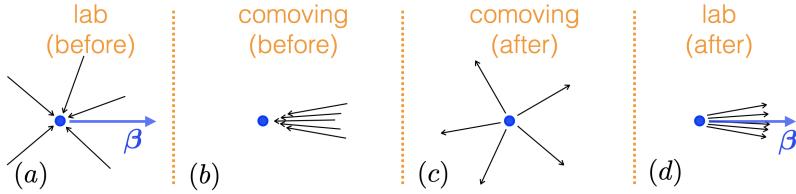


Figure 7.6: Radiation field seen in different frames before and after Compton scattering.

off. It turns out that this fractional energy gain is missing a factor of $4/3$ due to the fact that there are more head-on and rear-end collisions (see eq. 7.50 below).

Below we demonstrate that $\langle \beta \cdot \hat{k}'_s \rangle$ vanishes. In the electron comoving frame, without losing generality, let us consider that \hat{k}' is along the \hat{z}' direction and that the electron velocity vector is in the \hat{x}' - \hat{z}' plane given by $\beta = (\sin \theta', 0, \cos \theta')$ in Cartesian components, as shown in the panel (c) of Fig. 7.5. The angle between β and \hat{k}' is given by Lorentz transformation (although we will not need this for our purpose here)

$$\cos \theta' = \frac{\cos \theta - \beta}{1 - \beta \cos \theta}, \quad (7.34)$$

where $\theta = \arccos(\hat{\beta} \cdot \hat{k})$ is the angle between β and \hat{k} . The wavevector of the scattered photon is in the direction \hat{k}'_s that is described by angles θ'_s and ϕ'_s in spherical coordinates, and its Cartesian components are

$$\hat{k}'_s = (\sin \theta'_s \cos \phi'_s, \sin \theta'_s \sin \phi'_s, \cos \theta'_s), \quad (7.35)$$

and hence

$$d\Omega' = d\phi'_s d\cos \theta'_s, \quad \frac{d\sigma_T}{d\Omega'} = \frac{r_e^2}{2} (1 + \cos^2 \theta'_s). \quad (7.36)$$

The Doppler factor for the scattered photon is $\mathcal{D}_s = \gamma(1 + \beta \cdot \hat{k}'_s)$ and the projection of the velocity onto the direction of the scattered wavevector is given by

$$\beta \cdot \hat{k}'_s = \beta (\sin \theta' \sin \theta'_s \cos \phi'_s + \cos \theta' \cos \theta'_s). \quad (7.37)$$

When carrying out the averaging, the first term in $\langle \beta \cdot \hat{k}'_s \rangle$ is equal to zero because $\int_0^{2\pi} \cos \phi'_s d\phi'_s = 0$, and the second term also vanishes since $\int_{-1}^1 \cos \theta'_s d\cos \theta'_s = 0$.

7.3.2 Emitting power (*isotropic* photon/electron distribution)

Let us consider a single electron with velocity β moving through an isotropic radiation field, as shown in Fig. 7.6. Our goal is to find the energy loss rate from the electron

due to IC scattering of radiation. Due to symmetry, some of the results also apply to the case where electrons have isotropic velocity distribution whereas the radiation field may be arbitrary.

The photon momentum ($\mathbf{p} = \hbar\mathbf{k}$) distribution is described by

$$f(k)d^3\mathbf{k} = f(k)k^2dkd\Omega, \quad (7.38)$$

which is normalized such that the number density of photons and radiation energy density are given by

$$n_{\text{ph}} = \int f(k)d^3\mathbf{k} = 4\pi k^2 f(k)dk, \quad (7.39)$$

and

$$U_{\text{rad}} = \int f(k)\epsilon d^3\mathbf{k} = 4\pi k^2 \epsilon f(k)dk, \quad (7.40)$$

respectively. Since the electron has a cross-section σ_T , the photon-electron scattering rate R (unit: sec^{-1}) is given by

$$R_{\text{sca}} = \int d^3\mathbf{k} f(k) \sigma_T c (1 - \boldsymbol{\beta} \cdot \hat{\mathbf{k}}), \quad (7.41)$$

where the $1 - \boldsymbol{\beta} \cdot \hat{\mathbf{k}}$ factor (known as the *Møller factor*) is due to the fact that photons have to catch up with the moving electron. For an isotropic radiation field, the $\boldsymbol{\beta} \cdot \hat{\mathbf{k}}$ term does not contribute to the above integral (this also holds for an isotropic electron velocity distribution with an arbitrary radiation field), so we obtain

$$R_{\text{sca}} = n_{\text{ph}} \sigma_T c, \quad (7.42)$$

as we would have guessed. We see that the Møller factor is unimportant for obtaining the scattering rate in an isotropic radiation field or an isotropic electron velocity distribution. However, it is shown below that the Møller factor is important for obtaining the correct total emitting power as the mean energy for the scattered photons brings in another factor of $1 - \boldsymbol{\beta} \cdot \hat{\mathbf{k}}$ (see eq. 7.45).

Since each photon has energy $\epsilon = \hbar k c$, the power (unit: erg s^{-1}) of the incident photons onto the electron is given by

$$\frac{dW_0}{dt} = \int d^3\mathbf{k} f(k) \epsilon \sigma_T c (1 - \boldsymbol{\beta} \cdot \hat{\mathbf{k}}). \quad (7.43)$$

Again, for an isotropic radiation field *or* isotropic electron distribution, the $\boldsymbol{\beta} \cdot \hat{\mathbf{k}}$ term does not contribute and hence we obtain

$$dW_0/dt = U_{\text{rad}} \sigma_T c, \quad (7.44)$$

as we would have guessed. If all the scattered photons had identical energies with the incident ones in the lab frame, then the power of the scattered photons would be equal to dW_0/dt and we would not be interested in this topic in the first place.

In fact, due to Doppler effects, the scattered photons have an average energy of $\langle \epsilon_s \rangle = \gamma^2 \epsilon (1 - \boldsymbol{\beta} \cdot \hat{\mathbf{k}})$, as given by eq. (7.33). Therefore, the power of scattered photons is given by the following integral

$$\begin{aligned} \frac{dW}{dt} &= \int d^3\mathbf{k} f(k) \langle \epsilon_s \rangle \sigma_{TC} (1 - \boldsymbol{\beta} \cdot \hat{\mathbf{k}}) \\ &= \gamma^2 \sigma_{TC} c \int dk k^2 \epsilon f(k) \int d\Omega (1 - \boldsymbol{\beta} \cdot \hat{\mathbf{k}})^2. \end{aligned} \quad (7.45)$$

For an isotropic radiation field *or* isotropic electron distribution, the integral over solid angle gives $\int d\Omega (1 - \boldsymbol{\beta} \cdot \hat{\mathbf{k}})^2 = 2\pi \int_{-1}^1 d\mu (1 + \beta^2 \mu^2) = 4\pi(1 + \beta^2/3)$. Making use of eq. (7.40) for the radiation energy density, we obtain

$$dW/dt = \gamma^2 (1 + \beta^2/3) U_{\text{rad}} \sigma_{TC} c, \quad (7.46)$$

which is independent of the spectrum of the radiation field (as long as it does not contain very high energy photons that would violate our assumption of $\gamma\epsilon/m_e c^2 \ll 1$).

We could have derived the above result by calculating the power of photons striking the electron in its comoving frame

$$\left(\frac{dW}{dt} \right)' = \sigma_T \int d\Omega' \int d\nu' I'_{\nu'} = \sigma_T \gamma^2 \int d\nu I_{\nu} \int d\Omega (1 - \boldsymbol{\beta} \cdot \hat{\mathbf{k}})^2, \quad (7.47)$$

where we have made use of the Doppler factor \mathcal{D} to carry out the Lorentz transformations of $d\Omega' = \mathcal{D}^2 d\Omega$, $d\nu' = d\nu/\mathcal{D}$, and $I'_{\nu'} = I_{\nu}/\mathcal{D}^3$ (see [Chapter 3]). Again making use of $\int d\Omega (1 - \boldsymbol{\beta} \cdot \hat{\mathbf{k}})^2 = 4\pi(1 + \beta^2/3)$ for an isotropic radiation field and $\int d\nu I_{\nu} = U_{\text{rad}} c / (4\pi)$, we obtain $(dW/dt)' = \gamma^2 (1 + \beta^2/3) U_{\text{rad}} \sigma_{TC} c$. Since the scattering is coherent in the electron comoving frame ($\epsilon'_s = \epsilon'$), we know that $(dW/dt)'$ also equals to the power of the scattered photons — this is the emitting power of the electron in its comoving frame. From the discussion in [Chapter 3], we know that the emitting power is Lorentz invariant, provided that the radiation does not carry net linear momentum in the comoving frame of the particle (which is the case for Thomson scattering). Therefore, we obtain $dW/dt = (dW/dt)'$, in agreement with eq. (7.46).

Finally, the net inverse-Compton (IC) power of the electron, measured in the lab frame, is given by the difference between the incident and scattered powers

$$P_{\text{IC}} = \frac{dW}{dt} - \frac{dW_0}{dt} = \frac{4}{3} \gamma^2 \beta^2 U_{\text{rad}} \sigma_{TC} c. \quad (7.48)$$

The above IC power applies to arbitrary electron speeds (relativistic or non-relativistic), as long as *either* the radiation field *or* electron distribution is isotropic. It should be noted that P_{IC} is not Lorentz invariant (only the dW/dt part is).

The IC power in eq. (7.48) leads to Compton cooling timescale of

$$t_{\text{IC}} = \frac{(\gamma - 1)m_e c^2}{P_{\text{IC}}} = \frac{3m_e c}{4(\gamma + 1)U_{\text{rad}}\sigma_T} = 0.98 \text{ yr} \frac{\text{erg cm}^{-3}}{(\gamma + 1)U_{\text{rad}}}, \quad (7.49)$$

which scales as $t_{\text{IC}} \propto \gamma^{-1}$ in the relativistic limit and approaches a constant in the non-relativistic limit.

In each scattering considered in the lab frame, a photon undergoes a positive angle-average fractional energy *gain* of

$$\frac{\langle \Delta\epsilon \rangle}{\epsilon} = \frac{P_{\text{IC}}}{dW_0/dt} = \frac{4}{3}\gamma^2\beta^2, \quad (7.50)$$

which applies to any Lorentz factor γ . It can be shown that this is a generic consequence of any forward-backward symmetric scattering in the comoving frame of the scatterer. In the relativistic limit, this gives rise to an energy boost by a factor of $(4/3)\gamma^2$. In the non-relativistic limit, the fractional energy gain is equal to $(4/3)\beta^2$. The behavior of $\langle \Delta\epsilon \rangle / \epsilon = (4/3)\beta^2$ is also well-known in particle acceleration by turbulent motion of magnetized fluids. In the so-called *2nd-order Fermi acceleration* process, each time a charged particle is scattered by fluid turbulence (here β is the speed of the scattering clouds or regions of density fluctuations) it would undergo an average fractional energy gain of $(4/3)\beta^2$. Repetitive scatterings will keep accelerating the particle until either it escapes from the system or when cooling effects become important.

If the seed photons are sufficiently hard such that electron recoil effects are not negligible, the IC power P_{IC} will be reduced. In the non-relativistic limit ($\gamma \approx 1$), the average fractional energy gain/loss in each scattering is given by (making use of eq. 7.30)

$$\frac{\langle \Delta\epsilon \rangle}{\epsilon} = \frac{4}{3}\beta^2 - \frac{\epsilon}{m_e c^2}. \quad (7.51)$$

Thus, the IC power for a non-relativistic³ electron with recoiling is given by

$$\begin{aligned} P_{\text{IC}} &= \int d\epsilon \langle \Delta\epsilon \rangle \frac{dn_{\text{ph}}}{d\epsilon} \sigma_T c \\ &= \left[\frac{4}{3}\beta^2 - \frac{\langle \epsilon^2 \rangle / \langle \epsilon \rangle}{m_e c^2} \right] U_{\text{rad}} \sigma_T c, \end{aligned} \quad (7.52)$$

³In the relativistic limit ($\gamma \gg 1$), the IC power in eq. (7.48) is reduced by a factor of $1 - (63/10)\gamma \langle \epsilon^2 \rangle / (m_e c^2 \langle \epsilon \rangle)$ (Blumenthal and Gould 1970)

where $dn_{\text{ph}}/d\epsilon$ describes the energy spectrum of the seed photon number density n_{ph} , $\langle \epsilon \rangle = U_{\text{rad}}/n_{\text{ph}}$ is the mean energy of the seed photons, and $\langle \epsilon^2 \rangle$ is the mean squared photon energy. We see that, when recoiling is important, the inverse-Compton power depends on the spectrum of the seed photons.

Furthermore, for a non-relativistic Maxwellian electron velocity distribution with temperature T , the mean-squared velocity⁴ is given by

$$\langle \beta^2 \rangle_M = 3k_B T/m_e. \quad (7.53)$$

Thus, after averaging over the velocity distribution, the mean fractional energy gain for the scattered photon is given by

$$\frac{\langle \Delta\epsilon \rangle_M}{\epsilon} = \frac{4k_B T - \langle \epsilon^2 \rangle / \langle \epsilon \rangle}{m_e c^2}, \quad \text{for non-relativistic Maxwellian.} \quad (7.54)$$

Thus, the velocity-averaged Compton cooling/heating rate per electron is given by

$$\langle P_{\text{IC}} \rangle_M = \frac{4k_B T - \langle \epsilon^2 \rangle / \langle \epsilon \rangle}{m_e c^2} U_{\text{rad}} \sigma_T c. \quad (7.55)$$

For example, the ratio between mean-squared energy $\langle \epsilon^2 \rangle$ and mean energy $\langle \epsilon \rangle$ for optically thin free-free emission is given by

$$\left(\frac{\langle \epsilon^2 \rangle}{\langle \epsilon \rangle} \right)_{\text{ff}} = \frac{\int j_\nu h\nu d\nu}{\int j_\nu d\nu} \approx \frac{\Gamma(2)}{\Gamma(1)} k_B T = k_B T, \quad (7.56)$$

where $j_\nu \propto e^{-h\nu/k_B T}$ is the free-free emissivity (ignoring the Gaunt factor⁵). We see that, since $4k_B T > (\langle \epsilon^2 \rangle / \langle \epsilon \rangle)_{\text{ff}}$, the free-free photons can be further upscattered by the emitting electrons.

7.3.3 Emitting spectrum (*isotropic* photon distribution)

An isotropic seed radiation field is described by an angle-independent specific intensity function $I_\nu(\nu)$ in the lab frame. Suppose the number density of electrons is n_e in the lab frame and all of them are moving with the same velocity $\beta \parallel \hat{x}$. In this subsection, we calculate the volume emissivity of these electrons $j_{\nu_s}(\theta_s)$ at frequency ν_s and along the line of sight that is at an angle θ_s away from the \hat{x} direction. Hereafter, we use the short-hand

⁴The mean kinetic energy for a Maxwellian distribution is $\langle m_e v^2/2 \rangle = 3k_B T/2$, because each of the three dimensions (or translational degrees of freedom) contributes a mean energy of $k_B T/2$ according to the *equipartition theorem*.

⁵If we were to include the Gaunt factor, then $j_\nu \propto \nu^{-0.1} e^{-h\nu/k_B T}$ would give $\langle \epsilon^2 \rangle / \langle \epsilon \rangle \approx [\Gamma(1.9)/\Gamma(0.9)]k_B T = 0.9k_B T$.

notations $\mu = \cos \theta$ and $\mu_s = \cos \theta_s$ and the primed versions of these are in the electron's comoving frame.

From the fact that j_ν/ν^2 = Lorentz invariant (see §3.7), we obtain

$$j_{\nu_s}(\mu_s) = \mathcal{D}_s^2 j'_{\nu'_s}, \quad (7.57)$$

and then the goal is to find the volume emissivity $j'_{\nu'_s}$ at a given frequency $\nu'_s = \nu_s/\mathcal{D}_s$ along a given direction $\mu'_s = (\mu_s - \beta)/(1 - \beta\mu_s)$. To obtain this, we make the following two simplifying assumptions:

- Scattering is in the Thomson limit such that $\nu'_s = \nu'$, which means that we are ignoring electron recoiling.
- The differential cross-section in the electron rest frame is isotropic $d\sigma_{T,\text{iso}}/d\Omega' = \sigma_T/(4\pi)$, which means that we are ignoring the $(1 + \cos^2 \theta'_s)$ dependence in the scattering probability.

Our results under these two assumptions are in excellent agreement with those from the exact treatment. When relevant, we mention the exact results as well. Under these two assumptions, we obtain

$$j'_{\nu'_s} = \frac{n'_e \sigma_T}{4\pi} \int I'_{\nu'=\nu'_s}(\mu') d\Omega', \quad (7.58)$$

where $n'_e = n_e/\gamma$ is the electron number density in the comoving frame, $I'_{\nu'=\nu'_s}$ only selects the comoving-frame specific intensity at frequency $\nu' = \nu'_s = \nu_s/\mathcal{D}_s$ (for elastic scattering), and the integral over $d\Omega'$ sums over the photons at all possible incoming angles $-1 < \mu' < 1$ (and they have equal probability of being scattered to the direction of μ'_s).

It is more convenient to carry out the solid angle integral using lab-frame quantities, so we make use of the following Lorentz transformations $I'_{\nu'} = I_\nu/\mathcal{D}^3$, $d\Omega' = \mathcal{D}^2 d\Omega$, $\nu' = \nu/\mathcal{D}$, $\mathcal{D} = [\gamma(1 - \beta\mu)]$, and then obtain

$$\begin{aligned} j'_{\nu'_s} &= \frac{(n_e/\gamma)\sigma_T}{4\pi} \int I_{\nu=\nu_s\mathcal{D}/\mathcal{D}_s} \frac{d\Omega}{\mathcal{D}}, \\ &= \frac{n\sigma_T}{2} \int_{-1}^1 I_{\nu=\nu_s\mathcal{D}/\mathcal{D}_s}(1 - \beta\mu) d\mu, \end{aligned} \quad (7.59)$$

where we have used $d\Omega = 2\pi d\mu$, and $I_{\nu=\nu_s\mathcal{D}/\mathcal{D}_s}$ only selects the lab-frame specific intensity at frequency $\nu = \nu_s\mathcal{D}/\mathcal{D}_s$. For a given intensity spectrum $I_\nu(\nu)$ (with no angular dependence since the radiation field is assumed to be isotropic), the above integral can be carried out in the following way: for each angle μ , we first calculate the frequency $\nu = \nu_s\mathcal{D}/\mathcal{D}_s$ at which the seed photons contribute to the observed flux (at given frequency ν_s and angle μ_s), then evaluate I_ν at this frequency, and finally multiply the result by a factor of $(1 - \beta\mu)$ before adding up the contribution from μ to $\mu + d\mu$ to the integral.

Once we have obtained j'_{ν_s} , we immediately know $j_{\nu_s}(\mu_s) = \mathcal{D}_s^2 j'_{\nu_s}$ from eq. (7.57). Then, the *angular averaged* emitting spectrum is given by

$$\bar{j}_{\nu_s} = \frac{1}{4\pi} \int j_{\nu_s}(\mu_s) d\Omega_s = \frac{1}{2} \int_{-1}^1 j_{\nu_s}(\mu_s) d\mu_s. \quad (7.60)$$

One should be careful that the *emitting* power in general is not equal to the *observed* power. For a single narrow beam of electrons, one can imagine that the observed flux is much higher if the observer's line of sight is within the $1/\gamma$ beaming cone of the electrons than outside the beaming cone. For the special case of an isotropic electron distribution, the observer has equal probability of seeing an electron at any angle randomly drawn from the entire 4π sphere, so the observed power is simply given by $4\pi \bar{j}_{\nu_s}$ multiplied by the total volume of the gas.

In the following, we consider the simplest case of an isotropic electron distribution as well as a δ -function incident spectrum given by the specific intensity

$$I_\nu = I_0 \delta(\nu - \nu_0). \quad (7.61)$$

In this case, the integral in eq. (7.59) gives the lab-frame emissivity

$$\begin{aligned} j_{\nu_s}(\mu_s) &= \mathcal{D}_s^2 \frac{n_e \sigma_T I_0}{2} \int_{-1}^1 (1 - \beta\mu) \delta\left(\nu_s \frac{\mathcal{D}}{\mathcal{D}_s} - \nu_0\right) d\mu \\ &= \begin{cases} \frac{n \sigma_T I_0 \nu_s^2}{2 \gamma^2 \beta \nu_0^3}, & \text{if } \mu_{s,\min} < \mu_s < \mu_{s,\max}, \\ 0, & \text{otherwise,} \end{cases} \end{aligned} \quad (7.62)$$

where

$$\mu_{s,\min} = \max \left[-1, \frac{1 - (1 + \beta)\nu_0/\nu_s}{\beta} \right], \quad \mu_{s,\max} = \min \left[1, \frac{1 - (1 - \beta)\nu_0/\nu_s}{\beta} \right]. \quad (7.63)$$

When carrying out the integral over a δ -function, the following rule applies $\int f(x) \delta(g(x)) dx = f(x_0)/|g'(x_0)|$, where x_0 is the root of $g(x) = 0$ (if there are multiple roots, then we simply add up the contribution from each of them). Adopting this rule, we first obtain the root by solving $g(\mu) = 0 \Rightarrow 1 - \beta\mu_0 = (1 - \beta\mu_s)\nu_s/\nu_0$ and then $g'(\mu_0) = \nu_0^2 \beta / [\nu_s(1 - \beta\mu_s)]$.

Eq. (7.62) means that the emissivity at a given frequency ν_s is only non-zero along some viewing angles $\mu_{s,\min} < \mu_s < \mu_{s,\max}$ that is constrained by the requirements of $\nu_s = \nu_0 \mathcal{D}_s / \mathcal{D}$ as well as $-1 < \mu_s < 1$. The minimum ($\mu_{s,\min}$) and maximum ($\mu_{s,\max}$) viewing angles for different observing frequencies ν_s are shown in Fig. 7.7.

The next step is to calculate the angular average of the emissivity using eq. (7.60). From eq. (7.62), we know that the j_{ν_s} only depends on μ_s through the minimum and maximum

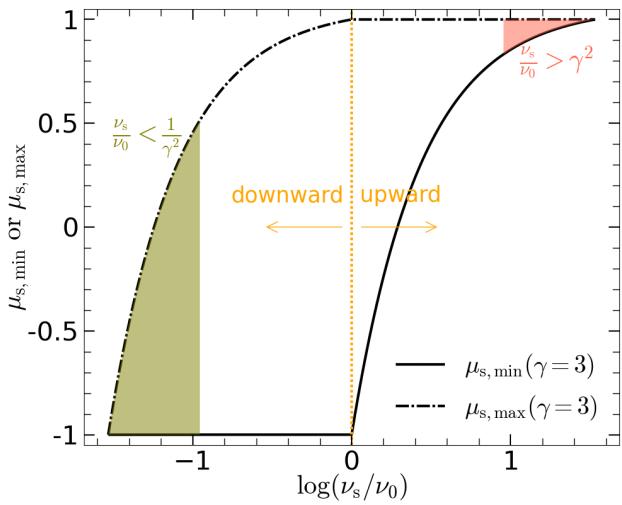


Figure 7.7: Minimum and maximum viewing angles $\mu_s = \cos \theta_s$ between the line of sight and electron's velocity, for different observing frequencies ν_s . The seed radiation field is assumed to be isotropic and monochromatic at frequency ν_0 . Upward (or Downward) Compton scatterings have $\nu_s/\nu_0 > 1$ (or < 1). The range of viewing angles for the highest (or lowest) frequency scattered photons with $\nu_s/\nu_0 > \gamma^2$ (or $< \gamma^{-2}$) are shown by red (or green) shaded regions.

viewing angles $\mu_{s,\min}$ and $\mu_{s,\max}$. Thus, we obtain

$$\bar{j}_{\nu_s} = \frac{1}{2} \int_{\mu_{s,\min}}^{\mu_{s,\max}} j_{\nu_s}(\mu_s) d\mu_s = \frac{n_e \sigma_T I_0 \nu_s^2}{4\gamma^2 \beta \nu_0^3} [\mu_{s,\max}(\nu_s) - \mu_{s,\min}(\nu_s)], \quad (7.64)$$

and, under the isotropic Thomson (isoT) scattering approximation, the final result is

$$\bar{j}_{\nu_s}^{\text{isoT}} = \frac{\nu_s}{\nu_0} \frac{1+\beta}{4\gamma^2 \beta^2} \frac{n_e \sigma_T I_0}{\nu_0} \begin{cases} \nu_s/\nu_0 - \Lambda^{-1}, & \text{if } \Lambda^{-1} < \nu_s/\nu_0 < 1, \\ 1 - \Lambda^{-1} \nu_s/\nu_0, & \text{if } 1 < \nu_s/\nu_0 < \Lambda, \\ 0, & \text{otherwise,} \end{cases} \quad (7.65)$$

where we have used the Compton boost factor $\Lambda \equiv (1+\beta)/(1-\beta)$. This can be written in a more compact, but less physically intuitive, way

$$\bar{j}_{\nu_s}^{\text{isoT}} = \frac{n_e \sigma_T I_0 \nu_s}{4\gamma^2 \beta^2 \nu_0^2} \left[\beta \left(1 + \frac{\nu_s}{\nu} \right) - \left| 1 - \frac{\nu_s}{\nu_0} \right| \right], \text{ for } \Lambda^{-1} < \frac{\nu_s}{\nu_0} < \Lambda. \quad (7.66)$$

The above spectrum is shown in Fig. 7.8, where we show $\nu_s \bar{j}_{\nu_s}$ in units of $n_e \sigma_T I_0$ as a function of observing frequency ν_s . It is straightforward to show that the average frequency of the scattered photons is given by

$$\frac{\langle \nu_s \rangle}{\nu_0} = \frac{\int \nu_s (j_{\nu_s}/\nu_s) d\nu_s}{\int (j_{\nu_s}/\nu_s) d\nu_s} = \frac{4}{3} \gamma^2 \beta^2 + 1, \quad (7.67)$$

where we have made use of the fact that the number of photons per frequency bin⁶ $dN_{\text{ph}}/d\nu_s$ is proportional to j_{ν_s}/ν_s . This shows that scattered photons, on average, receive a fractional energy gain of $(4/3)\gamma^2 \beta^2$, in agreement with eq. (7.50).

In the relativistic limit ($\gamma \gg 1$), the spectrum is a broken power-law with $j_{\nu_s} \propto \nu_s$ for up-scattered photons with $1 < \nu_s/\nu_0 < \Lambda$ (here $\Lambda \approx 4\gamma^2$) and $j_{\nu_s} \propto \nu_s^2$ for down-scattered photons in the frequency range of $\Lambda^{-1} < \nu_s/\nu_0 < 1$. This means that nearly all the energy of the scattered photons is contained near the maximum frequency $\approx \Lambda \nu_0$. Therefore, the down-scattered photons are often ignored in practical treatment of relativistic inverse-Compton scattering.

If we are only interested in the up-scattered photons in the frequency range of $1 < \nu_s/\nu_0 < \Lambda \approx 4\gamma^2$, then the average emissivity can be written as

$$\bar{j}_{\nu_s} \approx \frac{3n_e \sigma_T I_0}{\nu_0} x g(x), \quad g_{\text{iso}}(x) = \frac{2}{3}(1-x), \quad x \equiv \frac{\nu_s}{4\gamma^2 \nu_0}, \quad (\text{for } \gamma \gg 1) \quad (7.68)$$

⁶Note that the derivation of the IC spectrum in Rybicki & Lightman's book is based on photon number per frequency bin $dN_{\text{ph}}/d\nu_s$ instead of the emissivity j_{ν_s} , and these two approaches are equivalent.

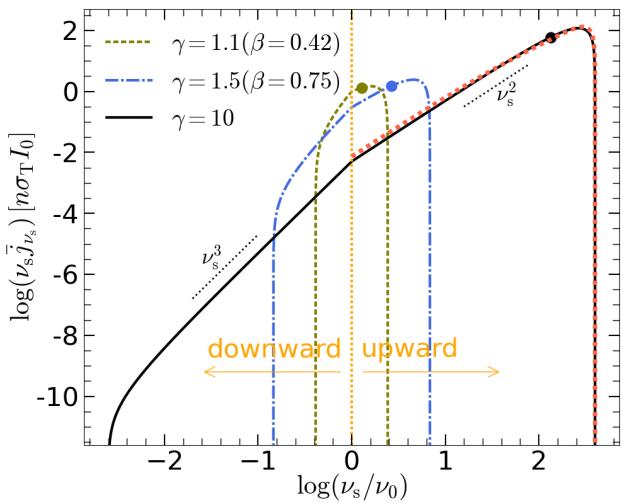


Figure 7.8: Angular averaged inverse-Compton spectral emissivity $\bar{\nu}_s \bar{j}_{\nu_s}$ (eq. 7.65) from a beam of electrons with Lorentz factor γ and lab-frame number density n_e interacting with an isotropic monochromatic radiation field, under the isotropic Thomson scattering approximation. The dot on each line shows the average frequency of the scattered photons (eq. 7.67). The thick red dotted line shows the exact result (eq. 7.72) which applies for $\nu_s/\nu_0 > 1$ in the limit $\gamma \gg 1$.

which clearly shows the scaling of $j_{\nu_s} \propto \nu_s$ at $x \ll 1$. The peak frequency is at where $[x(1-x)]' = 0 \Rightarrow x = 0.5$ or $\nu_{s,\text{peak}} = 2\gamma^2\nu_0$, and the emissivity at this peak frequency is given by

$$\bar{j}_{\nu_{s,\text{peak}}} = \frac{n_e \sigma_T I_0}{2\nu_0}. \quad (7.69)$$

The spectral power per electron is given by

$$\nu_{s,\text{peak}} P_{\nu_{s,\text{peak}}} = \frac{4\pi \nu_{s,\text{peak}} \bar{j}_{\nu_{s,\text{peak}}}}{n_e} = \gamma^2 U_{\text{rad}} \sigma_{\text{TC}} c, \quad (7.70)$$

where we have used the radiation energy density $U_{\text{rad}} = 4\pi I_0/c$. This is very close to the total power $P_{\text{IC}} = (4/3)\gamma^2 U_{\text{rad}} \sigma_{\text{TC}} c$, which means that nearly all the IC power is emitted near the peak frequency $\nu_{s,\text{peak}} = 2\gamma^2\nu_0$. Sometimes, for back-of-the-envelope calculations, the IC spectrum of a given electron may be roughly considered to be a δ -function at the peak frequency given by

$$P_{\nu_s} \simeq P_{\text{IC}} \delta(\nu_s - 2\gamma^2\nu_0), \quad (7.71)$$

where the delta-function is normalized by $\int_{-\infty}^{\infty} \delta(x) dx = 1$.

When the angular dependence of Thomson scattering $d\sigma_T/d\Omega'_s \propto (1 + \cos^2 \theta'_s)$ is included, [Blumenthal and Gould \(1970\)](#) calculated the exact functional form of $g(x)$ to be

$$g_T(x) = 1 + 2x \ln x + x - 2x^2, \quad x = \frac{\nu_s}{4\gamma^2\nu_0}, \quad (7.72)$$

which applies for $0 \approx (4\gamma^2)^{-1} < x < 1$ and $\gamma \gg 1$. The peak is at where $[xg(x)]' = 0 \rightarrow x \simeq 0.6$. The emissivity using the exact $g(x)$ is shown as a red dotted line in Fig. 7.8 for the $\gamma = 10$ case, and we see that the results under the isotropic Thomson scattering approximation is pretty accurate.

Finally, if the radiation field has a broad spectrum (instead of a δ -function), then the angle-averaged IC emissivity can be easily obtained by integrating over the spectrum of seed photons

$$\bar{j}_{\nu_s}(\gamma) \approx 3n \sigma_T \int_0^\infty d\nu \frac{I_\nu}{\nu} x g(x), \quad x = \frac{\nu_s}{4\gamma^2\nu}, \quad (7.73)$$

with the only caveat that for sufficiently high seed photon energies, the assumption of Thomson scattering breaks down and the cross-section is Klein-Nishina suppressed (then a 0th order approximation would be to ignore the IC scattering of those high energy seed photons). Suppose the seed spectrum is sufficiently sharply peaked at frequency ν_p , then most of the IC power will be near frequency $\nu_s \sim 2\gamma^2\nu_p$.

[Blumenthal and Gould \(1970\)](#) also considered the general case with Klein-Nishina differential cross-section and provided the following result for the function g in the relativistic

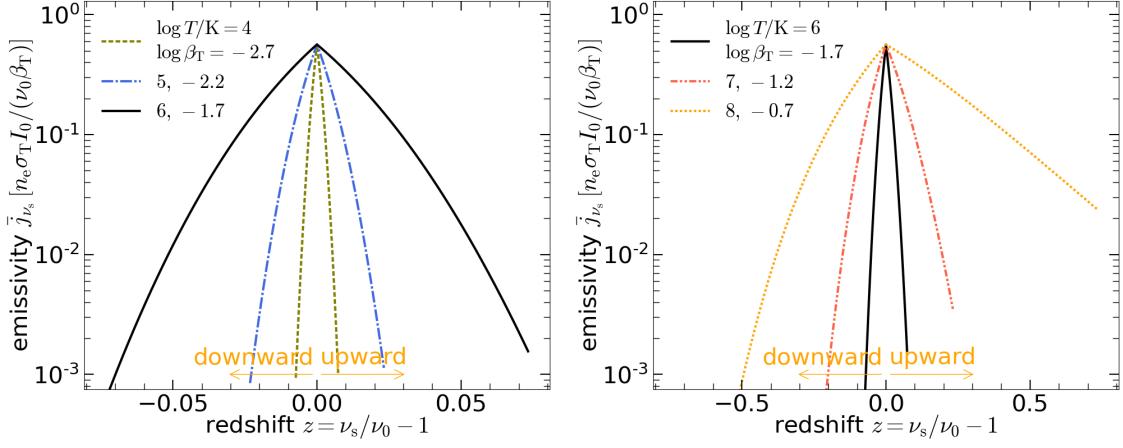


Figure 7.9: Angular averaged inverse-Compton emissivity \bar{j}_{ν_s} (eq. 7.77) from non-relativistic thermal electrons with temperature T interacting with an isotropic monochromatic radiation field, under the isotropic Thomson scattering approximation. For each T , the range of blueshifts shown is $|z| \leq 4\beta_T$. The asymmetry on the wings is caused by the slight preference of head-on collisions as compared to rear-end collisions.

limit $\gamma \gg 1$,

$$g_{\text{KN}}(x, q_s) = 1 + 2\tilde{x} \ln \tilde{x} + \tilde{x} - 2\tilde{x}^2 + \frac{q_s^2(1 - \tilde{x})}{2(1 - q_s)}, \quad \tilde{x} \equiv \frac{x}{1 - q_s}, \quad q_s \equiv \frac{h\nu_s}{\gamma m_e c^2}, \quad x = \frac{\nu_s}{4\gamma^2 \nu_0}, \quad (7.74)$$

where the factor q_s determines the regime we are in: $q_s \ll 1$ (and hence $\tilde{x} \approx x$) recovers the Thomson regime (eq. 7.72); and $q_s \lesssim 1$ gives Klein-Nishina regime. The above functional form of $g_{\text{KN}}(x, q_s)$ can be plugged into eq. (7.73) to calculate the IC emissivity, but under the following two constraints: $(4\gamma^2)^{-1} < x < 1$ (required by the Doppler boosts) and $q_s < 1$ (such that the energy of the scattered photon is always less than $\gamma m_e c^2$).

7.3.4 Thermal and power-law electron distributions

Let us then consider that electrons have a given isotropic momentum distribution $f_e(p)$, which is normalized such that $\int_0^\infty f_e(p) dp = 1$. To obtain the emissivity, we need to replace the electron number density n_e in eq. (7.66) by the electron distribution function $n_e \rightarrow n_e \int f_e(p) dp$. Let us consider two limiting cases in the following.

The first case is a non-relativistic ($\gamma \approx 1$) thermal electron distribution given by

$$f_e(\beta) d\beta = \frac{4}{\sqrt{\pi}} \frac{\beta^2}{\beta_T^3} e^{-\beta^2/\beta_T^2} d\beta, \quad \beta_T = \frac{\sigma_v}{c} = \sqrt{\frac{2k_B T}{m_e c^2}}, \quad (7.75)$$

where T is the electron temperature and β_T is the dimensionless velocity dispersion. Defining a *blueshift* factor⁷

$$z \equiv (\nu_s - \nu_0)/\nu_0, \quad (7.76)$$

we write the angle-averaged emissivity under the isotropic Thomson scattering

$$\begin{aligned} \bar{j}_{\nu_s} &= \frac{n_e \sigma_T I_0 (1+z)}{4\nu_0} \int_{\beta_{\min}}^{\infty} d\beta \beta^{-2} f_e(\beta) [\beta(2+z) - |z|] \\ &= \frac{n_e \sigma_T I_0}{\nu_0 \beta_T} (1+z) \left[\frac{1+z/2}{\sqrt{\pi}} e^{-\beta_{\min}^2/\beta_T^2} - \frac{|z|}{2\beta_T} \operatorname{erfc}(\beta_{\min}/\beta_T) \right], \end{aligned} \quad (7.77)$$

where $\operatorname{erfc}(x) \equiv 1 - \operatorname{erf}(x) = (2/\sqrt{\pi}) \int_x^{\infty} e^{-t^2} dt$ is the complementary error function and the minimum velocity β_{\min} for a given blueshift z is set by $\Lambda^{-1} < \nu_s/\nu_0 < \Lambda$,

$$\beta_{\min}(z) = \frac{|z|}{2+z}. \quad (7.78)$$

The resulting spectrum is shown in Fig. 7.9 for a single scattering. If the seed photons are from a narrow spectral line, then \bar{j}_{ν_s} would give a broadened line profile after a single scattering — the profile from a single scattering is neither a Lorentzian or Gaussian. The amount of broadening due to a single scattering can be characterized by the mean-squared blueshift per scattered photon

$$\langle z^2 \rangle = \frac{\int dz z^2 \bar{j}_{\nu_s}/(1+z)}{\int dz \bar{j}_{\nu_s}/(1+z)} = \beta_T^2 = \frac{2k_B T}{m_e c^2}, \quad (7.79)$$

where we have used $\bar{j}_{\nu_s}/(1+z)$ for the frequency distribution in *photon count* because \bar{j}_{ν_s} describes the frequency distribution in *energy*, although the factor of $(1+z)$ makes a negligible practical difference in the non-relativistic limit. We see that the root-mean-squared blueshift per scattering is simply given by the velocity dispersion $\beta_T = \sigma_v/c = \sqrt{2k_B T/m_e c^2}$. This means that, after $N \gg 1$ scatterings, the cumulative root-mean-squared blueshift will be $\sqrt{N}\beta_T = \sqrt{2Nk_B T/m_e c^2}$. In each scattering, due to the tendency of upward scattering, the profile is slightly asymmetric with a mean blueshift of

$$\langle z \rangle = \frac{\int dz z \bar{j}_{\nu_s}/(1+z)}{\int dz \bar{j}_{\nu_s}/(1+z)} = 2\beta_T^2 = \frac{4k_B T}{m_e c^2}, \quad (\text{ignoring recoil}) \quad (7.80)$$

which gives rise to a positive mean frequency drift in each scattering (in agreement with eq. 7.54). If the effects of electron recoil (eq. 7.30) is included, the mean blueshift is given by

$$\langle z \rangle = \frac{4k_B T}{m_e c^2} - \frac{h\nu_0}{m_e c^2}. \quad (\text{including recoil}) \quad (7.81)$$

⁷According to this definition, the wavelength-based, conventional redshift is $-z = \lambda/\lambda_0$. Apologies to the readers for the potential confusion — it might create more confusion if we were to directly use the conventional redshift.

The results of $\langle z \rangle$ and $\langle z^2 \rangle$ were first obtained by [Dirac \(1925\)](#).

We note that, in the situation of a large number of repeated scatterings, the final frequency distribution can be described by a random walk which depends on the first and second moments⁸, $\langle z \rangle$ (causing drift) and $\langle z^2 \rangle$ (causing diffusion), of the single scattering kernel in eq. (7.77). For instance, for a scattering slab of optical depth $\tau_s \gg 1$, a photon undergoes an average number of $N \sim \tau_s^2$ scatterings, during which the cumulative frequency drift of the line center is $\Delta\nu_c/\nu_0 = 2N\beta_T^2 \sim \tau_s^2\beta_T^2$ and the cumulative frequency diffusion is $\sigma_\nu/\nu_0 = \sqrt{N} \sim \tau_s\beta_T$. The drift term only dominates over the the diffusion term when $\tau_s\beta_T \gtrsim 1$ or when the Compton- y parameter exceeds unity (see §7.4.1). In such a case the spectral line is completely smeared into a continuum spectrum anyway — a full numerical treatment is needed.

It should be pointed out that eq. (7.77) based on “isotropic Thomson scattering” is reasonably accurate for practical purposes (see e.g., Fig. 2 of [Rybicki & Hummer 1994](#)). The precise analytic result based on the (anisotropic) Thomson differential cross-section is given by [Hummer & Mihalas \(1967\)](#) who ignored the small frequency drift. [Sazonov & Sunyaev \(2000\)](#) provided expressions with more high-order terms retained.

The second case is a relativistic ($\gamma \gg 1$) power-law electron distribution given by

$$f_e(\gamma)d\gamma = \frac{p-1}{\gamma_{\min}}(\gamma/\gamma_{\min})^{-p}d\gamma, \quad \gamma_{\min} < \gamma < \gamma_{\max}, \quad (7.82)$$

which has been normalized such that $\int_{\gamma_{\min}}^{\gamma_{\max}} f_e(\gamma)d\gamma = 1$, under the conditions of $\gamma_{\max} \gg \gamma_{\min}$ and the power-law index $p > 1$. Here the maximum Lorentz factor may be understood as the physical cutoff in the electron distribution or the Lorentz factor $\gamma_{\max} \sim m_e c^2 / (h\nu_0)$ above which the scattering cross-section is strongly Klein-Nishina suppressed⁹. The emissivity at a given frequency ν_s (here we only consider the up-scattered photons¹⁰ at $\nu_s > \nu_0$) only has contribution from electrons with Lorentz factors greater than γ_1 at which $\nu_s = 4\gamma_1^2\nu_0$, which means that

$$\gamma_1(\nu_s) = \max \left(\gamma_{\min}, \sqrt{\nu_s/(4\nu_0)} \right). \quad (7.83)$$

⁸The 0th moment of the frequency distribution follows the conservation of photon numbers, i.e., $\int dz \bar{j}_{\nu_s}/(1+z) = n_e \sigma_T I_0/\nu_0$.

⁹The KN suppression leads to a cutoff in the IC spectrum above energy $h\nu_{s,\max} \simeq m_e^2 c^4 / (4h\nu_0)$. At first sight, this might be a bit surprising since it says that a lower seed photon energy corresponds to a higher maximum energy for the scattered photons. You need to convince yourself by going to the comoving frame of the electron and comparing the seed photon energy $\gamma_{\max} h\nu_0$ with $0.5m_e c^2$ (because $\sigma_{\text{KN}}/\sigma_T \sim 0.5$ at this photon energy). The consequence is that the very highest energy photons from a non-thermal source may come from IC scattering of the CMB photons by electrons with extremely high Lorentz factors — the mean photon energy of the CMB $\langle h\nu_0 \rangle \simeq 3k_B T_{\text{CMB}} \simeq 7 \times 10^{-4}$ eV, which corresponds to $h\nu_{s,\max} \simeq 100$ TeV. We expect the IC spectrum above 100 TeV to soften significantly (with a steeper spectral slope or smaller $d\ln F_\nu/d\ln \nu$).

¹⁰The down-scattered photons are unimportant because the flux is usually fainter than that of the seed photons, but if you are still curious, the spectrum is given by $\bar{j}_{\nu_s} \propto \nu_s^2$ for $\nu_0 \gamma_{\min}^{-2} \lesssim \nu_s \lesssim \nu_0$ (see Fig. 7.8).

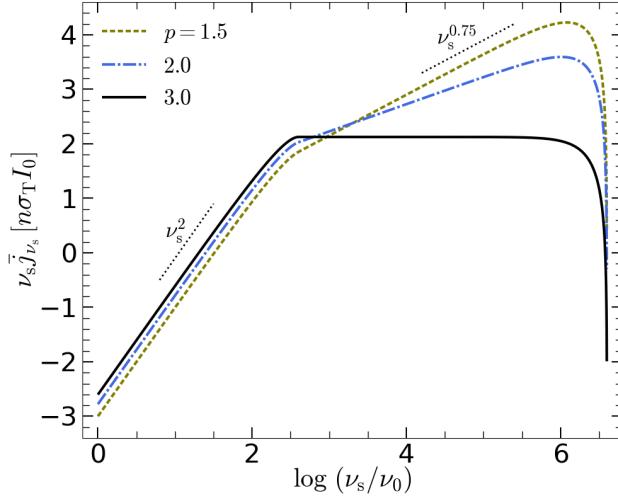


Figure 7.10: Angular averaged inverse-Compton spectral emissivity $\nu_s \bar{j}_{\nu_s}$ (eq. 7.84) from relativistic electrons with a power-law Lorentz factor distribution $f(\gamma)d\gamma \propto \gamma^{-p}d\gamma$ for $\gamma_{\min} < \gamma < \gamma_{\max}$. Here, we fix $\gamma_{\min} = 10$ and $\gamma_{\max} = 10^3$, and the results for three different electron power-law indices $p = 1.5, 2, 3$ are shown.

We take the IC spectrum kernel function $g_{\text{iso}}(x)$ based on the isotropic Thomson scattering approximation, and then the angular averaged emissivity is given by

$$\begin{aligned}\bar{j}_{\nu_s} &= \frac{3n_e \sigma_T I_0}{\nu_0} \int_{\gamma_1(\nu_s)}^{\gamma_{\max}} d\gamma f_e(\gamma) x g_{\text{iso}}(x) \\ &= \frac{n_e \sigma_T I_0}{2\nu_0} \frac{\nu_s}{\nu_0} \frac{p-1}{\gamma_{\min}^{1-p}} \int_{\gamma_1(\nu_s)}^{\gamma_{\max}} \frac{d\gamma}{\gamma^{p+2}} \left(1 - \frac{\nu_s/(4\nu_0)}{\gamma^2}\right) \\ &= \frac{n_e \sigma_T I_0}{2\nu_0} \frac{\nu_s}{\nu_0} \frac{p-1}{\gamma_{\min}^{1-p}} \left[\frac{\gamma_1^{-p-1} - \gamma_{\max}^{-p-1}}{p+1} - \frac{\nu_s}{4\nu_0} \frac{\gamma_1^{-p-3} - \gamma_{\max}^{-p-3}}{p+3} \right].\end{aligned}\quad (7.84)$$

This is shown in Fig. 7.10.

In the limit $\nu_s \ll 4\gamma_{\max}^2\nu_0$ (much below the maximum frequency of IC scattered photons), we can ignore the terms involving γ_{\max} and hence the above result can be simplified into the following broken power-laws

$$\bar{j}_{\nu_s} \approx \frac{4n_e \sigma_T I_0}{\nu_0} \frac{p-1}{(p+1)(p+3)} \min \left\{ x(\gamma_{\min}), [x(\gamma_{\min})]^{-\frac{p-1}{2}} \right\}, \quad x(\gamma_{\min}) = \frac{\nu_s}{4\gamma_{\min}^2\nu_0}. \quad (7.85)$$

Therefore, the IC spectrum from power-law electron distribution is $F_{\nu_s} \propto \nu_s^{-(p-1)/2}$ at high frequencies. This can be understood by the following simple argument. The number

of electrons at near Lorentz factor γ is $\sim n_e \gamma f_e(\gamma)$ and they scatter the seed photons to frequencies of the order $\nu_s \sim \gamma^2 \nu_0$, and this leads to a photon number spectrum $dN_{\text{ph}}/d\nu_s \sim N_{\text{ph}}/\nu_s \propto n_e \gamma f_e(\gamma)/(\gamma^2 \nu_0) \propto \gamma^{-p-1} \propto \nu_s^{-(p+1)/2}$, or $\bar{j}_{\nu_s} \propto \nu_s dN_{\text{ph}}/d\nu_s \propto \nu_s^{-(p-1)/2}$.

Although the discussion above is only for monochromatic seed photons, it is straightforward to convolve our results with an arbitrary seed spectrum — one simply take the following replacements $I_0 \rightarrow I_\nu d\nu$ and $\nu_0 \rightarrow \nu$ and then integrate over the seed spectrum $\int d\nu \bar{j}_{\nu_s}(\nu)$ (an example is given in eq. 7.73).

7.4 Repeated scatterings and Comptonization

Compton interactions between radiation field and electrons in a hot ionized gas can lead to significant change in the radiation spectrum — this process is called *Comptonization*. This can be captured by solving the radiative transfer equation, including attenuation of the intensity near a given direction due to scattering away from the beam and emissivity due to scattering from other directions into the given beam. Another more commonly adopted method is to solve the Boltzmann transport equation for the photon distribution function by tracking the flow of photons in the phase space. Of course, these two approaches are equivalent to each other because, in the end, the radiative transfer equation is derived from the Boltzmann transport equation. In this section, we discuss several important physical situations of Comptonization and their consequences.

7.4.1 Compton- y parameter

For a given soft photon of energy ϵ (in the limit $\gamma\epsilon \ll m_e c^2$, ignoring recoil effects), the average fractional energy change per scattering is $\langle \Delta\epsilon \rangle / \epsilon = (4/3) \langle \gamma^2 \beta^2 \rangle$, where $\langle \gamma^2 \beta^2 \rangle$ means averaging over the electron distribution function. To propagate through a gas cloud of scattering optical depth τ_s , a photon undergoes an average number of scatterings of $N_{\text{sca}} \simeq \tau_s + \tau_s^2$, where the τ_s^2 behavior (which applies when $\tau_s \gg 1$) is due to random walk of the photon. Therefore, we define the Compton y parameter as follows

$$y = \frac{4}{3} \langle \gamma^2 \beta^2 \rangle \times (\tau_s + \tau_s^2). \quad (7.86)$$

For a non-relativistic Maxwellian distribution, it is easy to show that $4 \langle \gamma^2 \beta^2 \rangle / 3 = 4k_B T / (m_e c^2)$. For an ultra-relativistic Maxwell-Jüttner distribution $f_e(\gamma) d\gamma \propto \gamma^2 e^{-\gamma m_e c^2 / (k_B T)} d\gamma$, then the mean squared Lorentz factor is given by

$$\langle \gamma^2 \rangle = \frac{\int f_e(\gamma) \gamma^2 d\gamma}{\int f_e(\gamma) d\gamma} = \frac{\Gamma(5)}{\Gamma(3)} \left(\frac{k_B T}{m_e c^2} \right)^2 = 12 \left(\frac{k_B T}{m_e c^2} \right)^2, \quad (7.87)$$

where $\Gamma(x)$ is the Gamma function. This then gives $4 \langle \gamma^2 \beta^2 \rangle / 3 = 16 (k_B T / m_e c^2)^2$. Note that $\langle \gamma^2 \rangle$ is different from $\langle \gamma \rangle^2$, because the mean Lorentz factor is given by

$$\langle \gamma \rangle = \frac{\int f_e(\gamma) \gamma d\gamma}{\int f_e(\gamma) d\gamma} = \frac{\Gamma(4)}{\Gamma(3)} \frac{k_B T}{m_e c^2} = \frac{3k_B T}{m_e c^2}. \quad (7.88)$$

Combining these two regimes, we write the Compton y parameter for thermal electron distribution at an arbitrary temperature

$$y = \frac{4k_B T}{m_e c^2} \left(1 + \frac{4k_B T}{m_e c^2} \right) \times (\tau_s + \tau_s^2). \quad (7.89)$$

If $y \gtrsim 1$, we know that the average photon energy will be changed significantly. This change may occur in the following two ways. (1) If $\tau_s \ll 1$ (in this case $y \gtrsim 1$ is only possible for a relativistic electron population with $\langle \gamma^2 \rangle \gg 1$), then only a small fraction τ_s of the photons are scattered, but each scattered photon gains energy by a factor of the order $\langle \gamma^2 \rangle \gg 1$. In this case, the scattered photons are no longer in the same frequency band as the seed photons. (2) If $\tau_s \gg 1$ (in this case $y \gtrsim 1$ is possible for non-relativistic electron population), then all photons undergo a large number of scatterings, each of which may only cause a small fractional energy gain.

A group of seed photons of an average initial energy of $\langle \epsilon \rangle_i$ would emerge from the cloud with an average final energy of

$$\langle \epsilon \rangle_f \simeq \min [\langle \epsilon \rangle_i e^y, (\langle \gamma \rangle - 1)m_e c^2], \quad (7.90)$$

where the first term captures the exponential growth of the photon energy due to the energy gain in each scattering and the second term is due to the saturation effect when the electron recoil becomes important. The above result only gives the average photon energy, and one must solve the Boltzmann transport equation to obtain the full spectrum from Comptonization — the non-relativistic version is the *Kompaneets equation* (see §7.4.5).

7.4.2 Optically thin thermal electrons

Ignoring recoil effects, the average energy amplification factor per scattering for each photon is given by

$$A = \frac{\langle \epsilon_s \rangle}{\epsilon} = 1 + 4k_B T / m_e c^2 (1 + 4k_B T / m_e c^2), \quad (7.91)$$

where we have combined the non-relativistic and relativistic regimes for electrons' kinetic temperature.

For a slab with modest scattering optical depth $\tau < 1$, the probability of having k scatterings is τ^k . The photons that have undergone k scatterings reaches an average energy

of $\langle \epsilon_k \rangle = A^k \epsilon_0$, where ϵ_0 is the energy of the seed photons. This leads to a power-law spectrum at $\epsilon_s \gg \epsilon_0$

$$\frac{dN}{d\epsilon_s} \propto \frac{\tau^k}{\epsilon_k} \propto \epsilon_k^{\log \tau / \log A - 1}, \quad (7.92)$$

which can be obtained by letting $\tau^k = A^{k\alpha} = (\epsilon_k/\epsilon_0)^\alpha$ and taking the log (base-10) on both sides gives $k \log \tau = k\alpha \log A$ and hence $\alpha = \log \tau / \log A$ is independent of k . For instance, for $\tau = 0.3$ and $A = 1.5$ (mildly relativistic electrons), one expect a power-law spectrum of $dN/d\epsilon \propto \epsilon^{-2.7}$, i.e., the photon index in this case is 2.7.

It should be noted that the power-law created by upscatterings must cut off at photon energies corresponding to severe recoil losses,

$$\epsilon_{s,\max} \sim (\langle \gamma \rangle - 1)m_e c^2. \quad (7.93)$$

7.4.3 Non-relativistic Comptonization and Compton temperature

For non-relativistic Maxwellian electrons, if we include recoil effects, the average change in the photon energy per scattering is given by (following eq. 7.81)

$$\langle \Delta \epsilon \rangle = \frac{\epsilon}{m_e c^2} (4k_B T - \epsilon). \quad (\text{for non-relativistic thermal electrons}) \quad (7.94)$$

Let us consider an isotropic and uniform optically thick radiation-gas mixture with only Compton scattering. When the system reaches equilibrium, the following condition must be satisfied

$$4k_B T = \langle \epsilon^2 \rangle / \langle \epsilon \rangle, \quad (7.95)$$

where $\langle \epsilon \rangle$ is the mean photon energy and $\langle \epsilon^2 \rangle$ is the mean-squared photon energy for the radiation spectrum. Note that eq. (7.95) is a *necessary condition* for Compton equilibrium. It is convenient to define the Compton temperature¹¹ as

$$4k_B T_C \equiv \langle \epsilon^2 \rangle / \langle \epsilon \rangle, \quad (7.96)$$

and the necessary condition for Compton equilibrium can then be written as $T = T_C$.

¹¹Our definition for the Compton temperature is the zeroth-order result based on the Thomson differential cross-section in the limit of $\epsilon \ll m_e c^2$. Including the first-order corrections due to Klein-Nishina effects, Sazonov et al. (2004) provided a more accurate expression for the Compton temperature

$$T_C = \frac{m_e c^2 \int x^2 [1 - 21x/5] (dn_{ph}/dx) dx}{4k_B \int x [1 - 47x/8] (dn_{ph}/dx) dx},$$

where $x = \epsilon/(m_e c^2)$ is the dimensionless photon energy and dn_{ph}/dx describes the photon energy distribution. If a significant fraction of the radiation energy is carried by photons with $\epsilon \gtrsim 10 \text{ keV}$ (but $\epsilon \ll m_e c^2$), the zeroth-order result $4k_B T_C = \langle \epsilon^2 \rangle / \langle \epsilon \rangle$ has a fractional error of $\gtrsim 10\%$ and the Sazonov expression should be used.

This condition is indeed satisfied for the *Wien spectrum* (written in terms of the specific intensity I_ν or the photon occupation number¹² $\eta_\gamma \propto I_\nu/\nu^3$ of the radiation field)

$$I_\nu \propto \epsilon^3 e^{-\epsilon/k_B T} \quad \text{or} \quad \eta_\gamma \propto e^{-\epsilon/k_B T}, \quad (7.97)$$

because we can make use of the Gamma function to obtain the following mean and mean-squared photon energies

$$\langle \epsilon \rangle = 3k_B T, \quad \langle \epsilon^2 \rangle = 12k_B T. \quad (7.98)$$

Note that the condition in eq. (7.97) only tells us the shape of the radiation spectrum, whereas the overall normalization is determined by the number density of photons as given by photon number conservation (since scatterings do not create or destroy photons).

It is possible to show that the Wien spectrum is indeed the unique solution for Compton equilibrium (but without considering stimulated emission). Let us consider an isotropic radiation field. Here, we have a stochastic process that involves drift and diffusion of photons in 1D energy space, and such a problem can be described by the following Fokker-Planck equation (see eq. 2.218 in §2.5.3)

$$\partial_t n'_{\text{ph}}(\epsilon, t) = -\partial_\epsilon [V n'_{\text{ph}}(\epsilon, t)] + \partial_\epsilon^2 [D n'_{\text{ph}}(\epsilon, t)], \quad (7.99)$$

where $n'_{\text{ph}} \equiv dn_{\text{ph}}/d\epsilon \propto I_\nu/\nu \propto \eta_\gamma/\epsilon^2$ describes the energy spectrum of photons (n_{ph} being the photon number density and η_γ being the photon occupation number), the drift velocity V and diffusion coefficient D are given by

$$V = R_{\text{sca}} \langle \Delta\epsilon \rangle, \quad D = R_{\text{sca}} \langle \Delta\epsilon^2 \rangle / 2, \quad (7.100)$$

and $R_{\text{sca}} = n_e \sigma_{\text{Tc}}$ is the rate of photon-electron scatterings (as given by eq. 7.42).

Here, we need the mean energy change $\langle \Delta\epsilon \rangle$ and mean-squared energy change $\langle \Delta\epsilon^2 \rangle$ per scattering. The former is given by eq. (7.81) and the latter is related to the mean-squared blueshift (eq. 7.79) as follows

$$\langle \Delta\epsilon^2 \rangle = \langle z^2 \rangle \epsilon^2 = \frac{2k_B T}{m_e c^2} \epsilon^2. \quad (7.101)$$

With V and D in hand, the Fokker-Planck equation comes down to the following form

$$\partial_t n'_{\text{ph}} = \frac{R_{\text{sca}}}{m_e c^2} \partial_\epsilon [(\epsilon - 4k_B T)\epsilon n'_{\text{ph}} + k_B T \partial_\epsilon (\epsilon^2 n'_{\text{ph}})]. \quad (7.102)$$

¹²Reminding ourselves that the photon occupation number is related to the phase-space distribution function f , the specific intensity I_ν , and the brightness temperature T_b : $\eta_\gamma = f/(2h^3) = I_\nu c^2/(2h\nu^3) = k_B T_b/(h\nu)$.

It is often more convenient to work with the (dimensionless) photon occupation number $\eta_\gamma(\epsilon, t) \propto n'_{\text{ph}}(\epsilon, t)/\epsilon^2$, and this leads to the following simpler form

$$\partial_t \eta_\gamma = \frac{n_e \sigma_T c}{\epsilon^2 m_e c^2} \frac{\partial}{\partial \epsilon} \left[\epsilon^4 \left(k_B T \frac{\partial}{\partial \epsilon} \eta_\gamma + \eta_\gamma \right) \right]. \quad (7.103)$$

At Compton equilibrium, we require $\partial_t \eta_\gamma = 0$ and hence

$$k_B T \frac{d}{d\epsilon} \eta_\gamma + \eta_\gamma = 0, \quad (7.104)$$

which holds at all photon energies ϵ . This leads to $d \ln \eta_\gamma = -d(\epsilon/k_B T)$ and we obtain the Wien spectrum

$$\eta_\gamma \propto e^{-\epsilon/k_B T}, \quad (7.105)$$

where the normalization constant is determined by the total photon number (Compton equilibrium only constrains the shape of the spectrum).

We conclude that the Wien spectrum with a normalization that conserves the total photon number should be the equilibrium solution for a radiation-gas mixture with only Compton scattering. However, this is *inconsistent* with the fundamental result that the equilibrium solution should take the shape of the *thermal* or Bose-Einstein spectrum

$$\eta_\gamma = \frac{1}{e^{(\epsilon-\mu)/(k_B T)} - 1}, \quad (7.106)$$

where μ is the chemical potential. It turns out that the Compton equilibrium condition in eq. (7.95) does not take into account stimulated emission! Suppose $A(\mathbf{k} \rightarrow \mathbf{k}_s)$ (in units of sec^{-1}) denotes the rate at which photons are spontaneously¹³ scattered from the phase-space near wavevector \mathbf{k} to the phase-space near another wavevector \mathbf{k}_s . If the end state near wavevector \mathbf{k}_s has photon occupation number $\eta_\gamma(\mathbf{k}_s)$, then the total rate of $\mathbf{k} \rightarrow \mathbf{k}_s$ quantum transitions must be given by $[1 + \eta_\gamma(\mathbf{k}_s)]A(\mathbf{k} \rightarrow \mathbf{k}_s)$. We will show in §7.4.5 that the correct equilibrium solution is indeed given by the Bose-Einstein spectrum when the effect of stimulated emission is included. When $\mu/(k_B T)$ is a very large negative number such that the photon occupation number is very small (and hence stimulated emission is negligible), the Bose-Einstein spectrum reduces to the Wien spectrum.

7.4.4 *Stimulated Compton scattering for cold electrons

In this subsection, we considered the effects of stimulated emission in the Comptonization process but limit ourselves to the case of cold electrons with zero kinetic temperature

¹³In the classical picture where scattering is due to the re-emission by an electron driven into oscillation by the incoming EM wave, the scattered photon is produced from spontaneous emission by the oscillating electron. The Compton scattering we have been studying so far is in fact the *spontaneous scattering*.

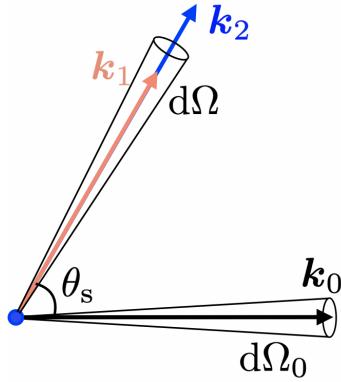


Figure 7.11: Geometry for stimulated Compton scattering. We consider the time evolution of the photon occupation number $\eta_\gamma(\mathbf{k}_0, t)$ near wavevector \mathbf{k}_0 due to Compton scattering outwards from the beam $\mathbf{k}_0 \rightarrow \mathbf{k}_1$ (or $d\Omega_0 \rightarrow d\Omega$) and inwards to the beam $\mathbf{k}_2 \rightarrow \mathbf{k}_0$ (or $d\Omega \rightarrow d\Omega_0$). The scattering angle is denoted as θ_s . Note that the wavenumber changes due to electron recoil, and we have $k_2/k_0 \approx k_0/k_1 \approx 1 + x(1 - \cos \theta_s)$.

$T = 0$. Stimulated emission associated with spontaneous Compton scattering is also called *induced Compton scattering*. We will begin with a generally anisotropic radiation field and then study the isotropic case. The radiation field is assumed to be spatially uniform and is described by the photon occupation number as a function of wavevector and time $\eta_\gamma(\mathbf{k}, t)$. Our goal is to study how Comptonization re-distributes photons in the momentum space (note that photon momentum is given by $\mathbf{p} = \hbar\mathbf{k}$).

Stimulated Compton scattering is important in the vicinity of sources that produce radiation with extremely high photon occupation numbers (e.g., pulsar radio emission and fast radio bursts). These are called “coherent sources” because a large number of photons are emitted into the same quantum state by different particles that undergo coherent motion.

The geometry of the problem is shown in Fig. 7.11. The time evolution of $\eta_\gamma(\mathbf{k}_0)$ due to *outward* spontaneous and stimulated transitions from the initial state \mathbf{k}_0 to other states \mathbf{k}_1 is given by

$$\partial_t \eta_\gamma(\mathbf{k}_0)|_{d\Omega_0 \rightarrow d\Omega} = -\eta_\gamma(\mathbf{k}_0) \sum_{\mathbf{k}_1} (1 + \eta_\gamma(\mathbf{k}_1)) A(\mathbf{k} \rightarrow \mathbf{k}_1), \quad (7.107)$$

where the sum goes over all possible final states \mathbf{k}_1 for outward transitions and $A(\mathbf{k} \rightarrow \mathbf{k}_1)$ (in units of sec^{-1}) is the Einstein-A coefficient. Hereafter, we keep the time dependence of $\eta_\gamma(\mathbf{k}, t)$ implicit to simplify the notation. The angle between \mathbf{k}_0 and \mathbf{k}_1 is denoted as θ_s .

For a given direction $\hat{\mathbf{k}}_1$, the wavenumber of the scattered photon is given by

$$k_1 = \frac{k_0}{1 + x(1 - \cos \theta_s)}, \quad x \equiv \hbar k_0 / m_e c. \quad (7.108)$$

Note that the fact that $k_1 < k_0$ is due to electron recoiling. At the same time, photons scattered from solid angle $d\Omega$ back into $d\Omega_0$ can also cause time evolution of $\eta_\gamma(\mathbf{k}_0)$. To reach the state \mathbf{k}_0 from the initial direction $\hat{\mathbf{k}}_2 = \hat{\mathbf{k}}_1$, these *inward* scattered photons must have an initial wavenumber

$$k_2 = \frac{k_0}{1 - x(1 - \cos \theta_s)}, \quad (7.109)$$

which means $k_2 > k_0$. In the following, we will take $x \ll 1$ such that the recoil is weak, but it is essential to the effects of induced Compton scattering we are interested in.

From the Klein-Nishina differential cross-section $d\sigma_{\text{KN}}/d\Omega$ (given by eq. 7.10)

$$\left. \frac{d\sigma_{\text{KN}}}{d\Omega} \right|_{\mathbf{k}_0 \rightarrow \mathbf{k}_1} = \frac{3\sigma_T}{16\pi} \left(\frac{k_1}{k_0} \right)^2 \left(\frac{k_1}{k_0} + \frac{k_0}{k_1} - \sin^2 \theta_s \right), \quad \theta_s \equiv \arccos(\mathbf{k}_0 \cdot \mathbf{k}_1 / k_0 k_1), \quad (7.110)$$

the spontaneous scattering rate (per photon) into the solid angle $d\Omega$ near the $\hat{\mathbf{k}}_1$ direction is given by

$$cn_e d\Omega \left. \frac{d\sigma_{\text{KN}}}{d\Omega} \right|_{\mathbf{k}_0 \rightarrow \mathbf{k}_1}, \quad (7.111)$$

which corresponds to the Einstein-A coefficient $A(\mathbf{k} \rightarrow \mathbf{k}_1)$. Thus, eq. (7.107) can be written as an integral

$$\partial_t \eta_\gamma(\mathbf{k}_0)|_{d\Omega_0 \rightarrow d\Omega} = -cn_e \eta_\gamma(\mathbf{k}_0) \int d\Omega (1 + \eta_\gamma(\mathbf{k}_1)) \left. \frac{d\sigma_{\text{KN}}}{d\Omega} \right|_{\mathbf{k}_0 \rightarrow \mathbf{k}_1}. \quad (7.112)$$

Multiplying both sides of the above equation by $d\Omega_0 k_0^2 dk_0$ (the differential volume in wavevector space), then we obtain the number of photons scattered from $d\Omega_0$ to $d\Omega$ per unit time, which causes the number of photons in the volume of $d\Omega k_1^2 dk_1$ (where dk_1/dk_0 is given by differentiating eq. 7.108) to increase at a rate

$$\begin{aligned} \partial_t \eta_\gamma(\mathbf{k}_1)|_{d\Omega_0 \rightarrow d\Omega} d\Omega k_1^2 dk_1 &= -\partial_t \eta_\gamma(\mathbf{k}_0)|_{d\Omega_0 \rightarrow d\Omega} d\Omega_0 k_0^2 dk_0 \\ &= (\eta_\gamma(\mathbf{k}_0) d\Omega_0 k_0^2 dk_0) cn_e d\Omega (1 + \eta_\gamma(\mathbf{k}_1)) \left. \frac{d\sigma_{\text{KN}}}{d\Omega} \right|_{\mathbf{k}_0 \rightarrow \mathbf{k}_1}. \end{aligned} \quad (7.113)$$

Following the same logic, we consider inward $d\Omega \rightarrow d\Omega_0$ (or $\mathbf{k}_2 \rightarrow \mathbf{k}_0$) scatterings, which cause the number of photons in the volume $d\Omega_0 k_0^2 dk_0$ to increase at a *differential* rate

$$\begin{aligned} \partial_t \eta_\gamma(\mathbf{k}_0)|_{d\Omega \rightarrow d\Omega_0} d\Omega_0 k_0^2 dk_0 \\ &= (\eta_\gamma(\mathbf{k}_2) d\Omega_0 k_0^2 dk_0) cn_e d\Omega_0 (1 + \eta_\gamma(\mathbf{k}_0)) \left. \frac{d\sigma_{\text{KN}}}{d\Omega} \right|_{\mathbf{k}_2 \rightarrow \mathbf{k}_0}, \end{aligned} \quad (7.114)$$

Note that the rate of stimulated emission is proportional to the photon occupation number at the final state in each transition, so we should use $1 + \eta_\gamma(\mathbf{k}_0)$ (instead of $1 + \eta_\gamma(\mathbf{k}_1)$). The *total* rate is given by an integral over $d\Omega$, i.e.

$$\partial_t \eta_\gamma(\mathbf{k}_0)|_{d\Omega \rightarrow d\Omega_0} = cn_e \int d\Omega \eta_\gamma(\mathbf{k}_2) \frac{k_2^2 dk_2}{k_0^2 dk_0} (1 + \eta_\gamma(\mathbf{k}_0)) \left. \frac{d\sigma_{\text{KN}}}{d\Omega} \right|_{\mathbf{k}_2 \rightarrow \mathbf{k}_0}. \quad (7.115)$$

Finally, we are ready to sum up the contributions from outward (eq. 7.112) and inward (eq. 7.115) transitions to obtain the net time evolution of $\eta_\gamma(\mathbf{k}_0)$,

$$\begin{aligned} \partial_t \eta_\gamma(\mathbf{k}_0) &= cn_e \int d\Omega \left[\eta_\gamma(\mathbf{k}_2)(1 + \eta_\gamma(\mathbf{k}_0)) \left. \frac{k_2^2 dk_2}{k_0^2 dk_0} \frac{d\sigma_{\text{KN}}}{d\Omega} \right|_{\mathbf{k}_2 \rightarrow \mathbf{k}_0} \right. \\ &\quad \left. - \eta_\gamma(\mathbf{k}_0)(1 + \eta_\gamma(\mathbf{k}_1)) \left. \frac{d\sigma_{\text{KN}}}{d\Omega} \right|_{\mathbf{k}_0 \rightarrow \mathbf{k}_1} \right]. \end{aligned} \quad (7.116)$$

This is the Master equation for Compton scattering by cold electrons including stimulated emission, and we have not made any approximations.

Hereafter, two approximations are used to simplify the system. First, since we are interested in the effect of induced Compton scattering, it is assumed that the photon occupation number is very large such that $1 + \eta_\gamma \approx \eta_\gamma$ (i.e., spontaneous scattering can be ignored). Second, we consider low-energy photons with $x = \hbar k_0/m_e c \ll 1$, so any 2nd-order small terms $O(x^2)$ can be ignored. Thus, the wavenumber shift due to electron recoiling can be written as

$$k_2/k_0 \approx k_0/k_1 \approx 1 + x(1 - \cos \theta_s), \quad (7.117)$$

and we also obtain

$$\frac{k_2^2 dk_2}{k_0^2 dk_0} = [1 - x(1 - \cos \theta_s)]^{-4} \approx [1 + 4x(1 - \cos \theta_s)]. \quad (7.118)$$

The Klein-Nishina differential cross-section is approximately given by

$$\left. \frac{d\sigma_{\text{KN}}}{d\Omega} \right|_{\mathbf{k}_0 \rightarrow \mathbf{k}_1} \approx \left. \frac{d\sigma_{\text{KN}}}{d\Omega} \right|_{\mathbf{k}_2 \rightarrow \mathbf{k}_0} \approx \frac{3\sigma_T}{16\pi} (1 + \cos^2 \theta_s) [1 - 2x(1 - \cos \theta_s)]. \quad (7.119)$$

We plug these approximations into eq. (7.116), take $1 + \eta_\gamma \approx \eta_\gamma$, and then obtain

$$\begin{aligned} \partial_t \eta_\gamma(\mathbf{k}_0) &\approx \frac{3n_e \sigma_T c}{16\pi} \frac{\eta_\gamma(\mathbf{k}_0)}{k_0^2} \int d\Omega (1 + \cos^2 \theta_s) [\eta_\gamma(\mathbf{k}_2) k_2^2 - \eta_\gamma(\mathbf{k}_1) k_1^2] \\ &\approx \frac{3\hbar n_e \sigma_T}{8\pi m_e} \eta_\gamma(\mathbf{k}_0) \int d\Omega (1 + \cos^2 \theta_s) (1 - \cos \theta_s) \left. \frac{\partial(\eta_\gamma(\mathbf{k}) k^2)}{\partial k} \right|_{\mathbf{k}=\mathbf{k}_0 \hat{\Omega}}, \end{aligned} \quad (7.120)$$

where the partial derivative $\partial_k(\eta_\gamma(\mathbf{k}) k^2)$ should be evaluated for the photon occupation number near wavenumber $k = k_0$ along the direction of each differential solid angle $\hat{\Omega}$ and

then we sum up the contributions from all the differential solid angles $d\Omega$. Eq. (7.120) is the kinetic equation for induced Compton scattering, which applies to the general case of an anisotropic radiation field in the limits of $\eta_\gamma \gg 1$ and $x = \hbar k/m_e c \ll 1$.

Let us then discuss the special case of an isotropic radiation field. In the case, the angular integral is given by $\int(1 + \cos^2 \theta_s)(1 - \cos \theta_s)d\Omega = 16\pi/3$, so the time evolution for the photon occupation number near arbitrary photon energy $\epsilon = \hbar k c$ is given by

$$\partial_t \eta_\gamma = \frac{2n_e \sigma_T c}{m_e c^2} \eta_\gamma \frac{\partial}{\partial \epsilon} (\epsilon^2 \eta_\gamma) = \frac{n_e \sigma_T c}{\epsilon^2 m_e c^2} \frac{\partial}{\partial \epsilon} (\epsilon^4 \eta_\gamma^2), \quad \text{for } \eta_\gamma \gg 1. \quad (7.121)$$

Stimulated Compton scattering changes the photon occupation number on a timescale

$$t_{\text{stm}} = \frac{\eta_\gamma}{|\partial_t \eta_\gamma|} \sim \frac{m_e c^2 / \epsilon}{\eta_\gamma R_{\text{sca}}} \sim \frac{m_e c^2}{k_B T_b} \frac{1}{R_{\text{sca}}}, \quad (7.122)$$

where $R_{\text{sca}} = n_e \sigma_T c$ is the spontaneous scattering rate and we have expressed the photon occupation number using the brightness temperature $k_B T_b = \eta_\gamma / \epsilon$. We see that, if the brightness temperature of the radiation field greatly exceeds $m_e c^2 / k_B \simeq 6 \times 10^9$ K, then the radiation field will quickly be modified by stimulated Compton scattering on a timescale t_{stm} that is much shorter than R_{sca}^{-1} (classically, one would naïvely expect that most photons have not even undergone one scattering yet!). The radiation spectrum tends to evolve towards a power-law steady-state solution $\eta_\gamma \propto \epsilon^{-2}$ or $I_\nu \propto \nu$. For a scattering medium of size L , we can define the effective optical depth for stimulated Compton scattering as

$$\tau_{\text{stm}} = \frac{L}{c t_{\text{stm}}} \sim \frac{k_B T_b}{m_e c^2} \tau_T, \quad (7.123)$$

where $\tau_T = n_e \sigma_T L$ is the Thomson optical depth of the medium. We see that the effective optical depth would greatly exceed τ_T at very high brightness temperatures.

On the other hand, if the photon occupation number is not much greater than unity, then one cannot ignore spontaneous scattering. We go back to the Master equation (7.116) and include all the $1 + \eta_\gamma$ factors. For an isotropic radiation field, we obtain the following equation

$$\partial_t \eta_\gamma = \frac{n_e \sigma_T c}{\epsilon^2 m_e c^2} \frac{\partial}{\partial \epsilon} [\epsilon^4 (1 + \eta_\gamma) \eta_\gamma], \quad \text{for any } \eta_\gamma, \quad (7.124)$$

which is the kinetic equation for Compton scattering in the limit of cold electrons $T \rightarrow 0$.

7.4.5 Kompaneets equation and Compton equilibrium

In the previous two subsections, we have described how the radiation field is modified due to Compton scattering by non-relativistic warm electrons in the absence of stimulated

emission (eq. 7.103) and by cold electrons including the effect of stimulated emission (eq. 7.124). Combining these two results, we arrive at the famous *Kompaneets equation*

$$\partial_t \eta_\gamma = \frac{n_e \sigma_T c}{\epsilon^2 m_e c^2} \frac{\partial}{\partial \epsilon} \left\{ \epsilon^4 \left[k_B T \frac{\partial}{\partial \epsilon} \eta_\gamma + \eta_\gamma + \eta_\gamma^2 \right] \right\}, \quad (7.125)$$

The above kinetic equation applies under the following conditions:

- (i) isotropic radiation field (so photons undergo drift and diffusion in 1D energy space),
- (ii) smooth radiation spectrum (as required by the Taylor expansion used in the Fokker-Planck equation),
- (iii) non-relativistic electron temperature $k_B T \ll m_e c^2$ (for weak Doppler broadening),
- (iv) low photon energies $\epsilon \ll m_e c^2$ (for weak electron recoiling), and
- (v) photon number conservation.

The system would become more complicated if one of the conditions (i)–(iv) is relaxed — usually Monte-Carlo simulations are required to obtain the time evolution of the system. As for condition (v), one can include source and sink terms on the RHS of the equation to account for photon emission and absorption.

The steady-state solution of the Kompaneets equation can be obtained from the requirement of $\partial_t \eta_\gamma(\epsilon, t) = 0$ at all photon energies ϵ , which gives the following result

$$k_B T \frac{d}{d\epsilon} \eta_\gamma + (\eta_\gamma + 1) \eta_\gamma = 0. \quad (7.126)$$

By changing the variables $x = \epsilon/k_B T$ and $y = (\eta_\gamma + 1)/\eta_\gamma$, we find $\eta_\gamma = 1/(y - 1)$ and hence $d\eta_\gamma = -dy/(y - 1)^2$ and hence

$$\frac{dy}{dx} (y - 1)^{-2} = \frac{y}{(y - 1)^2} \Rightarrow \frac{dy}{y} = dx \Rightarrow y(x) = C e^x, \quad (7.127)$$

where C is a integration constant (and we require $C > 1$ as $y > 1$ for all x). We then define a *negative* chemical potential $\mu = -k_B T \ln C$ as the integration constant, we then write $y = e^x e^{-\mu/k_B T} = e^{(\epsilon-\mu)/k_B T}$. Thus, the photon occupation number is given by

$$\eta_\gamma(\epsilon) = \frac{1}{y(x) - 1} = \frac{1}{e^{(\epsilon-\mu)/k_B T} - 1}. \quad (7.128)$$

This is the solution of *Compton equilibrium* and the spectrum is indeed given by the Bose-Einstein distribution (eq. 7.106) with a negative chemical potential.

In the limit of $|\mu|/k_B T \gg 1$, we can ignore the “ -1 ” term in the denominator of the Bose-Einstein distribution and then obtain the *Wien spectrum*

$$\eta_\gamma(\epsilon) \approx e^{\mu/k_B T} e^{-\epsilon/k_B T} \propto e^{-\epsilon/k_B T}. \quad (7.129)$$

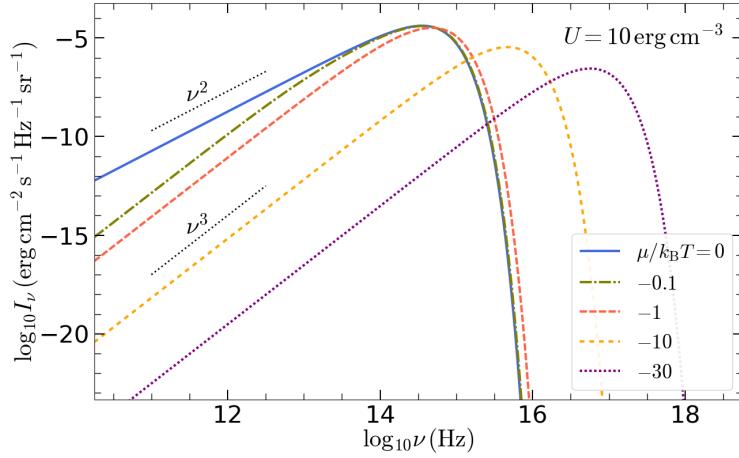


Figure 7.12: The specific intensity spectra of Bose-Einstein distributions with different chemical potentials $\mu/k_B T \leq 0$ but at the same radiation energy density $U = 10 \text{ erg cm}^{-3}$ (similar to that near the solar photosphere). The blue solid line shows the Planck spectrum with $\mu = 0$ and photon occupation number $\eta_\gamma(\epsilon) = 1/(e^{\epsilon/k_B T} - 1)$, whereas the solutions with $\mu/k_B T \lesssim -1$ asymptotes to the Wien spectrum with photon occupation number $\eta_\gamma(\epsilon) \approx e^{\mu/k_B T} e^{-\epsilon/k_B T}$.

In the Wien spectrum limit ($|\mu|/k_B T \gg 1$), the radiation energy density is given by

$$U \approx 0.924 aT^4 e^{\mu/k_B T} \ll aT^4, \quad (7.130)$$

and the photon number density is given by

$$n_{\text{ph}} \approx 0.924 \frac{aT^4}{3k_B T} e^{\mu/k_B T} \ll \frac{aT^4}{3k_B T}, \quad (7.131)$$

where a is the radiation density constant and the factor of $0.924 = 6/6.494$ comes from §2.1.4. The average photon energy is given by the ratio between the two

$$\langle \epsilon \rangle_W = U/n_{\text{ph}} = 3k_B T, \quad (7.132)$$

which is only slightly different from the blackbody case of $\langle \epsilon \rangle_{\text{BB}} \approx 2.70k_B T$ (see eq. 2.52). We also note that the peak frequency ν_{pk} of the intensity spectrum $I_\nu^{\text{Wien}} \propto \nu^3 e^{-h\nu/k_B T}$ is also located at $h\nu_{\text{pk}} = 3k_B T$, identical to the average photon energy. The small difference between $\langle \epsilon \rangle_W$ and $\langle \epsilon \rangle_{\text{BB}}$ means that it the average photon energy for any Bose-Einstein distribution is near $3k_B T$.

The Wien spectrum is “photon starved” in that the ratio between the existing photon number density n_{ph} and that of a blackbody at the same temperature $n_{\text{BB}} \approx aT^4/(3k_B T)$

is much less than unity, i.e.,

$$\xi \equiv e^{\mu/k_B T} \approx \frac{3k_B T}{aT^4} n_{\text{ph}} \approx \frac{n_{\text{ph}}}{n_{\text{BB}}(T)} \ll 1. \quad (7.133)$$

The quantity ξ may be called the *photon starvation parameter*. The physical reason for the occurrence of photon starvation is that Compton scattering conserves photon number. If we allow for photon emission, the system will eventually reach LTE such that $\xi \rightarrow 1$ and then the radiation field will have a Planck spectrum. In the following, we illustrate how an isolated system of radiation-gas mixture reaches LTE.

Let us assume that the energy density of the system is dominated by radiation, and in the absence of dynamical expansion/compression (which may occur on a timescale much longer than the duration of interest to us), the radiation energy density U is conserved. Suppose the frequent Compton scatterings between photons and electrons enforce the radiation spectrum to be in Compton equilibrium at temperature T and the electrons to be in Maxwellian distribution at the same temperature. Let us then consider that new photons are being produced at a constant¹⁴ rate of \dot{n}_{ph} . If we ignore photon absorption for now, the photon number density at time t is given by $n_{\text{ph}} = \dot{n}_{\text{ph}}t + n_{\text{ph},0} \approx \dot{n}_{\text{ph}}t$, where $n_{\text{ph},0}$ is a very small initial photon number density (at $t = 0$) that quickly gets overwhelmed by the production of new photons. The increasing photon number density reduces the average energy per photon $\langle \epsilon \rangle = U/n_{\text{ph}}$, so the radiation temperature $T \approx \langle \epsilon \rangle / (3k_B)$ also drops. Thus, the photon starvation parameter $\xi \propto n_{\text{ph}}/T^3 \propto n_{\text{ph}}^4 \propto (\dot{n}_{\text{ph}}t)^4$ (for $U = \text{const}$) increase rapidly with time. As shown in Fig. 7.12, the peak of the radiation spectrum moves towards lower and lower energies.

Associated with photon emission, there must be absorption as well. In the following, we provide a rough estimate of the timescale for absorption to become important. The photon production rate can be written in terms of the emissivity j_ν or absorption coefficient α_ν (they are connected by the Kirchhoff's law)

$$\begin{aligned} \dot{n}_{\text{ph}} &= 4\pi \int \frac{j_\nu}{h\nu} d\nu = 4\pi \int \frac{\alpha_\nu B_\nu}{h\nu} d\nu \equiv 4\pi \bar{\alpha} \int \frac{B_\nu}{h\nu} d\nu \\ &\approx \bar{\alpha} c \frac{aT^4}{3k_B T}, \end{aligned} \quad (7.134)$$

where $\bar{\alpha}$ defined above is the *photon-averaged absorption coefficient* and we have used the approximation $\int B_\nu/(h\nu) d\nu = acT^4/(2.70k_B T) \approx acT^4/(3k_B T)$ (see eq. 2.52). If we only consider photon emission over a duration of t , the photon number density is given by

$$n_{\text{ph}} \approx \dot{n}_{\text{ph}}t \approx \bar{\alpha} c t \frac{aT^4}{3k_B T}, \quad (7.135)$$

¹⁴In reality, as the electron temperature evolves with time, the photon production rate will not stay constant, but our qualitative conclusions are unaffected by a variable \dot{n}_{ph} .

which means that the photon starvation parameter (eq. 7.133) is given by

$$\xi \approx \bar{\alpha}ct = ct/\bar{\ell}_{\text{abs}}, \quad (7.136)$$

where we have defined an averaged absorption mean free path $\bar{\ell}_{\text{abs}} \equiv \bar{\alpha}$. We then define a *thermalization timescale* t_{th} during which photon emission alone would lead to $\xi = 1$ (or zero chemical potential), i.e.,

$$t_{\text{th}} = 1/(\bar{\alpha}c) = \bar{\ell}_{\text{abs}}/c. \quad (7.137)$$

The motivation for our definition of the thermalization timescale is that t_{th} is equal to the mean free time for absorption.

On very short timescales $t \ll t_{\text{th}}$, the effect of absorption is unimportant because an average photon has not propagated for a length of $\bar{\ell}_{\text{abs}}$ yet. On longer timescales $t \gg t_{\text{th}}$, the effect of absorption must be taken into account, and the photon number density then reaches an equilibrium such that only the photons emitted in a recent time window $t - t_{\text{th}}$ to t are retained, and we obtain

$$n_{\text{ph}} = n_{\text{BB}}(T) \approx \frac{aT^4}{3k_B T}, \quad (7.138)$$

Therefore, the system eventually reaches LTE at $t \gg t_{\text{th}}$ and then the equilibrium radiation field is described by the Planck distribution at temperature $T = (U/a)^{1/4}$ such that the chemical potential is zero and that the entropy of the system is maximized (see §2.1.4).

Finally, we briefly discuss a subtle detail in the range of frequencies that should be considered in the integral in eq. (7.134). The readers are referred to [Weaver \(1976\)](#) for more information. If the emission process (e.g., free-free) is broad-band, the frequency integral must include a lower limit ν_{min} , because only the photons that can be Compton-upscattered to energies $\sim k_B T$ will effectively participate in the Compton equilibrium. For non-relativistic electrons and soft photons $h\nu < k_B T$, the process of Compton-upscattering is characterized by the Compton- y parameter $y = N_{\text{sca}}4k_B T/(m_e c^2)$ (eq. 7.89), where $N_{\text{sca}} = ct/\ell_s$ is the average number of scatterings a photon undergoes during time t since the photon is emitted. A given low-energy seed photon will gain energy by a factor of e^y , so only photons with $h\nu \gtrsim k_B T e^{-y}$ will be upscattered to $\sim k_B T$, and this criterion gives rise to a minimum photon energy

$$h\nu_{\text{min}}^{(1)} \simeq k_B T \exp\left(-\frac{ct_{\text{th}}}{\ell_s} \frac{4k_B T}{m_e c^2}\right), \quad (7.139)$$

where we have taken $t \sim t_{\text{th}}$ (eq. 7.137) as we expect a typical photon near $k_B T$ to survive for a thermalization timescale due to absorption. If the system undergoes expansion on a dynamical timescale t_{dy} , then we should take $t \sim t_{\text{dy}}$, because the photon emission rate would drop significantly at $t \gtrsim t_{\text{dy}}$.

Another important consideration is that very low energy photons at $h\nu \ll k_B T$ usually have much shorter absorption mean free paths than those with energies near $k_B T$, i.e., $\ell_{\text{abs}}(\nu) \gg \bar{\ell}_{\text{abs}}$ for $h\nu \ll k_B T$. For this reason, we require that the mean free time for absorption $t_{\text{abs}}(\nu) = 1/(\alpha_\nu c) = \ell_{\text{abs}}(\nu)/c$ must be shorter than the time it takes for it to double a photon's energy due to Comptonization t_{dbl} — the latter is given by

$$t_{\text{dbl}} \simeq \frac{\ell_s}{c} \frac{m_e c^2}{4k_B T}. \quad (7.140)$$

We define a critical frequency ν_{coh} below which Compton scatterings are coherent,

$$\frac{\ell_{\text{abs}}(\nu_{\text{coh}})}{\ell_s} = \frac{m_e c^2}{4k_B T}, \quad (7.141)$$

and this gives rise to another criterion for the minimum photon frequency

$$\nu_{\text{min}}^{(2)} = \nu_{\text{coh}}. \quad (7.142)$$

Therefore, the minimum photon frequency that sets the lower limit of the integral in eq. (7.134) is given by the more stringent one of eqs. (7.139) and (7.142). We note that, if the absorption opacity is dominated by the free-free process, $\kappa_{a,\nu} = \kappa_{ff,\nu}$, the critical frequency for coherent scatterings is given by eq. (6.113).

7.5 Homework

Prob. 20. Fig. 7.13 shows the spectral energy distribution of the Crab Nebula. In this figure, the highest energy photons are near 100 TeV ($\nu \sim 2.4 \times 10^{28}$ Hz) and a large fraction of these are believed to be produced by inverse-Compton scattering of the CMB photons (a blackbody with temperature $T = 2.7$ K and mean photon energy $\approx 3k_B T$) by highly relativistic electrons.

- (1) Estimate the Lorentz factor of the electrons responsible for 100 TeV photons based on IC scattering of CMB photons.
- (2) Based on the Lorentz factor you obtained in (1), is the scattering in the Thomson or Klein-Nishina regime? Make a judgment based on the CMB photon energy in the electron's comoving frame.
- (3) Again based on the Lorentz factor you obtained (1), estimate the Compton cooling timescale due to scattering of CMB photons. Compare your cooling time with the age of the Crab Nebula, which is about 1000 yrs old.
- (4) The distance to the Crab Nebula is about 2 kpc. Based on the IC emitting power and the measured flux at 100 TeV, estimate the total number of electrons emitting at this energy.
- (5) The entire IC emission component peaks near frequency 10^{25} Hz. The emission near 10^{25} Hz are produced IC scattering of the seed photons produced by synchrotron emission — this is called

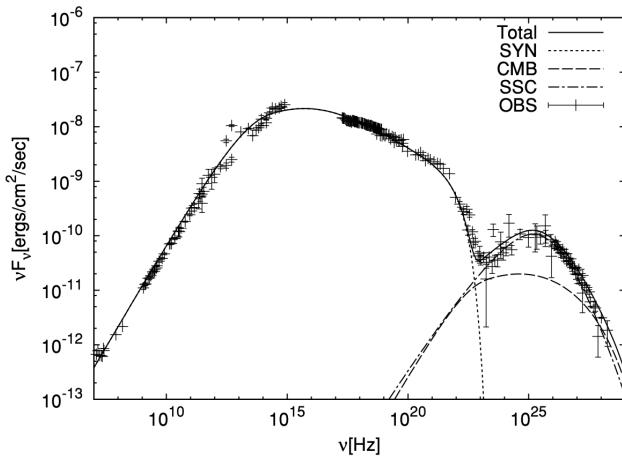


Figure 7.13: Broad-band spectrum of the Crab Nebula as measured by various instruments as summarized by [Tanaka & Takahara 2010](#). The curves are based on models: dotted line = synchrotron emission, dashed line = inverse-Compton scattering of CMB photons, dash-dotted line = inverse-Compton scattering of synchrotron photons (this is the synchrotron self-Compton component). The fitting is not very good near $\nu \sim 10^{13}$ Hz because there is an additional component due to thermal emission from radiatively heated dust grains.

synchrotron self-Compton (SSC) emission. Most of the synchrotron seed photons are near 10^{15} Hz (optical/UV bands). Based on this, estimate the Lorentz factor of the electrons responsible for emission of the 10^{25} Hz photons. Are these scatterings in the Thomson or Klein-Nishina regime?

- (6) The [LHAASO Collaboration](#) has detected PeV photons ($\nu \sim 2.4 \times 10^{29}$ Hz) from the Crab Nebula. Answer questions (1) and (2) for these PeV photons.

References

- Aharonian et al., 2004, ApJ, 614, 897
 Blumenthal & Gould, 1970, Reviews of Modern Physics, 42, 2
 P. A. M. Dirac, 1925, MNRAS, 85, 825
 Rybicki & Lightman, *Radiative processes in astrophysics* (A Wiley-Interscience Publication, New York, 1979).

Chapter 8

Synchrotron and Curvature Emission

In this chapter, we discuss the emission from electrons undergoing gyro-motion around magnetic fields. In the non-relativistic limit, the electron produces *cyclotron emission* at the frequency of the orbital gyration — this produces a spectral line. In the relativistic limit, we have *synchrotron emission*, for which the emitting power is much higher than the cyclotron power by a factor of the electron Lorentz factor squared. The synchrotron spectrum is broad-band. Since astrophysical plasmas are generally magnetized, synchrotron emission is produced in all high-energy astrophysical sources with a relativistic electron population. The synchrotron power from a proton is much weaker than that from an electron of the same kinetic energy by a factor of $(m_e/m_p)^4$, so we will focus on electrons in this chapter. Some well-known synchrotron sources are supernova remnants, termination shocks of stellar winds, pulsar wind nebulae, relativistic jets from black holes, Galactic radio synchrotron emission, virial shocks in galaxy clusters, etc.

8.1 Emitting power from an electron in a helical orbit

8.1.1 Helical orbit

The general orbit of an electron in a uniform magnetic field $\mathbf{B} = B\hat{\mathbf{z}}$ is described by a helix as shown in Fig. 4.8 (see §4.2.2)

$$\dot{\beta}_\perp = -\frac{\omega_B}{\gamma} \beta_\perp \times \hat{\mathbf{z}}, \quad \beta_\parallel = \beta_\parallel \hat{\mathbf{z}} = \text{const}, \quad (8.1)$$

where $\gamma = (1 - \beta^2)^{-1/2}$ is the constant Lorentz factor, the magnitudes of the perpendicular and parallel velocity components $\beta_\perp = |\beta_\perp|$ and $\beta_\parallel = |\beta_\parallel|$ are also constants, and we have

defined the *non-relativistic electron cyclotron frequency* by

$$\omega_B \equiv \frac{eB}{m_e c} \Rightarrow \nu_B = \frac{\omega_B}{2\pi} = 2.80 \times 10^6 (B/G) \text{ Hz.} \quad (8.2)$$

The angle between the velocity and the B-field vector is called the *pitch angle* α , which stays constant and is given by

$$\tan \alpha = \beta_{\perp}/\beta_{\parallel}. \quad (8.3)$$

The angular frequency of the gyro-motion, called the *relativistic cyclotron frequency*, is ω_B/γ independent of the pitch angle α . The electron's trajectory projected in the x - y plane (perpendicular to the B-field) is a circle whose radius is called the *Larmor radius*. The Larmor radius is given by the velocity in the x - y plane $\beta_{\perp}c$ divided by the gyro-frequency

$$r_L = \frac{\beta_{\perp}c}{\omega_B/\gamma} = \frac{\gamma\beta_{\perp}m_e c^2}{eB} = 1.70 \times 10^3 \text{ cm} \gamma \beta \sin \alpha (B/G)^{-1}. \quad (8.4)$$

8.1.2 Emitting power and cooling timescale

The total emitting power is given by the relativistic Larmor formula

$$P_{\text{syn}} = \frac{2e^2}{3c^3} \vec{a} \cdot \vec{a} = \frac{2e^2}{3c^3} \left[-(a^0)^2 + \sum_i (a^i)^2 \right], \quad (8.5)$$

where the 4-acceleration is given by $\vec{a} = d\vec{u}/d\tau$, the 4-velocity is $\vec{u} = \gamma c(1, \beta)$, and the proper time τ is related to the lab-frame time by $d\tau = dt/\gamma$ due to time dilation. Here, our Cartesian coordinate system has the z -axis along the direction of the uniform B-field \mathbf{B} . Going back to our helical orbit (with constant γ and β_z), we find the 4-acceleration

$$\vec{a} = \gamma^2 c(0, \dot{\beta}_x, \dot{\beta}_y, 0) = -\gamma \omega_B c(0, \beta_y, \beta_x, 0), \quad (8.6)$$

where $\dot{\beta}_i = d\beta_i/dt$ (eq. 8.1). The “Minkowski square” of the 4-acceleration is given by

$$\vec{a} \cdot \vec{a} = \eta_{\mu\nu} a^\mu a^\nu = \gamma^4 c^2 (\dot{\beta}_{\perp})^2 = \gamma^2 \beta_{\perp}^2 c^2 \omega_B^2. \quad (8.7)$$

Then, the synchrotron/cyclotron power of the electron is

$$P_{\text{syn}} = -\dot{\gamma} m_e c^2 = \frac{2}{3} \gamma^2 \beta_{\perp}^2 B^2 e^4 / (m_e^2 c^3) = \frac{2}{3} \gamma^2 \beta^2 B^2 r_e^2 c \sin^2 \alpha, \quad (8.8)$$

where $r_e = e^2/(m_e c^2)$ is the classical electron radius. The above expression applies to electrons with arbitrary velocities and we see that relativistic electrons emit at a much higher power than non-relativistic ones. If electrons have an isotropic velocity distribution (under the assumption of efficient pitch-angle scatterings), then we can take the average of the pitch angle dependence $\langle \sin^2 \alpha \rangle = (4\pi)^{-1} \int d\Omega_\alpha \sin^2 \alpha = 2/3$. We further use the

Thomson cross-section $\sigma_T = 8\pi r_e^2/3$, so the pitch-angle-averaged emitting power can be written as

$$\langle P_{\text{syn}} \rangle_\alpha = \frac{4}{3} \gamma^2 \beta^2 U_B \sigma_T c, \quad (\text{for isotropic pitch angle distribution}) \quad (8.9)$$

where $U_B = B^2/8\pi$ is the magnetic energy density. The electron's kinetic energy is $(\gamma - 1)m_e c^2$, so the synchrotron cooling timescale is

$$t_{\text{syn}} = \frac{(\gamma - 1)m_e c^2}{P_{\text{syn}}} = \frac{6\pi m_e c}{(\gamma + 1)B^2 \sigma_T} = 24.5 \text{ yr} (\gamma + 1)^{-1} (B/\text{G})^{-2}. \quad (8.10)$$

The synchrotron cooling rate is identical to the Compton cooling rate for an isotropic radiation field, except for the replacement of $U_{\text{rad}} \rightarrow U_B$ (see §7.3.2). Combing both mechanisms, we write the total electron cooling timescale as

$$t_{\text{cool}} = \frac{(\gamma - 1)m_e c^2}{P_{\text{syn}} + P_{\text{Comp}}} = 0.98 \text{ yr} \frac{\text{erg cm}^{-3}}{(\gamma + 1)(U_B + U_{\text{rad}})}. \quad (8.11)$$

8.2 Synchrotron/curvature emission

In this section, we discuss the emission from an ultra-relativistic ($\gamma \gg 1$) particle on a helical trajectory. We first estimate the characteristic frequency and peak spectral power by looking at the special case of a circular orbit (pitch angle $\alpha = \pi/2$). Then, we compute the detailed emission spectrum using Fourier transformation.

8.2.1 Characteristic frequency and peak spectral power

We will consider the simplest case of pitch angle $\alpha = \pi/2$ such that the electron takes a circular orbit in the x - y plane while the B-field is along the z -axis. The observer is located on the x -axis (with viewing angle $\theta = 0$). In the limit $\gamma \gg 1$, due to relativistic beaming, the synchrotron emission from the electron is beamed along a small cone of half opening angle $1/\gamma$. Thus, the observer only sees bright emission when the electron's retarded positions are within an angle $1/\gamma$ away from the $-\hat{y}$ direction. Since the orbital angular frequency is ω_B/γ , this corresponds to a retarded time separation of $\Delta t_{\text{ret}} = (2/\gamma)(\omega_B/\gamma)^{-1} = 2/\omega_B$. The observed pulse lasts for a duration

$$\Delta t_{\text{obs}} = (1 - \beta)\Delta t_{\text{ret}} \approx \frac{\Delta t_{\text{ret}}}{2\gamma^2} = \frac{1}{\gamma^2 \omega_B}. \quad (8.12)$$

Therefore, the characteristic frequency of the observed emission is

$$\omega_{\text{syn}} = \Delta t_{\text{obs}}^{-1} \simeq \gamma^2 \omega_B = \frac{\gamma^2 e B}{m_e c}, \quad (8.13)$$

or

$$\nu_{\text{syn}} \simeq \frac{\gamma^2 e B}{2\pi m_e c}. \quad (8.14)$$

The energy of a typical synchrotron photon is given by

$$\epsilon_{\text{syn}} = h\nu_{\text{syn}} \simeq \gamma^2 \hbar\omega_B, \quad (8.15)$$

where we recognize $\hbar\omega_B$ as the *Landau energy* — the energy separation between adjacent Landau levels for a non-relativistic electron¹. The electron cannot emit more than its kinetic energy in one photon, so for our classical description to be valid, we require

$$\epsilon_{\text{syn}} \ll \gamma m_e c^2 \Leftrightarrow B \ll B_{\text{QED}}/\gamma, \quad (8.16)$$

where the quantum critical B-field strength is defined as $\hbar\omega_B(B_{\text{QED}}) \equiv m_e c^2$, or

$$B_{\text{QED}} = \frac{m_e^2 c^3}{e\hbar} = 4.41 \times 10^{13} \text{ G}. \quad (8.17)$$

Near the surface of a neutron star, one often encounters the situation of $B \gtrsim B_{\text{QED}}/\gamma$ and hence quantum effects must be considered — this leads to a cutoff in spectral power near the maximum photon energy $\gamma m_e c^2$ (see §5.2 of Harding & Lai, 2006).

The spectral power reaches the peak value near the characteristic frequency, so we expect $\max(P_\nu)$ to be of the order $P_{\text{syn}}/\nu_{\text{syn}}$, which is independent of the electron Lorentz factor because $P_{\text{syn}} \propto \gamma^2$ and $\nu_{\text{syn}} \propto \gamma^2$. More detailed calculation of the Fourier spectrum (see later, eq. 8.36) shows that the characteristic frequency is given by

$$\nu_c = \frac{3}{2} \gamma^2 \nu_B = 4.2 \times 10^6 \text{ Hz} \gamma^2 (B/\text{G}), \quad (8.18)$$

and this gives a better estimate of the peak spectral power

$$P_{\nu_p} \simeq \frac{P_{\text{syn}}}{\nu_c} \simeq \frac{\sqrt{3} e^3 B}{m_e c^2}. \quad (8.19)$$

Suppose an observer measures spectral luminosity L_ν at a given frequency ν . We can estimate the Lorentz factor of the electrons emitting near this frequency by $\nu \simeq \nu_c(\gamma, B)$, which means

$$\gamma(\nu, B) \simeq \sqrt{\frac{4\pi m_e c \nu}{3eB}} = 15.4 (\nu/\text{GHz})^{1/2} (B/\text{G})^{-1/2}. \quad (8.20)$$

¹For relativistic electrons with mass γm_e , the energy separation between adjacent Landau levels is given by $\Delta E_B \approx \hbar\omega_B/\gamma$. We see that $\epsilon_{\text{syn}}/\Delta E_B \approx \gamma^3 \gg 1$ and hence the emission can be considered as a continuum process, provided that $\epsilon_{\text{syn}} \ll \gamma m_e c^2$.

Electrons with Lorentz factors much less or much larger than this value do not contribute appreciably to the observed flux at frequency ν , so we obtain

$$L_\nu \sim N_e(\gamma) P_{\nu_p}, \quad (8.21)$$

where $N_e(\gamma)$ is the number of electrons near Lorentz factor $\gamma(\nu, B)$. The above expression applies when the system is optically thin at frequency ν . This means that one can easily estimate the number of emitting electrons at a given frequency by $N_e \sim L_\nu / P_{\nu_p} \propto L_\nu / B$, which is degenerate with the (usually unknown) B-field strength. To measure the B-field strength, one usually relies on the observable effects of synchrotron cooling, which will be discussed in §8.3.2.

Suppose the system has a power-law electron distribution of $dN_e/d\gamma \propto \gamma^{-p}$, then the optically-thin specific luminosity is given by

$$L_\nu \sim \gamma (dN_e/d\gamma) P_{\nu_p} \propto B \gamma^{1-p} \propto \nu^{(1-p)/2} B^{(p+1)/2}. \quad (8.22)$$

This gives a power-law spectrum with a slope of $d \ln L_\nu / d \ln \nu = (1-p)/2 \in (-1, -0.5)$ for typical values of $p \in (2, 3)$, and the system is brighter for stronger B-fields.

8.2.2 Emission from instantaneous circular orbit

For an ultra-relativistic charged particle moving in a general curved trajectory, one can show (see below) that the emission is dominated by the perpendicular component of the acceleration. This means that the emitting power and spectrum at a given time only depends on the local curvature radius of the trajectory and the particle's Lorentz factor. Thus, an arbitrarily curved trajectory can be divided into many segments each of which is part of an instantaneous circular orbit². For this reason, our consideration of a circular orbit in §8.2.1 is in fact general and we only need to be careful that the curvature radius ρ is not always given by the Larmor radius. Here, our task is to calculate the emission spectrum of a particle on such an instantaneous circular orbit.

The emission spectrum from a charge q with a known trajectory can be obtained by first calculating the time-dependent electric field $\mathbf{E}(t, \mathbf{r})$ at the observer's position \mathbf{r} (far from the source) and time t and then taking the Fourier transform of $\mathbf{E}(t)$. The observed electric field depends on the particle's velocity $\beta = \dot{\mathbf{x}}/c$ and acceleration $\dot{\beta}$ at the retarded time t_{ret} and is given by (see Ch. 4)

$$\mathbf{E}(t, \mathbf{r}) = \frac{q}{Rc} \left[\frac{\hat{\mathbf{n}} \times ((\hat{\mathbf{n}} - \beta) \times \dot{\beta})}{(1 - \beta \cdot \hat{\mathbf{n}})^3} \right]_{\text{ret}}, \quad \mathbf{R} = \mathbf{r} - \mathbf{x}(t_{\text{ret}}), \quad \hat{\mathbf{n}} = \frac{\mathbf{R}}{R}, \quad t_{\text{ret}} = t - R/c. \quad (8.23)$$

²Readers who are not convinced can look at the “Circle of Curvature” demonstration on [this website](#) hosted by Wolfram.

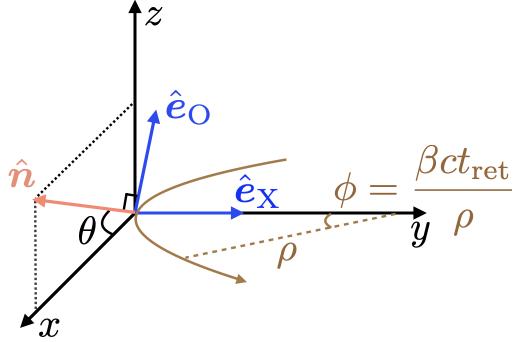


Figure 8.1: Geometry for curvature emission from an instantaneous circular orbit in the x - y plane. The observer's line of sight is along $\hat{\mathbf{n}}$ in the x - z plane, and the observed electric field has components along $\hat{\mathbf{e}}_X$ (the X-mode) and $\hat{\mathbf{e}}_O$ (the O-mode).

Here, we are mainly interested in the ultra-relativistic limit with $\gamma \gg 1$ and $\beta \approx 1$. Due to relativistic beaming, the emission is only strong within a small cone of half opening angle of the order $1/\gamma$ near the velocity vector. This means that $\hat{\mathbf{n}}$ is nearly parallel to β . One can decompose the acceleration into two components that are parallel and perpendicular to the velocity vector, i.e., $\dot{\beta} = \dot{\beta}_{\parallel} + \dot{\beta}_{\perp}$. We immediately see that $|(\hat{\mathbf{n}} - \beta) \times \dot{\beta}_{\parallel}| \ll |(\hat{\mathbf{n}} - \beta) \times \dot{\beta}_{\perp}|$ and that the ratio between the magnitude of these two terms is of the order $\dot{\beta}_{\parallel}/(\gamma \dot{\beta}_{\perp}) \ll 1$. The emitting power is always dominated by the instantaneous acceleration *perpendicular* to the velocity vector. Therefore, we can calculate the emission spectrum by considering a small segment of the particle's trajectory as part of an *instantaneous circular orbit*. Such an orbit is specified by the curvature radius ρ and angular frequency $\beta c/\rho$.

Let us place the instantaneous circular orbit to be inside the x - y plane. Without losing generality, we consider the observer to be at position \mathbf{r} that is in the x - z plane at an angle θ from $\hat{\mathbf{x}}$

$$\mathbf{r} = r(\cos \theta \hat{\mathbf{x}} + \sin \theta \hat{\mathbf{z}}). \quad (8.24)$$

The orbital motion is described by the position as a function of the retarded time

$$\mathbf{x}(t_{\text{ret}}) = \rho(\sin \phi \hat{\mathbf{x}} + (1 - \cos \phi) \hat{\mathbf{y}}), \quad \phi = \beta c t_{\text{ret}} / \rho, \quad (8.25)$$

where we have taken the orbital center to be at position $\rho \hat{\mathbf{y}}$ and $t_{\text{ret}} = 0$ to be when the particle is at the origin of the coordinate system. The particle's velocity is

$$\beta(t_{\text{ret}}) = \dot{\mathbf{x}}/c = \beta(\cos \phi \hat{\mathbf{x}} + \sin \phi \hat{\mathbf{y}}). \quad (8.26)$$

The vector pointing from the particle's retarded position to the observer is $\mathbf{R} = \mathbf{r} - \mathbf{x}$, and since the observer is far away ($r \gg |\mathbf{x}|$), we have

$$R = |\mathbf{R}| = r - \mathbf{x} \cdot \hat{\mathbf{n}}, \quad \hat{\mathbf{n}} = \hat{\mathbf{r}}. \quad (8.27)$$

Making use of $dt = (1 - \beta \cdot \hat{\mathbf{n}})dt_{\text{ret}}$, we calculate the Fourier transform of the observed electric field at angular frequency ω by changing the integration variable to the retarded time t_{ret} ,

$$\begin{aligned}\tilde{\mathbf{E}}(\omega) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{E}(t) e^{i\omega t} dt \\ &= \frac{q}{2\pi r c} e^{i\omega r/c} \int_{-\infty}^{\infty} \left[\frac{\hat{\mathbf{n}} \times ((\hat{\mathbf{n}} - \beta) \times \dot{\beta})}{(1 - \beta \cdot \hat{\mathbf{n}})^2} \right]_{\text{ret}} e^{i\omega(t_{\text{ret}} - \mathbf{x} \cdot \hat{\mathbf{n}}/c)} dt_{\text{ret}}.\end{aligned}\quad (8.28)$$

Our circular-orbit approximation is only applicable for $|t_{\text{ret}}| < \rho/c$, which should constrain the limits in the integral above. However, since the observed emission mainly comes from a short segment of the orbit within $|t_{\text{ret}}| \lesssim \rho/(\gamma c)$, we are making very little error in extending the limit from $-\infty$ to ∞ . Miraculously, one can show that the term in the square bracket is a total derivative³

$$\frac{\hat{\mathbf{n}} \times ((\hat{\mathbf{n}} - \beta) \times \dot{\beta})}{(1 - \beta \cdot \hat{\mathbf{n}})^2} = \frac{d}{dt_{\text{ret}}} \left[\frac{\hat{\mathbf{n}} \times (\hat{\mathbf{n}} \times \beta)}{1 - \beta \cdot \hat{\mathbf{n}}} \right].\quad (8.29)$$

This allows us to carry out the integral by parts and we then obtain

$$\tilde{\mathbf{E}}(\omega) = \frac{i\omega q}{2\pi r c} e^{i\omega r/c} \int_{-\infty}^{\infty} \hat{\mathbf{n}} \times (\hat{\mathbf{n}} \times \beta) e^{i\omega(t_{\text{ret}} - \mathbf{x} \cdot \hat{\mathbf{n}}/c)} dt_{\text{ret}}.\quad (8.30)$$

Since $\tilde{\mathbf{E}}(\omega)$ is in the plane perpendicular to the observer's line of sight (or $\hat{\mathbf{n}}$), it has two components.

The component along the $\hat{\mathbf{y}}$ direction is called the extraordinary-mode or *X-mode*, because the electric field is perpendicular to the $\hat{\mathbf{n}}\text{-}\mathbf{B}$ plane for the case of curvature emission (where $\beta \parallel \mathbf{B}$). The other component along the $\hat{\mathbf{n}} \times \hat{\mathbf{y}}$ direction is called the ordinary-mode or *O-mode*, because the electric field is inside the $\hat{\mathbf{n}}\text{-}\mathbf{B}$ plane for the curvature emission geometry. The reason why we decompose the emitted waves into these two modes is that a strongly magnetized plasma is birefringent such that these two polarization eigenmodes propagate differently (see Ch 5). Adopting this nomenclature, we use the following unit vectors

$$\hat{\mathbf{e}}_X = \hat{\mathbf{y}}, \quad \hat{\mathbf{e}}_O = \hat{\mathbf{n}} \times \hat{\mathbf{y}} = -\sin \theta \hat{\mathbf{x}} + \cos \theta \hat{\mathbf{z}}.\quad (8.31)$$

It turns out that for the synchrotron emission case, the B-field is along $\hat{\mathbf{z}}$ and hence the X-mode defined in the way above still has its electric field perpendicular to the $\hat{\mathbf{n}}\text{-}\mathbf{B}$ plane, and the electric field of the O-mode is inside the $\hat{\mathbf{n}}\text{-}\mathbf{B}$ plane. This further motivates our choice of the base vectors as in eq. (8.31).

³An intermediate step is $d/dt_{\text{ret}} [\hat{\mathbf{n}} \times \beta / (1 - \beta \cdot \hat{\mathbf{n}})] = \{(\hat{\mathbf{n}} - \beta) \times \dot{\beta} + [\hat{\mathbf{n}} \cdot (\beta \times \dot{\beta})] \hat{\mathbf{n}}\} / (1 - \beta \cdot \hat{\mathbf{n}})^2$.

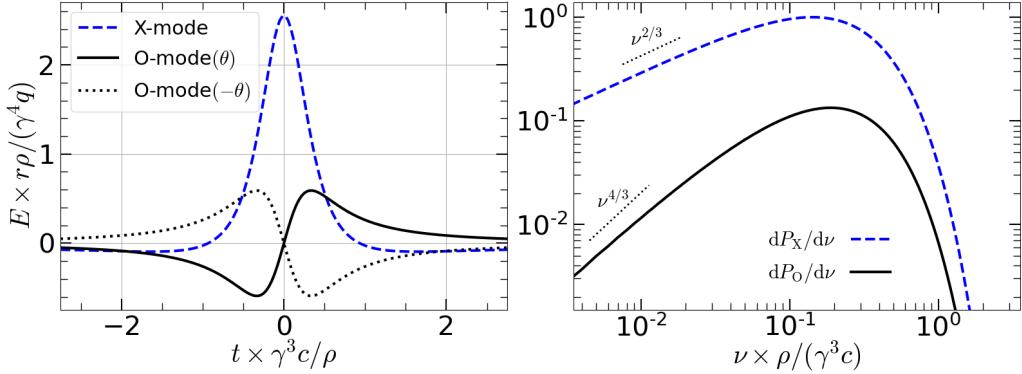


Figure 8.2: The observed E-field (upper panel) and spectrum (lower panel, normalized) for curvature emission, for an observer located at a latitude angle $\theta = \pm\gamma^{-1}/2$ from the orbital plane. The right panel shows the isotropic equivalent power viewed from a *given viewing angle*. Note that the viewing angle-averaged emitting powers in the X- and O-modes both have $P_\omega \propto \omega^{1/3}$ in the low-frequency limit.

The observed E-field components, expressed in \hat{e}_X and \hat{e}_O bases, are shown in the upper panel of Fig. 8.2, for two viewing angles $\theta = (2\gamma)^{-1}$ and $-(2\gamma)^{-1}$. The X-mode components in these two cases are identical and are an even function of observer's time, whereas the O-mode components flip signs between the two cases and are odd functions of observer's time. This shows that the single particle curvature emission is in general elliptically polarized, and the sense of rotation of the polarization depends on whether the observer's line of sight is above or below the orbital plane. The numerically computed energy spectra of these two components are shown in the lower panel of Fig. 8.2, and in the following, we show that the spectra can be obtained analytically.

Then, we use the fact that $|\phi|, \theta \ll 1$ and $\beta \approx 1$ to show that

$$\hat{n} \times (\hat{n} \times \beta) = \beta - \sin \phi \hat{e}_X + \sin \theta \cos \phi \hat{e}_O \approx \frac{ct_{\text{ret}}}{\rho} \hat{e}_X + \theta \hat{e}_O, \quad (8.32)$$

and that

$$t_{\text{ret}} - \mathbf{x} \cdot \hat{n} / c = t_{\text{ret}} \left(1 - \frac{\rho}{ct_{\text{ret}}} \sin \phi \cos \theta \right) \approx \frac{t_{\text{ret}}}{2\gamma^2} \left(1 + \gamma^2 \theta^2 + \frac{\gamma^2 c^2 t_{\text{ret}}^2}{3\rho^2} \right). \quad (8.33)$$

The two components of $\tilde{\mathbf{E}}(\omega)$ come down to the following two integrals

$$\begin{aligned} A_X &= \frac{c}{\rho} \int_{-\infty}^{\infty} t_{\text{ret}} \exp \left[\frac{i\omega t_{\text{ret}}}{2\gamma^2} \left(1 + \gamma^2 \theta^2 + \frac{\gamma^2 c^2 t_{\text{ret}}^2}{3\rho^2} \right) \right] dt_{\text{ret}} \\ &= \frac{2\rho(1+\gamma^2\theta^2)}{\gamma^2 c} \int_0^{\infty} x \sin \left[\frac{3\xi}{2} (x + x^3/3) \right] dx \\ &= \sqrt{3} \frac{\gamma}{\omega_c} (1 + \gamma^2 \theta^2) K_{2/3}(\xi), \end{aligned} \quad (8.34)$$

and

$$\begin{aligned} A_O &= \theta \int_{-\infty}^{\infty} \exp \left[\frac{i\omega t_{\text{ret}}}{2\gamma^2} \left(1 + \gamma^2 \theta^2 + \frac{\gamma^2 c^2 t_{\text{ret}}^2}{3\rho^2} \right) \right] dt_{\text{ret}} \\ &= \frac{2\rho\theta(1+\gamma^2\theta^2)^{1/2}}{\gamma c} \int_0^{\infty} \cos \left[\frac{3\xi}{2} (x + x^3/3) \right] dx \\ &= \sqrt{3} \frac{\gamma}{\omega_c} \gamma \theta (1 + \gamma^2 \theta^2)^{1/2} K_{1/3}(\xi), \end{aligned} \quad (8.35)$$

where we have defined the characteristic frequency

$$\omega_c = \frac{3\gamma^3 c}{2\rho}, \quad (8.36)$$

and used the following dimensionless variables

$$x = \frac{\gamma c t_{\text{ret}}}{\rho(1 + \gamma^2 \theta^2)^{1/2}}, \quad \xi = \frac{\omega}{2\omega_c} (1 + \gamma^2 \theta^2)^{3/2}, \quad (8.37)$$

and $K_a(\xi)$ is the modified Bessel function of order a . We note that the characteristic frequency for curvature emission, ω_c , is a factor of γ^3 times the local angular frequency c/ρ . One factor of γ comes from the fact that the emission is only observable when the (fixed) line of sight is within the beaming cone of the particle, and the observable window corresponds to a retarded time interval of the order $\Delta t_{\text{ret}} \sim \gamma^{-1} \rho/c$. The other two factors of γ comes from the difference between the observer's time and the retarded time $\Delta t/\Delta t_{\text{ret}} \approx 1 - \beta \sim \gamma^{-2}$, and hence $\Delta t \sim \gamma^{-3} \rho/c$. Finally, the characteristic frequency is $\omega_c \sim \Delta t^{-1} \sim \gamma^3 c/\rho$.

The total fluence received by the observer is $F = r^{-2} dW/d\Omega = c \int_0^\infty |\tilde{\mathbf{E}}(\omega)|^2 d\omega$, and so the energy radiated per unit angular frequency interval per solid angle is given by

$$\begin{aligned} \frac{dW}{d\omega d\Omega} &= cr^2 |\tilde{\mathbf{E}}(\omega)|^2 = \frac{\omega^2 q^2}{4\pi^2 c} (-A_X \hat{\mathbf{e}}_X + A_O \hat{\mathbf{e}}_O)^2 \\ &= \frac{3\gamma^2 q^2 \omega^2}{4\pi^2 c \omega_c^2} (1 + \gamma^2 \theta^2)^2 \left[\underbrace{K_{2/3}^2(\xi)}_{\text{X-mode}} + \underbrace{\frac{\gamma^2 \theta^2}{1 + \gamma^2 \theta^2} K_{1/3}^2(\xi)}_{\text{O-mode}} \right], \end{aligned} \quad (8.38)$$

The Bessel functions have asymptotic scalings

$$K_z(\xi) \approx \begin{cases} \Gamma(z) 2^{z-1} \xi^{-z}, & \text{for } \xi \ll 1, \\ \sqrt{\frac{\pi}{2\xi}} e^{-\xi} \left(1 + \frac{4z^2-1}{8\xi}\right), & \text{for } \xi \gg 1, \end{cases} \quad (8.39)$$

and hence $A_X \propto \omega^{-2/3}$ and $A_O \propto \omega^{-1/3}$ at low frequencies. This means that at a fixed viewing angle θ , the observed power is dominated by the X-mode at low frequencies and the spectrum is $dW/(d\omega d\Omega) \propto \omega^{2/3}$ (in agreement with Fig. 8.2).

Importantly, the emission is slightly more widely beamed at low frequencies $\omega \ll \omega_c$, because the emission comes from a longer orbital segment. In fact, the spectral power only drops substantially at $\xi \gtrsim 1$, and we can define a critical viewing angle θ_c by $\xi(\theta_c) = 1$, which gives

$$\theta_c \simeq \gamma^{-1} (2\omega_c/\omega)^{1/3}, \quad \text{for } \omega \ll \omega_c. \quad (8.40)$$

For a circular orbit, the emission from the particle is beamed into a disk-like region with vertical thickness of about $2\theta_c$. This means the total emitting solid angle is roughly $\Omega \sim 4\pi\theta_c$. Thus, in the low-frequency limit ($\omega \ll \omega_c$), the total emitting solid angle scales with frequency as $\Omega \propto \omega^{-1/3}$. The angle-integrated spectrum is given by

$$\frac{dW}{d\omega} \simeq 4\pi \int_0^{\theta_c} \frac{dW}{d\omega d\Omega} d\theta. \quad (8.41)$$

In the low-frequency limit, $\gamma\theta_c \gg 1$, and it can be shown that the power spectra of both X-mode and O-mode scale as $dW/d\omega \propto \omega^{1/3}$.

For a circular orbit with orbital period $T = 2\pi\rho/c$ (taking $\beta \approx 1$), the time-averaged emitting power per angular frequency per solid angle is

$$\frac{dP}{d\omega d\Omega} = \frac{1}{T} \frac{dW}{d\omega d\Omega} = \frac{q^2}{4\pi^3 c} \frac{\omega^2}{\gamma\omega_c} (1 + \gamma^2\theta^2)^2 \left[K_{2/3}^2(\xi) + \frac{\gamma^2\theta^2}{1 + \gamma^2\theta^2} K_{1/3}^2(\xi) \right]. \quad (8.42)$$

The total emitted energy per angular frequency bin in a full orbit is given by $\int_{4\pi} d\Omega (dW/d\omega d\Omega)$, where $d\Omega = 2\pi \cos\theta d\theta$ (note that θ is the latitude angle). Then, the time-averaged angle-integrated spectral power is given by

$$P_\omega = T^{-1} \frac{dW}{d\omega} = \frac{c}{2\pi\rho} \int_{-\pi/2}^{\pi/2} \frac{dW}{d\omega d\Omega} d\Omega \approx \frac{\omega^2 q^2}{2\pi^2 \rho} \int_0^\infty (A_X^2 + A_O^2) d\theta = P_{X,\omega} + P_{O,\omega}, \quad (8.43)$$

and it can be shown that (Westfold 1959)

$$P_{X,\omega} = \frac{\sqrt{3}\gamma q^2}{4\pi\rho} [F(x) + G(x)], \quad P_{O,\omega} = \frac{\sqrt{3}\gamma q^2}{4\pi\rho} [F(x) - G(x)], \quad (8.44)$$

where

$$F(x) = x \int_x^\infty K_{5/3}(\xi) d\xi, \quad G(x) = x K_{2/3}(x), \quad x = \frac{\omega}{\omega_c}. \quad (8.45)$$

These two functions F and G are shown in Fig. 8.4, where we see that, indeed, $P_{X,\omega} \propto \omega^{1/3}$ and $P_{O,\omega} \propto \omega^{1/3}$ at $\omega \ll \omega_c$.

Rybicki & Lightman provided the following integrals of $F(x)$ and $G(x)$

$$\begin{aligned} \int_0^\infty x^\mu F(x) dx &= \frac{2^{\mu+1}}{\mu+2} \Gamma\left(\frac{\mu}{2} + \frac{7}{3}\right) \Gamma\left(\frac{\mu}{2} + \frac{2}{3}\right) \\ \int_0^\infty x^\mu G(x) dx &= 2^\mu \Gamma\left(\frac{\mu}{2} + \frac{4}{3}\right) \Gamma\left(\frac{\mu}{2} + \frac{2}{3}\right). \end{aligned} \quad (8.46)$$

Based on these, we obtain the ratio between the frequency-integrated powers in the X-mode and O-mode

$$\frac{P_X}{P_O} = \frac{\Gamma(7/3) + \Gamma(4/3)}{\Gamma(7/3) - \Gamma(4/3)} = 7, \quad (8.47)$$

which means that the emission is mostly in the X-mode.

8.2.3 From circular to helical orbits

We wish to apply the above results for curvature emission from a circular orbit to the case of a helical orbit (for a uniform B-field) with pitch angle given by $\tan \alpha = p_\perp/p_\parallel$. The geometry is shown in Fig. 8.3. There are two ways of doing this.

The first way is to divide the helical trajectory into many short segments each of which is locally part of a circular orbit. To do this, one would like to know the local curvature radius of a helix, which is given by (a non-trivial task, see [this page](#) while noticing there $\kappa = \rho^{-1}$ and $r/c = \tan \alpha$)

$$\rho = r_L / \sin^2 \alpha, \quad (8.48)$$

where r_L is the Larmor radius (eq. 8.4) for the projected circular motion in the plane perpendicular to the B-field. Since $r_L = \gamma \beta_\perp c / \omega_B \approx \gamma c \sin \alpha / \omega_B$ (taking $\beta \approx 1$), we obtain the characteristic frequency for the curvature emission from a helical orbit

$$\tilde{\omega}_c = \frac{3}{2} \frac{\gamma^3 c}{\rho} = \frac{3}{2} \gamma^2 \omega_B \sin \alpha \Rightarrow \tilde{\nu}_c = \frac{3}{2} \gamma^2 \nu_B \sin \alpha, \quad (8.49)$$

where $\nu_B = \omega_B/(2\pi) = qB/(2\pi mc)$ is the non-relativistic cyclotron frequency. The emission spectrum can then be obtained by plugging the characteristic frequency $\tilde{\omega}_c$ into eq. (8.42). We also notice that the angle θ there (as shown in Fig. 8.1) must be replaced by the angle between the line of sight and the plane of the instantaneous circular orbit

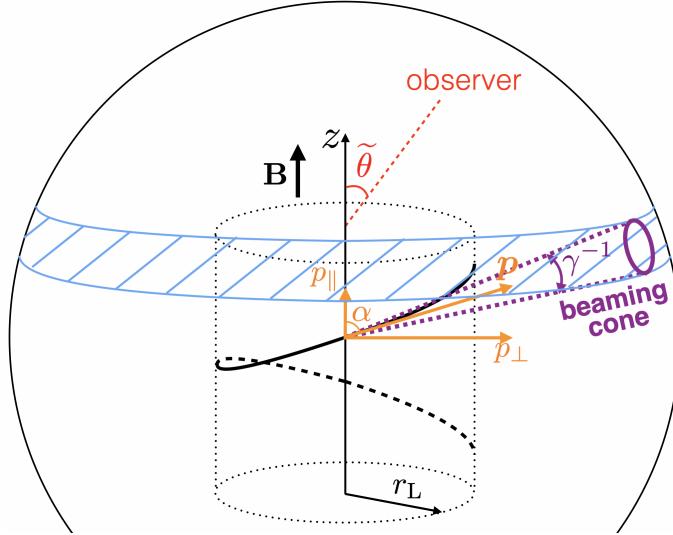


Figure 8.3: Helical orbit of a particle in a uniform B-field and its relativistic beaming cone. The observer only sees bright emission when the viewing angle $\tilde{\theta}$ (between line of sight and B-field) is within $1/\gamma$ away from the pitch angle α , i.e. $|\Delta\theta| = |\tilde{\theta} - \alpha| \lesssim \gamma^{-1}$.

$\Delta\theta = \tilde{\theta} - \alpha$, where $\tilde{\theta}$ is the viewing angle wrt. the B-field (see Fig. 8.3). The result from this procedure is

$$\frac{dP}{d\omega d\Omega}(\tilde{\theta}) = \frac{q^2}{4\pi^3 c} \frac{\omega^2}{\gamma \tilde{\omega}_c} (1 + \gamma^2 \Delta\theta^2)^2 \left[K_{2/3}^2(\xi) + \frac{\gamma^2 \Delta\theta^2}{1 + \gamma^2 \Delta\theta^2} K_{1/3}^2(\xi) \right], \quad (8.50)$$

where $\xi = (\omega/2\tilde{\omega}_c)(1 + \gamma^2 \Delta\theta^2)^{3/2}$.

The second strategy, the one we will take below, is to carry out a Lorentz transformation from the comoving frame of the velocity component β_{\parallel} (parallel to the B-field) to the lab frame. The particle's orbit is circular in this comoving frame and helical in the lab frame. We denote all quantities in the comoving frame of velocity β_{\parallel} with a prime ('), and the unprimed versions of them are measured in the lab frame. During the Lorentz transformation, the momentum in the perpendicular direction $p_{\perp} = \gamma \beta_{\perp} m_e c$ stays unchanged, the B-field is unchanged, and the Larmor radius $r_L \propto p_{\perp}/B$ is also unchanged.

In the comoving frame of the parallel velocity component, the curvature radius is $\rho' = r_L$ and the particle's Lorentz factor is given by

$$\gamma' = \sqrt{1 + (\gamma \beta_{\perp})^2} = \sqrt{1 + (\gamma^2 - 1) \sin^2 \alpha} \approx \gamma \sin \alpha, \quad (8.51)$$

where the approximation applies for $\gamma \sin \alpha \gg 1$ as is the case for typical pitch angles that are not extremely close to zero. The particle is still highly relativistic ($\gamma' \gg 1$) even in

the comoving frame of the parallel velocity component. We then obtain the characteristic frequency of curvature emission in the comoving frame (curvature radius = ρ')

$$\omega'_c = \frac{3}{2}(\gamma')^3 \frac{c}{\rho'} \approx \frac{3}{2}\gamma^2 \omega_B \sin^2 \alpha. \quad (8.52)$$

In the lab frame, we denote the angle between the line of sight and the B-field or the parallel velocity $\beta_{||}$ as $\tilde{\theta}$ (see Fig. 8.3). Note that $\tilde{\theta}$ is the polar angle but not the latitude angle from the equatorial plane. We expect that the observer only sees bright emission when the viewing angle $\tilde{\theta}$ is close to the pitch angle α . The viewing angle in the comoving frame is denoted as $\tilde{\theta}'$, which is related to $\tilde{\theta}$ via Lorentz transformation

$$\cos \tilde{\theta}' = \frac{\cos \tilde{\theta} - \beta_{||}}{1 - \beta_{||} \cos \tilde{\theta}} = \frac{\cos \tilde{\theta} - \beta \cos \alpha}{1 - \beta \cos \alpha \cos \tilde{\theta}}, \quad (8.53)$$

where we have used $\beta_{||} = \beta \cos \alpha$.

In the earlier discussion in §8.2.2 (and Fig. 8.1), we have been using the latitude angle θ' from the equatorial plane (but without a prime), as there we have a circular orbit in the equatorial plane. Here, we use the polar angle θ' (or the co-latitude angle) to denote the observer's line of sight, and these two angles θ' and $\tilde{\theta}'$ are related to each other by

$$\theta' = \pi/2 - \tilde{\theta}' \Rightarrow \sin \theta' = \cos \tilde{\theta}', \quad (8.54)$$

and $\tilde{\theta}'$ is related to $\tilde{\theta}$ via Lorentz transformation above. Note that, due to relativistic beaming, the observer only sees bright emission when $|\theta'| \ll 1$ rad relativistic beaming and hence $\tilde{\theta}'$ is close to $\pi/2$. We then Taylor expand $\cos \tilde{\theta}$ near $\tilde{\theta} = \alpha$ in terms of the small quantity

$$\Delta\theta \equiv \tilde{\theta} - \alpha, \quad |\Delta\theta| \ll 1 \text{ rad}, \quad (8.55)$$

and obtain

$$\cos \tilde{\theta} \approx \cos \alpha - \sin \alpha \Delta\theta + \mathcal{O}(\Delta\theta^2). \quad (8.56)$$

This can be plugged in eq. (8.53) and, in the limit $\gamma \gg 1$ and $|\Delta\theta| \ll 1$ rad, we obtain the latitude angle θ' in the comoving frame

$$\theta' \approx \sin \theta' = \cos \tilde{\theta}' = \frac{1}{\sin \alpha} \left[-\Delta\theta + \frac{\cot \alpha}{2\gamma^2} + \mathcal{O}(\Delta\theta^2) \right] \approx -\frac{\Delta\theta}{\sin \alpha}. \quad (8.57)$$

Using the Lorentz factor in the comoving frame $\gamma' \approx \gamma \sin \alpha$ (eq. 8.51), we further obtain the following linear mapping between the viewing angles in the two frames

$$\gamma' \theta' \approx -\gamma \Delta\theta = -\gamma(\tilde{\theta} - \alpha). \quad (8.58)$$

We see that, indeed, the observer only sees bright emission when the viewing angle $\tilde{\theta}$ is within an angle $1/\gamma$ away from the pitch angle α (such that $\gamma' \theta' \lesssim 1$). If we were to fix

the observer's viewing angle $\tilde{\theta}$ and consider particles with different pitch angles α , then the observer only sees the emission from particles with pitch angles within an angle $1/\gamma$ away from the viewing angle $\tilde{\theta}$.

From $\beta_{\parallel} = \beta \cos \alpha$, let us define a parallel Lorentz factor

$$\gamma_{\parallel} \equiv \frac{1}{\sqrt{1 - \beta_{\parallel}^2}} = \frac{1}{\sqrt{1 - \beta^2 \cos^2 \alpha}} \approx \frac{1}{\sin \alpha}. \quad (8.59)$$

The Doppler factor for the Lorentz transformation is given by

$$\mathcal{D} = \frac{1}{\gamma_{\parallel}(1 - \beta_{\parallel} \cos \tilde{\theta})} = \gamma_{\parallel}(1 + \beta_{\parallel} \cos \tilde{\theta}). \quad (8.60)$$

Since $\theta' \ll 1$ (otherwise the emitting power is exponentially suppressed), we know that $\tilde{\theta}' = \pi/2 - \theta' \approx \pi/2$ and hence

$$\mathcal{D} = \gamma_{\parallel}(1 + \beta \cos \alpha \sin \theta') \approx \gamma_{\parallel} \approx \frac{1}{\sin \alpha}. \quad (8.61)$$

The observer's frequency is given by

$$\omega = \mathcal{D}\omega' \approx \omega'/\sin \alpha. \quad (8.62)$$

The energy spectrum per solid angle in the lab frame is given by

$$\frac{dW}{d\omega d\Omega} = \frac{\mathcal{D}dW'}{(\mathcal{D}d\omega)(\mathcal{D}^{-2}d\Omega')} = \mathcal{D}^2 \frac{dW'}{d\omega' d\Omega'} \approx \frac{1}{\sin^2 \alpha} \frac{dW'}{d\omega' d\Omega'}, \quad (8.63)$$

where $dW'/d\omega' d\Omega'$ is given by eq. (8.38).

The orbital period in the lab frame is $T = \gamma_{\parallel} T' \approx T'/\sin \alpha$, and the emitting spectral power per solid angle along viewing angle $\tilde{\theta} = \arccos(\hat{n} \cdot \hat{B})$ is given by

$$\begin{aligned} \frac{dP}{d\omega d\Omega}(\tilde{\theta}) &= T^{-1} \frac{dW}{d\omega d\Omega} = \frac{1}{\sin \alpha} \left(\frac{dP}{d\omega d\Omega} \right)' \\ &= \frac{q^2}{4\pi^3 c} \frac{\omega^2}{\gamma \omega_c \sin \alpha} (1 + \gamma^2 \Delta\theta^2)^2 \left[K_{2/3}^2(\xi) + \frac{\gamma^2 \Delta\theta^2}{1 + \gamma^2 \Delta\theta^2} K_{1/3}^2(\xi) \right], \end{aligned} \quad (8.64)$$

where the dimensionless argument is

$$\xi = \frac{\omega}{2\omega_c \sin \alpha} (1 + \gamma^2 \Delta\theta^2)^{3/2}. \quad (8.65)$$

The angle-integrated time-averaged *emitting spectral power* per particle is

$$P_{\omega} = \int d\Omega \frac{dP}{d\omega d\Omega} = \frac{1}{\gamma_{\parallel} T'} \frac{\mathcal{D}dW'}{\mathcal{D}d\omega'} \approx P'_{\omega} \sin \alpha = \frac{\sqrt{3}q^3 B \sin \alpha}{2\pi m_e c^3} F(x), \quad x = \frac{\omega}{\omega_c \sin \alpha}. \quad (8.66)$$

The transformation of $P_\omega = P'_{\omega'} \sin \alpha$ also applies to each of the individual X-mode and O-mode, i.e.,

$$P_{X/O,\omega} = P'_{X/O,\omega'} \sin \alpha = \frac{\sqrt{3}q^3 B \sin \alpha}{2\pi m_e c^3} [F(x) \pm G(x)], \quad (8.67)$$

where $P'_{X/O,\omega'}$ are given by eq. (8.44).

It should be noted that the observed power is different because the observer's time interval between adjacent pulses⁴ is $T_{\text{obs}} = T'/\mathcal{D}$. Thus, the *observed spectral power* is

$$P_{\text{obs},\omega} = T_{\text{obs}}^{-1} \frac{dW}{d\omega} = \frac{1}{\sin^2 \alpha} P_\omega. \quad (8.68)$$

Note that $P_{\text{obs},\omega}$ is the spectral power for a particle with a fixed pitch angle. Practically, for a plasma without bulk motion in the lab frame, an average particle spends half of its time moving closer to the observer and the rest of the time moving away from the observer. Under an equilibrium momentum distribution, the long-term averaged spectral power per particle is given by P_ω , instead of $P_{\text{obs},\omega}$.

Finally, if the emitting particles have Lorentz factor and pitch angle distributions given by $d^2 n_e / (d\gamma d\alpha)$, then the emissivity along a given viewing angle is given by

$$j_\nu(\tilde{\theta}) = 2\pi \int d\gamma \int d\alpha \frac{d^2 n_e}{d\gamma d\alpha} \frac{dP}{d\omega d\Omega}(\gamma, \alpha). \quad (8.69)$$

Since the $dP/(d\omega d\Omega)$ is only non-zero for pitch angles within a narrow range between $\tilde{\theta} \pm \gamma^{-1}$ and $d\alpha = -d\Delta\theta = -d\Omega/(2\pi \sin \tilde{\theta})$, the above integral over the pitch angle α (while fixing the viewing angle $\tilde{\theta}$) can be converted into an integral over viewing angle (while fixing the pitch angle) as follows

$$\begin{aligned} j_\nu(\tilde{\theta}) &= \frac{1}{\sin \tilde{\theta}} \int d\gamma \left. \frac{d^2 n_e}{d\gamma d\alpha} \right|_{\alpha=\tilde{\theta}} \int d\Omega \frac{dP}{d\omega d\Omega}(\gamma, \alpha = \tilde{\theta}) \\ &= \frac{1}{2\pi \sin \tilde{\theta}} \int d\gamma \left[\frac{d^2 n_e}{d\gamma d\alpha} P_\nu(\gamma, \alpha) \right]_{\alpha=\tilde{\theta}}, \end{aligned} \quad (8.70)$$

where we have used $P_\nu = 2\pi P_\omega$. It is easy to show that the above result applies to individual X-mode and O-mode polarizations,

$$j_{X/O,\nu}(\tilde{\theta}) = \frac{1}{2\pi \sin \tilde{\theta}} \int d\gamma \left[\frac{d^2 n_e}{d\gamma d\alpha} P_{X/O,\nu}(\gamma, \alpha) \right]_{\alpha=\tilde{\theta}}, \quad (8.71)$$

where $P_{X/O,\nu} = 2\pi P_{X/O,\omega}$ are given by eq. (8.67).

⁴You can think that the particle emits a train of $N = \omega' T'/(2\pi)$ wave periods in each full orbit, and the observer will measure $N = \omega T_{\text{obs}}/(2\pi) = \mathcal{D}\omega' T_{\text{obs}}/(2\pi)$ wave periods because the wave phase is Lorentz invariant. This leads to $T_{\text{obs}} = T'/\mathcal{D}$.

For an isotropic pitch-angle distribution, we have $dn_e/d\alpha = (1/2) \sin \alpha n_e$, so the emissivity is given by

$$j_\nu(\tilde{\theta}) = \frac{1}{4\pi} \int d\gamma \frac{dn_e}{d\gamma} P_\nu(\gamma, \alpha = \tilde{\theta}), \text{ for isotropic pitch angles and uniform B-field, } (8.72)$$

The emission is the strongest in the direction perpendicular to the B-field at $\tilde{\theta} = \pi/2$, and there is no emission along the B-field at $\tilde{\theta} = 0$ or π .

On the other hand, in a plasma where the B-fields are randomly oriented, the viewing angle $\tilde{\theta}$ is drawn from a uniform distribution in 4π , and the emissivity is given by

$$j_\nu = \frac{1}{4\pi} \int d\gamma \frac{dn_e}{d\gamma} \langle P_\nu \rangle_\alpha, \text{ for isotropic pitch angles and random B-fields, } (8.73)$$

where $\langle P_\nu \rangle_\alpha = \frac{1}{2} \int_0^\pi P_\nu(\gamma, \alpha) \sin \alpha d\alpha$ is the pitch-angle-averaged spectral power.

8.3 Synchrotron spectrum

In this section, we discuss the numerical synchrotron spectrum.

8.3.1 Single and isotropic pitch angles

The total synchrotron emitting spectrum for an electron with pitch angle α is

$$P_\nu = \frac{\sqrt{3}e^3 B \sin \alpha}{m_e c^2} F(x), \quad x = \frac{\nu}{\nu_c \sin \alpha}, \quad \nu_c = \frac{3\gamma^2 \nu_B}{2}, \quad \nu_B = \frac{eB}{2\pi m_e c}, \quad (8.74)$$

where we have defined the characteristic ν_c without the $\sin \alpha$ factor (this definition is more convenient for the widely used case of isotropic pitch-angle distribution). Asymptotic forms of $F(x)$ are given by ([Ginzburg & Syrovatskii 1965](#))

$$F(x) = \begin{cases} \frac{4\pi}{2^{1/3} 3^{1/2} \Gamma(1/3)} x^{1/3} \approx 2.150 x^{1/3}, & \text{for } x \ll 1, \\ \sqrt{\pi/2} x^{1/2} e^{-x}, & \text{for } x \gg 1. \end{cases} \quad (8.75)$$

The above expressions are only useful under the following two conditions: (1) all electrons in a given system have the same pitch angle α , and (2) the local B-field orientations are randomly distributed. It turns out that neither of them are strictly satisfied in realistic astrophysical environments. Specifically, based on the non-zero linear polarizations of many synchrotron sources, we know that the B-field orientations are not completely random. In such a situation, one must use the emitting powers for individual polarizations along the projected B-field direction $P_{\parallel,\nu}$ and perpendicular to the projected B-field $P_{\nu,\perp}$. This will be discussed in §8.5. For now, we assume that the B-fields are randomly oriented.

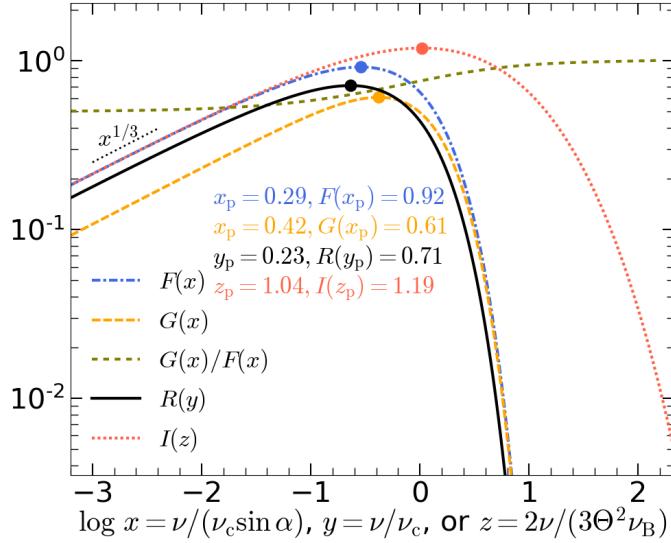


Figure 8.4: The numerical functions $F(x)$ (eq. 8.45), $G(x)$ (eq. 8.45), $R(y)$ (eq. 8.78), and $I(z)$ (eq. 8.93) relevant for the synchrotron spectrum. The peak values are indicated by dots. Frequency is normalized by $\nu_c = 3\gamma^2\nu_B/2$ and $\nu_B = eB/(2\pi m_e c)$.

However, the pitch-angle distribution is still an issue. The discussion below is based on the assumption of an isotropic pitch-angle distribution, which is more realistic than a δ -function distribution at a certain angle, but one should keep in mind that the realistic pitch-angle distribution depends on small-scale plasma physics and is usually not known *a priori*.

For an isotropic pitch-angle distribution, we would like to know the pitch-angle averaged spectral power

$$\langle P_\nu \rangle_\alpha = \frac{1}{2} \int_0^\pi P_\nu(\alpha) \sin \alpha d\alpha = \int_0^{\pi/2} P_\nu(\alpha) \sin \alpha d\alpha = \frac{\sqrt{3}e^3 B}{m_e c^2} R(y), \quad (8.76)$$

where

$$R(y) = \int_0^1 F\left(\frac{y}{\xi}\right) \frac{\xi^2}{\sqrt{1-\xi^2}} d\xi, \quad y = \frac{2\nu}{3\gamma^2\nu_B}, \quad \nu_B = \frac{eB}{2\pi m_e c}. \quad (8.77)$$

The function $R(y)$ can be analytically written as (see [Crusius & Schlickeiser 1986](#) and [Ghisellini et al. 1988](#))

$$R(y) = \frac{y^2}{2} \left\{ K_{4/3}(y/2)K_{1/3}(y/2) - \frac{y}{10} \left[K_{4/3}^2(y/2) - K_{1/3}^2(y/2) \right] \right\}, \quad (8.78)$$

and the asymptotic forms are

$$R(y) = \begin{cases} (2^{1/3}/5)\Gamma^2(1/3)y^{1/3} = 1.808y^{1/3}, & \text{for } y \ll 1, \\ (\pi/2)[1 - 99/(162y)]e^{-y}, & \text{for } y \gg 1. \end{cases} \quad (8.79)$$

Numerical values of the two functions $F(x)$ and $R(y)$ are shown in Fig. 8.4. It is sometimes desirable to have analytical approximations for these numerical functions, which are given as follows:

$$F(x) \approx \frac{\sqrt{\pi/2} [1 + (5/11)x^{5/29}] (1 + 0.8x^{1/2})e^{-x}}{0.8 + 0.583x^{-1/3} + (5/11)x^{5/29}}, \text{ fractional error} < 4\%, \forall x \in (0, \infty), \quad (8.80)$$

and

$$R(y) \approx \frac{(\pi/2)[1 + (7/4)y^{5/12}]e^{-y}}{1 + 0.869y^{-1/3} + (7/4)y^{5/12}}, \text{ fractional error} < 2\%, \forall y \in (0, \infty). \quad (8.81)$$

With these numerical functions, the specific emissivity for an arbitrary distribution of relativistic electrons $f_e(\gamma)d\gamma$ in randomly oriented magnetic fields is given by

$$j_\nu = \frac{n_e}{4\pi} \frac{\sqrt{3}e^3 B}{m_e c^2} \int_2^\infty d\gamma f_e(\gamma) R(y), \quad y = \frac{2\nu}{3\gamma^2 \nu_B}, \quad (8.82)$$

where n_e is the number density of all electrons, the distribution function is normalized by $\int_1^\infty f_e(\gamma)d\gamma = 1$, and the lower limit $\gamma_{\min} \sim 2$ for the integral is set by the requirement that only relativistic electrons produce broad-band synchrotron emission. At sufficiently high frequencies such that the radiating electrons are highly relativistic, the above expression can be simplified by changing the variable $d\gamma \rightarrow dy$ and extending the range of integration to $y \in (0, \infty)$, meaning that

$$j_\nu \approx \frac{n_e}{4\pi} \frac{\sqrt{3}e^3 B}{m_e c^2} \int_0^\infty \frac{dy}{2y} \gamma(y) f_e[\gamma(y)] R(y), \quad \gamma(y) = \left(\frac{2\nu}{3\nu_B y}\right)^{1/2}, \text{ for } \gamma(y) \gg 1. \quad (8.83)$$

8.3.2 Power-law and thermal electron distributions

Let us first consider a power-law electron distribution

$$f_e(\gamma)d\gamma = \frac{p-1}{\gamma_{\min}} (\gamma/\gamma_{\min})^{-p} d\gamma, \quad \gamma_{\min} < \gamma < \gamma_{\max}, \quad (8.84)$$

where we have assumed that $\gamma_{\max} \gg \gamma_{\min}$ and $p > 1$. It is convenient to define the minimum and maximum characteristic frequencies

$$\nu_{\min/\max} = \frac{3\nu_B}{2} \gamma_{\min/\max}^2, \quad \nu_B = \frac{eB}{2\pi m_e c}. \quad (8.85)$$

Plugging $f_e(\gamma)$ into eq. (8.83) and using $\gamma/\gamma_{\min} = y^{-1/2}(\nu/\nu_{\min})^{1/2}$, we see that the emissivity is given by

$$\begin{aligned} j_\nu &= \frac{n_e}{4\pi} \frac{\sqrt{3}e^3 B}{m_e c^2} C_1 (\nu/\nu_{\min})^{(1-p)/2}, \\ C_1(p, \nu) &= \frac{p-1}{2} \int_{y_{\min}}^{y_{\max}} dy y^{(p-3)/2} R(y), \quad y_{\min} = \frac{\nu}{\nu_{\max}}, \quad y_{\max} = \frac{\nu}{\nu_{\min}}. \end{aligned} \quad (8.86)$$

In the limit $\nu_{\min} \ll \nu \ll \nu_{\max}$, we have $y_{\min} \ll 1$ and $y_{\max} \gg 1$, so the limits in the C_1 integral can be extended to $C_1 \approx [(p-1)/2] \int_0^\infty dy (\dots)$, meaning that C_1 is only a function of p given by (see Crusius & Schlickeiser 1986)

$$C_1(p, \nu_{\min} \ll \nu \ll \nu_{\max}) = \pi^{\frac{1}{2}} 2^{\frac{p-7}{2}} \frac{(p-1)(p+7/3)}{p+1} \frac{\Gamma\left(\frac{p+5}{4}\right) \Gamma\left(\frac{3p-1}{12}\right) \Gamma\left(\frac{3p+7}{12}\right)}{\Gamma\left(\frac{p+7}{4}\right)}. \quad (8.87)$$

At high frequencies $\nu \gg \nu_{\max}$, we have $y_{\min} \gg 1$ and $y_{\max} \rightarrow \infty$, and hence

$$C_1(p, \nu \gg \nu_{\max}) = \frac{\pi(p-1)}{4} \left[\Gamma_{\text{inc}}\left(\frac{p-1}{2}, \frac{\nu}{\nu_{\max}}\right) - \frac{99}{162} \Gamma_{\text{inc}}\left(\frac{p-3}{2}, \frac{\nu}{\nu_{\max}}\right) \right], \quad (8.88)$$

where the upper incomplete Gamma function is given by

$$\Gamma_{\text{inc}}(s, a) = \int_a^\infty x^{s-1} e^{-x} dx. \quad (8.89)$$

At low frequencies $\nu \ll \nu_{\min}$, we have $y_{\min} \approx 0$ and $y_{\max} \ll 1$, and hence

$$C_1(p, \nu \ll \nu_{\min}) \left(\frac{\nu}{\nu_{\min}}\right)^{(1-p)/2} = \frac{3}{5} 2^{1/3} \Gamma^2(1/3) \frac{p-1}{3p-1} \left(\frac{\nu}{\nu_{\min}}\right)^{1/3}. \quad (8.90)$$

For frequencies near $\nu_{\min/\max}$, one must rely on either interpolation of these asymptotic results or numerical integration of eq. (8.86).

Next, we consider the case of an ultra-relativistic thermal electron distribution (aka the *Maxwell-Jüttner distribution*) at temperature T , as described as follows

$$f_e(\gamma) = \frac{\gamma^2 \beta}{\Theta K_2(1/\Theta)} e^{-\gamma/\Theta} \approx \frac{\gamma^2}{2\Theta^3} e^{-\gamma/\Theta}, \quad \Theta = \frac{k_B T}{m_e c^2}, \quad (8.91)$$

where $K_n(x)$ is the modified Bessel function, and we have taken $\beta \approx 1$ and $K_2(1/\Theta) \approx 2\Theta^2$ for relativistic temperatures $\Theta \gg 1$. Again, we plug this into eq. (8.83) and obtain

$$j_\nu = \frac{n_e}{8\pi} \frac{\sqrt{3}e^3 B}{m_e c^2} I(z), \quad z = \frac{2\nu}{3\Theta^2 \nu_B}, \quad (8.92)$$

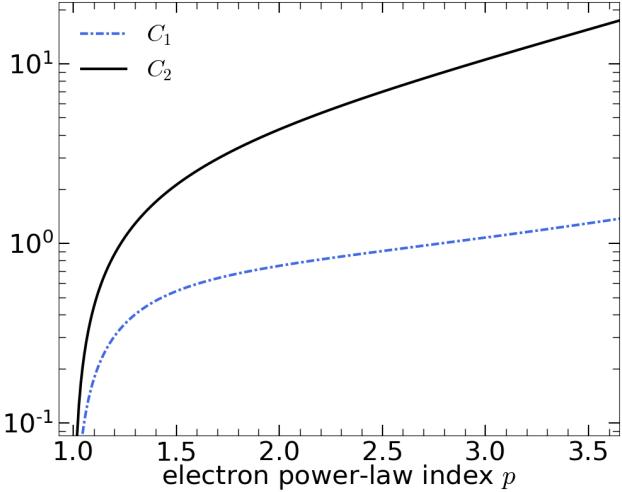


Figure 8.5: The constants $C_1(p)$ (eq. 8.87) and $C_2(p)$ (eq. 8.134) relevant for synchrotron emissivity and absorption coefficient, for frequencies $\nu_{\min} \ll \nu \ll \nu_{\max}$.

where the function $I(z)$ is given by

$$I(z) = \int_0^\infty d\xi \xi^2 e^{-\xi} R(z/\xi^2), \quad \xi = \gamma/\Theta = (z/y)^{1/2}. \quad (8.93)$$

This numerical function is shown⁵ in Fig. 8.4. It can be analytically approximated by (see [Mahadevan et al. 1996](#))

$$I(z) = 4.05 \left(z^{5/6} + 0.40 z^{7/12} + 0.532 z^{1/3} \right) \exp \left(-1.890 z^{1/3} \right), \quad \text{for } \Theta \gg 1, \quad (8.94)$$

which has a fractional error $< 2.3\%$ for any z . An interesting property of the result here is that the synchrotron spectrum from a thermal electron population does not cut off exponentially as e^{-z} , whereas it declines a bit slower as $I(z) \propto z^{5/6} e^{-1.890 z^{1/3}}$. The high-frequency logarithmic slope is given by

$$\frac{d \ln j_\nu}{d \ln \nu} = \frac{d \ln I(z)}{d \ln z} = 5/6 - 0.63 z^{1/3}, \quad \text{for } z \gg 1. \quad (8.95)$$

8.3.3 Effects of electron cooling

Now we consider the dynamical evolution of the Lorentz factor distribution function $dn_e/d\gamma$, which is denoted as n'_e (no Lorentz transformation has been made here) for brevity. Due

⁵We see that the normalization for $I(z)$ at $z \ll 1$ seems to be the same as that for the function $F(x)$ at $x \ll 1$, but this is just a coincidence since the accurate values of these two normalization constants are slightly different (by 0.004).

to a certain particle acceleration mechanism (e.g., shocks), relativistic electrons are added to the system at a rate described by the *injection function* (in units of $\text{cm}^{-3} \text{s}^{-1}$)

$$S(\gamma) = \frac{n_{e,0}(p-1)}{t_{\text{dyn}}\gamma_{\min}} (\gamma/\gamma_{\min})^{-p}, \text{ for } \gamma > \gamma_{\min}, \quad (8.96)$$

where t_{dyn} is the dynamical timescale, $n_{e,0}$ is the number of newly injected electrons per unit volume per dynamical timescale, and γ_{\min} is the minimum Lorentz factor of the newly injected electron population. A physical system must also have a maximum injection Lorentz factor γ_{\max} , but here we assume that γ_{\max} is sufficiently large such that its effect on the observed part of the spectrum can be ignored. At the same time, each electron cools due to synchrotron and inverse-Compton emission⁶ and hence

$$\dot{\gamma} = -\frac{\gamma}{t_{\text{cool}}} = -\frac{P_{\text{syn}} + P_{\text{Comp}}}{m_e c^2} = -\gamma^2 \frac{4\sigma_T(U_B + U_{\text{rad}})}{3m_e c}, \quad (8.97)$$

where $t_{\text{cool}}(\gamma)$ is the cooling timescale. The time evolution of the Lorentz factor distribution $n'_e(\gamma) = dn_e/d\gamma$ is governed by the following continuity equation

$$\partial_t n'_e = -\partial_\gamma(\dot{\gamma}n'_e) + S(\gamma), \quad (8.98)$$

The first term on the right-hand side is the *advection term* which moves electrons from high- γ to low- γ regions according to the advection speed $\dot{\gamma}$. The second term $S(\gamma)$ is called the *source term*, which tries to increase the normalization of the distribution function at all γ 's.

We define a *cooling Lorentz factor* γ_c that is given by

$$t_{\text{cool}}(\gamma_c) = t_{\text{dyn}}. \quad (8.99)$$

It is easy to see that there are two regimes. Since most electrons have Lorentz factors near γ_{\min} (as long as $p > 1$), there are two regimes depending on whether electrons near γ_{\min} are cooling significantly on a dynamical timescale. In the *slow cooling regime*, we have $\gamma_{\min} \ll \gamma_c$ and electrons injected near Lorentz factor γ_{\min} retain their energies for a dynamical time — the radiative efficiency is low in this case. In the *fast cooling regime*, we have $\gamma_{\min} \gg \gamma_c$ and electrons lose nearly all their energy in radiation — the radiative efficiency is close to 100%.

Let us first consider the slow cooling regime. At $\gamma \gg \gamma_c$, the electrons are cooling rapidly with $t_{\text{cool}} \ll t_{\text{dyn}}$, and n'_e quickly reach an equilibrium solution on a timescale of t_{cool} . The

⁶Here the effect of adiabatic loss is crudely accounted for by imposing a finite dynamical timescales t_{dyn} , during which the system expands significantly such that the electrons older than a dynamical time t_{dyn} are no longer dominating the emission. This means that the emitting electrons at any given moment are dominated by those injected within the one (current) dynamical time.

consequence of the equilibrium is that the distribution function at $\gamma \gg \gamma_c$ approaches a steady state with $\partial_t n'_e \approx 0$ and hence n'_e is only a function of γ , as given by

$$\frac{d}{d\gamma}(\dot{\gamma}n'_e) \approx S(\gamma) \rightarrow n'_e(\gamma \gg \gamma_c) \approx \frac{1}{|\dot{\gamma}|} \int_{\gamma}^{\infty} S(\gamma) d\gamma = \frac{n_{e,0}}{|\dot{\gamma}| t_{\text{dyn}}} (\gamma/\gamma_{\min})^{1-p} \propto \gamma^{-p-1}. \quad (8.100)$$

A mathematical way of understanding the above equilibrium solution is that the $\partial_t n'_e$ term on the LHS of the continuity equation is smaller than the other two terms by a factor of the order $t_{\text{cool}}/t_{\text{dyn}} \ll 1$.

On the other hand, at low Lorentz factors $\gamma \ll \gamma_c$, the cooling term can be ignored and, since the injection lasts for a dynamical time, we simply have

$$n'_e(\gamma \ll \gamma_c) \approx t_{\text{dyn}} S(\gamma) = \frac{n_{e,0}(p-1)}{\gamma_{\min}} (\gamma/\gamma_{\min})^{-p}. \quad (8.101)$$

Sometimes, when the B-fields and/or the radiation field are strong, the system may be in the fast cooling regime. In this case, electrons can cool down to Lorentz factors $\sim \gamma_c \ll \gamma_{\min}$. At $\gamma_c \ll \gamma < \gamma_{\min}$, since no electrons are injected ($S(\gamma) = 0$) while n'_e can quickly reach equilibrium ($\partial_t n'_e \approx 0$), so we have

$$\frac{d}{d\gamma}(\dot{\gamma}n'_e) \approx 0 \rightarrow n'_e \approx \frac{n_{e,0}}{|\dot{\gamma}| t_{\text{dyn}}} \propto \gamma^{-2}, \quad (8.102)$$

where we have used the following as normalization (from eq. 8.100)

$$(\dot{\gamma}n'_e)|_{\gamma_{\min}} = \frac{n_{e,0}}{|\dot{\gamma}(\gamma_{\min})| t_{\text{dyn}}}. \quad (8.103)$$

To summarize, we have obtained the following power-law scalings

$$\frac{dn_e}{d\gamma} \propto \begin{cases} \gamma^{-2}, & \text{for } \gamma_c \ll \gamma < \gamma_{\min}, \\ \gamma^{-p}, & \text{for } \gamma_{\min} < \gamma \ll \gamma_c, \\ \gamma^{-p-1}, & \text{for } \gamma \gg \gamma_c \text{ and } \gamma > \gamma_{\min}. \end{cases} \quad (8.104)$$

We define a characteristic frequency corresponding to the cooling Lorentz factor γ_c as $\nu_c = 3\gamma_c^2 \nu_B / 2$. Then, the broad-band synchrotron emissivity spectrum has the following power-law scalings

$$j_{\nu} \propto \begin{cases} \nu^{1/3}, & \text{for } \nu \ll \min(\nu_{\min}, \nu_c), \\ \nu^{-1/2}, & \text{for } \nu_c \ll \nu \ll \nu_{\min}, \\ \nu^{-(p-1)/2}, & \text{for } \nu_{\min} \ll \nu \ll \nu_c, \\ \nu^{-p/2}, & \text{for } \nu \gg \max(\nu_{\min}, \nu_c). \end{cases} \quad (8.105)$$

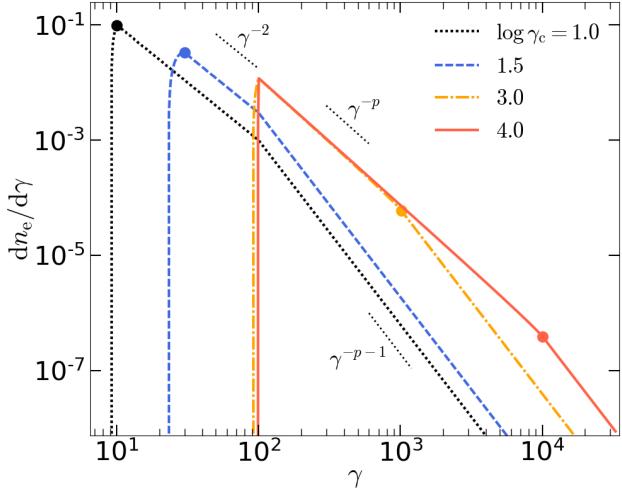


Figure 8.6: The Lorentz factor distributions after cumulative injection for a dynamical time, under different cooling rates as specified by the cooling Lorentz factors γ_c (positions marked by dots). In all cases, the minimum Lorentz factor in the injection function is fixed at $\gamma_{\min} = 100$ and the power-law index is $p = 2.2$. The distribution functions are normalized by $\int (dn_e/d\gamma) d\gamma = 1$.

This is still not the end of the story, because the observed spectrum is further affected by synchrotron self-absorption, which will be discussed in §8.4.

Sometimes it is desirable to accurately solve eq. (8.98) for the electron distribution function. This can be done using the *method of characteristics*. This is a standard method in solving partial differential equations involving advection.

The first step is to obtain the cooling trajectory $\gamma(t)$ of a given electron injected at time t_0 and Lorentz factor γ_0 — this trajectory is called the *characteristic* of the partial differential equation. Since $\dot{\gamma} = -\gamma/t_{\text{cool}} = -\gamma^2/(t_{\text{dyn}}\gamma_c)$, we can easily integrate this equation from time t_0 to t , and this gives

$$\gamma_0(\gamma, t, t_0) = \left[\gamma^{-1} - \frac{(t - t_0)}{\gamma_c t_{\text{dyn}}} \right]^{-1}, \quad (8.106)$$

where we must restrict the injection time to be in the range

$$\max \left(0, 1 - \frac{\gamma_c t_{\text{dyn}}}{\gamma} \right) \equiv t_{0,\min} < \frac{t_0}{t} < 1. \quad (8.107)$$

The meaning of the above expression is that, if at the current time t we have an electron of Lorentz factor γ , then its initial Lorentz factor was γ_0 at injection time t_0 . Differentiating

γ_0 with respect to γ , we obtain

$$d\gamma_0 = \left[1 - \frac{\gamma}{\gamma_c} \left(1 - \frac{t_0}{t_{\text{dyn}}} \right) \right]^{-2} d\gamma. \quad (8.108)$$

The electrons within a Lorentz factor bin of $\gamma \rightarrow \gamma + d\gamma$ at the current time t were injected from a Lorentz factor bin of $\gamma_0 \rightarrow \gamma_0 + d\gamma_0$, provided that we fix the injection time t_0 . Of course, these electrons may have a range of injection time t_0 within the allowed range in eq. (8.107). Thus, the number density of electrons in the $\gamma \rightarrow \gamma + d\gamma$ bin at time t is given by an integral over the injection time

$$dn_e = \int_{t_{0,\min}}^t dt_0 S(\gamma_0) d\gamma_0 \rightarrow \frac{dn_e}{d\gamma}(\gamma, t) = \int_{t_{0,\min}}^t dt_0 \frac{S[\gamma_0(\gamma, t, t_0)]}{\left[1 - \frac{\gamma}{\gamma_c} \left(1 - \frac{t_0}{t_{\text{dyn}}} \right) \right]^2}. \quad (8.109)$$

The solution we are interested in is the electron distribution function at $t = t_{\text{dyn}}$ (i.e., after the cumulative injection for a dynamical time), and this is given by

$$\frac{dn_e}{d\gamma}(\gamma, t_{\text{dyn}}) = \frac{n_{e,0}(p-1)}{\gamma_{\min}} \int_{x_{\min}}^1 dx \left[\frac{\gamma_0(\gamma, x)}{\gamma_{\min}} \right]^{-p} \left[1 - \frac{\gamma}{\gamma_c} (1-x) \right]^{-2}, \quad x \equiv t_0/t_{\text{dyn}}, \quad (8.110)$$

where $x_{\min} = \max(0, 1 - \gamma_c/\gamma)$ and, additionally, we must also restrict the integral to $\gamma_0(\gamma, x) > \gamma_{\min}$ since no electrons are injected below γ_{\min} . The results are shown in Fig. 8.6 and are in agreement with the anticipated broken power-law scalings.

8.3.4 *Synchrotron burn-off limit — maximum photon energy

Since $P_{\text{syn}} = -\dot{\gamma}m_e c^2 \propto B^2 \gamma^2$, the most energetic electrons in a given system cool off extremely rapidly due to synchrotron emission. If the synchrotron cooling timescale $t_{\text{cool}} \propto \gamma/P \propto B^{-2} \gamma^{-1}$ becomes shorter than the inverse of the gyro-frequency $(\omega_B/\gamma)^{-1} \propto \gamma B^{-1}$, the picture of a helical orbit in our treatment of synchrotron emission breaks down due to the importance of radiative drag forces. In fact, in such a situation, we do not expect a given physical system to be able to accelerate electrons to such high energies in the first place. This is because particle acceleration relies on the electric field, which is typically comparable to or weaker than the B-field (in Gauss units) in a typical system, and this means that the timescale for particle acceleration is comparable to or longer than the inverse of the gyro-frequency. From $t_{\text{syn}}(\gamma_{\max}) = (\omega_B/\gamma_{\max})^{-1}$, we obtain the maximum electron Lorentz factor

$$\gamma_{\max} = \sqrt{\frac{6\pi e}{B\sigma_T}} \sim 1.2 \times 10^8 (B/G)^{-1/2}, \quad (8.111)$$

which corresponds to a maximum synchrotron photon energy (aka the *synchrotron burn-off limit*, [de Jager et al. 1996](#))

$$\epsilon_{\max} = h\nu_{\text{syn}}(\gamma_{\max}) \simeq \frac{9m_e c^2}{4\alpha} \simeq 160 \text{ MeV}, \quad (8.112)$$

where $\alpha = e^2/(\hbar c) = 1/137$ is the fine-structure constant. Note that ϵ_{\max} is purely given by fundamental constants (independent of the B-field strength), so it should be a universal number. It should also be noted that if the emitting plasma has bulk motion with Lorentz factor Γ , then it is possible to Doppler boost the observed maximum synchrotron photon energy up to $2\Gamma\epsilon_{\max}$ provided that the plasma is moving towards the observer.

8.4 Synchrotron absorption

For any radiative processes, once we know the single-particle emissivity as well as the particles' distribution function, it is straightforward to use the generalized Kirchhoff's law (or the Einstein relations) to calculate the absorption coefficient of the system.

8.4.1 Applying the generalized Kirchhoff's law

Let the phase-space distribution function be $f(\mathbf{r}, \mathbf{p}, t)$, which describes the number of electrons per phase space volume $d^3\mathbf{r} d^3\mathbf{p}$ near momentum \mathbf{p} , position \mathbf{r} and time t . Here, we consider a small physical volume within which the distribution function is spatially uniform and we also consider a small time interval during which the distribution function does not change appreciably. Since we only care about the dependence of f on the momentum \mathbf{p} , we use $f(\mathbf{p})$ to denote the *momentum distribution only* and normalize it such that $\int f(\mathbf{p})d^3\mathbf{p} = 1$.

Because the 6-dimensional phase-space is quantized with individual volume quanta of h^3 and each volume element contains two spin states, the electron occupation number⁷ near momentum \mathbf{p} is given by

$$\eta(\mathbf{p}) = \frac{f(\mathbf{p})}{2h^3n_e}, \quad (8.113)$$

where n_e is the total electron number density.

In this section, we will consider an isotropic momentum distribution for simplicity, although the method below can be generalized for any distribution. In this isotropic case, the occupation number only depends on the magnitude of momentum $\eta(\mathbf{p}) = \eta(p) = f(p)/(2h^3n_e)$. For ultra-relativistic electrons $p \approx \gamma m_e c$, we write

$$f(\mathbf{p})d^3\mathbf{p} = f(p)4\pi p^2 dp = 4\pi(m_e c)^3 f(p)\gamma^2 d\gamma. \quad (8.114)$$

The Lorentz factor distribution is given by $(dn_e/d\gamma)d\gamma = n_e 4\pi(m_e c)^3 f(p)\gamma^2 d\gamma$, and hence

$$f(p) = \frac{dn_e/d\gamma}{n_e 4\pi(m_e c)^3 \gamma^2}. \quad (8.115)$$

⁷The electron occupation number is defined as the average number of electrons in a given distinct quantum state (with no degeneracy).

One can verify the normalization $\int f(p)4\pi p^2 dp = n_e^{-1} \int (dn_e/d\gamma) d\gamma = 1$.

We first calculate the ratio between the electron occupation numbers at two different Lorentz factors γ_1 and γ_2 that are closely separated by $\gamma_2 - \gamma_1 = h\nu/m_e c^2 \ll \gamma_1$. This ratio is given by

$$\frac{\eta(\gamma_1)}{\eta(\gamma_2)} = \frac{\left(\gamma^{-2}\frac{dn_e}{d\gamma}\right)_{\gamma_1}}{\left(\gamma^{-2}\frac{dn_e}{d\gamma}\right)_{\gamma_2}} = 1 - \frac{\frac{d}{d\gamma} \left(\gamma^{-2}\frac{dn_e}{d\gamma}\right)_{\gamma_1}}{\left(\gamma^{-2}\frac{dn_e}{d\gamma}\right)_{\gamma_1}} (\gamma_2 - \gamma_1), \quad (8.116)$$

where we have Taylor-expanded the function $\gamma^{-2}dn_e/d\gamma$ in the denominator since $\gamma_2 - \gamma_1$ is small. An equivalent way of describing the electron occupation number ratio is the excitation temperature, which is defined as

$$\eta(\gamma_1)/\eta(\gamma_2) \equiv \exp(h\nu/k_B T_{\text{exc}}). \quad (8.117)$$

For $\eta(\gamma_1)/\eta(\gamma_2) > 1$, i.e. the occupation number of the lower energy state is higher than that of the higher energy state, the excitation temperature is positive. In the opposite case of $\eta(\gamma_1)/\eta(\gamma_2) < 1$, the excitation temperature is negative, which is called *population inversion* — this is a necessary (but not sufficient) condition for a negative absorption coefficient. Under population inversion, we will potentially have exponentially growing intensity leading to bright maser emission, which will be discussed in §8.4.3.

Since synchrotron absorption may only be significant at low-frequency bands, we consider the Rayleigh-Jeans limit of $h\nu \ll k_B|T_{\text{exc}}|$ and hence can further use Taylor expansion to obtain

$$\frac{\eta(\gamma_1)}{\eta(\gamma_2)} \approx 1 + \frac{h\nu}{k_B T_{\text{exc}}}. \quad (8.118)$$

Therefore, the excitation temperature of the system near any Lorentz factor γ is given by the local slope of the Lorentz factor distribution,

$$\frac{k_B T_{\text{exc}}(\gamma)}{m_e c^2} = -\frac{\left(\gamma^{-2}\frac{dn_e}{d\gamma}\right)}{\frac{d}{d\gamma} \left(\gamma^{-2}\frac{dn_e}{d\gamma}\right)} = -\left[\frac{d}{d\gamma} \ln \left(\gamma^{-2}dn_e/d\gamma\right)\right]^{-1}. \quad (8.119)$$

Now we pause and verify that for the Maxwell-Jüttner distribution in eq. (8.91) in the relativistic limit gives $\gamma^{-2}dn_e/d\gamma \propto \exp(-\gamma/\Theta)$ and hence $k_B T_{\text{exc}}/(m_e c^2) = \Theta$ (independent of the Lorentz factor) or

$$T_{\text{exc}} = T, \text{ for thermal distribution.} \quad (8.120)$$

For the power-law distribution in eq. (8.84), we obtain $\gamma^{-2}dn_e/d\gamma \propto \gamma^{-2-p}$ and hence

$$\frac{k_B T_{\text{exc}}(\gamma)}{m_e c^2} = \gamma/(p+2), \text{ for power-law distribution.} \quad (8.121)$$

The linear dependence on the Lorentz factor, $T_{\text{exc}} \propto \gamma$, is expected because the electron's kinetic energy is $\gamma m_e c^2$ and this corresponds to an effective electron temperature $k_B T$ of the order $\gamma m_e c^2$.

We recall that *pitch-angle-averaged* single-particle emissivity (in units of $\text{erg s}^{-1} \text{Hz}^{-1} \text{sr}^{-1}$) for electrons with Lorentz factor γ is given by (cf. eq. 8.76)

$$\xi_\nu(\gamma) = \frac{1}{4\pi} \frac{\sqrt{3}e^3 B}{m_e c^2} R(y), \quad y \equiv \frac{2\nu}{3\gamma^2 \nu_B}. \quad (8.122)$$

The generalized Kirchhoff's law [Chapter 2] tells us that the *net* absorption cross-section is proportional to the single-particle emissivity and that the ratio between the two is the Planck function evaluated at the excitation temperature. This gives the (net) absorption cross-section of an average electron at Lorentz factor γ ,

$$\sigma_\nu(\gamma) = \frac{\xi_\nu(\gamma)}{B_\nu(T_{\text{exc}}(\gamma))}. \quad (8.123)$$

The (net) absorption coefficient α_ν at a given frequency ν by all electrons is given by the net absorption cross-section multiplied by the total number density, but since electrons with different Lorentz factors do not have the same cross-section, the result is an integral over the Lorentz factor distribution

$$\alpha_\nu = \int d\gamma \frac{dn_e}{d\gamma} \sigma_\nu(\gamma) = \int d\gamma \frac{dn_e}{d\gamma} \frac{\xi_\nu(\gamma)}{B_\nu(T_{\text{exc}}(\gamma))}. \quad (8.124)$$

The above result applies to any Lorentz factor distribution $dn_e/d\gamma$ and angle-averaged single-particle emissivity $\xi_\nu(\gamma)$. The only assumption here is that the electron momentum distribution is isotropic.

Practically, since we would only care about synchrotron self-absorption at low frequencies $h\nu \ll k_B T_{\text{exc}} \sim \gamma m_e c^2$, the Planck function is in the Rayleigh-Jeans limit $B_\nu(T_{\text{exc}}) = 2k_B T_{\text{exc}}/\lambda^2$ (where $\lambda = c/\nu$), and hence the absorption coefficient is given by

$$\begin{aligned} \alpha_\nu &= \frac{\lambda^2}{2} \int d\gamma \frac{dn_e}{d\gamma} \frac{\xi_\nu(\gamma)}{k_B T_{\text{exc}}} = \frac{\lambda^2}{2m_e c^2} \int d\gamma \xi_\nu(\gamma) \frac{d}{d\gamma} \left(\frac{dn_e/d\gamma}{\gamma^2} \right) \\ &= \frac{\lambda^2}{2m_e c^2} \int d\gamma \frac{\xi_\nu(\gamma)}{\gamma^3} \left(\gamma \frac{d^2 n_e}{d\gamma^2} - \frac{dn_e}{d\gamma} \right). \end{aligned} \quad (8.125)$$

The above expression is identical to the result (eq. 6.50) in Rybicki & Lightman's book.

For the general case of an anisotropic electron momentum distribution $f(\mathbf{p})d^3\mathbf{p}$ and an angle-dependent emissivity $\xi_\nu(\mathbf{p}, \mathbf{k})$ along the direction specified by the wavenumber \mathbf{k} , the absorption coefficient at frequency ν along the direction of \mathbf{k} is given by (see Ch 2)

$$\alpha_\nu(\mathbf{k}) = \int d^3\mathbf{p} f(\mathbf{p}) \frac{\xi_\nu(\mathbf{p}, \mathbf{k})}{B_\nu(T_{\text{exc}}(\mathbf{p}))}. \quad (8.126)$$

We will return to this point later in §8.4.3 when considering the conditions for maser.

For now, we still consider the simpler case of an isotropic momentum distribution. Let us then specify the shape of the Lorentz factor distribution.

Thermal distribution

The first case is a thermal distribution, for which we have $T_{\text{exc}} = T$, independent of the electron's Lorentz factor. Since the specific volume emissivity is $j_\nu = \int d\gamma (dn_e/d\gamma) \xi_\nu(\gamma)$, we simply have $\alpha_\nu = j_\nu/B_\nu(T)$, as expected. This leads to

$$\alpha_\nu = \frac{3^{3/2}}{(8\pi)^2} \frac{n_e B \lambda^2 \sigma_T}{2e\Theta} I(z) = 5.12 \times 10^{-15} \text{ cm}^{-1} \nu_{\text{GHz}}^{-2} \frac{B}{\text{G}} \frac{n_e}{\text{cm}^{-3}} \frac{I(z)}{\Theta}, \quad z = \frac{2\nu}{3\Theta^2 \nu_B}, \quad (8.127)$$

where σ_T is the Thomson cross-section and $\lambda = c/\nu$ is the wavelength. The average absorption cross-section per electron is given by $\bar{\sigma}_\nu = \alpha_\nu/n_e$ and hence

$$\frac{\bar{\sigma}_\nu}{\sigma_T} \approx \frac{3^{3/2}}{128\pi^2} \frac{B \lambda^2}{e\Theta} I(z) = 7.7 \times 10^9 \nu_{\text{GHz}}^{-2} (B/\text{G}) \frac{I(z)}{\Theta}, \quad \Theta = \frac{k_B T}{m_e c^2} = 1.69 \frac{T}{10^{10} \text{ K}}. \quad (8.128)$$

This can be compared with the free-free absorption cross-section per electron for a thermal plasma of temperature T (see Chapter 6)

$$\frac{\bar{\sigma}_{\text{ff},\nu}}{\sigma_T} \simeq 2.7 \times 10^{-2} \nu_{\text{GHz}}^{-2} (T/10^4 \text{ K})^{-3/2} \frac{\sum_i g_{\text{ff},i} Z_i^2 n_i}{\text{cm}^{-3}}. \quad (8.129)$$

The synchrotron absorption cross-section for relativistic electrons is typically much higher than σ_T for radio frequencies and it scales as the electron temperature as $\sigma_\nu \propto \Theta^{-1}$.

The absorption cross-section in eq. (8.128) roughly applies to other electron distributions as long as Θ is replaced by⁸ $\gamma/3$ (γ being the electron Lorentz factor in consideration), since the excitation temperature is roughly given by $3k_B T_{\text{exc}} \simeq \gamma m_e c^2$ (cf. eq. 8.121). For a given observer's frequency ν , the Lorentz factor is related to the B-field strength by $\gamma \simeq 15 \nu_{\text{GHz}}^{1/2} (B/\text{G})^{-1/2}$ (eq. 8.20), so we obtain

$$\bar{\sigma}_\nu/\sigma_T \simeq 1.5 \times 10^9 \nu_{\text{GHz}}^{-5/2} (B/\text{G})^{3/2}, \quad \text{for a given frequency } \nu, \quad (8.130)$$

where we have taken $I(z) \simeq 1$ near the characteristic frequency ($z \simeq 1$). On the other hand, if we would like to know whether most of the synchrotron emission from electrons near Lorentz factor γ are absorbed or not, then we use the characteristic frequency $\nu \simeq 4.2 \times 10^{-3} \text{ GHz} \gamma^2 (B/\text{G})$ (eq. 8.18, where $z \simeq 1$ and $I(z) \simeq 1$) and obtain

$$\bar{\sigma}_\nu/\sigma_T \simeq 1.3 \times 10^{15} \gamma^{-5} (B/\text{G})^{-1}, \quad \text{for a given Lorentz factor } \gamma. \quad (8.131)$$

We see that synchrotron self-absorption quickly becomes unimportant for electrons with sufficiently high Lorentz factors.

⁸The factor of 3 here is motivated by the fact that the average Lorentz factor of a relativistic Maxwellian distribution is $\langle \gamma \rangle = 3k_B T$ (eq. 7.88).

Power-law distribution

The next case we consider is a power-law electron distribution given by (cf. eq. 8.84)

$$dn_e/d\gamma = \frac{n_e(p-1)}{\gamma_{\min}} (\gamma/\gamma_{\min})^{-p}. \quad (8.132)$$

The results can be easily applied to the case of a broken power-law distribution (e.g., as in eq. 8.104), as long as we keep track of each of the power-law segments. We plug the single-particle emissivity (eq. 8.122) and excitation temperature (eq. 8.121) into eq. (8.124) for the absorption coefficient, and obtain

$$\begin{aligned} \alpha_\nu &= \int_{\gamma_{\min}}^{\gamma_{\max}} d\gamma \frac{p-1}{\gamma_{\min}} \left(\frac{\gamma}{\gamma_{\min}} \right)^{-p} \frac{c^2}{2\nu^2} \frac{p+2}{\gamma m_e c^2} \frac{n_e}{4\pi} \frac{\sqrt{3}e^3 B}{m_e c^2} R(y) \\ &= \frac{3^{3/2}}{128\pi^2} \frac{B\lambda^2 n_e \sigma_T}{e\gamma_{\min}} C_2 \left(\frac{\nu}{\nu_{\min}} \right)^{-p/2} \\ &= 5.12 \times 10^{-15} \text{ cm}^{-1} \nu_{\text{GHz}}^{-2} \frac{B}{G \text{ cm}^{-3}} \frac{C_2}{\gamma_{\min}} \left(\frac{\nu}{\nu_{\min}} \right)^{-p/2}, \\ C_2(p, \nu) &= (p-1)(p+2) \int_{y_{\min}}^{y_{\max}} dy y^{(p-2)/2} R(y), \quad y_{\min} = \frac{\nu}{\nu_{\max}}, \quad y_{\max} = \frac{\nu}{\nu_{\min}}. \end{aligned} \quad (8.133)$$

where $y = 2\nu/(3\gamma^2\nu_B)$ and $\nu_{\min/\max} = \gamma_{\min/\max}^2 (3\nu_B/2)$.

Similar to earlier calculation of the specific emissivity for a power-law distribution, we discuss various frequency regimes in the following.

At intermediate frequencies satisfying $\nu_{\min} \ll \nu \ll \nu_{\max}$, we have $y_{\min} \ll 1$ and $y_{\max} \gg 1$, so the limits in the C_2 integral can be extended to $C_2 \approx (p-1)(p+2) \int_0^\infty dy (\dots)$, and

$$C_2(p, \nu_{\min} \ll \nu \ll \nu_{\max}) = \pi^{\frac{1}{2}} 2^{p/2-2} (p-1)(p+10/3) \frac{\Gamma\left(\frac{p+6}{4}\right) \Gamma\left(\frac{3p+2}{12}\right) \Gamma\left(\frac{3p+10}{12}\right)}{\Gamma\left(\frac{p+8}{4}\right)}, \quad (8.134)$$

which only depends on p . This function $C_2(p)$ is shown in Fig. 8.5. At high frequencies $\nu \gg \nu_{\max}$, we have $y_{\min} \gg 1$ and $y_{\max} \rightarrow \infty$, and hence

$$C_2(p, \nu \gg \nu_{\max}) = \frac{\pi(p-1)(p+2)}{2} \left[\Gamma_{\text{inc}}\left(\frac{p}{2}, \frac{\nu}{\nu_{\max}}\right) - \frac{99}{162} \Gamma_{\text{inc}}\left(\frac{p-2}{2}, \frac{\nu}{\nu_{\max}}\right) \right], \quad (8.135)$$

which means a rapid cutoff at $\nu \gg \nu_{\max}$. The upper incomplete Gamma function Γ_{inc} is given by eq. (8.89). At low frequencies $\nu \ll \nu_{\min}$, we have $y_{\min} \approx 0$ and $y_{\max} \ll 1$, and

$$C_2(p, \nu \ll \nu_{\min}) \left(\frac{\nu}{\nu_{\min}} \right)^{-p/2} = \frac{6}{5} 2^{1/3} \Gamma^2(1/3) \frac{(p-1)(p+2)}{3p+2} \left(\frac{\nu}{\nu_{\min}} \right)^{1/3}. \quad (8.136)$$

Finally, we obtain the following source function scalings for a power-law distribution

$$S_\nu = \frac{j_\nu}{\alpha_\nu} \propto \begin{cases} \nu^2, & \text{for } \nu \ll \nu_{\min}, \\ \nu^{5/2}, & \text{for } \nu_{\min} \ll \nu \ll \nu_{\max}. \end{cases} \quad (8.137)$$

For a thermal distribution, we have $S_\nu \propto \nu^2$ in the low-frequency limit $\nu \ll \nu_\Theta \equiv 3\Theta^2\nu_B/2$. In fact, the source functions at low frequencies ($\nu \ll \nu_\Theta$ and $\nu \ll \nu_{\min}$) would be similar for the cases of thermal and power-law distributions, provided that one does the following replacement $\Theta \leftrightarrow \gamma_{\min}$. This is because the emission and absorption are both mainly due to electrons with Lorentz factors $\gamma \sim \Theta$ or $\sim \gamma_{\min}$ in these two cases.

8.4.2 Combining emission and absorption

Finally, we are ready to discuss the broad-band spectrum of a plasma slab accounting for both emission and absorption. Let us assume that the slab has a spatially uniform source function $S_\nu = j_\nu/\alpha_\nu$. Solving the radiative transfer equation $dI_\nu/d\tau_\nu = -I_\nu + S_\nu$, we obtain

$$I_\nu(\tau_\nu) = S_\nu + [I_\nu(0) - S_\nu]e^{-\tau_\nu}, \quad (8.138)$$

where $I_\nu(0)$ is the incident intensity, $\tau_\nu = \int_0^{s_0} \alpha_\nu ds$ is the optical depth of the slab along the line of sight, and s_0 is the geometrical thickness. For simplicity, let us ignore the incident intensity by taking $I_\nu(0) = 0$, and hence

$$I_\nu(\tau_\nu) = S_\nu(1 - e^{-\tau_\nu}), \quad (8.139)$$

We expect the slab to be optically thick at low frequencies, and hence the emerging intensity is given by $I_\nu(\tau_\nu \gg 1) \approx S_\nu$. At sufficiently high frequencies, the slab is optically thin, and the intensity is given by $I_\nu(\tau_\nu \ll 1) \approx j_\nu s_0$.

Let us take the case of a single power-law electron distribution in eq. (8.84) as an example. The key is to find the *self-absorption frequency* ν_a at which $\tau_{\nu_a} = \alpha_{\nu_a} s_0 = 1$. After determining ν_a , we know that the spectrum will be a broken power-law with the following scalings

$$I_\nu \begin{cases} \propto \nu^2, & \text{for } \nu \ll \min(\nu_a, \nu_{\min}), \\ \propto \nu^{5/2}, & \text{for } \nu_{\min} \ll \nu \ll \nu_a, \\ \propto \nu^{1/3}, & \text{for } \nu_a \ll \nu \ll \nu_{\min}, \\ \propto \nu^{(1-p)/2}, & \text{for } \nu \gg \max(\nu_{\min}, \nu_a), \\ \approx 0, & \text{for } \nu \gg \nu_{\max}. \end{cases} \quad (8.140)$$

If we include electron cooling effects, then the broken power-laws are a bit more complicated. The numerical results for broad-band spectral intensity (νI_ν) for different electron column densities are shown in Fig. 8.7.

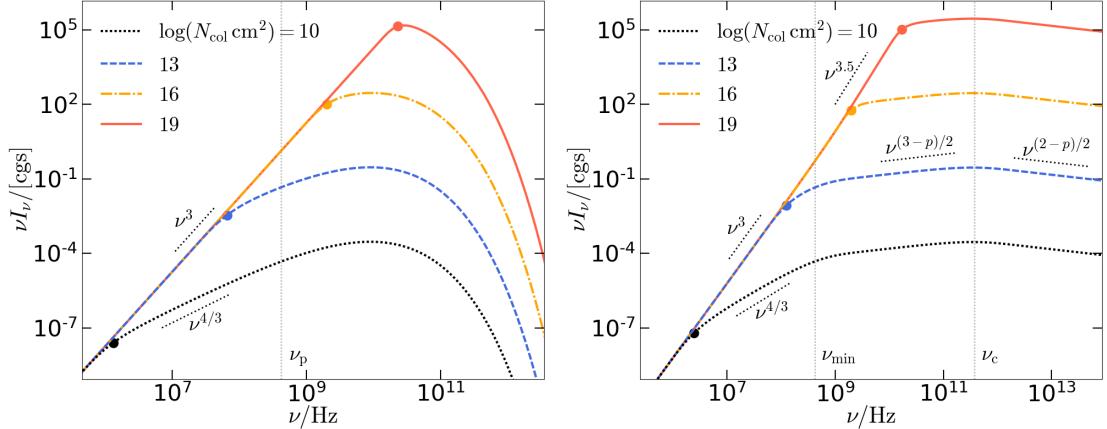


Figure 8.7: The synchrotron spectrum for a slab with different electron column densities N_{col} and fixed $B = 1 \text{ G}$. The self-absorption frequencies ν_a ($\tau_{\nu_a} = 1$) are marked by dots. *Left panel:* Thermal electron Lorentz factor distribution with $\Theta = 10$. The grey vertical dotted lines mark the characteristic frequency $\nu_p = 3\Theta^2\nu_B/2$ (or “peak frequency” in the ν vs. I_ν plot). *Right panel:* Broken power-law distribution with indices $-p$ and $-p-1$ below and above γ_c (respectively), and the parameters are: $\gamma_{\min} = 10$, $\gamma_c = 300$, and $p = 2.5$. The grey vertical dotted lines mark $\nu_{\min} = 3\gamma_{\min}^2\nu_B/2$ and $\nu_c = 3\gamma_c^2\nu_B/2$.

In fact, we can roughly estimate the self-absorption frequency ν_a using the generalized Kirchhoff’s law. We first define the self-absorption Lorentz factor γ_a by

$$\gamma_a = \max \left[\sqrt{\frac{2\nu_a}{3\nu_B}}, \gamma_{\min} \right], \quad (8.141)$$

where we have restricted $\gamma_a \geq \gamma_{\min}$. The number density of electrons near γ_a is roughly $(\gamma d n_e / d\gamma)|_{\gamma_a} \sim n_e(\gamma_a/\gamma_{\min})^{1-p}$, so the specific emissivity near ν_a is roughly given by

$$j_{\nu_a} \sim \left(\gamma \frac{d n_e}{d\gamma} \right) \Big|_{\gamma_a} \frac{P_{\nu_p}}{4\pi} \min \left[1, (\nu_a/\nu_{\min})^{1/3} \right], \quad P_{\nu_p} \simeq \frac{\sqrt{3}e^3 B}{m_e c^2}, \quad (8.142)$$

where the factor of $(\nu_a/\nu_{\min})^{1/3}$ would be important if $\nu_a < \nu_{\min}$ (and in this case $\gamma_a = \gamma_{\min}$). The excitation temperature for electrons with Lorentz factors near γ_a is roughly given by $3k_B T_{\text{exc}} \simeq \gamma_a m_e c^2$, so we obtain $B_{\nu_a}(T_{\text{exc}}) = (2/3)(\nu_a/c)^2 \gamma_a m_e c^2$ (in the Rayleigh-Jeans limit). Finally, from $1 = \alpha_{\nu_a} s_0 \sim j_{\nu_a} s_0 / B_{\nu_a}(T_{\text{exc}})$, we write

$$N_{\text{col}}(\gamma_a/\gamma_{\min})^{1-p} \frac{\sqrt{3}e^3 B}{4\pi m_e c^2} \min \left[1, (\nu_a/\nu_{\min})^{1/3} \right] \simeq \frac{2}{3} \nu_a^2 \gamma_a m_e, \quad (8.143)$$

where $N_{\text{col}} = n_e s_0$ is the total electron column density along the line of sight. Knowing γ_{\min} , B and N_{col} , one can solve the above equation first for γ_a and then for ν_a . If $\nu_a < \nu_{\min}$,

the solution is given by

$$\nu_a \simeq \left[\frac{N_{\text{col}}}{\gamma_{\min} \nu_{\min}^{1/3}} \frac{3\sqrt{3}e^3 B}{8\pi m_e^2 c^2} \right]^{3/5} \simeq 6 \text{ GHz} \gamma_{\min}^{-1} \left(\frac{B}{G} \right)^{2/5} \left(\frac{N_{\text{col}}}{10^{14} \text{ cm}^{-2}} \right)^{3/5}, \text{ for } \nu_a < \nu_{\min}. \quad (8.144)$$

If $\nu_a > \nu_{\min}$, the solution is given by

$$\gamma_a \simeq \left[\frac{2\pi e N_{\text{col}}}{\sqrt{3} B} \gamma_{\min}^{p-1} \right]^{\frac{1}{p+4}} \rightarrow \nu_a = \frac{3}{2} \gamma_a^2 \nu_B \simeq 4 \text{ MHz} \gamma_{\min}^{\frac{2(p-1)}{p+4}} \left(\frac{2\pi e N_{\text{col}}}{\sqrt{3} B} \right)^{\frac{2}{p+4}} \frac{B}{G}, \text{ for } \nu_a > \nu_{\min}. \quad (8.145)$$

Due to unknown power-law index p , it is difficult to put numbers in the above expression. Combining these two regimes, the final solution can be written as

$$\nu_a \simeq \min \left[6 \text{ GHz} \gamma_{\min}^{-1} \left(\frac{B}{G} \right)^{\frac{2}{5}} \left(\frac{N_{\text{col}}}{10^{14} \text{ cm}^{-2}} \right)^{\frac{3}{5}}, 4 \text{ MHz} \gamma_{\min}^{\frac{2(p-1)}{p+4}} \left(\frac{2\pi e N_{\text{col}}}{\sqrt{3} B} \right)^{\frac{2}{p+4}} \frac{B}{G} \right]. \quad (8.146)$$

The above estimate is very close to the numerical results shown in Fig. 8.7.

8.4.3 *Conditions for synchrotron maser (under construction)

Synchrotron maser is possible when the electrons' pitch-angle distribution is far from isotropic. (under construction)

8.5 *Polarization

Earlier in this chapter, we have considered the case of randomly oriented B-fields, and in that limiting case the synchrotron emission is completely unpolarized. In this section, we consider the opposite limit of an ordered, uniform B-field.

The observer's line of sight \hat{n} is fixed at an angle $\tilde{\theta}$ away from the B-field direction, i.e. $\cos \tilde{\theta} = \hat{n} \cdot \hat{B}$. As we see in Fig. 8.2, the emission from a single particle with a given pitch angle α is generally elliptically polarized. The sense of polarization (left or right) depends on the sign of the viewing angle in the comoving frame of the parallel velocity $\theta' \approx (\alpha - \tilde{\theta}) / \sin \alpha$ (cf. eq. 8.57), i.e., for $\tilde{\theta} > \alpha$, we get one direction of E-field rotation; and for $\tilde{\theta} < \alpha$ we get the opposite direction of rotation. Since nearly all the observed emission comes from particles within a narrow range of pitch angles $|\alpha - \tilde{\theta}| \lesssim 1/\gamma \ll 1$, for any reasonably broad pitch-angle distribution, there will be roughly equal number of emitting particles with $\theta' > 0$ and $\theta' < 0$. Therefore, there will be no net elliptical polarization⁹.

⁹If the emitting electrons are only mildly relativistic, the cancellation of left and right elliptical polarizations will not be perfect. Roughly speaking, we expect the degree of circular polarization to be a factor of γ^{-1} less than that of linear polarization.

Because the strengths of X-mode and O-mode field components are rather different, there will be net linear polarization. Suppose our telescope has two antennae arms: one is aligned with the X-mode unit vector \hat{e}_X , and the other one is aligned with the O-mode unit vector $\hat{e}_O = \hat{n} \times \hat{e}_X$ which is parallel to the B-field projected on the plane of the sky. In an actual observation, it is unlikely that the two antennae arms are arranged like this (before knowing the projected B-field direction), but it is straightforward to carry out a coordinate rotation to achieve the above setup.

Then, the frequency-integrated Stokes parameters are given by

$$I = \langle E_X^2 \rangle + \langle E_O^2 \rangle, \quad Q = \langle E_X^2 \rangle - \langle E_O^2 \rangle, \quad U = 0, \quad V = 0, \quad (8.147)$$

where $E_{j,\text{tot}}(t)$ are real field components for $j = X$ or O and $\langle \dots \rangle$ means averaging on a timescale $T_{\text{avg}} \gg 1/\nu$. Typically, particles' gyro-orbits are not temporally correlated, so each of the terms above is given by the incoherent sum of the contributions from all particles

$$\langle E_j^2 \rangle = \sum_k \langle E_{j,k}^2 \rangle = \frac{1}{T_{\text{avg}}} \sum_k \int_{-\infty}^{\infty} E_{j,k}^2(t) dt, \quad j = X \text{ or } O, \quad k = 1, 2, \dots, N, \quad (8.148)$$

where N is the number of pulses received in a time interval of T_{avg} (each emitting particle emits one short-duration pulse per orbital period). For each pulse k , we use Parseval's theorem

$$\int_{-\infty}^{\infty} E_{j,k}^2(t) dt = 4\pi \int_0^{\infty} |\tilde{E}_{j,k}(\omega)|^2 d\omega, \quad (8.149)$$

where $\tilde{E}_{j,k}(\omega) = \int_{-\infty}^{\infty} E_{j,k}(t) e^{i\omega t} dt$ is the Fourier transform of an individual component. Therefore, the frequency spectrum of the Stokes parameters are

$$I_{\omega} = \frac{dI}{d\omega} = \frac{4\pi}{T_{\text{avg}}} \sum_k \left[|\tilde{E}_{X,k}|^2 + |\tilde{E}_{O,k}|^2 \right], \quad Q_{\omega} = \frac{dQ}{d\omega} = \frac{4\pi}{T_{\text{avg}}} \sum_k \left[|\tilde{E}_{X,k}|^2 - |\tilde{E}_{O,k}|^2 \right]. \quad (8.150)$$

This shows that the Stokes-I (or intensity) spectrum is simply given by the sum of the spectrum of the radiation emitted by all particles along the observer's line of sight, whereas the Stokes-Q spectrum is given by the difference between the X-mode and O-mode spectral powers. Therefore, we obtain

$$I_{\nu} \propto j_{X,\nu}(\tilde{\theta}) + j_{O,\nu}(\tilde{\theta}), \quad Q_{\nu} \propto j_{X,\nu}(\tilde{\theta}) - j_{O,\nu}(\tilde{\theta}), \quad (8.151)$$

where the X-mode and O-mode emissivities are given by eq. (8.70) for arbitrary Lorentz factor and pitch angle distributions. For the special case of an isotropic pitch angle distribution (and a fixed viewing angle $\tilde{\theta}$), we obtain

$$I_{\nu} \propto \int d\gamma \frac{dn_e}{d\gamma} F(x), \quad Q_{\nu} \propto \int d\gamma \frac{dn_e}{d\gamma} G(x), \quad (8.152)$$

where $x = \nu/(\nu_c \sin \tilde{\theta})$ and $\nu_c = 3\gamma^2 eB/(4\pi m_e c)$.

For monoenergetic electron Lorentz factor distribution, we obtain the frequency-dependent polarization fraction

$$\Pi(x) = \frac{Q_\nu}{I_\nu} = \frac{G(x)}{F(x)} \in (0.5, 1), \quad (8.153)$$

which is a monotonic function that increases from 50% (at $x \ll 1$) to 100% at $x \gg 1$ (see the green dashed line in Fig. 8.4). The frequency-integrated polarization fraction is

$$\Pi_{\text{intgr}} = \frac{\int_0^\infty G(x)dx}{\int_0^\infty F(x)dx} = \frac{\Gamma(4/3)}{\Gamma(7/3)} = \frac{3}{4} = 75\%. \quad (8.154)$$

For a power-law electron Lorentz factor distribution $dn_e/d\gamma \propto \gamma^{-p}$ (for $\gamma_{\min} < \gamma < \gamma_{\max}$), then at intermediate frequencies $\nu_{\min} \ll \nu \ll \nu_{\max}$ (where $\nu_{\min/\max} = (3/2)\gamma_{\min/\max}^2 \nu_B \sin \tilde{\theta}$), the polarization fraction is given by

$$\Pi(\nu) = \frac{\int x^{(p-3)/2} G(x)dx}{\int x^{(p-3)/2} F(x)dx} = \frac{p+1}{p+7/3}. \quad (8.155)$$

For $2 < p < 3$, the polarization fraction is between 70% and 75%. The B-fields in realistic astrophysical sources are unlikely to be uniform, so the observed polarization fractions are usually smaller than 70%.

8.6 Synchrotron self-Compton

Coexisting synchrotron photons and relativistic electrons unavoidably lead to inverse-Compton emission called *synchrotron self-Compton* (SSC) process. To calculate the SSC spectrum, one must convolve the single-electron Klein-Nishina differential scattering cross-section with both the electron Lorentz factor distribution and the synchrotron seed photon spectrum. This is straightforward. The only complication is that if the radiation energy density is sufficiently high, then inverse-Compton cooling must be self-consistently included when calculating the electron Lorentz factor distribution (as given by the solution of the continuity equation). Here, we discuss the simplified case where the electron Lorentz factor distribution $dn_e/d\gamma$ is known.

8.6.1 Compton- y parameter again

In each inverse-Compton scattering, the photon gains energy by a factor of $(4/3)\gamma^2$, if we are in the Thomson limit. Thus, we define the following *Compton- y parameter*

$$y \equiv s_0 \int_{\gamma_{\min}}^{\gamma_{\text{KN}}} \frac{4}{3} \gamma^2 \frac{dn_e}{d\gamma} d\gamma, \quad (8.156)$$

where s_0 is the the geometrical thickness of the emitting plasma along the line of sight, $\gamma_{\text{KN}} \simeq m_e c^2 / (h\nu_{\text{p,syn}})$, and $\nu_{\text{p,syn}}$ is the peak frequency of the νL_ν spectrum of the synchrotron emission (provided that the synchrotron spectrum is reasonably sharply peaked).

The above definition is a generalization of the $y = (4/3)\gamma^2\pi_T$ case for monoenergetic electrons considered in 7. The Compton-y parameter gives the ratio between the SSC and synchrotron luminosities as well as ratio between the energy densities of SSC and synchrotron photons inside the emitting plasma, i.e.

$$y = \frac{L_{\text{ssc}}}{L_{\text{syn}}} = \frac{U_{\text{ssc}}}{U_{\text{syn}}}. \quad (8.157)$$

In the broad-band νL_ν spectral energy distribution (SED) of a given non-thermal source, the synchrotron component is usually well separated from the SSC component in frequency. In such a case, we can measure the peak frequencies of both components $\nu_{\text{p,syn}}$ and $\nu_{\text{p,ssc}}$. The ratio between them tells us the characteristic Lorentz factor γ_{ssc} that dominates the integral in the Compton-y parameter (eq. 8.156)

$$\gamma_{\text{ssc}} \sim \sqrt{\nu_{\text{p,ssc}} / \nu_{\text{p,syn}}}. \quad (8.158)$$

The column density of electrons with Lorentz factors near γ_{ssc} is then given by

$$N_{\text{col}}(\gamma_{\text{ssc}}) \sim \frac{y}{\gamma_{\text{ssc}}^2} \sim \frac{L_{\text{ssc}}}{\gamma_{\text{ssc}}^2 L_{\text{syn}}}. \quad (8.159)$$

By jointly fitting the broad-band SED, one can determine the B-field strength, electron Lorentz factor distribution, size of emitting region, and the bulk Lorentz factor of the source.

The SSC photons can act as the seed photons for the 2nd-order inverse-Compton scattering. This means that the luminosity of the 2nd-order scattered photons is given by $L_{\text{ssc2}} = yL_{\text{ssc}} = y^2L_{\text{syn}}$, provided that the 2nd-order scattering is in the Thomson regime ($\gamma_{\text{ssc}}^2 h\nu_{\text{p,ssc}} \ll m_e c^2$). One can see that, in a source with $y \gtrsim 1$, the electrons with Lorentz factors near γ_{ssc} may undergo run-away cooling, because the inverse-Compton emission they produce causes stronger cooling which then causes more inverse-Compton emission. We expect that the electrons that are responsible for $y \gtrsim 1$ would be quickly removed from the system (by emitting high energy gamma-ray photons). Typically, realistic sources either have $y \ll 1$ or the 2nd-order scattering is Klein-Nishina suppressed, and hence L_{ssc2} is usually weak.

8.6.2 Maximum brightness temperature

The brightness temperature of a spatially resolved source with broad-band intensity spectrum I_ν is defined as

$$T_b(\nu) = \frac{I_\nu c^2}{2\nu^2 k_B}, \quad (8.160)$$

which means that the brightness temperature is the highest at the self-absorption frequency ν_a where $\tau_{\nu_a} = 1$ (because at higher frequencies $\nu > \nu_a$, we have $d \ln I_\nu / d \ln \nu < 2$). The electrons emitting at ν_a are near the characteristic Lorentz factor $\gamma_a \simeq \sqrt{2\nu_a/3\nu_B}$ (here we assume that γ_a is above the minimum Lorentz factor γ_{\min} of the non-thermal electron population).

Since the synchrotron absorption optical depth is directly related to the electron column density $N_{\text{col}}(\gamma_a)$ and hence the Thomson scattering optical depth $\tau_T(\gamma_a)$ of the electrons near Lorentz factor γ_a , it is easy to Compton-y parameter for these electrons

$$y(\gamma_a) \simeq \frac{4}{3} \gamma_a^2 \tau_T(\gamma_a). \quad (8.161)$$

At frequency ν_a , the source is marginally optically thick, so the intensity is given by the Planck function at the local excitation temperature $I_{\nu_a} = B_\nu(T_{\text{exc}}) \approx 2\nu_a^2 k_B T_{\text{exc}}(\nu_a)/c^2$, where the excitation temperature is given by $3k_B T_{\text{exc}} \simeq \gamma_a m_e c^2$ and it is also equal to the brightness temperature $T_b = T_{\text{exc}}$. On the other hand, the source is marginally optically thin, so the intensity is also given by the line-of-sight-integrated emissivity $I_{\nu_a} = P_{\nu_p} N_{\text{col}}(\gamma_a)/(4\pi)$, where $P_{\nu_p} = \sqrt{3}e^3 B/(m_e c^2)$ is the peak spectral power for a single electron. Therefore, we obtain

$$\tau_T(\gamma_a) = N_{\text{col}} \sigma_T \simeq \frac{2m_e c \sigma_T}{\sqrt{3}e^2} \nu_a \gamma_a^3. \quad (8.162)$$

Then, the Compton-y parameter is given by

$$y(\gamma_a) \simeq \frac{8m_e c \sigma_T}{3\sqrt{3}e^2} \nu_a \gamma_a^5 = 4 \frac{\nu_a}{\text{GHz}} \left(\frac{T_b(\nu_a)}{10^{12} \text{K}} \right)^5. \quad (8.163)$$

This means that if the brightness temperature (typically measured in the radio band) exceeds 10^{12} K even by a modest factor, we have $y(\gamma_a) \gg 1$ and hence these electrons will rapidly cool due to their SSC emission and higher-order inverse-Compton emission. This is called the *inverse-Compton catastrophe* ([Kellermann & Pauliny-Toth 1969](#)). Realistic sources must have brightness temperature $T_b < 10^{12}$ K in the *comoving frame of their bulk motion*¹⁰. In fact, the constraint based on $y(\gamma_a) \lesssim 1$ is rather conservative, because electrons with higher Lorentz factors than γ_a may dominate the total Compton-y in eq. (8.156). Based on equipartition between the energy density of electrons near γ_a and that of the B-fields, [Readhead \(1994\)](#) proposed that the maximum brightness temperature in the comoving frame is closer to 10^{11} K.

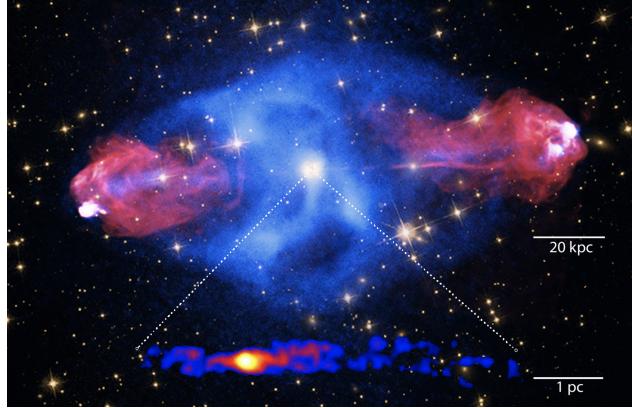


Figure 8.8: Radio (red), X-ray (blue) and optical image of the radio galaxy Cygnus A (taken from Figure 4 of the review paper by [Blandford, Meier & Readhead \(2019\)](#)). The three bright hot spots are located where the jets are strongly decelerated by shock interactions. The electrons in the hot spots produce bright synchrotron emission as well as X-ray emission by inverse-Compton scattering of the CMB and synchrotron photons. The giant X-ray “cocoon” surrounding the jet is produced by free-free emission from the shock-heated hot plasma. The Zoom-in image shows the nearly symmetric (due to a viewing angle close to 90°) jets launched from the accreting supermassive black hole at the center of the galaxy.

8.7 Homework

Prob. 21. Cygnus A (or 3C 405) is one of the brightest radio galaxies on the sky, as shown in Fig. 8.8. This galaxy is at a distance of about 200 Mpc. Accretion onto the supermassive black hole at its center launches a pair of powerful jets, which propagate to a distance of about 50 kpc into the circum-galactic medium. The ambient gas is shock-heated, and electrons are accelerated to very high Lorentz factors. Radio observations show that the flux density is $F_\nu \simeq 10^3$ Jy at $\nu \simeq 1$ GHz, which is also the peak frequency of the νF_ν spectrum (see [McKean et al. 2016](#)). The emitting region is spatially resolved and based on that, we infer a total emitting volume of $V \sim (30\text{ kpc})^3 \simeq 8 \times 10^{68} \text{ cm}^3$. Our goal is to obtain rough estimates for the properties of the system, including the total energy, B-field strength, cooling timescale, etc.

- (1) Estimate the total radio luminosity by calculating νL_ν at $\nu \simeq 1$ GHz. Note that $1\text{ Jy} = 10^{-23} \text{ erg cm}^{-2} \text{ s}^{-1} \text{ Hz}^{-1}$.
- (2) The number of electrons radiating near 1 GHz is denoted as N_e , and these electrons have Lorentz factor γ , so their total energy is $\gamma m_e c^2 N_e \simeq U_e V$, where U_e is the energy density of the radiating

¹⁰If the source is moving towards the observer with a relativistic bulk motion, then the observed brightness temperature as determined by the intensity (eq. 8.160) is higher than that in the comoving frame by a factor of the Doppler factor \mathcal{D} (since I_ν/ν^3 is Lorentz invariant and $\nu = \mathcal{D}\nu'$).

electrons. The total radio luminosity can then be expressed by

$$\nu L_\nu \sim \gamma^2 U_B \sigma_T c N_e \simeq \gamma \frac{\sigma_T}{m_e c} U_B U_e V. \quad (8.164)$$

Making use of $\nu \sim \gamma^2 \nu_B$ and $B = \sqrt{8\pi U_B}$, we can roughly estimate the Lorentz factors of the electrons radiating at $\nu \simeq 1 \text{ GHz}$,

$$\gamma \simeq (8\pi)^{-1/4} U_B^{-1/4} \sqrt{\frac{m_e c \nu}{e}}. \quad (8.165)$$

Let us further define a dimensionless *equipartition parameter* $f = U_e/U_B$, and then the radio luminosity can be written as

$$\nu L_\nu \sim (8\pi)^{-1/4} \sqrt{\frac{\nu \sigma_T^2}{em_e c}} f U_B^{7/4} V. \quad (8.166)$$

The total energy contained in electrons and B-fields is given by

$$E = (U_e + U_B)V = (1 + f)U_B V \rightarrow U_B V = E/(1 + f). \quad (8.167)$$

Then, the energy can be expressed in terms of observables and the unknown equipartition parameter f in the following way

$$E(f) \simeq (8\pi)^{1/7} \left(\frac{em_e c}{\sigma_T^2} \right)^{2/7} \nu^{2/7} L_\nu^{4/7} V^{3/7} \frac{1 + f}{f^{4/7}}. \quad (8.168)$$

It can be easily shown (by differentiating E wrt. f) that a minimum energy is achieved when $f = 4/3$, i.e. $U_e = 4U_B/3$. Your task is to calculate this minimum energy $E_{\min} = E(f = 4/3)$ based on the observed flux at $\nu \simeq 1 \text{ GHz}$ for the radio lobes of Cygnus A.

- (3) Based on your E_{\min} and $f = 4/3$, calculate the B-field strength inside the radio lobes. Compare the magnetic energy density to that of the CMB, $U_{\text{CMB}} = aT^4$ with $T = 2.7 \text{ K}$.
- (4) Using the B-field strength obtained in part (3), estimate the Lorentz factor of the electrons radiating near $\nu \simeq 1 \text{ GHz}$. What is the synchrotron cooling timescale for these electrons? (hint: your answer should be comparable to $E_e/\nu L_\nu$, because the frequency we picked is close to the peak of the νL_ν spectrum, i.e. the synchrotron cooling frequency is $\nu_c \sim 1 \text{ GHz}$.) The energy density of radio photons in the synchrotron emitting region is given by $U_{\text{radio}} \sim \nu L_\nu V^{-2/3}/c$ (since the energy is $\nu L_\nu V^{1/3}/c$). Is the electron cooling dominated by synchrotron or inverse-Compton emission?
- (5) Suppose an exact copy of Cygnus A exists at redshift $z = 6$ where the CMB energy density is much higher than in the nearby Universe by a factor of $(1+z)^4$, do you expect the electron cooling to be dominated by synchrotron emission or inverse-Compton scattering?

References

Blandford, Meier & Readhead, 2019, ARA& A, 57, 467

- F. F. Chen, *Introduction to Plasma Physics and Controlled Fusion* (Springer International Publishing, 3rd Edition, Switzerland, 2016)
- Crusius & Schlickeiser, 1986, A&A, 164, 16
- de Jager et al., 1996, ApJ, 457, 253
- Ghisellini, Guilbert, & Svensson, ApJL, 1988, 334, 5
- Ginzburg & Syrovatskii, ARA&A, 3, 297
- Harding & Lai, 2006, Reports on Progress in Physics, 69, 2631
- Jackson, *Classical Electrodynamics* (John Wiley & Sons, New York, 1975).
- Kellermann & Pauliny-Toth, 1969, ApJL, 155, 71
- Mahadevan, Narayan, & Yi, 1996, ApJ, 465, 327
- McKean, Godfrey, et al., 2016, MNRAS, 463, 3143
- Readhead, 1994, ApJ, 426, 51
- Rybicki & Lightman, *Radiative processes in astrophysics* (A Wiley-Interscience Publication, New York, 1979).
- Westfold, 1959, ApJ, 130, 241