MPhys Radiation and Matter 2016–2017

John Peacock

Room C20, Royal Observatory; jap@roe.ac.uk

http://www.roe.ac.uk/japwww/teaching/ram5.html

I. Physical Introduction

1 Review of Classical Radiation
- 1.1 Maxwell's equations, field energy and potentials
- 1.2 Potentials in the wave zone – classical radiant energy
- 1.3 The Hertzian dipole and the Larmor radiation formula
- 1.4 Scattering cross-sections and the damped oscillator
- 1.5 The radiation field as a set of oscillators

2 Quantum Radiation
- 2.1 Quantized SHM
- 2.2 Quantizing the radiation oscillators

3 Quantum interaction of radiation with matter
- 3.1 Fermi's Golden Rule
- 3.2 The interaction Hamiltonian
- 3.2 Semiclassical transitions
- 3.3 Quantum radiative transitions
- 3.5 Electric dipole transitions and selection rules
- 3.6 Electric quadrupole and magnetic dipoles

4 Radiative transfer
- 4.1 Specific intensity and the equation of radiative transfer
- 4.2 Optical depth and a solution to the equation of RT
- 4.3 Line profiles

5 Collisions
- 5.1 Collision rate and cross-section
- 5.2 Critical density

II. Astronomical applications

- 6 Lyman alpha line
- 7 Photoionization and recombination
- 8 Fine structure and forbidden diagnostic lines
- 9 HI 21 cm line
- 10 CO

Bibliography

Standard texts on EM & QM (Jackson; Landau & Lifshitz; etc.)
Emerson, Interpreting Astronomical Spectra. Wiley (1996)
Rybicki & Lightman, Radiative processes in astrophysics. Wiley (1979)
Osterbrock, Astrophysics of Gaseous Nebulae and AGN. University Science (1989)
Shu, The physics of astrophysics, Volume 1. University Science (1991)

Radiation and Matter

I. Physical Introduction

1 Review of Classical Radiation

This section reviews the properties of the electromagnetic field in free space and shows

(1) that they are derivable from four potentials (ϕ, A_x, A_y, A_z), which can be changed (gauge transformations) without changing the physics i.e. *the resulting fields;*

(2) that at distances large compared with source size and wavelength (the wave zone) the potentials and fields can be expressed as transverse waves with $E = cB$ and $\mathbf{E}$ *perpendicular to* $\mathbf{B}$*;*

(3) that the fields due to a point charge lead – via the Poynting vector – to Larmor's formula for classical radiation (power proportional to acceleration squared);

(4) that radiation in an otherwise empty box can be expressed as a superposition of wave modes, for each of which there is a dynamical similarity between the amplitude of the vector potential and the displacement of a simple harmonic oscillator. This view of radiation as oscillators will be the basis for quantizing the EM field.

1.1 Maxwell's equations, field energy and potentials

1.1.1 Maxwell's equations

Maxwell's equations *in vacuo* are

$$\boldsymbol{\nabla}\wedge\mathbf{E} = -\dot{\mathbf{B}} \qquad \boldsymbol{\nabla}\cdot\mathbf{B} = 0 \tag{1}$$

$$\boldsymbol{\nabla}\cdot\mathbf{E} = \rho/\epsilon_0 \qquad \boldsymbol{\nabla}\wedge\mathbf{B} = \mu_0(\mathbf{j} + \epsilon_0\dot{\mathbf{E}}), \tag{2}$$

where a dot denotes partial time derivative. The term $\epsilon_0\dot{\mathbf{E}}$ in the last equation is the 'displacement current', needed so that the equations are consistent with conservation of charge:

$$\dot{\rho} + \boldsymbol{\nabla}\cdot\mathbf{j} = 0, \tag{3}$$

where ρ is the charge density and $\mathbf{j}$ is the current density. These equations for the fields should be supplemented by the *Lorentz force* acting on a charge q:

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v}\wedge\mathbf{B}). \tag{4}$$

1.1.2 SI vs cgs

A practical difficulty with electromagnetism is that not everyone agrees on the units. The above equations in the SI or MKS system are made messy by the constants ϵ_0 and μ_0, necessary to reconcile empirical units of charge with electromagnetic forces. Only one of these is independent, because

$$\epsilon_0\mu_0 = 1/c^2, \tag{5}$$

as will be verified shortly. The remaining independent constant can be set to unity, because its value depends on the units of charge that we adopt; but changing the units of charge

also changes the units of the electromagnetic field, because of the form of the Lorentz force law. One other simplification is to realise that it would be more appealing if the units of E and B were the same, which suggests redefining B by a power of c so that the Lorentz force becomes $\mathbf{F} = q(\mathbf{E} + \mathbf{v}\wedge\mathbf{B}/c)$. This gives the *Heaviside-Lorentz* system. But what is more commonly encountered (especially in American textbooks and research papers) is the *Gaussian or cgs* system. This differs from the Heaviside-Lorentz system in being 'unrationalized' – *i.e.* with a factor of 4π retained in Gauss's law:

$$\boldsymbol{\nabla}\wedge\mathbf{E} = -\dot{\mathbf{B}}/c \qquad\qquad \boldsymbol{\nabla}\cdot\mathbf{B} = 0 \tag{6}$$

$$\boldsymbol{\nabla}\cdot\mathbf{E} = 4\pi\rho \qquad \boldsymbol{\nabla}\wedge\mathbf{B} = (4\pi\mathbf{j} + \dot{\mathbf{E}})/c. \tag{7}$$

We got here with the help of $\epsilon_0 \to 1$; how do we invert this? Remember all units have changed, so write $q_{\rm cgs} = a_q q$, $E_{\rm cgs} = a_E E$, $B_{\rm cgs} = a_B B$. Comparing the two versions of Maxwell's equations plus the Lorentz force requires $(a_q^2, a_E^2, a_B^2) = (1/4\pi\epsilon_0, 4\pi\epsilon_0, 4\pi\epsilon_0 c^2)$. Thus, for example, the dimensionless fine-structure constant is

$$\alpha \simeq \frac{1}{137} = \frac{e_{\rm cgs}^2}{\hbar c} = \frac{e_{\rm SI}^2}{4\pi\epsilon_0\hbar c}. \tag{8}$$

Which convention is being used can be guessed by dimensions: with no explicit field, e cannot occur in SI without ϵ_0 also appearing. The other conversion most commonly needed in the astronomy literature is in numerical units for magnetic field (Tesla in SI, Gauss in cgs): $1\,\mathrm{T} = 10^4\,\mathrm{G}$.

A note in passing: $\hbar \equiv h/2\pi$ should be called 'h-cross'. An regrettable growing transatlantic trend is to use the name 'h-bar'. But this means $\bar{h}$.

1.1.3 Field energy and Poynting vector

By manipulating Maxwell's equations (see appendix) we can arrive at the result

$$-\int \mathbf{j}\cdot\mathbf{E}\,dV = \frac{\partial}{\partial t}\int\left(\frac{\epsilon_0}{2}\mathbf{E}^2 + \frac{1}{2\mu_0}\mathbf{B}^2\right)\,dV + \int_{\rm surface}\mathbf{E}\wedge\mathbf{B}/\mu_0\cdot d\mathbf{S}. \tag{9}$$

The left-hand side is minus the work done by the field on moving charges (only by $\mathbf{E}$: force is perpendicular to $\mathbf{B}$). This must equal the change in energy that is somehow stored in the field. We see that this energy change is a combination of a change in energy density and a surface term, which must represent a transport of energy across the boundary of the region being considered. The energy flux density in the surface integral is the *Poynting vector*:

$$\mathbf{S} = \mathbf{E}\wedge\mathbf{B}/\mu_0. \tag{10}$$

1.1.4 Electromagnetic potentials

The first pair of Maxwell's equations express constraints on the fields rather than saying how they arise. Using the general vector identities $\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}\wedge\mathbf{A}) = \mathbf{0}$, $\boldsymbol{\nabla}\wedge(\boldsymbol{\nabla}\phi) = \mathbf{0}$, the fields $\mathbf{E}, \mathbf{B}$ can be seen to be derivatives of potential functions $\mathbf{A}$ and ϕ:

$$\mathbf{B} = \boldsymbol{\nabla}\wedge\mathbf{A} \qquad\qquad \mathbf{E} = -\partial\mathbf{A}/\partial t - \boldsymbol{\nabla}\phi. \tag{11}$$

The potentials $\mathbf{A}, \phi$ can be changed by the gradient of an arbitrary function $\chi(\mathbf{r}, t)$, without changing the values of $\mathbf{E}, \mathbf{B}$. This is called *gauge invariance*.

$$\mathbf{A} \to \mathbf{A}' = \mathbf{A} + \boldsymbol{\nabla}\chi, \qquad\qquad \phi \to \phi' = \phi - \partial\chi/\partial t. \tag{12}$$

This freedom allows some interesting choices. For example, we can always set $\phi = 0$ and describe EM purely with a vector potential; but this is not the simplest option.

Substituting the potentials in the second pair of Maxwell's equations, we see the wave magic in Maxwell's equations:

$$\nabla^2 \mathbf{A} - \frac{\partial^2 \mathbf{A}}{c^2 \partial t^2} - \boldsymbol{\nabla}\left(\boldsymbol{\nabla}\cdot\mathbf{A} + \frac{\partial \phi}{c^2 \partial t}\right) = -\mu_0 \mathbf{j} \tag{13}$$

and

$$\nabla^2 \phi - \frac{\partial^2 \phi}{c^2 \partial t^2} + \partial/\partial t \left(\boldsymbol{\nabla}\cdot\mathbf{A} + \frac{\partial \phi}{c^2 \partial t}\right) = -\rho/\epsilon_0. \tag{14}$$

These are *nearly* wave equations, and can be made so using the gauge invariance of the fields: It is possible to choose χ so that the transformed potentials satisfy the *Lorenz gauge condition*:

$$\boldsymbol{\nabla}\cdot\mathbf{A} + \partial\phi/c^2\partial t = 0 \tag{15}$$

(note the spelling: this is not due to Lorentz, although the error is commonly made). The potentials then satisfy *wave equations*:

$$\Box \mathbf{A} = \mu_0 \mathbf{j} \qquad \Box \phi = \rho/\epsilon_0, \tag{16}$$

where $\Box$ is the Lorentz invariant wave operator, or d'Alembertian:

$$\Box \equiv \partial^\mu \partial_\mu = \partial^2/c^2\partial t^2 - \nabla^2. \tag{17}$$

Here, $\partial_\mu \equiv \partial/\partial x^\mu = (\partial/\partial ct, \boldsymbol{\nabla})$ and $\partial^\mu = (\partial/\partial ct, -\boldsymbol{\nabla})$ – the usual relativistic distinction of up and down indices being related by a change of sign in spatial parts of 4-vectors.

In these terms, charge conservation is simply

$$\partial_\mu J^\mu = 0; \qquad J^\mu = (\rho\, c, \mathbf{j}). \tag{18}$$

Physically, we can see that ρ transforms like the time-component of a 4-vector since $dq = \rho\, dx\, dy\, dz$, the charge element, is invariant as is $dt\, dx\, dy\, dz$. Thus, if we write a 4-potential as $A^\mu = (\phi/c, \mathbf{A})$, the wave equations for the potentials take the pleasingly covariant form

$$\Box A^\mu = \mu_0 J^\mu. \tag{19}$$

1.1.5 Solution of the wave equation

The solution to the wave equation is

$$A^\mu(\mathbf{r}, t) = \left(\frac{\mu_0}{4\pi}\right) \int \frac{J^\mu(\mathbf{r}', t_{\rm ret})}{|\mathbf{r} - \mathbf{r}'|}\, d^3 r' \,, \tag{20}$$

where $t_{\rm ret}$, the 'retarded time', is given by

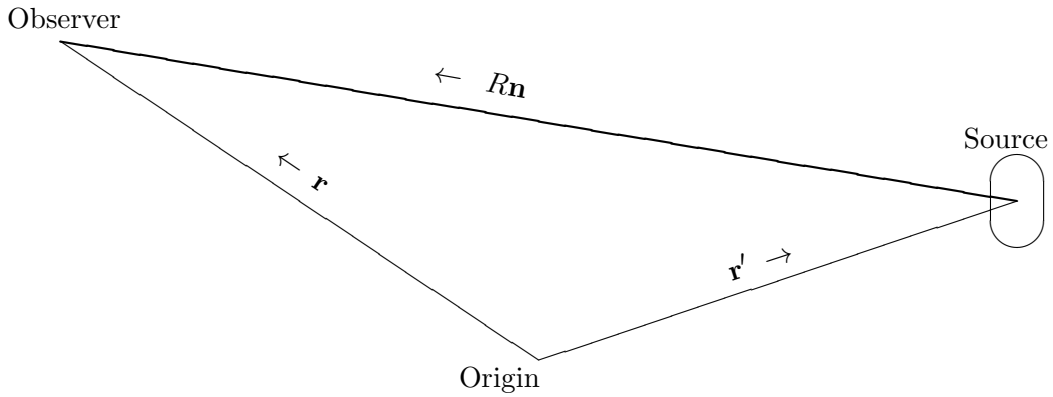
$$t_{\rm ret} = t - |\mathbf{r} - \mathbf{r}'|/c. \tag{21}$$

This is proved in the appendix. Less formally, we can note that the result is obviously correct in the time-independent case: the wave equation is then Poisson's equation, whose solution is a superposition of $1/r$ potentials from the different parts of the source. Where the source varies, the retarded-time recipe just says that information propagates at the speed of light: the potential is as if the source were static with the properties it had at the appropriate time in the past. The fields obey 'what you see is what you get'. It is common to use the following shorthand:

$$[J^\mu(\mathbf{r})] \equiv J^\mu(\mathbf{r}, t_{\rm ret}). \tag{22}$$

1.2 Potentials in the wave zone – radiant energy

In studying the EM fields received from a source, let the source be confined within a region of size d, and let the typical wavelength of radiation be λ ($= c\tau$, where τ is a typical timescale of variability). The expression for the field can be simplified in one limit, called the *wave zone* or *radiation zone*, where the source is far away ($R \gg d$, where R is the typical distance to the source).



Now, $|\mathbf{r} - \mathbf{r}'|$ is the dot product of $(\mathbf{r} - \mathbf{r}')$ with a unit vector in the same direction – call this $\mathbf{n}$. When the angle subtended by the source is small, we can treat $\mathbf{n}$ as if it were a constant vector – so that the position vector of the observer with respect to the centre of mass can be written just as $R\mathbf{n}$: see the above diagram. The retarded time can therefore be written as

$$t_{\rm ret} = t - \mathbf{r}\cdot\mathbf{n}/c + \mathbf{r}'\cdot\mathbf{n}/c. \tag{23}$$

If the source is distant in the above sense, we can replace $|\mathbf{r} - \mathbf{r}'|$ *in the denominator* by the constant distance R. In the function $\mathbf{j}(\mathbf{r}', t_{\rm ret})$ which is now the integrand, the variable $\mathbf{r}'$ is the (dummy) integration variable on integration over the source, and so the vector potential is a function of the non-integrated parameters, $t - \mathbf{r}\cdot\mathbf{n}/c$:

$$\mathbf{A}(\mathbf{r}, t) \simeq \frac{\mu_0}{4\pi R}\int \mathbf{j}(\mathbf{r}', [t - \mathbf{r}\cdot\mathbf{n}/c] + \mathbf{r}'\cdot\mathbf{n}/c)\, d^3r' \equiv \mathbf{F}(t - \mathbf{r}\cdot\mathbf{n}/c)\,/\,R\,. \tag{24}$$

This expression for $\mathbf{A}$ is that of an outgoing wave, amplitude decreasing as $1/R$, travelling with velocity c.

Knowing $\mathbf{A}$, we can immediately get the magnetic field, $\mathbf{B} = \boldsymbol{\nabla}\wedge\mathbf{A}$. We can use (see appendix) the following relations for a vector $\mathbf{F}$ (or scalar f) whose argument is of the form $t - \mathbf{n}\cdot\mathbf{r}/c$:

$$\boldsymbol{\nabla}\cdot\mathbf{F} = -\frac{\mathbf{n}}{c}\cdot\frac{\partial}{\partial t}\mathbf{F}, \qquad \boldsymbol{\nabla}\wedge\mathbf{F} = -\frac{\mathbf{n}}{c}\wedge\frac{\partial}{\partial t}\mathbf{F}, \qquad \boldsymbol{\nabla} f = -\frac{\mathbf{n}}{c}\frac{\partial}{\partial t} f\,. \tag{25}$$

Thus the magnetic field is

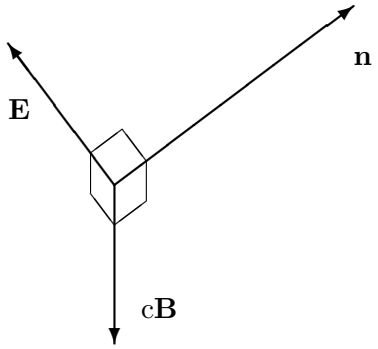
$$\mathbf{B} = -\frac{\mathbf{n}}{c}\wedge\dot{\mathbf{A}}. \tag{26}$$

This is clearly *transverse*: perpendicular to $\mathbf{n}$.

The electric field comes from the last Maxwell equation (with $\mathbf{j} = 0$): $\dot{\mathbf{E}} = \mathbf{c^2}\boldsymbol{\nabla}\wedge\mathbf{B}$. Because of the form of $\mathbf{B}$, $\boldsymbol{\nabla}\wedge\mathbf{B} = -\mathbf{n}\wedge\dot{\mathbf{B}}/c$, so integrating in time gives

$$\mathbf{E} = -c\mathbf{n}\wedge\mathbf{B}, \tag{27}$$

so both $\mathbf{E}$ and $\mathbf{B}$ are transverse and they are perpendicular to each other. Thus $\mathbf{E}$, $\mathbf{B}$, $\mathbf{n}$ form a right-handed triad (like x, y, z), and $E = cB$.



There is a little paradox here. We have derived the fields using only $\mathbf{A}$. But if ϕ was negligible (and apparently it vanishes if the source consists of currents only), then we would have $\mathbf{E} = \dot{\mathbf{A}}$, which is not transverse. In fact, the Lorenz condition shows that ϕ must be non-zero – and this would be the easiest way of calculating it. The scalar potential is non-zero even when ρ vanishes, which corresponds to adding a solution of $\Box\phi = 0$ to the above expression in terms of $[\rho]$.

The energy transported by the field is determined by the Poynting vector $\mathbf{S}$ which in the wave zone becomes

$$\mathbf{S} = -(\mathbf{n}\wedge\mathbf{B}c)\wedge(\mathbf{B}/\mu_0) = B^2 c\mathbf{n}/\mu_0 = \epsilon_0 c|\dot{\mathbf{A}}_\perp|^2\mathbf{n} \tag{28}$$

and the form derived in the wave zone for $\mathbf{A} \propto R^{-1}\mathbf{F}(t - \mathbf{r}\cdot\mathbf{n}/c)$ can be substituted. This expression makes perfect sense, as we would expect radiation to transport energy at speed c, so

$$S = cU = c(\epsilon_0 E^2 + B^2/\mu_0)/2. \tag{29}$$

This gives the correct answer since we have shown that the electric and magnetic energy densities are equal.

1.3 The Hertzian dipole and the Larmor radiation formula

The simplest example of this machinery is a radiating dipole, where we assume that the current density, $\mathbf{j}$, is uniform throughout the source. We also assume that the source is *coherent*: small compared to a wavelength. Real astronomical sources are almost always distant, but rarely coherent. What happens in the latter case is that we add the *intensities* from different parts of the source, rather than adding the fields.

In this case, the source integration is trivial: $\int[\mathbf{j}]\, dV = [\mathbf{j}]\, V$. We can view this in a number of ways. If the source is a wire of length ℓ, then $jV = I\ell$, where I is the current in the wire. More interestingly, a moving charge has the same effect. The current density must be $\mathbf{j} = \rho\mathbf{v}$, so $\mathbf{j}V = q\mathbf{v}$ (assuming the motion of the charge to be non-relativistic). Thus, the vector potential in the wave zone is

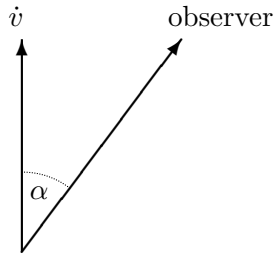
$$\mathbf{A}(\mathbf{r}, t) = \frac{\mu_0 q\mathbf{v}}{4\pi R}, \tag{30}$$

where henceforth we drop the explicit [ ] indication of retardation.

Using the earlier discussion, we have the Poynting vector:

$$\mathbf{S} = \frac{1}{4\pi R^2}\frac{q^2/c^3}{4\pi\epsilon_0}\dot{v}^2\sin^2\alpha\ \mathbf{n} \tag{31}$$

and we can integrate this over direction to find the total radiated power.



Since $d\Omega = \sin\alpha\, d\alpha\, d\phi$ and $\int \sin^3\alpha\, d\alpha\, d\phi = 8\pi/3$,

$$P = \int \mathbf{S}\cdot d = \int \mathrm{SR}^2\, \mathrm{d}\Omega = \frac{\mathrm{q}^2}{6\pi\epsilon_0 \mathrm{c}^3}\dot{v}^2, \tag{32}$$

which is Larmor's radiation formula. This says that the power radiated by a charge is proportional to the square of the acceleration, and that the radiation comes out predominantly perpendicular to the acceleration vector (with zero intensity in the direction of acceleration itself).

We note in passing that this result contains a big conceptual puzzle, which is often ignored: where does the radiated energy come from? It must result from work being done by the force that accelerates the charge, but it is possible to accelerate a charge without doing work: *e.g.* an electron gyrating in a magnetic field. This says that the Lorentz force $\mathbf{F} = q(\mathbf{E} + \mathbf{v}\wedge\mathbf{B})$ must be wrong: it needs to be supplemented by an extra force that slows the electron down when it radiates. This force is called *radiation reaction*; it is difficult to analyse exactly because one needs to consider the EM forces of a particle on itself. We will not try to solve this problem in full, but the next section shows that it can be tackled in simple cases, where the acceleration is low enough that the energy loss rate does not significantly disrupt the trajectory.

1.4 Scattering cross-sections and the damped oscillator

The Larmor formula has many interesting applications, depending on the source of the acceleration. One is *synchrotron radiation*, where the electron performs helical orbits in the presence of an external magnetic field; another is *bremsstrahlung*, where electrons in a plasma are accelerated by the electrostatic fields due to the ions. But a distinct case is when the fields that accelerate the charge are themselves due to radiation. In this case, energy is removed from the incoming beam of radiation, leading to *scattering*.

Thomson scattering The simplest case is a free electron. Consider incident radiation with linear polarization, so that $\mathbf{E}$ is along a fixed direction. Then $\dot{v} = eE/m$, and the Larmor formula is

$$P = \frac{e^4}{6\pi\epsilon_0 c^3 m^2}\, E^2. \tag{33}$$

But the energy flux density is $2\times c\epsilon_0 E^2/2$, so we can define an effective cross-sectional area that intercepts this radiation:

$$\sigma_\mathrm{T} = P/(\text{flux density}) = \frac{8\pi}{3}\left(\frac{e^2}{4\pi\epsilon_0 mc^2}\right)^2 = \frac{8\pi}{3}\alpha^2\left(\frac{\hbar}{mc}\right)^2, \tag{34}$$

where the 'T' stands for J.J. Thomson, α is the fine structure constant, and $\frac{\hbar}{mc}$ is the *Compton length* of an electron. So the electron behaves as if it was 'dressed' with the

area $\sigma_{\rm T}$, such that any photon that travels within a distance $\sqrt{\sigma_{\rm T}/\pi}$ will be removed from the beam of incident radiation and scattered. This idea of a cross-section as governing interaction rates is a powerful concept, which we will use many times in this course.

The angle-dependent cross-section is, from above

$$\frac{d\sigma}{d\Omega} = \frac{3}{8\pi}\,\sigma_{\rm T}\,\sin^2\alpha. \tag{35}$$

This is for polarized radiation. For the polarized case, there will be a similar dependence for the incoherent radiation associated with the perpendicular polarization, and we have to average these:

$$\sin^2\alpha \to \frac{(1-[\mathbf{n}\cdot\mathbf{e}_1]^2)+(1-[\mathbf{n}\cdot\mathbf{e}_2]^2)}{2}. \tag{36}$$

But $\mathbf{e}_1$, $\mathbf{e}_2$, and $\hat{\mathbf{k}}$ are a triplet of unit vectors, so this is $(2-[1-(\mathbf{n}\cdot\hat{\mathbf{k}})^2])/2$. Thus the unpolarized cross-section is

$$\frac{d\sigma}{d\Omega} = \frac{3}{16\pi}\,\sigma_{\rm T}\,(1+\cos^2\theta), \tag{37}$$

where θ is the scattering angle.

Damped oscillator An interesting generalization of the free electron is to place the electron in a potential with some damping, with the applied electric field as a driving term:

$$\ddot{\mathbf{x}} + \Gamma\dot{\mathbf{x}} + \omega_0^2\mathbf{x} = \frac{e}{m}\mathbf{E}. \tag{38}$$

As usual, we can look for oscillatory solutions where the driving force varies as $\propto \exp(i\omega t)$, and solve for the amplitude of the resulting acceleration:

$$|\ddot{\mathbf{x}}|^2 = (e/m)^2E^2\,\frac{\omega^4}{(\omega^2-\omega_0^2)^2+\Gamma^2\omega^2}. \tag{39}$$

Thus the total cross-section is

$$\sigma(\omega) = \sigma_{\rm T}\,\frac{\omega^4}{(\omega^2-\omega_0^2)^2+\Gamma^2\omega^2}. \tag{40}$$

For $\omega \gg \omega_0$, we just get the free electron result, whereas for low frequencies, $\sigma = \sigma_{\rm T}(\omega/\omega_0)^4$, which is *Rayleigh scattering*.

We can take this analysis a little further, because we can figure out what the damping Γ has to be. For no applied force and Γ small enough, the solution of the EOM is $x \propto \exp(i\omega_0 t - \Gamma t/2)$, so that the energy in the oscillator decays $\propto \exp(-\Gamma t)$. This energy loss must come from the radiation. If $x = A\sin\omega_0 t$, then the power averaged over one cycle is $\langle P\rangle = (q^2/6\pi\epsilon_0 c^3)A^2\omega_0^4/2$. The total energy must be the kinetic energy at $x=0$, $K = mA^2\omega_0^2/2$. Thus $\Gamma = \langle P\rangle/K = q^2\omega_0^2/6\pi\epsilon_0 c^3 m$. We can then write the cross-section in the following form:

$$\sigma(\omega) = \frac{\pi e^2}{2\epsilon_0 mc}\,\frac{\Gamma/2\pi}{(\omega-\omega_0)^2+(\Gamma/2)^2}, \tag{41}$$

where we have assumed that ω is close to ω_0, so that $(\omega^2-\omega_0^2) \simeq 2\omega_0(\omega-\omega_0)$. The reason for using this form is that the function of ω is a *Lorentzian*, which integrates to unity. This will be a useful reference case when we come to consider atomic transitions.

1.5 The radiation field as a set of oscillators

A different application of the vector potential is to give a useful description of the vacuum electromagnetic field, in a region free of sources. In this case, we can choose a gauge with

$\phi = \boldsymbol{\nabla} \cdot \mathbf{A} = 0$ (consistent with Maxwell's 2nd equation, since $\boldsymbol{\nabla} \cdot \mathbf{E} = \boldsymbol{\nabla} \cdot (-\partial \mathbf{A}/\partial t) = 0$). We now have all of the information about the electromagnetic field in just two components of the vector potential (one suppressed by requiring $\boldsymbol{\nabla} \cdot \mathbf{A} = 0$), instead of the four components of the complete set.

The Lorenz condition is still obeyed, so $\mathbf{A}$ obeys the source-free wave equation,

$$\nabla^2 \mathbf{A} - \ddot{\mathbf{A}}/c^2 = 0. \tag{42}$$

We now solve this equation in effect by Fourier analysis: imagine the radiation inside a very large cube of side L. Then the potential is expressible as a triple Fourier series with amplitudes $\mathbf{X}_\mathbf{k}$:

$$\mathbf{A}(\mathbf{r}, t) = \sum_\mathbf{k} \mathbf{X}_\mathbf{k}(t) e^{i\mathbf{k}\cdot\mathbf{r}}. \tag{43}$$

The allowed wave vectors are those that obey *periodic boundary conditions*, in which whole number of oscillations span the box:

$$\mathbf{k} = (k_x, k_y, k_z) = (2\pi/\lambda)\hat{\mathbf{k}} = (2\pi/L)(n_x, n_y, n_z), \tag{44}$$

where the n_i are positive or negative integers. The summation is over these integers. The volume is assumed to be so large that edge effects are negligible, and that the wavenumbers are very closely spaced, so that sums over these can be approximated by integrals when convenient. As usual, with a real function, the Fourier coefficients must be *Hermitian*: $\mathbf{X}_{-\mathbf{k}} = \mathbf{X}_\mathbf{k}^*$.

The wave equation gives

$$\ddot{\mathbf{X}}_\mathbf{k}(t) + \omega^2 \mathbf{X}_\mathbf{k}(t) = 0, \tag{45}$$

where $\omega = ck$ (because if a Fourier sum is zero, each term vanishes independently), so the possible time dependence is $\exp(\pm i\omega t)$: in combination with the $\exp(i\mathbf{k}\cdot\mathbf{r})$ dependence, the modes are of course travelling waves. At this point, we need a little care over the choice of conventions for the complex representation of travelling waves. A complex disturbance, ψ, can be written either as

$$\psi \propto e^{i(\mathbf{k}\cdot\mathbf{x} - \omega t)} \qquad \text{or} \qquad \psi \propto e^{i(\omega t - \mathbf{k}\cdot\mathbf{x})}. \tag{46}$$

Both the real and imaginary parts of these expressions represent waves moving in the direction of $\mathbf{k}$; the real parts are identical, and the imaginary parts just have a relative phase shift of π. Thus these are not independent waves. The conventional solution to this is to represent travelling waves so that the frequency is positive, so that a wave moving in the $\mathbf{k}$ direction is always represented as

$$\psi \propto e^{i(\mathbf{k}\cdot\mathbf{x} - \omega t)}, \tag{47}$$

with $\omega > 0$. We can therefore represent the vector potential as a sum of travelling waves by writing

$$\mathbf{A}(\mathbf{r}, t) = \sum_\mathbf{k} \mathbf{X}_\mathbf{k} e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)}. \tag{48}$$

Finally, in order to make the reality of $\mathbf{A}$ explicit, we can add this expression to its complex conjugate (easier than making $\mathbf{X}_\mathbf{k}$ Hermitian):

$$\mathbf{A}(\mathbf{r}, t) = \sum_\mathbf{k} \mathbf{X}_\mathbf{k} e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)} + \sum_\mathbf{k} \mathbf{X}_\mathbf{k}^* e^{-i(\mathbf{k}\cdot\mathbf{r} - \omega t)}. \tag{49}$$

The condition $\boldsymbol{\nabla} \cdot \mathbf{A} = 0$ leads to $\mathbf{k} \cdot \mathbf{X}_\mathbf{k}(t) = 0$, *i.e.* the waves are transverse. This transverse vector can be expressed as a superposition of two independent vectors (which we also choose to be orthogonal to each other), representing the polarization components of the field. We therefore define a right-handed triad involving $\mathbf{k}$ and the unit vectors $\mathbf{e}_{\mathbf{k},1}$

and $\mathbf{e}_{\mathbf{k},2}$. The orientation of the $\mathbf{e}_{\mathbf{k},1} - \mathbf{e}_{\mathbf{k},2}$ basis is arbitrary. Finally, then, we can express $\mathbf{A}$ as a sum of wave modes summing over both $\mathbf{k}$ and a polarization index, α. For future convenience, we choose to renormalize the field amplitudes by a factor $\sqrt{(\hbar/(2\epsilon_0 V\omega)}$ and take the exponential time dependence back into the time dependence of the field coefficients:

$$\mathbf{A}(\mathbf{r},t) = \sum_{\mathbf{k},\alpha} \sqrt{\frac{\hbar}{2\epsilon_0 V\omega}}\, \mathbf{e}_{\mathbf{k},\alpha} \left(a_{\mathbf{k},\alpha}(t) e^{i\mathbf{k}\cdot\mathbf{r}} \; + \; a^*_{\mathbf{k},\alpha}(t) e^{-i\mathbf{k}\cdot\mathbf{r}} \right) \,, \tag{50}$$

where the factor e^{iwt} has been absorbed into $a_{\mathbf{k},\alpha}$ and correspondingly e^{-iwt} into $a^*_{\mathbf{k},\alpha}$. This expression is a little messy (life would be easier if EM waves were unpolarized), but it can be written more transparently in the form

$$\mathbf{A}(\mathbf{r},t) = \sum_n (\mathbf{A}_n a_n + \mathbf{A}^*_n a^*_n) \,, \tag{51}$$

where n is a general index labelling the modes: $n \equiv (\mathbf{k},\alpha)$. This general form makes it clear that we could have used basis functions other than plane waves – for instance, spherical waves of given angular momentum.

1.5.1 The oscillator energies

The total field energy, or Hamiltonian, is

$$H = \int \frac{1}{2}(\epsilon_0 E^2 + B^2/\mu_0)\, dV = \int \frac{1}{2}(\epsilon_0 \dot{A}^2 + |\boldsymbol{\nabla}\wedge\mathbf{A}|^{\mathbf{2}}/\mu_0)\, dV. \tag{52}$$

At first sight, this looks a mess: the quadratic terms mix modes of different k and α. But in practice very little of this matters: the polarization states are orthogonal, and mixing modes will give terms with a spatial dependence $\propto \exp(i\boldsymbol{\Delta}\mathbf{k}\cdot\mathbf{r})$, which averages to zero over the box. The only way to get a non-zero term is by mixing the $\exp(i\mathbf{k}\cdot\mathbf{r})$ and $\exp(-i\mathbf{k}\cdot\mathbf{r})$ parts of a single mode. This is easy in the $\dot{A}^2$ term, where the effect of the time derivative is

$$\mathbf{A}_n a_n + \mathbf{A}^*_n a^*_n \to (i\omega)\mathbf{A}_n a_n + (-i\omega)\mathbf{A}^*_n a^*_n, \tag{53}$$

and the non-vanishing term is $\omega^2|\mathbf{A}|^2(a_n a^*_n + a^*_n a_n)$. The magnetic term is just a little more work – but it must give the same result as the electric one, since this is radiation. Allowing for the normalization constant, we get

$$H_n = \frac{1}{2}(a_n a^*_n + a^*_n a_n)\, \hbar\omega, \tag{54}$$

where we haven't commuted the a, a^* for future convenience.

What is the linear *momentum density* of the plane wave field? Consider a pulse of radiation in a given direction containing energy E, and for which special relativity requires $E = pc$ as it travels with the speed of light. By measuring the flow of this pulse across a surface we find $SA\Delta t = E$ where S is the magnitude of the Poynting vector, A is the cross section of the pulse, $\Delta t = \ell/c$ is the time for the pulse of length ℓ to pass. Hence $S = c^2 p/(A\ell)$, and is therefore c^2 times the momentum density in the pulse.

A calculation similar to that for energy then gives for total linear momentum

$$\mathbf{p} = c^{-2} \int \mathbf{S}\, dV = \sum \frac{1}{2} \left(a_{\mathbf{k},\alpha} a^*_{\mathbf{k},\alpha} + a^*_{\mathbf{k},\alpha} a_{\mathbf{k},\alpha} \right) \hbar\mathbf{k}. \tag{55}$$

Now introduce real linear combinations of a_n, a^*_n

$$q_n = \sqrt{\hbar/2\omega}(a_n + a^*_n), \quad p_n = \dot{q}_n = -i\omega\sqrt{\hbar/2\omega}(a_n - a^*_n). \tag{56}$$

Then the energy per mode can be written

$$H_n = \frac{1}{2}(p_n^2 + \omega^2 q_n^2). \tag{57}$$

Furthermore p_n and q_n obey

$$\partial H_n/\partial q_n = -\dot{p}_n, \qquad \partial H_n/\partial p_n = \dot{q}_n, \tag{58}$$

so behave as dynamically conjugate momentum and coordinate respectively, with respect to the *Hamiltonian* H; these are Hamilton's equations. Note that the p_n quantities are completely different from $\mathbf{p}$ above.

These equations are identical to the equations of motion of a harmonic oscillator of unit mass and natural frequency $\omega = kc$ (see next section). Each state is formally the same as such an oscillator, each independent of the others. The energy of the electromagnetic field in the box is identical with that of a set of independent oscillators, one for each mode $(\mathbf{k}, \alpha)$. This was the start of Dirac's breakthrough to quantum electrodynamics.

2 Quantum Radiation

This section
(a) Reviews the quantum mechanics of SHM, using the 'number representation' and ladder operators.

(b) Completes the analogy between radiation oscillations in a box and SHM, and carries across the quantization procedure to the waves, thus inventing photons or radiation quanta.

2.1 Quantized SHM

First of all the quantum mechanics of the simple harmonic oscillator will be reviewed from the point of view of annihilation and creation operators. Then the connection noted above with modes of the radiation field is used to quantize the radiation field (*Dirac's second step to quantum electrodynamics, soon after the invention of proper quantum mechanics*).

Assume for simplicity that the oscillating mass of the oscillator is unity. Then the energy equation is

$$H = \frac{1}{2}(p^2 + \omega^2 q^2), \tag{59}$$

where q is the displacement and p is the corresponding canonical momentum (in this case it is v) with respect to the Hamiltonian H. These obey Hamilton's equations

$$\partial H/\partial q = -\dot{p} \quad \partial H/\partial p = \dot{q}. \tag{60}$$

To make this a quantum calculation p, q and therefore H become operators and p, q obey

$$[q, p] = i\hbar. \tag{61}$$

New operators a, $a^\dagger$ are defined by

$$a = (2\hbar\omega)^{-1/2}(\omega q + ip) \quad a^\dagger = (2\hbar\omega)^{-1/2}(\omega q - ip). \tag{62}$$

Here we have used the † notation for the Hermitian conjugate of an operator, which in general means

$$\langle\psi|O^\dagger|\phi\rangle \equiv \langle\phi|O|\psi\rangle^*. \tag{63}$$

A Hermitian operator is self-conjugate, and $(iO)^\dagger = -iO^\dagger$. In these terms,

$$[a,\, a^\dagger] = 1 \quad \text{and} \quad H = \frac{\hbar\omega}{2}\left(aa^\dagger + a^\dagger a\right). \tag{64}$$

These two results allow the most convenient way of writing the Hamiltonian:

$$H = (N + 1/2)\hbar\omega \quad \text{where} \quad N \equiv a^\dagger a, \tag{65}$$

in terms of N, the *number operator*.

Here is Dirac's argument for the states of a harmonic oscillator. We want to find a set of energy eigenstates $|E\rangle$, which are eigenstates of H with eigenvalue E (*i.e.* $H|E\rangle = E|E\rangle$). From above, it is clear that this is equivalent to finding the eigenstates of N: $N|n\rangle = n|n\rangle$, where $E = (n + 1/2)\hbar\omega$. This notation presupposes that the eigenvalues of N are integers, which we now prove. The first step is to establish the raising/lowering properties of the $a^\dagger$ and a operators, which means considering their action on the state $|n\rangle$. This will make a different state: is it possible that this is related to one of the other eigenstates? To test for

this, we need to consider $N\,a|n\rangle$; if this is proportional to $a|n\rangle$, then our new state is indeed an eigenstate, and we can figure out which one by reading off the eigenvalue. To carry out this test, use the commutator:

$$N\,a|n\rangle = a^\dagger a\,a|n\rangle = (aa^\dagger - 1)\,a|n\rangle = a(a^\dagger a - 1)|n\rangle = (n-1)\,a|n\rangle. \tag{66}$$

Thus we see that the action of a is to lower the eigenvalue of N by one, i.e. to lower the energy eigenvalue by $\hbar\omega$. In exactly the same way, $a^\dagger$ raises n by one.

So given a single eigenstate, we can construct a ladder of larger and smaller n eigenvalues; but as yet there is no reason for this ladder to take integer values. The key to proving that it does so is to realise that the series must terminate at the negative end. This makes physical sense: an oscillator has positive kinetic and potential energies, so how can it have negative total energy? For this reason, we would be disturbed to encounter $n < -1/2$. But this result should arise in a more inevitable mathematical way, and it comes from the normalization of the eigenstates. Consider some state $|n\rangle$, and make the new state $a\,|n\rangle$. The integral under the modulus squared of the wavefunction cannot be negative, so that

$$\langle a\,n|a\,n\rangle \geq 0. \tag{67}$$

Now we can rewrite this using the definition of the Hermitian conjugate. Remember this was $\langle\psi|O^\dagger|\phi\rangle \equiv \langle\phi|O|\psi\rangle^*$. We can rewrite this by explicitly associating operators with states, and then realizing that the complex conjugate of a bracket of two wavefunctions just reverses the order:

$$\langle\psi|O\phi\rangle = \langle\phi|O^\dagger\psi\rangle^* = \langle O^\dagger\psi|\phi\rangle. \tag{68}$$

In words, we usually think of O as 'looking right' and operating on the wavefunction there: but this has the same effect as $O^\dagger$ 'looking left'. This is a very important result; if the proof using Dirac notation seems too abstract, write out the steps in explicit integrals.

For our current problem, this says the following:

$$\langle a\,n|a\,n\rangle = \langle n|a^\dagger a|n\rangle = n, \tag{69}$$

where the last step follows because $|n\rangle$ is a normalized eigenstate of $N = a^\dagger a$. But the initial bracket could not be negative, so we must have $n \geq 0$. The only way to prevent this is for it to be impossible to reach negative n, and this is why n has to be an integer. In that case, we find that attempting to lower the $|0\rangle$ state yields zero for the normalization: rather than $a|0\rangle$ generating $|-1\rangle$, it generates no state at all. Thus the allowed number eigenvalues are $n = 0, 1, 2, \cdots$, and the energies of the harmonic oscillator are

$$E_n = (n + 1/2)\hbar\omega; \qquad n = 0, 1, 2, \cdots, \tag{70}$$

so that the ground state has *zero-point energy* $\hbar\omega/2$. This can be seen as an illustration of the uncertainty principle: zero energy would require $x = 0$ and zero momentum.

The same argument as above also tells us that the a and $a^\dagger$ operators do change the normalization of the states. Suppose $\langle n|n\rangle = 1$ for some n, and assume that $a|n\rangle = \alpha|n-1\rangle$, where α is the normalization constant that we want to find. Now we have

$$\langle a\,n|a\,n\rangle = |\alpha|^2 = \langle n|a^\dagger a\,n\rangle = n, \tag{71}$$

so that $|\alpha| = \sqrt{n}$, although the phase is undetermined (but unobservable, in any case. The normalization of $a^\dagger$ is derived in the same way, although now we need to use the commutator to re-express $aa^\dagger = N + 1$.

In summary, we have derived the following properties for our operators:

$$a^\dagger|n\rangle = \sqrt{n+1}\,|n+1\rangle. \tag{72}$$

$$a\,|n\rangle = \sqrt{n}\,|n-1\rangle. \tag{73}$$

The operator $a^\dagger$ increases the energy by one quantum (creates a quantum) and amplifies the state by $\sqrt{n+1}$ and the operator a decreases the energy (annihilates a quantum) and amplifies the state by $\sqrt{n}$ (automatically preventing attempts to lower below $n = 0$). The operators are thus known as *raising and lowering operators* or *creation and annihilation operators* or collectively as *ladder operators*. The operator $N \equiv a^\dagger a$ has eigenvalues n, and is called the *number operator*.

2.2 Quantizing the radiation oscillators

The radiation oscillators are quantized identically to SHM. The classical values for the jth state defined in section 1.4 become non-commuting operators:

$$[q_j, p_j] = i\hbar, \qquad [a_j, a_j^\dagger] = 1 \tag{74}$$

and clearly the amplitudes $a_j, a_j^\dagger$ have become the annihilation and creation operators for this state, which, analogously with the harmonic oscillator case, can be denoted in terms of its number eigenstate $|n_j\rangle$. This explains the strange normalization constant we chose: it was so the complex coefficients $a, a*$ could be immediately replaced by the ladder operators $a_n, a_n^\dagger$ (and this is why we kept aa^* distinct from a^*a).

Therefore, the Hamiltonian of the radiation field is a sum over modes:

$$H = \sum_j (n_j + 1/2)\hbar\omega_j, \tag{75}$$

where we assume that the radiation field sits in an eigenstate $|n_j\rangle$ for each mode. A puzzling aspect of this expression is the *zero-point energy* $\hbar\omega/2$. This is present even in the vacuum state ($|0\rangle$ for all modes), and the sum is infinite. For field theory, this constant offset is generally ignored; but it is a big problem in cosmology, where the gravitational effects of the zero-point energy should be apparent. The lack of such an effect is a big problem.

Photons It is convenient to talk of the energy quanta in these states as *photons*, with *energy* $\hbar\omega$ and *linear momentum* $\hbar\mathbf{k}$ in the case of plane waves, or *angular momentum* $\ell\hbar$ in the case of spherical waves, as specified by the state. The same physical system is being described whether we specify the numbers of photons in each momentum and polarization state, or specify the energy (quantized) of all oscillators, identified by their momentum and polarization. The eigenvalue n_j is called the *occupation number*, to denote the fact that there are n_j photons in mode j. But this description conflicts with the common picture of photons as distinct particles: where are they? The best we can say is that the particles are delocalized over the size of the box we used to analyse the field. By the uncertainty principle, we should then expect a momentum uncertainty $\sim \hbar/L$, and this is indeed the spacing of the modes.

3 Quantum interaction of radiation with matter

The previous section sketched the quantum nature of free radiation. This section (a) introduces Fermi's Golden rule as one expression of time dependent perturbation theory, giving the rate of transition between two states under the action of a perturbation

(b) sketches a formal way in which the interaction between radiation and matter can be expressed as a perturbation on free particles and fields

(c) derives the form of the transition rate

(d) introduces the dipole approximation

(e) shows that the photon occupation numbers in thermodynamic equilibrium are those of the black body distribution

3.1 Fermi's Golden Rule

Time dependent perturbation of a stationary system results in a state function that varies in time in a non-trivial way (*i.e.* not just by a term $e^{-iEt/\hbar}$). The initial (stationary) state function Ψ would be represented as a linear superposition of the eigenstates of the system, the coefficients b_j being the amplitudes to be in each state. With the perturbation switched on, these amplitudes change:

$$\Psi = \sum b_j(t) u_j(\mathbf{r})\, e^{-iE_j t/\hbar}. \tag{76}$$

Schrödinger's equation is

$$i\hbar\, \partial\Psi/\partial t = H\Psi \tag{77}$$

where the Hamiltonian $H = H_0 + H'$ and H' is the perturbing Hamiltonian. If the initial state is 'i' this reduces to

$$i\hbar\, (db_f/dt) = H'_{fi}\, e^{-i(E_i - E_f)t/\hbar}. \tag{78}$$

We prove this using orthonormality:

$$\langle f|i\rangle \equiv \int u_f^* u_i\, dV = \delta_{if}, \tag{79}$$

where

$$H'_{fi} \equiv \langle f|H'|i\rangle \equiv \int u_f^* H' u_i\, dV. \tag{80}$$

The case of interest here is when the perturbation is oscillatory and switched on at time $t = 0$. We might consider $\delta H \to H' \exp(i\omega t)$, but the physical significance of a complex perturbation is unclear. It is therefore common to consider $\delta H \to 2H' \sin\omega t$, which gives the same mathematical result. Then

$$b(t) = \frac{2H'}{i\hbar} \int_0^t \sin\omega t\, \exp(-i\omega_0 t)\, dt = \frac{H'}{i\hbar} \int_0^t \exp[i(\omega - \omega_0)t] - \exp[i(\omega + \omega_0)t]\, dt, \tag{81}$$

where $\hbar\omega_0 = E_i - E_f$. The latter exponential oscillates rapidly and averages to zero, leaving

$$|b(t)|^2 = \frac{|H'|^2}{\hbar^2}\, t^2\, \frac{\sin^2(\Delta\omega t/2)}{(\Delta\omega t/2)^2}, \tag{82}$$

where $\Delta\omega = \omega - \omega_0$. The sinc function is sharply peaked around $\omega = \omega_0$, and the area under $\sin^2 \alpha x/(\alpha x)^2$ is π/α. Thus the transition rate between states i and f, $\Gamma \equiv |b|^2/t$, is approximately

$$\Gamma = \frac{2\pi}{\hbar^2}|H'_{fi}|^2 \, \delta(\omega - \omega_0) \tag{83}$$

If there are many possible final states, then the probabilities of transition add. If these matrix elements all have virtually identical values, and the final states are closely-spaced in energy with a density of $(dN/dE)_f$ states per unit energy interval, then the the result of the sum of rates is Fermi's Golden Rule (which was actually invented by Dirac):

$$\Gamma = \frac{2\pi}{\hbar}|H'_{fi}|^2 \, (dN/dE)_f. \tag{84}$$

Strictly, this common expression is not really applicable to our situation: f is a single distinct final state. It would apply if we were ionizing an atom: in *bound-free transitions* there is indeed a continuum of final electron states. But even for discrete atomic transitions, a similar expression ends up applying. Consider perturbations that cover a range of frequencies, but which are *incoherent* (unrelated in phase): then the transition rates from each $d\omega$ will add. We will now show that the matrix element for perturbations from the electromagnetic field obeys $|H'_{fi}|^2 \propto A^2$, *i.e.* to the energy density. If we integrate this over frequency, we will remove the delta function and end up with a transition rate proportional to the energy density of the radiation, $\Gamma \propto U$.

3.2 The interaction Hamiltonian from gauge invariance

The easiest way to see how electromagnetic fields perturb quantum mechanics is to make a deep and general argument from gauge invariance. In quantum mechanics, we are familiar with the phase of the wave function being unobservable: only $|\psi|^2$ matters. Therefore we can make a *global* phase transformation with no effect:

$$\psi \to e^{i\alpha}\psi. \tag{85}$$

But this makes no practical sense, as it applies everywhere: how is someone on a distant galaxy to know what value of α we chose? Clearly, this phase transformation ought to be *local*: $\alpha \to \alpha(\mathbf{x}, t)$. Such a local transformation is called a *gauge transformation*, as when we allowed $A^\mu \to A^\mu + \partial^\mu \Psi$.

But local phase changes mess up the Schrödinger equation. If we insert $e^{i\alpha}\psi$ for ψ into

$$i\hbar\frac{\partial\psi}{\partial t} = H\psi = \left(\frac{\mathbf{p}^2}{2m} + V\right)\psi = \left(-\frac{\hbar^2}{2m}\nabla^2 + V\right)\psi, \tag{86}$$

then unwanted derivatives of α are clearly going to appear. The way to fix this is to realise that α really just affects the derivatives:

$$\frac{\partial}{\partial t} \to \frac{\partial}{\partial t} + i\dot{\alpha}; \qquad \boldsymbol{\nabla} \to \boldsymbol{\nabla} + i\boldsymbol{\nabla}\alpha. \tag{87}$$

We can cure this problem by adding something to the derivatives that can 'eat' the unwanted evidence of α. The *covariant derivatives* are

$$\frac{\partial}{\partial t} \to \frac{\partial}{\partial t} + f(\mathbf{r}, t); \qquad \boldsymbol{\nabla} \to \boldsymbol{\nabla} + \mathbf{F}(\mathbf{r}, t), \tag{88}$$

where the *gauge fields* need to transform as

$$f \to f - i\dot{\alpha}; \qquad \mathbf{F} \to \mathbf{F} - i\boldsymbol{\nabla}\alpha. \tag{89}$$

This seems ridiculously contrived, and has just replaced the α problem by a new one. But these transformations of the gauge field should look familiar: they are the ones we have already used in electromagnetism. This means that the electromagnetic field will do the required trick for us. We can therefore write a gauge invariant Schrödinger equation, in which phase is locally observable:

$$i\hbar\frac{\partial\psi}{\partial t} = \left(\frac{(\mathbf{p} - q\mathbf{A})^2}{2m} + V\right)\psi + q\phi\psi. \tag{90}$$

This substitution (relativistically $p^\mu \to p^\mu - qA^\mu$) is known traditionally as *minimal coupling*. Note that we have set the arbitrary constant of proportionality to include the electric charge. This is clearly reasonable: in the case of just an electrostatic field, we would expect to add $q\phi$ to the potential, V.

A more powerful view of this manipulation is to say that *gauge invariance is why electromagnetism exists*. The same reasoning applies with other local symmetries of the wave function that correspond to the nuclear forces: they can only stay hidden if we introduce gauge fields to accomplish this. The electroweak bosons, W & Z, and the gluons of the strong force, are all gauge bosons – analogues of the photon. To make this analysis watertight, we should really demonstrate that the classical Hamiltonian of a particle in an electromagnetic field is indeed

$$H = \frac{(\mathbf{p} - q\mathbf{A})^2}{2m} + q\phi. \tag{91}$$

For non-examinable interest, this is done in the appendix.

To get the first-order perturbation to Schrödinger's equation, write $(\mathbf{p} - q\mathbf{A})^2 = \mathbf{p}^2 - q(\mathbf{p}\cdot\mathbf{A} + \mathbf{A}\cdot\mathbf{p})$, and neglect the $\mathbf{A}^2$ term (for now). Because $\boldsymbol{\nabla}\cdot\mathbf{A} = 0$ and $\mathbf{p} \propto \boldsymbol{\nabla}$, $(\mathbf{p}\cdot\mathbf{A})\psi = (\mathbf{A}\cdot\mathbf{p})\psi$, so

$$i\hbar\frac{\partial\psi}{\partial t} = \frac{\mathbf{p}^2}{2m}\psi - \frac{q}{m}\mathbf{A}\cdot\mathbf{p} \equiv H_0 + H_{\text{int}} \tag{92}$$

(remember we have $\phi = 0$). The interaction, or perturbation, Hamiltonian is thus

$$H' = -\frac{q}{m}\mathbf{A}\cdot\mathbf{p}\,. \tag{93}$$

3.3 Semiclassical transitions

As a first application of this key result, we consider the case where the electromagnetic field is treated as an applied external quantity. This is the *semiclassical approximation*, in which quantum mechanics is applied to the electron, but not to the electromagnetic field. We shall expose the paradox that this creates and then show how it is resolved with proper quantum electrodynamics.

To keep the discussion as simple as possible, consider the *dipole approximation*, in which we ignore any variation of the EM field over the atom. It's not obvious that this is correct: by dimensional analysis, the wavelength of radiation that would ionize an atom must be of order the size of atomic orbits. But in practice the dimensionless factors work in our favour. Consider the Bohr atom, where the orbital radii are $r = \alpha^{-1}(\hbar/mc)n^2$ and the energies are $E = -\alpha^2(mc^2/2)(1/n^2)$, where $\alpha = e^2/4\pi\epsilon_0\hbar c \simeq 1/137$ is the fine structure constant. Thus the wavelength of the radiation that would ionise an atom in the $n = 1$ state is $\lambda = (4\pi/\alpha)r_1 \simeq 1722r_1$, so the dipole approximation is good at the 0.1% level. In this case,

$$\langle f|\mathbf{A}\cdot\mathbf{p}|i\rangle \simeq \mathbf{A}\cdot\langle f|\mathbf{p}|i\rangle. \tag{94}$$

The latter matrix element can be simplified by using a useful commutator with the unperturbed Hamiltonian:

$$[H_0, x] = -\frac{\hbar^2}{2m}\left(\nabla^2 x - x\nabla^2\right) = -\frac{i\hbar}{m}\,p_x. \tag{95}$$

Applying this, and recalling $\langle\phi|AB|\psi\rangle = \langle A\phi|B\psi\rangle$ for any two Hermitian operators, $\langle[H_0, x]\rangle = (E_f - E_i)\langle x\rangle$, so that

$$\langle f|\mathbf{p}|i\rangle = im\omega_0\langle f|\mathbf{x}|i\rangle. \tag{96}$$

Hence, the transition rate (for radiation polarized in the $\mathbf{e}$ direction) is

$$\Gamma = \frac{2\pi}{\hbar^2}\omega^2 A^2 |\langle f|q\mathbf{e}\cdot\mathbf{x}|i\rangle|^2 \delta(\omega - \omega_0), \tag{97}$$

which depends on the dipole moment of the atom. We could have got this answer immediately by writing down an electrostatic perturbation $\delta H = q\mathbf{x}\cdot\mathbf{E}$, but this would have been a serious fudge. Note that the origin is irrelevant: adding a constant vector to $\mathbf{x}$ adds a term proportional to $\langle f||i\rangle$, which is zero through orthogonality. In practice, an origin at the centre of mass will be used, but this was not necessary.

Because the transition rate is proportional to E^2, we see that it is proportional to the energy density (or energy flux density) of the incident radiation. For a single plane wave, $U = (\epsilon_0 E^2 + B^2/\mu_0)/2 = \epsilon_0 E^2$. This oscillates, with an average over time (or space) of $U = \epsilon_0 E_{\max}^2/2$. Because we took $\delta H = H' \times 2\sin\omega t$, the previous expression becomes

$$\Gamma = \frac{2\pi}{\hbar^2}(E_{\max}/2)^2 |\langle f|q\mathbf{e}\cdot\mathbf{x}|i\rangle|^2 \delta(\omega - \omega_0) = \frac{\pi}{\epsilon_0\hbar^2} U |\langle f|q\mathbf{e}\cdot\mathbf{x}|i\rangle|^2 \delta(\omega - \omega_0). \tag{98}$$

Finally, if we integrate over ω, assuming a broad-band contribution to the energy density, $U = \int_0^\infty U_\omega\, d\omega$,

$$\Gamma = \frac{\pi}{\epsilon_0\hbar^2} U_\omega |\langle f|q\mathbf{e}\cdot\mathbf{x}|i\rangle|^2. \tag{99}$$

More realistically, we will not have radiation travelling in a single direction, but there will be a set of waves travelling in a variety of directions. The total transition rate is the sum of the rates due to these independent disturbances. Since energy densities add, the total rate is also proportional to the total energy density of radiation (with the caveat that the matrix element depends on direction, so we should really consider the energy density of the radiation travelling in a small range of solid angles about a given direction).

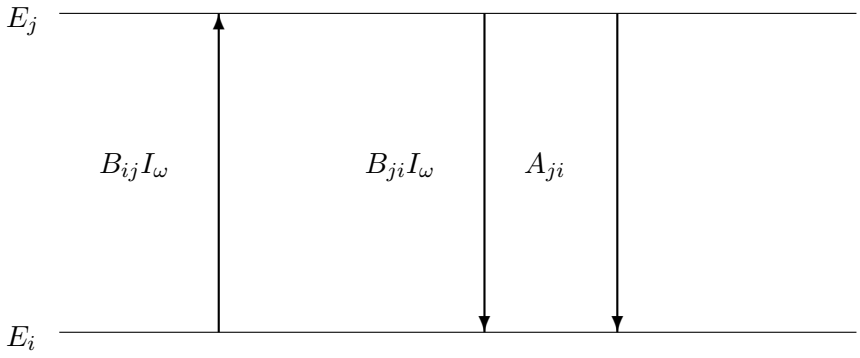
3.3.1 Spontaneous transitions

The semiclassical transition rate obeys *detailed balance*: the induced rate for $i \to f$ is the same as the reverse rate $f \to i$. Also, the transition rate vanishes if the applied field is zero. We know that both of these properties are incorrect. In the latter case, we observe that excited states of atoms have a finite lifetime and always decay to the ground state. In the former case, detailed balance would imply that the population of levels i and f must be equal in thermal equilibrium (where the radiation field is black-body); but in fact the level populations must be in the ratio dictated by a Boltzmann factor:

$$p_i/p_f = \exp[-(E_i - E_f)/kT]. \tag{100}$$

A cure for these problems was produced by Einstein, in a brilliant piece of improvisation. He accepted the need for spontaneous processes, but guessed that the semiclassical expression was correct in representing a term proportional to the radiation intensity. The transition rates between states i and j, with $E_j > E_i$, are then

$$\Gamma_{ij} = BI_\omega; \qquad \Gamma_{ji} = A + BI_\omega. \tag{101}$$



Invoking detailed balance plus Boltzmann occupation probabilities for the two levels gives

$$\frac{\Gamma_{ji}}{\Gamma_{ij}} = \exp(\hbar\omega/kT), \tag{102}$$

From which A can be deduced in terms of B. The usual textbook expression for A is completely obscure, but there is a better way to write this. Consider the usual expression for black-body radiation, which is

$$I_\nu = \frac{2h\nu^3/c^2}{\exp(h\nu/kT)-1} = (2\nu^2/c^2)\times h\nu \times N_\nu; \qquad N_\nu = (\exp(h\nu/kT)-1)^{-1} \tag{103}$$

i.e. a number of modes $\propto \nu^2\, d\nu$, times a photon energy, times the number of photons in that mode. In terms of this photon occupation number, the ratio of transition rates is

$$\frac{\Gamma_{ji}}{\Gamma_{ij}} = \frac{N+1}{N}. \tag{104}$$

In other words, if stimulated transitions proceed at a rate $\propto N$, the spontaneous transitions occur at a rate corresponding to a single additional photon being present. We now show how this very expression emerges when we treat the radiation field in a proper quantum manner.

3.4 Quantum radiative transitions

In discussing transitions properly, we need to specify initial and final states for both electron *and* radiation. Apart from small perturbations, these aspects are independent and so we write *product states*:

$$|i\rangle = |X\rangle\,|N_{\mathbf{k},\alpha}\rangle; \qquad |f\rangle = |Y\rangle\,|N'_{\mathbf{k},\alpha}\rangle, \tag{105}$$

where the atomic states are X & Y and $|N_{\mathbf{k},\alpha}\rangle$ is shorthand for $|N_{\mathbf{k}_1,\alpha_1}, N_{\mathbf{k}_2,\alpha_2}, \cdots\rangle$.

The matrix element of H' between an initial state and the final state is therefore

$$\langle Y|\langle N'_{\mathbf{k},\alpha}|\, H'\, |X\rangle|N_{\mathbf{k},\alpha}\rangle\,. \tag{106}$$

Since the first-order H' is linearly proportional to $\mathbf{A}$, this matrix element will be proportional to the radiation factor $\langle N'_{\mathbf{k},\alpha}|\mathbf{A}|N_{\mathbf{k},\alpha}\rangle$. It can immediately be seen that this places

strong restrictions on the allowed transitions, as follows. We have previously seen that the vector potential can be expressed as a sum of creation and annihilation operators:

$$\mathbf{A} = \sum_{\mathbf{k},\alpha} \sqrt{\frac{\hbar}{2\epsilon_0 \omega V}}\, \mathbf{e}_{\mathbf{k},\alpha} \left(a_{\mathbf{k},\alpha}(t) e^{i\mathbf{k}\cdot\mathbf{r}} \; + \; a^\dagger_{\mathbf{k},\alpha}(t) e^{-i\mathbf{k}\cdot\mathbf{r}} \right) . \tag{107}$$

The effect of $\mathbf{A}$ on an initial photon state $|N_i\rangle$, containing N_i photons, is to raise or lower N_i by one. Therefore the final state must be either $|N_i+1\rangle$ or $|N_i-1\rangle$, otherwise the matrix element will vanish. But each a or $a^\dagger$ applies only to a single mode, so all the occupation numbers in every other mode must stay the same.

Thus the proper quantum treatment of the radiation field guarantees that a single photon is either emitted or absorbed in just one of the modes. The emission is photon cloning: we increase N by one in the same mode that is doing the exciting. Thus the emitted photon has the same direction as the incident radiation, and is completely coherent with it. This is what makes the laser possible, and it is a property that is not obvious from the semiclassical analysis. The mode of interest is the one with $\hbar\omega$ $(= \hbar|\mathbf{k}|c) = \hbar\omega_0 = E_{\rm Y} - E_{\rm X}$. We know this has to be the case physically; mathematically, it arises from the delta function in the transition rate:

$$\Gamma = \frac{2\pi}{\hbar^2} |H'_{fi}|^2 \; \delta(\omega - \omega_0). \tag{108}$$

The radiation frequency is set by recalling that $a(t) \propto \exp(-i\omega t)$.

Non-examinable aside: This time dependence may possibly cause some confusion, since normally in quantum mechanics we are familiar with the *Schrödinger picture* in which the wave function depends on time and the operators are independent of time. But in dealing with radiation, it is more natural to use the opposite *Heisenberg picture*, which involves the unitary transformation of wave function and operators:

$$\begin{aligned} \psi_{\rm H} &= e^{iHt/\hbar} \psi_{\rm S} \\ O_{\rm H} &= e^{iHt/\hbar} O_{\rm S} e^{-iHt/\hbar}. \end{aligned} \tag{109}$$

In the Heisenberg representation, the equation of motion corresponding to the Schrödinger equation is now an equation for the operator: $i\hbar \dot{O}_{\rm H} = [O, H]$ (this is proved by considering the time derivative of the matrix element $\langle i|O|j\rangle$ and using the facts that $O_{\rm S}$ is time independent and $H_{\rm S} = H_{\rm H}$). In this way, the radiation eigenstates $|N_i\rangle$ can be time-independent statements about the occupation number of a given mode.

We can now distinguish two processes, depending on the sign of $E_{\rm Y} - E_{\rm X}$:

Absorption: Because of the normalization of the operators, $a_{\mathbf{k},\alpha}|N_{\mathbf{k},\alpha}\rangle = \sqrt{N_{\mathbf{k},\alpha}}|N_{\mathbf{k},\alpha}-1\rangle$, The radiation matrix element is

$$\sqrt{\frac{\hbar}{2\epsilon_0 \omega V}}\, \mathbf{e}_{\mathbf{k},\alpha} e^{i\mathbf{k}\cdot\mathbf{r}} \sqrt{N_{\mathbf{k},\alpha}}, \tag{110}$$

and the overall matrix element for the absorption of a photon by an atomic system is

$$H'_{fi} = \frac{q}{m} \sqrt{\frac{\hbar}{2\epsilon_0 \omega V}}\, \sqrt{N_{\mathbf{k},\alpha}}\; M_{\rm YX}(\mathbf{k}, \alpha), \tag{111}$$

where

$$M_{\rm YX}(\mathbf{k}, \alpha) \equiv \mathbf{e}_{\mathbf{k},\alpha} \cdot \int \psi^*_{\rm Y} e^{i\mathbf{k}\cdot\mathbf{r}} \mathbf{p} \psi_{\rm X} \; dV. \tag{112}$$

Emission: If Y had lower energy than X, leading to emission, the matrix element would be the nearly the same: $\sqrt{N_{\mathbf{k},\alpha}}$ would become $\sqrt{N_{\mathbf{k},\alpha}+1}$ and $e^{i\mathbf{k}\cdot\mathbf{r}}$ would become $e^{-i\mathbf{k}\cdot\mathbf{r}}$.

Density of states: To handle the delta-function in frequency, we need to recognise that this is only an approximation, and that modes of frequency in a small range around $\omega_0 = (E_Y - E_X)/\hbar$ can all cause transitions. We don't care which of the $N_{\mathbf{k},\alpha}$ changes: it could be any one of the modes that lie close to the transition frequency. Therefore, in order to get the atomic transition rate, we should add up the effect of all modes (even though only one of them will actually make the transition). This gives us an expression that is identical to the Golden Rule:

$$\Gamma = \frac{2\pi}{\hbar^2} |H'_{fi}|^2 \, \rho_\omega(\omega = \omega_0). \tag{113}$$

But now the mode density ρ_ω is a density of *perturbing* states, not of the initial or final electron states (which remain discrete) – so the meaning of this equation is quite different.

The number of photon states is given by the usual expression (for a single polarization):

$$dN_{\text{modes}} = \frac{V}{(2\pi)^3} \, k^2 dk \, d\Omega. \tag{114}$$

The required mode density is then (since $kc = \omega$)

$$\rho_\omega = \frac{V}{(2\pi)^3} \, (\omega^2/c^3) \, d\Omega. \tag{115}$$

This then gives (putting $q = -e$ for the electron) the transition rate for absorption

$$\Gamma_{\text{abs}} = \frac{\omega \, d\Omega}{2\pi\hbar c^3 m^2} \, N_{\mathbf{k},\alpha} \, \frac{e^2}{4\pi\epsilon_0} \, |M_{YX}(\mathbf{k},\alpha)|^2 \tag{116}$$

This is the probability per second that a photon of polarization α, frequency $\omega = |\mathbf{k}|c$ and direction within $\mathrm{d}\Omega$ of $\mathbf{k}$ will be absorbed, and the term obtained by replacing $N_{\mathbf{k},\alpha}$ by $N_{\mathbf{k},\alpha} + 1$ is the corresponding emission probability Γ_{em}.

3.5 Electric dipole transitions

We have previously seen the *dipole approximation*, which consists of putting $e^{i\mathbf{k}\cdot\mathbf{r}} = 1$ in the expression for $M_{YX}(\mathbf{k},\alpha)$, assuming that the perturbing 'potential' doesn't vary over the extent of the atomic system. We also saw how to re-express the atomic matrix element of $\mathbf{p}$ in terms of that of $\mathbf{x}$: $\langle f|\mathbf{p}|i\rangle = im\omega_0\langle f|\mathbf{x}|i\rangle$. In this approximation therefore,

$$\Gamma_{\text{abs}} = \frac{\omega^3 d\Omega}{2\pi\hbar c^3} \, N_{\mathbf{k},\alpha} \, \frac{e^2}{4\pi\epsilon_0} \, |\mathbf{r}_{YX} \cdot \mathbf{e}_{\mathbf{k},\alpha}|^2, \qquad \text{where } \mathbf{r}_{YX} = \int \psi_Y^* \, \mathbf{r} \, \psi_X \, dV \, . \tag{117}$$

To relate this to the previous semiclassical expression, consider the specific energy density U_ω due to modes in the frequency range $d\omega$ and the solid angle range $d\Omega$:

$$U_\omega = \frac{1}{V} \, N \, \hbar\omega \, \frac{dN_{\text{modes}}}{d\omega} = \frac{\hbar\omega^3}{(2\pi)^3 c^3} \, N. \tag{118}$$

Thus, the absorption rate is identical to the semiclassical expression:

$$\Gamma = \frac{\pi}{\epsilon_0 c\hbar^2} \, I_\omega \, d\Omega \, |\langle f|q\mathbf{e}\cdot\mathbf{x}|i\rangle|^2. \tag{119}$$

The *spontaneous* emission rate corresponds to replacing $N_{\mathbf{k},\alpha} + 1$ by 1. The expression still depends on polarization direction, which we don't care about when it comes to the total rate. We can eliminate this by noting that $\mathbf{e}_1, \mathbf{e}_2$ and $\mathbf{k}$ are all mutually perpendicular, so that $|\mathbf{e}_1 \cdot \mathbf{r}|^2 + |\mathbf{e}_2 \cdot \mathbf{r}|^2 = |\mathbf{r}|^2 - |\hat{\mathbf{k}} \cdot \mathbf{r}|^2 = |\mathbf{r}|^2 \sin^2\theta$ where θ is the angle between $\mathbf{k}$ and $\mathbf{r}$. Integrating $\sin^2\theta$ over solid angle gives $8\pi/3$. Finally, then, the rate of spontaneous emission of a photon of any polarization in any direction is

$$\Gamma_{\text{spon}} = \frac{4\omega}{3\hbar c^3 m^2} \, \frac{e^2}{4\pi\epsilon_0} \, |\mathbf{p}_{YX}|^2 = \frac{4\omega^3}{3\hbar c^3} \, \frac{e^2}{4\pi\epsilon_0} \, |\mathbf{r}_{YX}|^2 \, . \tag{120}$$

3.5.1 Dipole selection rules

Γ depends on the matrix element $\int \psi_{\rm Y}^* \, \mathbf{r} \, \psi_{\rm X} \, dV$. Typically, the single particle wave functions ψ can be expanded – as is the case for the hydrogen atom – in products of spherical harmonics $Y(\theta, \phi)$, representing eigenstates of orbital angular momentum. The space integral then factors into a radial integral, and an angle integral.

As an example, the z-component of $\mathbf{r}$ is $r\cos\theta$. The integration over $dV = r^2 dr \, d\Omega$ is separable and the resulting (θ, ϕ) integral in the matrix element is

$$\int Y^*_{\ell_{\rm Y}, m_{\rm Y}}(\theta, \phi) \; \cos\theta \; Y_{\ell_{\rm X}, m_{\rm X}}(\theta, \phi) \; d\Omega. \tag{121}$$

This integral is only non-zero if $\ell_{\rm Y} = \ell_{\rm X} \pm 1$ (because of the orthogonality properties of the $Y(\theta, \phi)$, and in general if state Y is of opposite parity to state X). The selection rules are complicated by the issue of spin angular momentum, which yields a total angular momentum that is different to the orbital angular momentum. In some multi-electron atoms, orbital and spin angular momentum may both be good quantum numbers (LS coupling), but this is not always the case (JJ coupling).

For reference, the selection rules for electric dipole transitions are

(1) $J_{\rm Y} = J_{\rm X} - 1, J_{\rm X}, J_{\rm X} + 1$ (J is total angular momentum quantum number)
(2) Parity change
(3) In the case of LS coupling, also
  (3a) no spin change
  (3b) same rules for orbital AM L as for total AM J.

3.6 Electric quadrupole and magnetic dipole transitions

The above selection rules often forbid astronomically interesting transitions, and we so have to improve on the dipole approximation. The details of this are not examinable.

The electric dipole approximation was $e^{i\mathbf{k}\cdot\mathbf{r}} \simeq 1$. But if the resulting matrix element vanishes, we must examine the next term in the expansion of $e^{i\mathbf{k}\cdot\mathbf{r}} = 1 + i\mathbf{k}\cdot\mathbf{r} + \cdots$. The transition rates from this next term are evidently smaller than the dipole term by roughly $(ka_0)^2$, where a_0 is the typical atomic size. The matrix element (*cf.* $M_{\rm YX}(\mathbf{k}, \alpha)$ in section 3.4) is

$$\mathbf{e}_{\mathbf{k},\alpha} \cdot \int \psi_{\rm Y}^* \; i(\mathbf{k}\cdot\mathbf{r})\mathbf{p}\psi_{\rm X} \, dV = im\mathbf{k} \cdot \int \psi_{\rm Y}^* \; \mathbf{r}\mathbf{v} \; \psi_{\rm X} \, dV \cdot \mathbf{e}_{\mathbf{k},\alpha}, \tag{122}$$

which involves the *dyadic* $\mathbf{rv}$. This is meaningless in conventional vector algebra, but it is a useful shorthand for a matrix: $(\mathbf{rv})_{ij} \equiv r_i v_j$. Thus $\mathbf{a}\cdot\mathbf{bc} = (\mathbf{a}\cdot\mathbf{b})\mathbf{c}$, $\mathbf{a}\wedge\mathbf{bc} = (\mathbf{a}\wedge\mathbf{b})\mathbf{c}$ etc.

We shall write the dyadic as the sum of symmetric and antisymmetric parts:

$$\mathbf{rv} = \frac{1}{2}(\mathbf{rv} + \mathbf{vr}) + \frac{1}{2}(\mathbf{rv} - \mathbf{vr}). \tag{123}$$

The matrix element of the first term can be rewritten as $\frac{1}{2}i\omega(\mathbf{rr})_{\rm YX}$. This is proved in the same way as we proved $\langle\mathbf{p}\rangle = im\omega\langle\mathbf{x}\rangle$, although an intuitive justification comes from treating the 'velocity' p/m literally: $\mathbf{rv} + \mathbf{vr} = d(\mathbf{rr})/dt$; just as the matrix element of $\mathbf{v} = d(\mathbf{r})/dt$ was $i\omega\mathbf{r}_{\rm YX}$, the matrix element of the first term is $\frac{1}{2}i\omega(\mathbf{rr})_{\rm YX}$. This symmetric tensor is analogous to the quadrupole tensor of a charge distribution, proportional to $\langle x_i x_j\rangle$. Normally that tensor would be defined as trace-free, and this could be done here as

well: if our matrix is isotropic, then the matrix element vanishes, because $\mathbf{e}_{\mathbf{k},\alpha}$ and $\mathbf{k}$ are perpendicular. Hence the name *quadrupole transition*. Such transitions are important for ground state transitions in molecular hydrogen, for example.

In the second term put $\mathbf{v} = \mathbf{p}/m$, and the expression for this part of the matrix element becomes (remember $\mathbf{k} = k\mathbf{n} = (\omega/c)\mathbf{n}$)

$$\frac{1}{2m} i\mathbf{k} \cdot (\mathbf{rp} - \mathbf{pr}) \cdot \mathbf{e}_{\mathbf{k},\alpha} = -\frac{1}{2m} i\mathbf{k} \wedge \mathbf{L} \cdot \mathbf{e}_{\mathbf{k},\alpha} = \frac{i\omega}{2mc} \mathbf{L} \wedge \mathbf{n} \cdot \mathbf{e}_{\mathbf{k},\alpha} = \frac{i\omega}{2mc} \mathbf{L} \cdot (\mathbf{e}_{\mathbf{k},\alpha} \wedge \mathbf{n}), \quad (124)$$

using $\mathbf{k} \cdot (\mathbf{rp} - \mathbf{pr}) = -\mathbf{k} \wedge (\mathbf{r} \wedge \mathbf{p}) = -\mathbf{k} \wedge \mathbf{L}$, $\mathbf{L}$ being the orbital angular momentum operator. Note that $(\mathbf{e}_{\mathbf{k},\alpha} \wedge \mathbf{n}) = \mathbf{e}_{\mathbf{k},\text{other } \alpha}$, which gives the direction of the $\mathbf{B}$ field. This term involves the **magnetic dipole** operator $e\mathbf{L}/2m$ (magnitude of order $e\hbar/2m$, the Bohr magneton, times an angular momentum quantum number). It thus couples the $\mathbf{B}$ field to the magnetic moment of the electric current associated with orbital motion in the atomic system (classically, the electric dipole energy is $\mathbf{E} \cdot \mathbf{d}$ while the magnetic dipole energy is $\mathbf{B} \cdot \boldsymbol{\mu}$).

The non-relativistic expression we used for the current density, $\mathbf{v}$, neglects electron spin angular momentum $\mathbf{S}$, and a more detailed treatment gives $\mathbf{L} \to \mathbf{L} + \mathbf{2S}$. Thus the transition rate for **the magnetic dipole** radiation is the same as for the electric dipole, but with electric dipole moment replaced by magnetic dipole moment: $e\mathbf{r} \to e(\mathbf{L} + \mathbf{2S})/2mc$.

3.6.1 Electric quadrupole and magnetic dipole selection rules

The selection rules for these types of transition – which evidently exclude each other at this level of approximation – are:

Electric quadrupole:
(1) $J_\text{Y} = (J_\text{X} \pm 2)$ *or* J_X $(J = 0 \not\to J = 0)$
(2) no parity change
(3) for LS coupling also: no orbital AM change, no spin change.

Magnetic dipole (incl. spin):
(1) $J_\text{Y} = J_\text{X} - 1, J_\text{X}, J_\text{X} + 1$
(2) no parity change
(3) for LS coupling: no spin change, so orbital AM change, only J may (and must) change, representing fine structure transitions in a single term.

4 Radiative Transfer

This section
(a) Revises specific intensity as a full classical description of radiation, and its properties

(b) relates the emission coefficient and opacity – required to derive the specific intensity – to the transition rate derived previously

(c) describes line profiles

4.1 Specific intensity and the equation of radiative transfer

Specific Intensity, I_ν, is the nearly complete classical description of the radiation field. It misses polarization, which requires another three closely related quantities: the Stokes parameters Q, U, V, which describe the fractional degree of circular and linear polarization, plus the orientation of linear polarization. We will generally neglect these complications, and handle polarization by saying that each mode can come in one of two polarization states – although usually we will consider unpolarized radiation.

I_ν is the rate at which energy crosses unit area, as carried by radiation in unit frequency range and in unit solid angle. This means it is related to the specific energy density we used previously:

$$cU_\nu \, d\nu = I_\nu \, d\nu \, d\Omega, \tag{125}$$

because cU is the rate of transport of energy per unit area. Note we have changed frequency unit from the physics ω to the more practical ν.

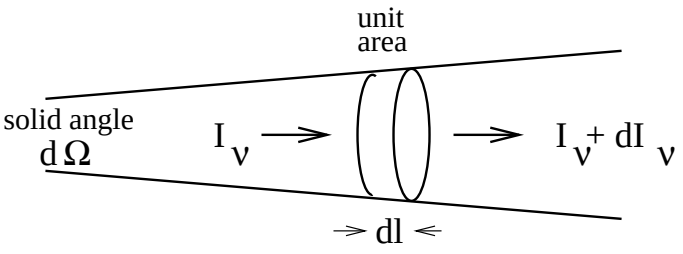
Specific intensity and photon occupancy I_ν is related to $N_{\mathbf{k},\alpha}$ in a straightforward manner. The energy flux density equals the number of photon states *per unit volume* in $d\nu$ times $E = h\nu$ times c times occupancy of these photon states, *i.e.*

$$I_\nu d\Omega \, d\nu = (k^2 d\, k d\Omega/(2\pi)^3) \times h\nu c \times (N_{\mathbf{k},1} + N_{\mathbf{k},2}). \tag{126}$$

Hence (supposing the radiation to be unpolarized: $N_{\mathbf{k},1} = N_{\mathbf{k},2}$), and using $k = 2\pi\nu/c$,

$$I_\nu = (2h\nu^3/c^2) \, N_{\mathbf{k},\alpha} \, , \tag{127}$$

where the factor 2 is because both polarizations are assumed equally populated.



Radiative transfer Consider for concreteness the transfer of radiation due to a particular bound/bound transition ($L \to U$, say, from *lower* to *upper*) of randomly oriented atoms. The number of absorbers per unit area in the elemental volume shown above is $n_{\rm L} \, d\ell$ and the number of emitters is $n_{\rm U} \, d\ell$. The photon occupancy is $N_{\mathbf{k},\alpha}$. The number of 'up' transitions per second due to photons in the beam $d\Omega$ is

$$n_{\rm L} \, d\ell \, \Gamma_{\rm abs}(d\Omega/4\pi) = n_{\rm L} \, d\ell \, \Gamma_{\rm spon} N_{\mathbf{k},\alpha}(d\Omega/4\pi) \tag{128}$$

while the number per second of 'down' transitions is

$$n_{\rm U}\, d\ell\, \Gamma_{\rm spon}(N_{\mathbf{k},\alpha}+1)(d\Omega/4\pi). \tag{129}$$

$(I_\nu d\Omega/h\nu)$ photons arrive at the volume per second, $(n_{\rm L}-n_{\rm U})\, d\ell\, \Gamma_{\rm spon} N_{\mathbf{k},\alpha}(d\Omega/4\pi)$ are removed and $n_{\rm U}\, d\ell\, \Gamma_{\rm spon}(d\Omega/4\pi)$ are added – all in the beam $d\Omega$.

In practice, absorption will occur at a range of frequencies near the central frequency, due to *e.g.* Doppler shifts from thermal motion. The total absorption rate is spread over frequency proportional to ϕ_ν, with $\int \phi_\nu d\nu = 1$. Then the change in flux at a given frequency as the beam passes through the volume can be written

$$dI_\nu d\Omega/h\nu = [-(n_{\rm L}-n_{\rm U})d\ell\, N_{\mathbf{k},\alpha}\Gamma_{\rm spon}(d\Omega/4\pi) + n_{\rm U} d\ell\, \Gamma_{\rm spon}(d\Omega/4\pi)]\, \phi_\nu \tag{130}$$

or by substituting $I_\nu = (2h\nu^3/c^2)\, N_{\mathbf{k},\alpha}$,

$$\frac{dI_\nu}{d\ell} = -(n_{\rm L}-n_{\rm U})\frac{\Gamma_{\rm spon}\lambda^2}{8\pi} I_\nu\phi_\nu + n_{\rm U}\frac{\Gamma_{\rm spon}h\nu}{4\pi}\phi_\nu. \tag{131}$$

This can be written

$$\frac{dI_\nu}{d\ell} = -\kappa_\nu I_\nu + \mathcal{E}_\nu \tag{132}$$

which is known as the **equation of radiative transfer**. Here

$$\kappa_\nu = (n_{\rm L}-n_{\rm U})\Gamma_{\rm spon}\lambda^2\phi_\nu/8\pi \tag{133}$$

is the *opacity* (units m^{-1}) and

$$\mathcal{E}_\nu = n_{\rm U}\Gamma_{\rm spon}h\nu\phi_\nu/4\pi \tag{134}$$

(units are $\mathrm{W\,m^{-3}\,sr^{-1}}$) is the *emissivity*.

Note that we have assumed that the frequency dependence of the emission and absorption processes will be the same. This can be justified by detailed balance: at high densities we expect to reach thermal equilibrium where I_ν is the Planck function, which doesn't change along the photon path. Then emission and absorption must balance, requiring ϕ_ν to be the same in both terms.

At times we may want to work with angular frequency, ω, instead of $\nu = \omega/2\pi$. The necessary conversions are $I_\omega\, d\omega = I_\nu\, d\nu$, so that $I_\omega = I_\nu/2\pi$. Similarly, we could have a normalized line profile in ω space: $\phi_\omega = \phi_\nu/2\pi$. Be aware of a potential confusion of the notation: the opacity is independent of our convention, but emissivity is a 'per unity bandwidth' quantity. Thus we would have

$$\kappa_\omega = (n_{\rm L}-n_{\rm U})\Gamma_{\rm spon}\lambda^2\phi_\omega/4; \qquad \mathcal{E}_\omega = n_{\rm U}\Gamma_{\rm spon}\hbar\omega\phi_\omega/4\pi. \tag{135}$$

Correction for stimulated emission. Note that the opacity represents the difference of two effects: removal of photons via upwards transitions versus stimulated creation of new photons (clones of the original, as can be seen by the quantum treatment in which the radiation perturbation contains creation and annihilation operators). These two effects are seen in $\kappa_\nu \propto \Gamma_{\rm spon}\,(n_{\rm L}-n_{\rm U})$, which is perhaps better written as $\Gamma_{\rm spon}\, n_{\rm L}(1-n_{\rm U}/n_{\rm L})$ – i.e. as a modification of our naive expectation in which the opacity would only correspond to absorption transitions form L to U.

The factor $(1-n_{\rm U}/n_{\rm L})$ can have a dominating effect. For example, if the states are Boltzmann populated (thermodynamic equilibrium), and if (Rayleigh-Jeans limit) $h\nu \ll kT$, the factor becomes $h\nu/kT$, a very small quantity. The states are so nearly equally populated that almost as much stimulated emission as absorption occurs, and the material appears

almost transparent. In certain cases, a non-thermal effect can tip the balance, and $n_{\rm U} > n_{\rm L}$ in which case the opacity is negative and the specific intensity *grows*. This is maser emission (Microwave – it usually is – Amplification by Stimulated Emission of Radiation).

Modification for multiple states per level. Often one requires the emissivity and opacity for transitions between two *energy levels* L and U – each level containing several states ($g_{\rm L}$ and $g_{\rm U}$ respectively). In this case, the opacities and emissivities for each pair of individual states are added, weighted by their occupation probabilities, to give the total.

The simplest case is when the individual states in a level are equally populated (*e.g.* by collisions). Then the fraction in each state is $1/g$. For upward transitions, each lower level now has $g_{\rm U}$ states to jump to, so the rate increases by a factor $g_{\rm U}$. Similarly, the downward rate increases by a factor $g_{\rm L}$. There is no factor from the degeneracy of the starting level: there are $g_{\rm L}$ choices for starting an upward jump, but the probability of being in any one of them is $1/g_{\rm L}$ so these factors cancel. In this case

$$\kappa_\nu = (g_{\rm U} n_{\rm L} - g_{\rm L} n_{\rm U}) \Gamma_{\rm spon} \lambda^2 \phi_\nu / 8\pi; \tag{136}$$

$$\mathcal{E}_\nu = g_{\rm L} n_{\rm U} \Gamma_{\rm spon} h\nu \phi_\nu / 4\pi. \tag{137}$$

Absorption cross-section. A powerful way of thinking about absorption is in terms of a cross-section, σ. Imagine a set of particles of number density n, each dressed with a circle of area σ, and assume that absorption takes place whenever a photon intersects one of these circles. Now consider a length L of a photon trajectory and dress it with the cross-section to make a cylinder of volume $L\sigma$. Absorption will occur if a particle lies within that cylinder, for which the probability is $p = nL\sigma$ (if this is small). The fractional reduction in specific intensity on passing through the cylinder will be pI on average, which must also be κL. Thus in general, $\kappa_\nu = n\sigma_\nu$. We have the general expression for κ_ν above. It simplifies in the common case that the absorbing material is all in the lower state, so that we can ignore stimulated processes:

$$\sigma_\nu^{\rm abs} = g_{\rm U} \Gamma_{\rm spon} \lambda^2 \phi_\nu / 8\pi. \tag{138}$$

4.2 Optical depth and a solution to the equation of radiative transfer

Before considering the detailed behaviour of specific intensity, note that in empty space opacity and emissivity are zero. As a consequence, $dI_\nu/d\ell$ is zero and *specific intensity is constant along a line of sight (or ray) in empty space*. This constancy is the crux of observational astronomy: I_ν is set local to a source and carries information about the conditions there unchanged to Earth.

This statement needs slight modification in cosmology, since the specific intensity is altered by redshifting. The simplest way to see this is to recall that $I_\nu \propto \nu^3 N$, so that I_ν/ν^3 is proportional to the occupation number, which must be a relativistic invariant. Therefore, if redshifting alters frequency, the Specific Intensity (at the redshifted frequency) is altered by the cube of the redshifting factor:

$$I_\nu^{\rm observed}(\nu) = (1+z)^{-3} I_\nu^{\rm emitted}([1+z]\nu)\ . \tag{139}$$

But if the propagation is not in vacuum, then we must allow for radiative transfer. Define **optical depth** τ_ν by $d\tau_\nu = \kappa_\nu\, d\ell$ to get

$$\frac{dI_\nu}{d\tau_\nu} + I_\nu = \frac{\mathcal{E}_\nu}{\kappa_\nu} = S_\nu \tag{140}$$

where the latter equality *defines* the **source function** S_ν. If κ_ν, S_ν are independent of position, the equation integrates to

$$I_\nu = I_\nu(0)e^{-\tau_\nu} + S_\nu(1 - e^{-\tau_\nu}). \tag{141}$$

The $e^{-\tau_\nu}$ factor in the first term has a simple interpretation in terms of scattering probabilities. Suppose the probability of a photon being scattered in traversing a certain pathlength is p; thus the probability of not being scattered is $1-p$, and the transmitted intensity would be multiplied by $1 - p$. What is the probability of not being scattered? Each scattering is a rare event, so these should obey a Poisson distribution. If the expected number of scattering events is μ, the probability of no events is $e^{-\mu}$. So we see that the optical depth is just the expected number of scattering events, which we previously wrote in terms of a cross-section, as $nL\sigma$.

If there is no background radiation, then

$$I_\nu = S_\nu(1 - e^{-\tau_\nu}) \tag{142}$$

and two extreme cases can be seen.

(1) $\tau_\nu \ll 1$ – optically thin emission:
expand $\exp(-\tau_\nu)$ to two terms and get

$$I_\nu = S_\nu\tau_\nu = \mathcal{E}_\nu\ell \tag{143}$$

(2) $\tau_\nu \gg 1$ – optically thick emission:
$\exp(-\tau_\nu) \to 0$ and

$$I_\nu = S_\nu = \mathcal{E}_\nu/\kappa_\nu = \mathcal{E}_\nu \times (\text{mean free path}) \tag{144}$$

4.3 Line profiles

The transitions $U \to L$ appear to involve exact energy differences, but this is not consistent with the energy-time uncertainty principle:

$$\Delta E\, \Delta t \gtrsim \hbar, \tag{145}$$

where Δt is the time available to measure the energy, and ΔE is the resulting uncertainty in the measurement. Since $\Delta t \simeq 1/\Gamma_{\rm spon}$, the energy uncertainty is $\hbar\Gamma_{\rm spon}$, giving a finite width to the line. The natural profile of a spectral line is the *Lorentzian* function, proportional to $1/[(\Gamma_{\rm spon}/2)^2 + (\omega - \omega_{\rm UL})^2]$. This is obtained by treating the probability of remaining in the upper state, which is $\exp(-\Gamma t)$, as corresponding to a field amplitude decaying as $\exp(-\Gamma t/2)$. The Lorentzian arises when we Fourier transform this and use Parseval's theorem to argue that the Intensity as a function of frequency is proportional to the modulus-squared of the FT (this brief outline argument is not examinable).

An effect normally much greater, at least in emission lines, is Doppler, or thermal, broadening. Each atomic system in a source in thermal equilibrium at temperature T is moving with a velocity associated with the energy distribution. In one dimension (along the line of sight) the velocities are distributed as $\exp(-E/kT) = \exp(-v^2/(2kT/m))$. The rms velocity of this distribution is $\sigma_v = \sqrt{kT/m}$. The frequency of the line is altered by $\delta\nu/\nu = v/c$, and the line has a Gaussian profile with rms frequency width σ_v/λ.

To express this frequency spread we modify the emissivity and opacity coefficients by multiplying each by a normalized profile (or frequency probability distribution). Put $\mathcal{E}_\nu = \mathcal{E}_{\rm UL}\phi_\nu$ and $\kappa_\nu = \kappa_{\rm LU}\tilde{\phi}_\nu$ with ϕ_ν, $\tilde{\phi}_\nu$ each satisfying $\int \phi_\nu d\nu = 1$. Then each, integrated over the

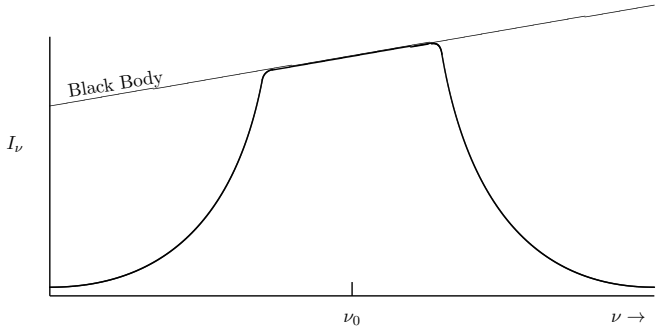
frequency range, gives the value consistent with the total transition rate. In the example of Doppler broadening the expression for ϕ_ν is often written

$$\phi_\nu = \frac{1}{\sqrt{\pi} b} e^{(\nu-\nu_0)^2/b^2}, \quad \text{where } b = \sqrt{2}\sigma_v/\lambda. \tag{146}$$

b is called the Doppler broadening parameter (note the non-standard factor of $\sqrt{2}$).

In the common case that the emissivity and the opacity have the same origin, the source function is independent of the profile shape since it is the ratio $\mathcal{E}_\nu/\kappa_\nu$. This will occur for instance if the emission and absorption processes are dominated by the same Doppler broadening, or generally are part of the same process.

Radiative transfer modifies this line profile in an interesting way. When the optical depth at the centre of the line approaches and exceeds one, the optically thick solution applies and the specific intensity there cannot exceed the source function – whereas the tails of the profile remain optically thin. If the occupation of the energy levels is set by thermal equilibrium (*e.g.* via collisions) then the source function must be black-body radiation, $\mathcal{E}_\nu/\kappa_\nu = S_\nu = B_\nu(T)$. The profile will be a Gaussian profile chopped off at the black body specific intensity characteristic of the kinetic temperature of the emitting material.

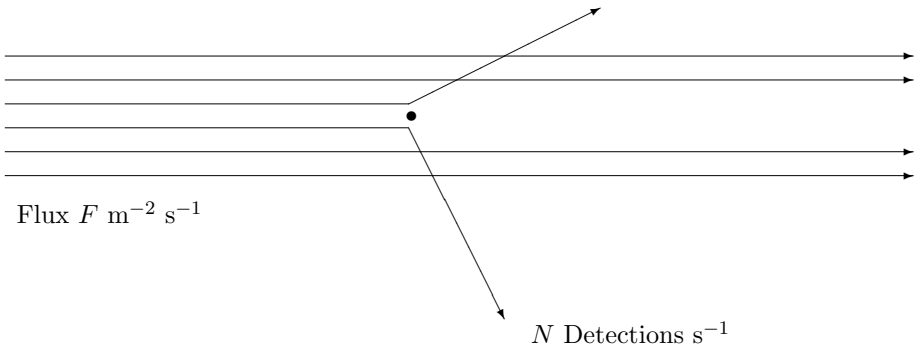


5 Collisions

This section
(a) describes cross section

(b) derives reciprocity between collisions causing upward and downward transitions

(c) defines critical density for a pair of levels



5.1 Cross section and collision rate

Collisional excitation is one of the two principal means by which atoms and ions in neutral or partly ionized gases become excited (the other being via absorption of radiation). The cross section for a collision process is defined above. The rate of detection of the process occurring, N s^{-1}, is proportional to the flux F. The constant of proportionality, σ, has dimensions m^2 and is called the **cross-section**.

In Maxwellian distributions, collisions between species '1' and '2' proceed at a rate

$$n_1 n_2 \langle \sigma v \rangle \quad \text{m}^{-3}\ \text{s}^{-1} \tag{147}$$

where n_1, n_2 are particle number densities (m^{-3}) and $\langle \sigma v \rangle$ is the Maxwell average of the product σv. (see appendix).

This also leads gives the definition of **mean free path**, which is the mean distance travelled by particle '2' before 'hitting' a particle '1' (a 'hit' is an elastic collision, an excitation, or any reaction of interest for which the cross section σ can be calculated or measured):

$$\text{MFP} = \frac{1}{n_1 \sigma}. \tag{148}$$

The **collision time** t_{coll} is the mean time for such a hit, and is evidently

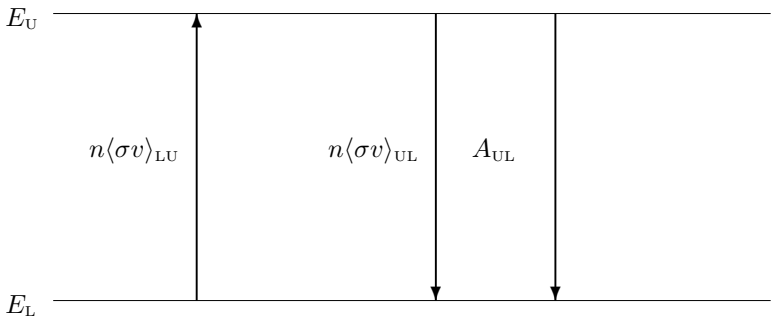
$$t_{\text{coll}} = \frac{1}{n_1 \langle \sigma v \rangle}. \tag{149}$$

5.2 Critical density

A useful concept in interpreting level populations is **critical density**, defined in a two-level system. The radiation field is presumed to be negligible, and collisions cause up and

down transitions. The particles colliding with the 2-level system have number density n. In statistical equilibrium (rate of ups equals rate of downs in an ensemble of such systems):

$$n_{\rm L}\, n\langle\sigma v\rangle_{\rm LU} = n_{\rm U}\left(n\langle\sigma v\rangle_{\rm UL} + g_{\rm L}\Gamma_{\rm spon}\right). \tag{150}$$



In the case where n is extremely large, we have

$$\frac{n_{\rm U}}{n_{\rm L}} = \frac{\langle\sigma v\rangle_{\rm LU}}{\langle\sigma v\rangle_{\rm UL}}\,. \tag{151}$$

We should also expect thermal equilibrium in this case, with Boltzmann level populations $(n_{\rm U}/g_{\rm U})/(n_{\rm L}/g_{\rm L}) = \exp(-\Delta E/kT)$; this gives an important **reciprocity relation**:

$$\langle\sigma v\rangle_{\rm LU} = (g_{\rm U}/g_{\rm L})\langle\sigma v\rangle_{\rm UL} e^{-\Delta E/kT}. \tag{152}$$

Although the de-excitation rate $\langle\sigma v\rangle_{\rm UL}$ is not independent of temperature, the exponential factor means that it is much less temperature sensitive than the excitation rate. One way of understanding this is to appreciate that excitation will only be achieved by electrons of energy $> \Delta E$, and their abundance is suppressed in the Maxwellian by exactly the exponential factor. Conversely, de-excitation can be achieved by electrons of any energy, so this rate varies only as a power of temperature. Since temperatures in nebular gas are often set by photoionization and are close to a few times 10^4 K, this variation is unimportant.

Now rearranging the original equation for $n_{\rm U}$ and $n_{\rm L}$, using the reciprocity relation between $\sigma_{\rm LU}$ and $\sigma_{\rm UL}$, and defining the **critical density** as

$$n_{\rm crit} = g_{\rm L}\Gamma_{\rm spon}/\langle\sigma v\rangle_{\rm UL}, \tag{153}$$

we get

$$\frac{n_{\rm U}}{n_{\rm L}} = \frac{(g_{\rm U}/g_{\rm L})e^{-x}}{1 + n_{\rm crit}/n}, \qquad x = \frac{E_{\rm U} - E_{\rm L}}{kT}. \tag{154}$$

There are two limiting cases.

(1) When $n \gg n_{\rm crit}$ the denominator is nearly unity, and the levels are Boltzmann populated. The emissivity is then independent both of collider density n and of $\langle\sigma v\rangle$:

$$\mathcal{E}_\nu = \frac{h\nu}{4\pi} n_{\rm U} g_{\rm L}\Gamma_{\rm spon}\phi_\nu \tag{155}$$

where

$$n_{\rm U} = n_{\rm total} g_{\rm U} e^{-E_{\rm U}/kT}/Q(T) \tag{156}$$

and $Q(T) = \sum g_n e^{-E_n/kT}$ is the *partition function*. In the thermodynamic equilibrium case, other things being equal, the emissivity is proportional to density.

(2) When $n \ll n_{\rm crit}$ the downward collision rate is dwarfed by radiative transitions, and rate of collisional transitions upward equals rate of radiative transitions downward:

$$\frac{n_{\rm U}}{n_{\rm L}} = \left(\frac{n}{n_{\rm crit}}\right)\left(\frac{n_{\rm U}}{n_{\rm L}}\right)_{\rm Boltzmann}. \tag{157}$$

The emissivity is independent of $A_{\rm UL}$ but depends on the collision rate:

$$\mathcal{E}_\nu = \frac{h\nu}{4\pi} n n_{\rm L} \langle\sigma v\rangle_{\rm LU} \phi_\nu = \frac{h\nu}{4\pi} n n_{\rm L} \frac{g_{\rm U}}{g_{\rm L}} \langle\sigma v\rangle_{\rm UL}\, e^{-x} \phi_\nu, \tag{158}$$

proportional to density squared.

The most important application of this analysis is to *forbidden lines*. These are lines corresponding to transitions that only occur at higher order than the dipole level, and hence are very slow. Some of the most important examples are the 3727Å line of once-ionized Oxygen (OII), or the 4959/5007Å doublet of twice-ionized Oxygen (OIII). These lines are very strong in clouds of interstellar photoionized plasma, but for some while their interpretation was confused and they were attributed to a new element: Nebulium. This was a sensible enough conclusion: Helium had first been found in this way, in the spectrum of the Sun. But in 1927 Ira Bowen gave the true interpretation in terms of Oxygen. The problem is that the transitions of interest are very slow, with a natural lifetime of order seconds. This can be contrasted with $\sim 10^{-9}$s for the $n = 2 \to n = 1$ transition in Hydrogen that generates Lyman α. Because of this slow spontaneous decay rate, the critical density for these Oxygen lines is very low (very roughly $10^{11}{\rm m}^{-3}$ – lower than any laboratory vacuum), and so they are suppressed under normal conditions. More strictly, they are suppressed *relative* to the $\mathcal{E} \propto n^2$ scaling that holds for collisional excitation below the critical density: above that density, the intensity still increases, but more slowly than for other lines, so the forbidden lines are dwarfed by other lines.

II. Astronomical Applications

6 The Lyman alpha transition 2p – 1s in hydrogen

We begin the set of selective astronomical applications of the quantum apparatus of atomic transitions by discussing the Lyman alpha line: this is the transition from the $n = 2, \ell = 1$ state to the ground $(n = 1, \ell = 0)$ state of atomic hydrogen, producing the Lyman α line at $\hbar\omega = (3/8)\alpha^2 mc^2$, a wavelength of 1216Å. As n changes, we look for the electric dipole moment. The selection rule requires ℓ to change by unity, and since the ground state has $\ell = 0$ the upper state must be the $\ell = 1$ state.

The upper state is $|Xm\rangle = R_{2,1}(r)Y_{1,m}(\theta,\phi)$ where $R_{2,1}$ is the hydrogen radial wave function for $n = 2,\ \ell = 1$ and $Y_{1,m}$ is the spherical harmonic corresponding to $\ell = 1$ and $m = 1, 0, -1$, which latter is the component of angular momentum (divided by $\hbar$) in the z direction. We do not know which value of m our initial state has, but it will turn out that it doesn't matter (as it shouldn't – the transition doesn't depend on which way is up). The lower state is $|Y\rangle = R_{1,0}(r)Y_{0,0}(\theta,\phi)$.

Matrix elements are

$$\langle Y|\mathrm{op}|X\rangle = \int R^*_{\mathrm{Y}}(r)Y^*_{\mathrm{Y}}(\theta,\phi)\ \mathrm{op}\ R_{\mathrm{X}}(r)Y_{\mathrm{X}}(\theta,\phi)\ r^2\,dr\,d\Omega \tag{159}$$

The radial and angular parts of the hydrogen wave functions are:

$$R_{1,0}(r) = \left(2/a^{3/2}\right)e^{-r/a}, \qquad R_{2,1}(r) = \left(r/\sqrt{24a^5}\right)e^{-r/2a} \qquad a = \alpha^{-1}(\hbar/mc) \tag{160}$$

$$Y_{0,0} = \sqrt{\frac{1}{4\pi}} \qquad Y_{1,\pm1} = \mp\sqrt{\frac{3}{8\pi}}\sin\theta e^{\pm i\phi} \qquad Y_{1,0} = \sqrt{\frac{3}{4\pi}}\cos\theta \tag{161}$$

Remember that for the spherical harmonics Y

$$\int Y^*_{\ell',m'}Y_{\ell,m}\,d\Omega = \delta_{\ell',\ell}\,\delta_{m',m}. \tag{162}$$

The Cartesian components of $\mathbf{r}$ can be written in polar coordinates as

$$x = r\ \sin\theta\cos\phi = r\sqrt{\frac{2\pi}{3}}\,(Y^*_{1,-1} - Y^*_{1,1}) \tag{163}$$

$$y = r\ \sin\theta\sin\phi = r\sqrt{\frac{2\pi}{3}}\,(Y^*_{1,-1} + Y^*_{1,1}) \tag{164}$$

$$z = r\ \cos\theta = r\sqrt{\frac{4\pi}{3}}\,Y^*_{1,0} \tag{165}$$

The second equalities in these expressions (obtained by inspection) will be used to calculate matrix elements, using the orthonormality of the spherical harmonics. Since $Y^*_{\mathrm{Y}} = Y^*_{0,0}$ is constant, the angle integrals simplify to different applications of the orthonormality conditions on the Y's. We want $|\langle Y|\mathbf{r}|X\rangle|^2 = |\langle Y|x|X\rangle|^2 + |\langle Y|y|X\rangle|^2 + |\langle Y|z|X\rangle|^2$.

For example,

$$\langle Y|x|Xm\rangle = \int r^3 R_{1,0}(r)R_{2,1}(r)\,dr \int Y^*_{0,0}(\theta,\phi)\frac{x}{r}Y_{1,m}(\theta,\phi)\,d\Omega = F\sqrt{\frac{1}{6}}(\delta_{m,-1} - \delta_{m,1}) \tag{166}$$

where F is the radial integral which is straightforward[1] to do and equals $(2^{15/2}/3^{9/2})a$. Also,

$$\langle Y|y|Xm\rangle = F\sqrt{\frac{1}{6}}(\delta_{m,-1}+\delta_{m,1}), \qquad \langle Y|z|Xm\rangle = F\sqrt{\frac{1}{3}}\delta_{m,0}. \tag{168}$$

It follows that $r_{\rm YX}^2 = |\langle Y|\mathbf{r}|Xm\rangle|^2 = F^2/3 = (2^{15}/3^{10})a^2 = 0.555a^2$ whatever the value of m.

Now we can write down the spontaneous transition rate (Einstein A value) which is

$$\Gamma_{\rm spon}(2p\to 1s) = \frac{4\omega^3}{3\hbar c^3}\frac{e^2}{4\pi\epsilon_0}r_{\rm YX}^2 = \frac{2^{17}}{3^{11}}\,\frac{\alpha\omega^3a^2}{c^2} = (2/3)^8\alpha^5\frac{mc^2}{\hbar} = 6.3\times 10^8\ \ {\rm s}^{-1} \tag{169}$$

(using $a=\hbar/\alpha mc$ and $\hbar\omega = (3/8)\alpha^2mc^2$).

6.1 Absorption cross-section and equivalent width

The spontaneous transition rate allows us to work out the cross-section for Lyman-alpha absorption by neutral hydrogen. We have seen that this is $\sigma = g_{\rm U}\Gamma_{\rm spon}\lambda^2\phi_\omega/4$, so with $g_{\rm U} = (2\ell+1)_{\rm U} = 3$ and $\lambda = 2\pi c/\omega$ and $\hbar\omega$ as above, this can be written as

$$\sigma = \frac{2^{13}}{3^9}\,2\pi^2\,\alpha\,\frac{\hbar}{m}\,\phi_\omega. \tag{170}$$

This is in the form of the dimensionless factor $f = 2^{13}/3^9 = 0.42$ times the classical expression for a damped oscillator. Thus f is known as the *oscillator strength*. Every atomic transition can be assigned such a dimensionless measure of its rate, which tend to be of order unity for permitted transitions.

For the classical oscillator, we saw that the frequency profile, ϕ_ω, is of Lorentzian form. The width of the Lorentzian is set by the uncertainty principle as $\Delta\omega\sim\Gamma_{\rm spon}$. This is an alternative argument to the one given above in terms of Parseval's theorem, saying that the *natural line profile* of a transition is Lorentzian. The core of the line will be Gaussian, via convolution by Doppler broadening, but the *damping wings* of the natural profile dominate far from the line centre.

The optical depth is $\tau = n\sigma\ell = \sigma N$, where N is the *column density* of neutral atoms (number of atoms in projection per unit area). Adopting the Lorentzian profile, we have

$$\phi_\omega = \frac{\Gamma/2\pi}{(\omega-\omega_0)^2+(\Gamma/2)^2}, \tag{171}$$

so that ϕ_ω at the line centre is $2/\pi\Gamma$. Thus we see that the optical depth at the line centre is independent of Γ:

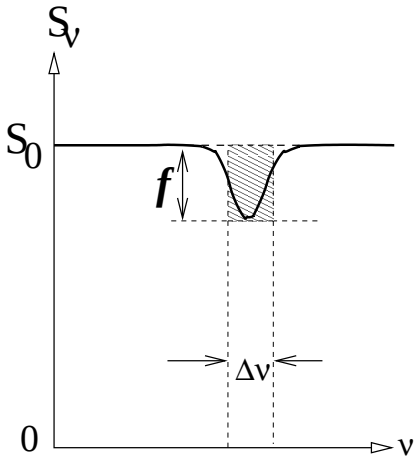
$$\tau(0) = N\sigma(0) = Ng_{\rm U}\Gamma\lambda^2\phi_\omega(0)/4 = N\,\frac{g_{\rm U}}{2\pi}\,\lambda^2. \tag{172}$$

For Lyman alpha, this is unity at a critical column density of $10^{14.15}\,{\rm m}^2$. But in practice there is always some Doppler broadening, so that $\phi_\omega(0) = (2\pi)^{-1/2}(\omega_0\,\sigma_v/c)^{-1}$. In this case,

$$\tau(0) = \left(N\ /\ 10^{16.27}{\rm m}^{-2}\right)\left(\sigma_v/\,{\rm km\,s}^{-1}\right)^{-1}. \tag{173}$$

[1]
$$\int_0^\infty r^n e^{-\alpha r}\,dr = \left(-\frac{\partial}{\partial\alpha}\right)^n\int_0^\infty e^{-\alpha r}\,dr = \left(-\frac{d}{d\alpha}\right)^n\left(\frac{1}{\alpha}\right) = \frac{n!}{\alpha^{n+1}} \tag{167}$$

Measuring column density from the area of lines: For noisy line profiles, measuring the maximum absorption exactly will be hard; nor may we know the line width very well. But a robust estimate of the absorbing column can be obtained from the total amount of spectrum eroded by absorption, as illustrated below:



If the flux density in the continuum on either side of the line is S_0, the line spectrum is $S_0\ e^{-\tau_\nu}$. The area between the spectrum and the extrapolated continuum (divided by the continuum value, and therefore measured in Hz) is the *equivalent width* $\mathcal{W}$:

$$\mathcal{W} = \int 1 - e^{-\tau_\nu}\ d\nu \tag{174}$$

Reference to the diagram shows the area in question as a shaded block with the same depth as the line, and therefore the same effective width. If the largest (central) value of τ_ν is much less than unity, then

$$\mathcal{W} \simeq \int \tau_\nu\ d\nu = N g_{\rm U} \frac{\lambda^2}{8\pi} \Gamma_{\rm spon} \tag{175}$$

where N is the column density of neutral hydtrogen in the ground state.

If the optical depth in the line centre is large, then the equivalent width is no longer linearly proportional to the column density. The line will become flat-bottomed and the equivalent width measures the range of frequencies over which $\tau > 1$. More precisely, for a Gaussian line profile, the flux is proportional to $1-\exp[-\tau_0 \exp(-x^2/2)]$, where τ_0 is the optical depth at the line centre (which *is* proportional to the column) and x is frequency offset in units of the Gaussian line-width, $\sigma_{\rm G}$. If in practice we measure the FWHM, this will be $2x$, where x is deduced by setting the flux equal to 0.5. In the limit that $\tau_0 \gg 1$, This gives

$$\mathrm{FWHM} = 2\sigma_{\rm G}\sqrt{2\ln(\tau_0/\ln 2)}, \tag{176}$$

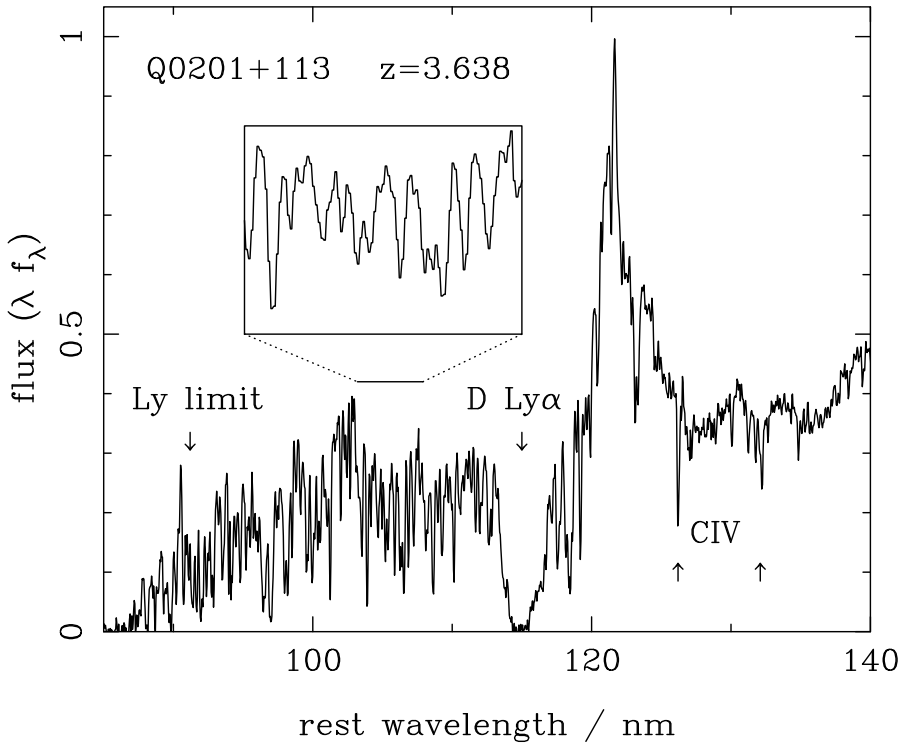
so that the line width grows logarithmically with the column. But at large frequency offsets, the line profile will be dominated by the damping wings. If we replace the Gaussian form, $\exp(-x^2/2)$, by a Lorentzian, $1/(1+x^2/2)$, the corresponding formula is

$$\mathrm{FWHM} = 2\sigma_{\rm L}\sqrt{2\tau_0/\ln 2}, \tag{177}$$

and now the line width grows much more rapidly with column. This yields characteristic *damped absorption lines* where systems of high column density can be recognised via their unusual widths.

6.2 The Lyman-alpha forest in quasars

The Lyman-alpha line, in the ultraviolet locally, is shifted into visibility at high redshift. The large value of transition rate means it is a very strong line in absorption, and hydrogen clouds in the line of sight to a distant ($z > 1.7$) quasar will show themselves as Lyman α lines at their own local redshift. An example is shown below.



What we see is a 'forest' of narrow absorption lines that eat away the ultraviolet continuum of the quasar. The fact that these lines are found uniformly at all redshifts shows that they represent intervening material, rather than material ejected from the quasar. The observed velocity widths of the Lyman-α lines are a few tens of $\mathrm{km\,s^{-1}}$, so the weakest known lines are sensitive to absorption by column densities of around $10^{16.5}\mathrm{m}^{-2}$. There is a wide distribution of column densities, going up to $\sim 10^{25}\mathrm{m}^{-2}$. For $N \gtrsim 10^{22}\mathrm{m}^{-2}$, the optical depth becomes unity at a point where the profile is dominated by the Lorentzian wings of the 'damped' profile rather than the Gaussian core. So we have characteristic damped absorption systems, such as the one seen in the example spectrum. These systems have columns similar to that in the disc of a galaxy like the Milky Way, so the remaining weaker lines represent very low-mass 'clouds' of gas in the intergalactic medium.

The Gunn–Peterson effect The absorption in the Lyα clouds clearly eats away the quasar continuum to the blue of the Lyα emission. This provides us with a very powerful way to estimate the general abundance of neutral hydrogen in the universe (Gunn & Peterson 1965; see also Scheuer 1965). Suppose we study radiation at some frequency that lies bluewards of the observed Lyman alpha in the quasar; this was emitted in the UV, but at some intermediate point between the quasar and us, it has been redshifted sufficiently to reach the rest wavelength of Lyman alpha – and can then be absorbed by the neutral hydrogen at that point in space.

It's easy to estimate the overall optical depth of such absorption by returning to our basic expression for the optical depth, including the definition of column density: $\tau = \sigma \int n \, d\ell$. Here, we need to recognise that the frequency changes along the path because of the cos-

mological redshift, and that the cross-section is frequency dependent. We deal with this by using Hubble's law for a change in velocity corresponding to $d\ell$: $dv = H\,d\ell$. The Doppler effect of this velocity is $d\nu/\nu = dv/c$, so $d\ell = (c/H)\,d\nu/\nu$. Since σ is proportional to ϕ_ν, the integration over frequency removes this function, leaving

$$\tau = n\,\frac{c}{H}\,g_{\rm U}\,\frac{\lambda^2}{8\pi\nu}\,\Gamma_{\rm spon}. \tag{178}$$

How large is this prediction? The mean cosmic density of baryonic material is about $10^{-27.38}$ SI, of which 75% is hydrogen, so that $n = 0.19\,{\rm m}^{-3}$ (from primordial nucleosynthesis arguments, but one can get the order of magnitude by counting stars and gas in galaxies). H is about $70\,{\rm km\,s^{-1}Mpc^{-1}}$ today, so this gives $\tau = 10^{4.1}$ as a prediction. This is actually an underestimate, as n will be larger at high redshifts by a factor $(1+z)^3$, so for the illustrated quasar at $z = 3.638$, we now have $\tau = 10^{6.1}$. In fact, H will be a little higher at these early times, but we still predict τ well above 10^5. However, just looking at the spectrum, we can see that the majority of the UV flux makes it through, so our estimate is too high by a factor 100,000 or so.

The resolution of this disagreement must be that the gas is highly ionized: $x \simeq 10^{-5}$, where x is the fractional ionization. The reason for this is photoionization: the very UV continuum from quasars that shows us the presence of absorbing hydrogen will tend to destroy it, as does UV radiation from galaxies (which probably dominates the ionizing budget). Neutral hydrogen only exists via the inverse process of recombination, where electrons and protons combine to make new Hydrogen atoms. This is more effective in high-density regions, so the Lyman-alpha clouds are dense clumps that are less ionized than the average.

The absorption of the UV continuum rises with increasing redshift, but at high z it rises more abruptly than expected purely from the $(1+z)^3$ scaling in density. At just beyond $z = 6$, the UV flux disappears altogether (Becker et al. 2001), indicating that we are coming near to the *reionization era*. A frontier area in cosmology is the study of this transition from the almost neutral post-recombination universe at $10 \lesssim z \lesssim 1000$ to one that becomes ionized by the first astronomical structures.

7 Photoionization and recombination

Lyman alpha is an example of a *recombination line*. This means that it arises naturally in highly ionized gas, where a free electron and a proton come together to make an excited Hydrogen atom, which can then spontaneously radiate emission lines to reach a lower level, and eventually the ground state. This process is known as the *recombination cascade*, and it can generate all the Hydrogen emission-line series (Lyman, Balmer, Paschen etc.). For highly ionized material, this process is more effective than collisional excitation.

7.1 Photoionization transition rate

It is interesting to start with the closely related process of photoionization, in which an energetic photon can remove an electron from an atom, causing a *bound–free transition*. We will work this out in the simplest case of ionization from the ground state, where the initial wave function is

$$\psi_{\rm X} = (\pi a^3)^{-1/2}\, e^{-r/a}; \qquad a = \alpha^{-1}(\hbar/mc) \qquad E_{\rm X} = -\chi = -\alpha^2 mc^2/2. \tag{179}$$

The final state is a free electron, so it is tempting to write

$$\psi_{\rm Y} = V^{-1/2} e^{-i\mathbf{K}\cdot\mathbf{x}}, \tag{180}$$

where the minus sign on the free electron wave vector is chosen for later convenience. This is only an approximation, however, since the electron is not free near to the proton – which is where we need to know the wave function. But the plane-wave form should be adequate for a highly energetic electron, whose trajectory will hardly be affected by the electrostatic field of the proton. We will stick with this simple case, which should give the correct result for ionizing radiation with photon energy well above the ionization potential of Hydrogen ($\chi = 13.6$ eV, or a wavelength of 912Å).

The energy of the liberated electron is the difference between photon energy and ground-state binding energy: $E = \hbar\omega - \chi$, but here we are assuming $\hbar\omega \gg \chi$ – so that $E \simeq \hbar\omega$. In order to keep within non-relativistic QM, we need $E \ll mc^2$. The analysis here will therefore be valid for the range $\alpha^2 \ll \hbar\omega/mc^2 \ll 1$.

The matrix element for the bound–free transition (without making the dipole approximation) is then

$$M_{\rm XY} = (\pi a^3 V)^{-1/2}\, \mathbf{e} \cdot \int \exp(-r/a)\ \exp(i\mathbf{k}\cdot\mathbf{x})\, \mathbf{p}\ \exp(-i\mathbf{K}\cdot\mathbf{x})\, dV. \tag{181}$$

Note that we have chosen to consider $M_{\rm XY}$ rather than $M_{\rm YX}$; the order doesn't change the modulus of the matrix element, because the momentum operator is Hermitian. Since $\mathbf{p} = -i\hbar\boldsymbol{\nabla}$, the effect of this operator is to bring down $-\mathbf{K}$ from the exponential, so that

$$M_{\rm XY} = -\hbar(\pi a^3 V)^{-1/2}\, (\mathbf{e}\cdot\mathbf{K}) \int \exp(-r/a) \exp[i(\mathbf{k}-\mathbf{K})\cdot\mathbf{x}]\, dV. \tag{182}$$

The integral can be written as $2\pi \int r^2 dr \exp(-r/a) \int \exp(iqr\mu)\, d\mu$, where q is short for $|\mathbf{k}-\mathbf{K}|$ and μ is $\cos\theta$ in terms of the normal polar angle. It is then easy to do the μ integral and show that the overall integral comes out to be $8\pi a^3/(1+a^2q^2)^2$.

We can simplify this expression in two ways. First there is the dependence on the orientation of the electron wavenumber relative to the polarization vector, $\mathbf{e}\cdot\mathbf{K}$. As before, if we want to sum over unpolarized radiation, we can use the trick that $(\mathbf{e_1}\cdot\mathbf{K})^2 + (\mathbf{e_2}\cdot\mathbf{K})^2 = K^2 - (\hat{\mathbf{k}}\cdot\mathbf{K})^2 = K^2\sin^2\theta$. Thus when we sum rates over the two independent polarizations

and integrate over solid angle, we just get a factor $(8\pi/3)K^2$. The other simplification is $1 + a^2|\mathbf{k} - \mathbf{K}|^2 \simeq a^2K^2$ – i.e. the electron wavenumber K dominates over the photon wavenumber k. This can be seen in two steps that follow from the assumption that the ionizing photons are of very high energy. Thus the kinetic energy of the ejected electron is very nearly the photon energy: $(\hbar K)^2/2m \simeq \hbar\omega = \hbar kc$. Hence $K^2/k^2 = 2mc/\hbar k$ – but this must be large if the ejected electron remains nonrelativistic. Similarly, $aK \simeq \alpha^{-1}(\hbar/mc)(2m\omega/\hbar)^{1/2} = \alpha^{-1}(2\hbar\omega/mc^2)^{1/2}$; we have seen that this combination must be $\gg 1$ for the energetic photons where this analysis applies.

Putting these two facts together, we can now get the ionization transition rate. Recall the basic expression:

$$\Gamma = \frac{2\pi}{\hbar^2}\,\delta(\omega - \omega_0)\,\frac{\hbar}{2\epsilon_0\omega V}\,\frac{e^2}{m^2}\,N\,|M_{\rm XY}|^2. \tag{183}$$

As usual, we have to integrate over a density of photon states, $\rho_\omega = [V/(2\pi)^3](\omega^2/c^3)\,d\Omega$, to get the transition rate to a single electron state at energy $E = \hbar\omega_0$ above the ground state. Since

$$\int |M_{\rm XY}|^2 d\Omega = \hbar^2(\pi a^3 V)^{-1}\,(8\pi K^2/3)\,(8\pi a^3)^2\,/\,(a^2K^2)^4, \tag{184}$$

this boils down to

$$\Gamma = \frac{8\hbar^4 e^2 N}{3\epsilon_0 c^3 m^5 a^5 V \omega^2}. \tag{185}$$

7.2 Photoionization cross-section

The expression for the transition rate is not very illuminating. A more practical way to express it is in terms of a cross section $\sigma_{\rm ion}$. So far, Γ is just the rate at which radiation excites a single free electron state. A more interesting question is the rate at which there is excitation into a band of electron momenta, in response to photons in some range of photon momenta. We would then write that the total rate of transitions is equal to the cross-section times the photon flux density (number density times c):

$$\Gamma\,\frac{g_e V\,d^3p_e}{(2\pi\hbar)^3} = \sigma_{\rm ion}\,\frac{g_\gamma V\,d^3p_\gamma}{(2\pi\hbar)^3}\,\frac{Nc}{V}. \tag{186}$$

The momenta are related in magnitude by $p_e^2/2m = p_\gamma c = \hbar\omega$. From this, and using $d^3p = 4\pi p^2\,dp$, the final expression for the cross-section follows:

$$\sigma_{\rm ion} = \frac{2^8}{3}\,\alpha\,(\pi a^2)\,\left(\frac{\chi}{\hbar\omega}\right)^{7/2}, \tag{187}$$

where α is the fine-structure constant, and $\chi = \alpha^2 mc^2/2$ is the ionization energy of the ground state. Our expression is only valid for photon energies $\gg \chi$, but the basic picture remains in a more detailed treatment: the cross section is of order the geometric size of the atom, πa^2 (which it must be, by dimensions), for photons that barely ionize the atom. But for much more energetic photons, the cross section is heavily suppressed. Thus, UV photons are very efficiently absorbed by neutral hydrogen, but gamma rays can pass relatively unscathed through large amounts of gas.

7.3 Recombination cross-section

The analysis for the inverse process of recombination proceeds in very much the same way. We just swap the initial and final states. As seen before in the case of single discrete transitions, the rate of stimulated emission and absorption are the same (because the squared

matrix elements $|M_{\rm XY}|^2$ and $|M_{\rm YX}|^2$ are identical). All that is different is in the definition of the cross-section, since now we are interested in the Hydrogen nucleus intercepting a flux of electrons with velocity $v = p_e/m$, rather than a neutral Hydrogen atom intercepting a flux of photons with velocity c.

If we let ρ denote the number of states, $\rho = V/(2\pi\hbar)^3 d^3p$ in terms of momentum, we can then write down the following two formulae for cross-sections of bound-free transitions:

(a) Ionization: $g_{\rm ion}\Gamma_s N_\gamma \rho_e = \sigma_{\rm ion}\rho_\gamma N_\gamma c/V$.
(b) Recombination: $g_{\rm atom}\Gamma_s N_e \rho_e = \sigma_{\rm rec} N_e \rho_e (p_e/m)/V$.

In the first of these, we have a stimulated transition (hence ΓN) and we add the rates to all final states (hence ρ_e). In the second, we have spontaneous decay of a free electron, but for the total rate we need to add the effects of all possible free electrons (hence $N_e\rho_e$). These rates also get a final-state g factor. These factors don't need to include spin corrections since spin doesn't appear in the Hamiltonian, and so it doesn't change in these transitions. Thus both g factors would be unity for recombination involving the hydrogen ground state, for example.

These expressions must now be manipulated to relate the two cross-sections. First note that the density of states is $\rho \propto p^2\, dp\, d\Omega$, so we need to relate dp_e and dp_γ: this is done using the energy relation $p_e^2/2m = p_\gamma c$. Finally, we need to integrate over directions, bearing in mind that the spontaneous rate Γ_s already implicitly contains a factor $d\Omega_\gamma$. When this is done, we are left with a rather simple conversion between the cross-sections for photoionization and recombination, known as the **Milne relation**:

$$\sigma_{\rm rec} = \left(\frac{g_{\rm atom}}{g_{\rm ion}}\right)\left(\frac{p_\gamma}{p_e}\right)^2 \sigma_{\rm ion}. \tag{188}$$

In our high-energy approximation, $p_e^2 = 2mcp_\gamma = 2m\hbar\omega$, so that $p_\gamma/p_e \propto p_e$. Previously, we had $\sigma_{\rm ion} \propto \omega^{-7/2}$, so now $\sigma_{\rm rec} \propto p_e^{-5}$. This looks like there will be a divergence for slowly-moving electrons, but our treatment breaks down there. At the level of these lectures, it is therefore hard to go beyond order-of-magnitude estimates. Typically, the spectra of ionizing sources fall quite rapidly, so that those photons that do exist above the threshold energy of 13.6 eV will mainly only be slightly more energetic than the threshold – giving a residual kinetic energy of a few eV to the liberated electron. The corresponding thermal energy is

$$kT\,/\,{\rm eV} = T\,/\,1.16\times10^4\,{\rm K}, \tag{189}$$

so photoionized plasmas naturally have a temperature around 10^4K. But if we ignore this distinction between the total and the residual photon energy, then $p_e^2/2m = p_\gamma c$, so $(p_\gamma/p_e)^2 = E_\gamma/2mc^2$. Since the ionization energy is $\chi = \alpha^2 mc^2/2$, the conversion factor is of order $(p_\gamma/p_e)^2 \sim \alpha^2$. We therefore have the following orders of magnitude for the cross-sections:

$$\sigma_{\rm ion} \sim \alpha\pi a^2; \qquad \sigma_{\rm rec} \sim \alpha^3\pi a^2. \tag{190}$$

In detail, because the recombination cross-section depends on electron momentum, it is necessary to integrate over a Maxwellian distribution to find the mean rate. Also, what is interesting is not so much the cross-section, but the recombination rate coefficient: $\mathcal{R} = \langle\sigma_{\rm rec} v\rangle$. In these terms, the total rate of recombinations per unit volume is

$$\text{rate} = \mathcal{R} n_e n_p, \tag{191}$$

in terms of the electron and proton densities. Since these are equal for ionized hydrogen, the process proceeds as the square of the density. Carrying out the Maxwellian integration over a more careful computation of the cross-section, the following useful approximation results:

$$\mathcal{R}/({\rm m}^3\,{\rm s}^{-1}) \simeq 1.4\times10^{-19}(kT/1\,{\rm eV})^{-1/2}. \tag{192}$$

7.3.1 Photoionization and HII regions

Consider a hot massive star with temperature $T_* = 30\,000$ K. The average energy of a photon from the star is

$$\langle \epsilon_* \rangle \simeq kT_* = 4 \times 10^{-19}\ \mathrm{J} = 0.2\chi, \tag{193}$$

so apparently even such an extreme star cannot ionise hydrogen. But the spectrum of a star is a black body, which has a long high energy tail. A hot O star with $L \simeq 20\,L_\odot$ will typically produce ionizing photons at a rate $S_* \simeq 10^{49}\ \mathrm{s}^{-1}$. As these are efficiently absorbed by neutral gas, the star will grow a bubble of fully ionized gas around it, with a rather sharp edge between the ionized gas and the surrounding neutral medium. In a time t, $N_\gamma = S_* t$ photons will be produced. These will be all used up in ionizing the surrounding gas. If n_H is the number density of hydrogen atoms (number of H atoms per unit volume), then the number of ionized atoms equals N_γ:

$$N_\gamma = S_* t = \frac{4\pi R^3}{3}\, n_\mathrm{H} \quad \Rightarrow \quad R = \left(\frac{3}{4\pi}\, \frac{S_* t}{n_\mathrm{H}} \right)^{1/3}. \tag{194}$$

So as time passes, an ever-increasing ionized sphere will surround the star. This is called an HII region.

But this growth will not continue indefinitely. Within the bubble, new neutral Hydrogen atoms will be produced via recombination, and some of the ionizing photons will be absorbed by these before they reach the edge of the bubble. Eventually, the HII region attains an equilibrium maximum size where the injection rate of ionizing photons can only just keep pace with the recombination rate. If the electron density in the HII region is n_e, the total balanced event rate is

$$S_* = \text{recombination rate} = \frac{4\pi}{3}\, R_\mathrm{S}^3 n_e\, \mathcal{R} n_p, \tag{195}$$

where R_S is called the *Strömgren radius*.

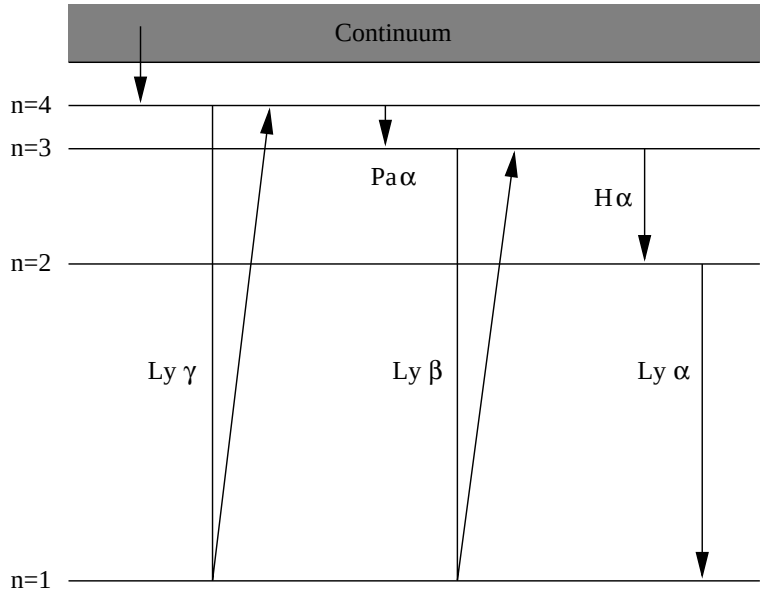
For a pure hydrogen nebula, $n_e = n_p$ by ionization balance, and this is $\simeq n_\mathrm{H}$, the original total H-atom density, since the gas is highly ionized. So

$$R_\mathrm{S} \equiv \left(\frac{3}{4\pi}\, \frac{S_*}{\mathcal{R} n_\mathrm{H}^2} \right)^{1/3}. \tag{196}$$

Example: Consider the O star again with $S_* = 10^{49}\ \mathrm{s}^{-1}$ in a cloud of density $n_\mathrm{H} = 10^7\ \mathrm{m}^{-3}$. Then $R_\mathrm{S} \simeq 4.3 \times 10^{17}\,\mathrm{m} = 14$ pc.

7.4 Recombination, resonant scattering, and cosmology

Our treatment has concentrated on transitions between a free electron and the Hydrogen ground state, but recombination can take place to all levels. Once captured, the electron can make its way down by spontaneous transitions to levels of lower energy, resulting in a complex cascade of transitions. Eventually, we want this process to terminate at the $n = 1$ ground state, but getting there is not so easy. Recombination direct to the ground state gets us nowhere: it generates a photon that is energetic enough to ionize an atom, which will quickly find an atom and ionize it, leaving us back where we started. The same applies to any photon in the Lyman series, which will be immediately re-absorbed by a nearby H atom. he result is that Lyα photons undergo *resonant scattering* within nebular gas.



Things may often be simplified somewhat by the use of the *case B approximation* (a prime example of impenetrable astronomical jargon). This refers to a situation where the density is sufficiently low that all lines are optically thin except the Lyman series. Thus the excited post-recombination atom radiates in various series, eventually reaching the $n = 2$ level (which must involve a Balmer-series photon). Thereafter a Lyα photon will be created, which can eventually diffuse out of the emitting region, leading to the case B rule: every hydrogen recombination yields one Lyα photon plus one Balmer photon. However, in a typical HII region, the odds are high that on its effectively long journey through the HII region, a Lyα photon will eventually hit a dust grain and be absorbed by it. As a consequence, the Lyα emissivity from HII regions can be well below the case-B prediction.

Some of these issues also arise in cosmology, where initially all matter is in the form of an extremely hot plasma. But as the universe expands and cools, the temperature drops to the point ($T \sim 1000$ K) where it is thermodynamically favourable for the ionized plasma to form neutral atoms. This process is known as cosmological recombination: a complete misnomer, as the plasma has always been completely ionized up to this time. Diffusive scattering is a particular issue in the early universe, since the region of ionized gas is infinite, so Lyα photons can never escape.

This means that the route by which the Universe becomes neutral is significantly different from what would be expected if we simply assumed thermal equilibrium. We first derive the *Saha equation*, which describes this situation. Consider a single hydrogen atom, for which we want to know the relative probabilities that the electron is bound or free. In equilibrium, with the probability of a given state being proportional to the Boltzmann factor, this must be given by the ratio of partition functions:

$$\frac{P_{\rm free}}{P_{\rm bound}} = \frac{\sum_{\rm free} g_i e^{-E_i/kT}}{\sum_{\rm bound} g_i e^{-E_i/kT}} = \frac{Z_{\rm free}}{Z_{\rm bound}}. \tag{197}$$

In the case of this *single* object (the electron) the degeneracy factor $g = 2$, from spin. Now, a good approximation is that $Z_{\rm bound} = g \exp(\chi/kT)$ where χ is the ionization potential of hydrogen ($\chi > 0$, but the energy of the electron is $-\chi$). This will be valid at low temperatures $kT \lesssim \chi$, where the contribution of excited states to Z is negligible. Now, in a box of volume V,

$$Z_{\rm free} = g \frac{V}{(2\pi)^3} \int e^{-p^2/2mkT} \frac{d^3p}{\hbar^3} = (2\pi mkT)^{3/2} \frac{gV}{(2\pi\hbar)^3}. \tag{198}$$

To use this expression, we have to decide what V is; do we set $V \to \infty$ in an infinite universe, and so conclude that the atom is always ionized, whatever T may be? Clearly not: the

volume 'available' to each electron is effectively $1/n_e$ – the reciprocal of the electron number density. Using this, and defining the fractional ionization x ($n_e = xn$, where n is the total nucleon number density, $n_{\rm H} + n_p$), we get the Saha equation

$$\frac{x^2}{1-x} = \frac{(2\pi m_e kT)^{3/2}}{n(2\pi\hbar)^3} e^{-\chi/kT}. \tag{199}$$

The Saha result is quite interesting: x is not simply a function of temperature and does depend on the density also. However, the $\exp(-\chi/kT)$ term ensures that x in practice always diverges from 1 when $kT \sim \chi$.

The problem with the Saha approach is that, although it describes the initial phase of the departure from complete ionization, the assumption of equilibrium rapidly ceases to be valid. This is because of the problem that the ground state can never be reached, since Lyman α can never escape. There is a way out, however, using *two-photon emission*. Quantum-mechanically, this arises from the $e^2A^2/2m$ term neglected from the perturbed Hamiltonian $(\mathbf{p} - e\mathbf{A})^2/2m$. Since A is an operator containing a and $a^\dagger$, A^2 can create (or destroy) two photons. These sum to the energy of Lyman α, but the energy of the more energetic of these can vary continuously between half this figure and just less than the full Lyman α energy. This gives the mechanism we need for transferring the ionization energy into photons with $\lambda > \lambda_{{\rm Ly}\alpha}$. Being a process of second order in perturbation theory, this is slow (lifetime $\simeq 0.1$ s); because recombination has to pass through this bottleneck, it actually proceeds at a rate completely different from the Saha prediction.

A highly stripped-down analysis of events simplifies the hydrogen atom to just two levels ($1S$ and $2S$). Any chain of recombinations that reaches the ground state can be ignored through the above argument: these reactions produce photons that are immediately re-absorbed elsewhere, so they have no effect on the ionization balance. The main chance of reaching the ground state comes through the recombinations that reach the $2S$ state, since some fraction of the atoms that reach that state will suffer two-photon decay before being re-excited. The rate equation for the fractional ionization is thus

$$\frac{d(nx)}{dt} = -\mathcal{R}\,(nx)^2\,\frac{\Gamma_{2\gamma}}{\Gamma_{2\gamma} + \Gamma_{\rm U}(T)}, \tag{200}$$

where n is the number density of protons, x is the fractional ionization, $\mathcal{R}$ is the recombination coefficient ($\mathcal{R} \simeq 3 \times 10^{-17} T^{-1/2}\,{\rm m^3 s^{-1}}$), $\Gamma_{2\gamma}$ is the two-photon decay rate, and $\Gamma_{\rm U}(T)$ is the stimulated transition rate upwards from the $2S$ state. This equation just says that recombinations are a two-body process, which create excited states that cascade down to the $2S$ level, from whence a competition between the upward and downward transition rates determines the fraction that make the downward transition.

An important point about the rate equation is that it is only necessary to solve it once, and the results can then be scaled immediately to some other cosmological model. Consider the rhs: both $\mathcal{R}$ and $\Gamma_{\rm U}(T)$ are functions of temperature, and thus of redshift only, so that any parameter dependence is carried just by n^2, which scales in proportion to the density of baryonic material, $\rho_{\rm B}$. Similarly, the lhs depends on the baryon density through n; the other parameter dependence comes from the relation between time and cosmic redshift. This is driven by the dimensional relation between the age of the universe and its total density (including dark matter), $t \sim (G\rho)^{-1/2}$. Putting these together, the fractional ionization must scale as

$$x(z) \propto \frac{\rho_{\rm tot}^{1/2}}{\rho_{\rm B}} \tag{201}$$

times a universal function of redshift. This is a very different scaling from the prediction of the Saha equation.

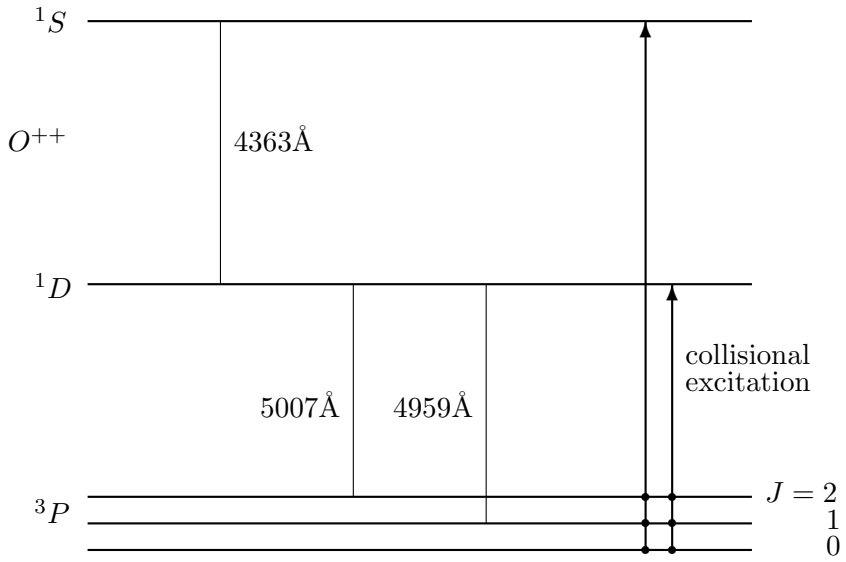
8 Fine structure excited by electron collisions: diagnostic lines

We now turn to showing how transitions in more complex ions can be used to give diagnostics of temperature and density in diffuse astrophysical plasmas. We will illustrate this using two common sets of transitions in ionized Oxygen. This is an abundant element, and also has the right degree of binding for its outer electrons, so that Oxygen exists in a partially ionized state at the sort of temperatures characteristic of many astrophysical situations (a few times 10^4 K).

The electronic structure of Oxygen is $1s^2 2s^2 2p^4$, so there are two closed inner shells (which we ignore). The interesting cases will be twice ionized OIII (two 2p electrons) and once ionized OII (three 2p electrons). As usual, the key factor in determining the energy levels in these multi-electron situations is to recall that the individual angular momenta are not conserved, so that we must produce total quantum numbers for angular momentum, and for total spin. This is LS coupling, which is a good approach until the spin-orbit coupling that causes the fine structure splittings becomes too large (as it does in larger atoms with $Z > 30$, where instead we have to use jj coupling, which works in terms of total angular momenta). Here, LS coupling applies, and the general theory of angular momenta (see the Appendix) tells us that the possible outcomes for OIII are $L = 2, 1, 0$ and $S = 1, 0$, whereas for OII we have $L = 3, 2, 1, 0$ and $S = 3/2, 1/2$. But not all combinations are allowed.

8.1 the O^{++} ground state: $2p^2$

For two electrons, the wavefunction can factorise into a function of angular momentum and one of spin. Since the overall wavefunction must be antisymmetric in exchange of electron labels, this means that the symmetries of these two parts must be opposite. By imagining the state of maximal z angular momentum, we can see that such a state must be symmetric (as all particles are in the same state). Thus for OIII, $L = 2$ pairs with $S = 0$, $L = 1$ with $S = 1$, and $L = 0$ with $S = 0$. There are thus three spectroscopic energy terms in the usual ${}^{2S+1}L$ format: ${}^3P$, ${}^1D$, ${}^1S$, and this is true for the isoelectronic sequence N^+, C, and for any other p^2 configuration.



The appendix recapitulates Hund's rules, which says that the highest spin gives the lowest energy, followed by a ranking in angular momentum. The levels with non-zero S are split

by spin-orbit interactions, as described in the appendix: $\mathbf{J} = \mathbf{L} + \mathbf{S}$, so that $\mathbf{L} \cdot \mathbf{S} = (|J|^2 - |L|^2 - |S|^2)/2$, and the splitting depends on the J quantum number.

Consider collisional excitation of the D term, and of the S term. The P term is treated as a single state. The critical densities (part I, section 5) are high enough that we use the low density limit (each collision leads to a photon emission).

$$\mathcal{E}_{\rm UL} = \frac{h\nu}{4\pi} n_{\rm L} n_e \langle \sigma v \rangle_{\rm LU} = \frac{h\nu}{4\pi} n_{\rm L} n_e (g_{\rm U}/g_{\rm L}) \langle \sigma v \rangle_{\rm UL} e^{-\Delta E/kT} \tag{202}$$

Applying this to ${}^3P \rightarrow {}^1D$ and ${}^3P \rightarrow {}^1S$ gives

$$\frac{\mathcal{E}_{4363\text{Å}}}{\mathcal{E}_{5007\text{Å}}} = \left(\frac{4}{3}\right) \frac{5007}{4363} \frac{\langle \sigma v \rangle_{(S,P)}}{\langle \sigma v \rangle_{(D,P)}} e^{-\Delta E_{4363}/kT} = 0.20\, e^{-32900/T}. \tag{203}$$

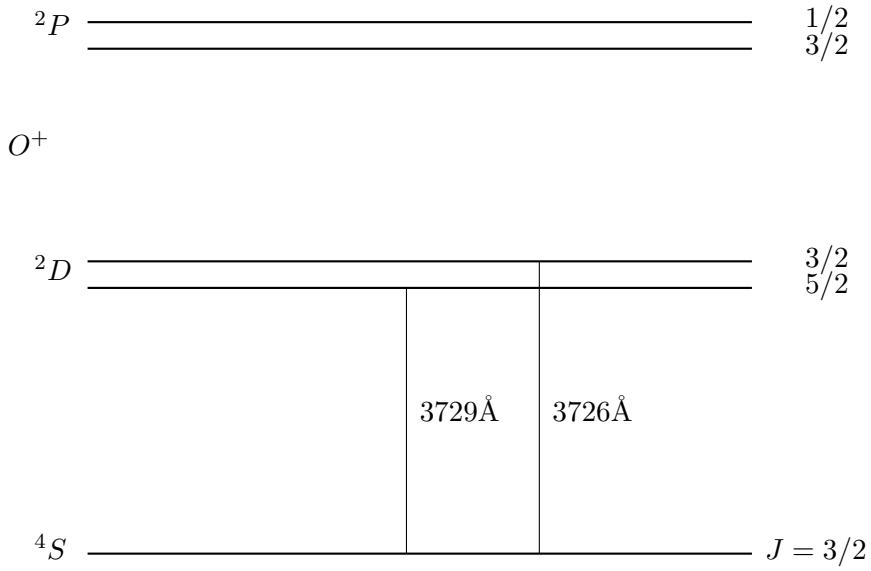
The factor (4/3) arises because approximately 3/4 of the transitions from the D level give the λ5007Å line. The final approximation uses the fact that the de-excitation rates are of the same order, as usual. The actual ratio (*Osterbrock, AGN*[2]) is $\langle \sigma v \rangle_{{}^1D \rightarrow {}^3P} / \langle \sigma v \rangle_{{}^1S \rightarrow {}^3P} = 7.75$.

Thus the intensity ratio of the 4363Å and 5007Å lines is highly temperature sensitive. Below 8000K the 4363Å line is hard to detect.

In very high density ($n_e \gg 10^{11}$ m^{-3}) HII regions, such as the very young 'ultra-compact HII regions', the critical density is exceeded, and a full level calculation is required.

8.2 the O^+ ground state: $2p^3$

(Also for N, S^+, etc)



The situation regarding the energy levels of this three-electron configuration is more complicated. Although addition of angular momentum indicates there should be F levels with $L = 3$, in fact these are impossible because the wavefunction no longer factorises. The Appendix discusses the situation in a little more detail, but we will simply accept the configuration shown in the above diagram.

Consider the equilibrium of the ${}^4S$ and ${}^2D$ terms. For each of the D levels separately

$$\frac{n_{\rm U}}{n_{\rm L}} = \frac{(g_{\rm U}/g_{\rm L}) e^{-(E_{\rm U} - E_{\rm L})/kT}}{1 + n_{\rm crit,e}/n_e} \tag{204}$$

and so the ratio gives

$$\frac{n_{\rm U}(3/2)}{n_{\rm U}(5/2)} = \frac{2}{3}\,\frac{1+n_{\rm crit,e(5/2)}/n_e}{1+n_{\rm crit,e(3/2)}/n_e}. \tag{205}$$

The latter equality has neglected the small difference in energy between the two upper levels (only 29K); and the ratio of statistical weights $(2J+1)$ of the upper levels is 2/3.

At high density, this number ratio is just set by the $2J+1$ factor, so that the ratio of the line intensities is

$$I_{3726}/I_{3729} = (2/3)\,\Gamma_{\rm spon}(3/2)/\Gamma_{\rm spon}(5/2) = 3.33, \tag{206}$$

Using the fact that the spontaneous rates can be calculated to be in a ratio of about 5:1.

Now, at low density, the intensity ratio is

$$I_{3726}/I_{3729} = (2/3)\,[n_{\rm crit,e(5/2)}/n_{\rm crit,e(3/2)}]\,[\Gamma_{\rm spon}(3/2)/\Gamma_{\rm spon}(5/2).] \tag{207}$$

The critical densities are

$$n_{\rm crit,e}(3/2) = (g_{\rm L}\Gamma_{\rm spon}/\langle\sigma v\rangle)_{^2D_{3/2}\rightarrow{}^4S} = 1.6\times10^{10}{\rm m}^{-3} \tag{208}$$

$$n_{\rm crit,e}(5/2) = (g_{\rm L}\Gamma_{\rm spon}/\langle\sigma v\rangle)_{^2D_{5/2}\rightarrow{}^4S} = 3.1\times10^{9}{\rm m}^{-3} \tag{209}$$

at 10^4 K. But the rate $\langle\sigma v\rangle_{^2D_{3/2}\rightarrow{}^4S}$ equals $\langle\sigma v\rangle_{^2D_{5/2}\rightarrow{}^4S}$, so the ratio of critical densities is just a ratio of spontaneous rates.

The ratio of the emissivity of the lines at 3726Å and 3729Å is $\Gamma_{\rm spon}(3/2)/\Gamma_{\rm spon}(5/2)$ times the number ratio above, and gives therefore at *low* density

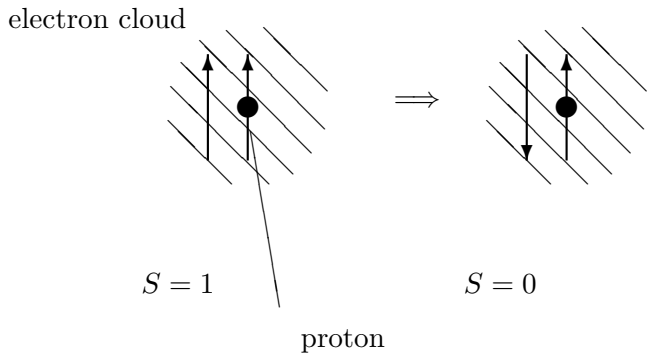
$$I_{3726}/I_{3729} = 0.67 \tag{210}$$

In retrospect this is obvious: the intensity is just proportional to the rate at which collisions populate the upper levels, which is the same apart from the $2J+1$ statistical weight.

In between, in the density range $10^9 - 10^{11}$ m^{-3}, the density dependence in the density ratio makes the intensity ratio a unique function of electron density, giving an excellent (and reddening-free) means of measuring density in ionized gas regions.

The $3p^2$ structure of S^+ is identical, producing the density sensitive line pair at 6716Å and 6732Å.

9 The HI 21cm line



Level splitting

This important transition arises from a hyperfine splitting of the ground state of the Hydrogen atom. This contains two levels (electron spin up and spin down), which would normally be completely degenerate, but this is broken by the interaction between the spins of the electron and proton. Each particle generates a magnetic moment proportional to its spin vector:

$$\boldsymbol{\mu}_p = \frac{ge}{2m_p}\,\mathbf{s}_p; \qquad \boldsymbol{\mu}_e = -\frac{2e}{2m_e}\,\mathbf{s}_e, \tag{211}$$

where $g \simeq 5.6$ is the proton's g-factor, as opposed to $g = 2$ for the electron, and $-e$ is the charge on the electron, where e is positive. It is not hard to derive an approximation for the interaction energy of these spins, using the Bohr picture of the atom. If the spin is perpendicular to the orbit (i.e. orthogonal to the separation), the interaction energy is

$$E = \mu_0\,\frac{\boldsymbol{\mu}_\mathbf{e}\cdot\boldsymbol{\mu}_\mathbf{p}}{4\pi r^3}, \tag{212}$$

which can be thought of as $-\mathbf{B}\cdot\boldsymbol{\mu}$, where one moment feels the magnetic field of the other. Thus

$$E = -\frac{2ge^2\mathbf{s}_e\cdot\mathbf{s}_p}{16\pi m_e m_p r^3}. \tag{213}$$

The difference in $\mathbf{s}_e\cdot\mathbf{s}_p$ between spins parallel and antiparallel sounds like it should be $2(\hbar/2)^2$, so the splitting is

$$E = -\frac{2ge^2\hbar^2\mathbf{s}_e\cdot\mathbf{s}_p}{32\pi m_e m_p r^3} = (g/4)(m_e/m_p)\alpha^4 m_e c^2, \tag{214}$$

where the final expression adopts the Bohr radius $r = \alpha^{-1}(\hbar/m_e c)$ and $\alpha = e^2/4\pi\epsilon_0\hbar c$ is the fine-structure constant. The proper quantum mechanics of this splitting is discussed in e.g. Section 6.5 of Griffiths's QM book, where he shows that the exact result is larger than our simple estimate by a factor 16/3. But we have reproduced perhaps the most important feature, which is that the splitting is $\propto \alpha^4 m_e c^2$. It might thus be objected that the term 'hyperfine' is inappropriate, since it is of the same α^4 order as the fine-structure splittings (as compared to $E = -\alpha^2 m_e c^2/2$ for the ground state). But the factor m_e/m_p makes the splitting much smaller: $6\,\mu$eV or 0.07 K, with the corresponding wavelength 21.1 cm or frequency 1420.4 MHz. This splitting was first found by Fermi in 1930, proposed as a probe of the neutral universe by Van de Hulst in 1944, and first detected in 1951, opening a new era in astronomy.

Transition rate and emissivity

We have treated the quantum mechanics of the energy level splitting only approximately, but given the splitting, the quantum tools of transition rates can be used to calculate exactly the spontaneous transition rate. This is a spin-changing transition, with no change in the electronic orbital wave functions, so the only perturbation to the Hamiltonian that needs to be considered is from the interaction of the electron spin with the electromagnetic field. We have already seen the required expression: $\Delta H = -\mathbf{B}\cdot\boldsymbol{\mu}$, and that the magnetic moment of an electron is one *Bohr Magneton*, $e\hbar/2m$, so that the electron magnetic moment vector is $\mu = -2\mathbf{s}(e/2m)$. Thus we have

$$\Delta H = \frac{e}{m}\,(\mathbf{B}\cdot\mathbf{s}). \tag{215}$$

We should include both the electron and nuclear spins, but since each is multiplied by e/m and $m_{\rm proton} \gg m_{\rm electron}$, the nuclear component is neglected.

The part of this calculation that needs a little more care is the question of the spin states. As usual, the good quantum number is S, which results from the addition of the electron and proton spins. We either have $S = 0$ (singlet), or $S = 1$ (triplet). Griffiths shows that $\mathbf{s}_p{\cdot}\mathbf{s}_e = \hbar^2/4$ for the triplet, or $-3\hbar^2/4$ for the singlet (prove this by squaring $\mathbf{S} = \mathbf{s}_e + \mathbf{s}_p$ and using $|\mathbf{S}|^2 = S(S+1)\hbar^2$ in terms of the spin quantum number S). A negative $\mathbf{s}_p{\cdot}\mathbf{s}_e$ for the singlet makes intuitive sense (spins antiparallel), and might then be expected to lead to parallel magnetic moments and higher energy. But in fact the ground state is the lowest spin state – the opposite of Hund's rule. The explanation for this is connected to the fact that ψ is non-zero at $r = 0$ for the Hydrogen ground state.

We can calculate the spontaneous transition rate for $|X(S=1)\rangle \to |Y(S=0)\rangle$ by using our standard formula, but changing the matrix element. Previously we had $\Delta H = e\mathbf{A}\cdot\mathbf{p}/m$. The magnetic field magnitude is $B = E/c = -\dot{A}/c = i\omega A/c$, and its direction is along the opposite polarization vector from that of the electric field. But if we are summing over polarizations, the change of direction is irrelevent. Thus we can consider our normal formula for the transition rate, but replace $\langle X|\mathbf{p}|Y\rangle$ by $(\omega/c)\langle X|\mathbf{s}|Y\rangle$. The spatial part of the integral in the transition matrix is unity, as the (purely spin) operator doesn't act on the orbital state which is the same before and after (*i.e.* $\psi(\mathbf{r})_{\rm Y} = \psi(\mathbf{r})_{\rm X}$).

The remaining spin matrix element is between the upper triplet state ($X : S = 1$) and the lower singlet state ($Y : S = 0$). The calculation is given in the appendix. Briefly, the states can be expressed in terms of individual or of combined spins. The initial state can be described as one of the three states $|+\rangle_e|+\rangle_p$, $(|+\rangle_e|-\rangle_p + |-\rangle_e|+\rangle_p)/\sqrt{2}$ and $|-\rangle_e|-\rangle_p$. These can be alternatively expressed as $|1, M_S\rangle$, $M_S = 1, 0, -1$. The final state, $|0,0\rangle$ is also expressed as $(|+\rangle_e|-\rangle_p - |-\rangle_e|+\rangle_p)/\sqrt{2}$.

The spin operator $\mathbf{s} = s_x\mathbf{e}_x + s_y\mathbf{e}_y + s_z\mathbf{e}_z$ can be written in terms of the 'raising' and 'lowering' operators $s_x \pm is_y$ and s_z to arrive at

$$|\mathbf{s}_{\rm YX}|^2 = (|s^+_{\rm YX}|^2 + |s^-_{\rm YX}|^2)/2 + |s_{z,\rm YX}|^2. \tag{216}$$

The electron spin operators operate only on the electron spin function, and the upshot is

$$|2\mathbf{s}_{\rm YX}|^2 = \hbar^2, \qquad \text{any } M_S \tag{217}$$

(the independence of M_S, as for the Lyman α calculation, says that the result of an angle average should not depend on the direction of the chosen z-axis). The transition rate is therefore

$$\Gamma_{\rm spon}(21{\rm cm}) = \frac{4\omega^3}{3\hbar c^3}\frac{e^2}{4\pi\epsilon_0}\left(\frac{\hbar}{2mc}\right)^2 = \frac{\alpha}{3}\,\omega\,\left(\frac{\hbar\omega}{mc^2}\right)^2 = 2.85\times10^{-15}\ \ {\rm s}^{-1} \ \simeq\ (10^7\ {\rm y})^{-1} \tag{218}$$

The critical density of this transition is $n_{\rm crit} = 4.7 \times 10^3 (T/{\rm K})^{0.77}$ m^{-3} (*Kulkarni & Heiles in Hollenbach & Thronson 'Interstellar Processes'*). We may therefore assume that the levels are thermally populated (at a temperature far greater than 0.07K). The statistical weights $(2S+1)$ are 3 and 1 for upper and lower energies, and we assume no other level is occupied, so 3/4 of H atoms are in the upper energy level. Thus the emissivity is

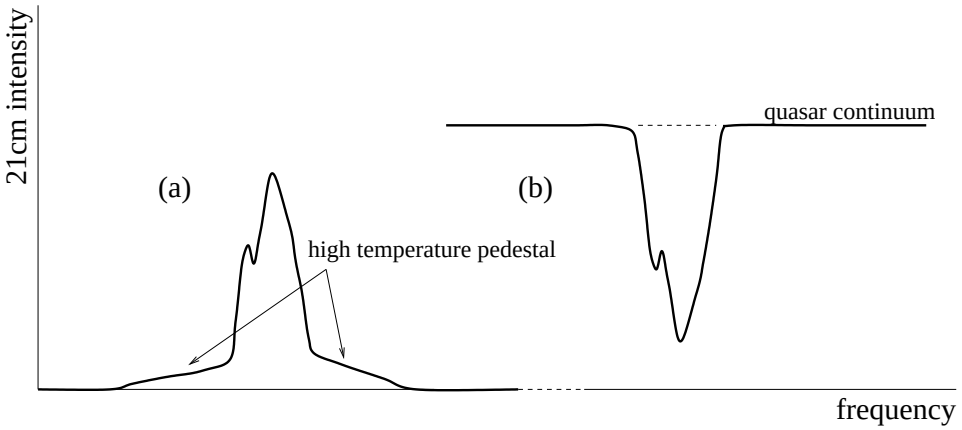
$$\mathcal{E}_\nu = \frac{h\nu}{4\pi}\Gamma_{\rm spon}\frac{3n_{\rm H}}{4}\phi_\nu = 1.6\times10^{-40}\,(n_{\rm H}/{\rm m}^{-3})\,{\rm W\,m^{-1}\,sr^{-1}}\phi_\nu. \tag{219}$$

The Rayleigh-Jeans limit ($h\nu \ll kT$) certainly applies in this case, and so

$\kappa_\nu = \mathcal{E}_\nu/B_\nu(T) = \mathcal{E}_\nu\lambda^2/2kT$, and

$$\tau_\nu = \frac{3hc^2\Gamma_{\rm spon}}{32\pi\nu kT}\,N_{\rm H}\phi_\nu = 2.6\times10^{-19}\left(\frac{N_{\rm H}/{\rm m}^{-2}}{T/{\rm K}}\right)\phi_\nu \tag{220}$$

The optical depth is inversely proportional to the gas kinetic temperature, while the emission is not. This has the following interesting consequence. If HI is observed in emission ($I_\nu = \mathcal{E}_\nu\ell$) in the plane of the galaxy, profile (a) might result. In absorption against a bright extragalactic continuum source in the same direction, profile (b) results, lacking the broad pedestal in profile (a). When first studied in detail (Radhakrishnan 1972, ApJ J suppl. 24) this indicated that there was a component, seen everywhere, that was at high temperature ($\geq$ 1000 K). This result led directly to a 2-phase cloudy model of the interstellar medium.



Incidentally, the fact that in radio astronomy the Rayleigh-Jeans limit so often applies has led to the use of *brightness temperature* T_b (and antenna temperature and a host of cousins) as a measure of specific intensity.

$$T_b = I_\nu/(2k\nu^2/c^2) \tag{221}$$

Similarly, since velocity is related to frequency by the Doppler effect, velocity v is often used as the variable in the line profile.

$$\int I_\nu\,d\nu = \int T_b\,dv \;\times\; 2k\nu^3/c^3 \tag{222}$$

and the line strength (area under the profile) is quoted in Kelvin km s^{-1}.

Other uses for 21cm radiation: mapping spiral arms in our Galaxy – mapping spiral arms in other galaxies – measuring velocity vs. radius in external spirals to detect dark halos by their gravitation – measuring velocity width in distant galaxies to estimate maximum rotational velocity, which correlates with luminosity (Tully-Fisher relation) to determine distances.

10 The CO molecule

As a final application of the machinery for calculating transition rates, we explore a new frequency regime, intermediate between the optical atomic transitions and the 21-cm radio regime: molecular transitions. In principle, these should be complex systems, considering all the electronic and nuclear degrees of freedom. What simplifies things is the *Born-Oppenheimer approximation*, which proposes that we should solve the electron wave functions with the nuclei considered as fixed, and then ask how the solution responds to slow changes of the nuclear configuration. The informal justification for this is that electrons move faster than the heavy nuclei (via the uncertainty principle, which suggests that they should have similar momenta, being confined in a similar space). For a given set of electronic states, the energy as a function of nuclear separation(s) defines an effective classical potential, $V(r)$, so that the nuclei can be thought of as if connected by springs. This must have a minimum around the equilibrium separation, where the cloud of electrons between the nuclei succeeds in cancelling the electrostatic repulsion of the nuclei, leading to a bound state. If the reduced mass of a pair of nuclei is $\mu = m_1 m_2/(m_1 + m_2)$, then around the minimum separation at $r = a$ we will have an approximate oscillator potential:

$$V = \mu\omega^2(r-a)^2/2. \tag{223}$$

When r changes by an amount of order a, the energy change will be of the order of the electronic energy $E_e \simeq p^2/2m$, where $p \sim \hbar/a$. Thus

$$E_{\rm vib} = \hbar\omega \sim \frac{\hbar^2}{(m\mu)^{1/2}a^2} \sim (m/\mu)^{1/2}E_e. \tag{224}$$

If the electronic transitions are in the optical as usual, these vibrational transitions will be in the IR: wavelengths of a few microns.

In addition, molecules can undergo rotational motion. The corresponding energy will be $|\mathrm{J}|^2/2I$, where I is the moment of inertia ($I = \mu a^2$), so these energies are $E_{\rm rot} \sim \hbar^2/\mu a^2$. Thus we have the approximate hierarchy

$$E_e \,:\, E_{\rm vib} \,:\, E_{\rm rot} = 1 \,:\, (m/\mu)^{1/2} \,:\, (m/\mu), \tag{225}$$

so the rotational transitions are in the regime of mm wavelengths. For this lecture, we will be content with considering only rotational transitions (although there are interesting phenomena where both oscillatory and rotational states change simultaneously). This assumption is justified at low temperatures, where the vibrational modes are not excited and the simplest case is pure rotation.

Consider one of the simplest diatomic molecules: CO At low excitation energies (collisions in a gas at a few 10's K, for instance), the molecule behaves like a rigid rotator (dumbbell). The Hamiltonian is

$$H = \frac{\mathbf{J}^2}{2I} \tag{226}$$

where $\mathbf{J}$ is the angular momentum operator, and the eigenstates are represented by the spherical harmonics $Y_{JM}(\theta,\phi)$. The operator which is the square of the AM operator, $\mathbf{J}^2$, has eigenvalues $J(J+1)\hbar^2$, and the z component, J_z, has eigenvalues $M\hbar$ where $M = J, J-1, J-2, \cdots -J$. The moment of inertia I about the centre of mass is $\sum m_n r_n^2 = \mu a^2$ where the reduced mass is $\mu = 6.8 m_{\rm H}$ for nuclear masses of 12 and 16, and the interparticle separation is $a = 1.13$Å. Thus

$$\frac{\hbar^2}{2I} = 2.7K \tag{227}$$

and the first few rotational levels are easily excited in cool gas.

The electric dipole moment $\mathbf{d} = e\overline{\mathbf{r}}$ is 0.11 Debye (4.8 Debye $\equiv |e| \times 1$Å, so $\overline{\mathbf{r}} = 0.0229$Å), and is so small because the molecule is so nearly symmetric (the actual separation of the C and O nuclei is 50 times larger). The dipole moment operator in this case is the sum of $e_i\mathbf{r}_i$ over all the particles in the atom (28 electrons and two nuclei), but as justified earlier we will ignore any changes in the electron wavefunctions: the initial and final states of the electrons will be the same, and the integral with respect to electron positions gives a mean (expectation) value for the vector $e_i\mathbf{r}_i$, say $e\overline{\mathbf{r}}$. The only change will be the (relatively) slow rotation in 3-space, for which the remaining wave functions are spherical harmonics.

The matrix elements $\langle Y|\mathbf{r}|X\rangle = \langle J'M'|\overline{\mathbf{r}}|JM\rangle$ are

$$\mathbf{r}_{\rm YX} = \int Y^*_{J',M'}(\theta,\phi)\,\overline{\mathbf{r}}\,Y_{J,M}(\theta,\phi)d\Omega = \overline{r}\int Y^*_{J',M'}(\theta,\phi)\,\mathbf{n}\,Y_{J,M}(\theta,\phi)d\Omega. \qquad (228)$$

These have exactly the same form as the angular part of the Lyman α calculation – at least if Y is a state with $J' = 0$ and X with $J = 1$, and so the squared modulus of the angle integral is $1/3$ in this case as well. For other J values, tables of integrals of spherical harmonics give the algebraic functions of J, J' that result.

The transition rate is therefore

$$\Gamma_{\rm spon}(J = 1 \to 0) = \frac{4\omega^3}{3\hbar c^3}\frac{e^2}{4\pi\epsilon_0}(\overline{r}^2/3) = 7.4 \times 10^{-8}\ \ {\rm s}^{-1}. \qquad (229)$$

This is the main transition, radiating at 2.6mm, by which molecular gas has been detected and mapped in the Universe. For interest, the rate for $J \to J-1$ is the above rate times $3J/(2J+1)$.

CO is of interest not so much in its own right as because it is a *tracer* of the main actor in the molecular universe: molecular hydrogen. A dominant aim in astronomy is to study star formation in all its contexts, and this is difficult because of extinction. Star formation requires gravitational collapse, which in turn requires cold neutral gas that lacks pressure support. But the presence of ionizing radiation means that such gas is preferentially located in dense and dusty regions that can shield themselves from ionization – and the dust extinction makes it hard to study such regions at optical wavelengths. A more direct indicator of the accumulation of the raw material for star formation would be molecular hydrogen, since such fragile molecules can only form in well-shielded regions (and, indeed, their formation is catalyzed on the surface of dust grains). But molecular hydrogen is not easy to study directly: its moment of inertia is low, so the spacing of rotational energy levels is larger. Also, it is a symmetric molecule, so there are no dipole transitions. The lowest rotational transition is $J = 2 - 0$ at $28\,\mu$m. CO lines are much stronger and easier to access, so it is assumed that the CO and H_2 abundances go in step. The main problem with this strategy is that CO is too strong a transition and therefore is often optically thick: the brightness temperature indicates the thermal temperature, rather than the column density. A better indicator of the molecular column is therefore obtained from the velocity width of the line: this indicates the frequency where the optical depth is unity, yielding an estimate of the column if the kinetic temperature is assumed to be of order 10 K.

Mathematical Appendix *(a non-examinable addition to the notes)*

A1 Vector Identities

Stokes's theorem

$$\int_{\text{area}} \boldsymbol{\nabla}\wedge\mathbf{F}\cdot\mathrm{d}\mathbf{A} = \oint_{\text{edge}} \mathbf{F}\cdot\mathrm{d}\ell \tag{230}$$

Gauss's divergence theorem

$$\int_{\text{volume}} \boldsymbol{\nabla}\cdot\mathbf{F}\;\mathrm{d}V = \int_{\text{area}} \mathbf{F}\cdot\mathrm{d}\mathbf{A} \tag{231}$$

Identities – often provable using $(\boldsymbol{\nabla}\wedge\mathbf{A})_i = \epsilon_{ijk}\nabla_j A_k$

$$\boldsymbol{\nabla}\wedge\boldsymbol{\nabla}\chi = 0, \text{ any } \chi(\mathbf{r}) \tag{232}$$

$$\boldsymbol{\nabla}\cdot\boldsymbol{\nabla}\wedge\mathbf{X} = 0, \text{ any } \mathbf{X}(\mathbf{r}) \tag{233}$$

$$\boldsymbol{\nabla}\wedge\boldsymbol{\nabla}\wedge\mathbf{X} = \boldsymbol{\nabla}\boldsymbol{\nabla}\cdot\mathbf{X} - \nabla^2\mathbf{X}, \text{ any } \mathbf{X}(\mathbf{r}) \tag{234}$$

$$\boldsymbol{\nabla}\cdot(\mathbf{X}\wedge\mathbf{Y}) = \mathbf{Y}\cdot\boldsymbol{\nabla}\wedge\mathbf{X} - \mathbf{X}\cdot\boldsymbol{\nabla}\wedge\mathbf{Y}, \text{ any } \mathbf{X}(\mathbf{r}),\mathbf{Y}(\mathbf{r}) \tag{235}$$

$$\boldsymbol{\nabla}(\mathbf{X}\cdot\mathbf{Y}) = \mathbf{X}\wedge\boldsymbol{\nabla}\wedge\mathbf{Y} + \mathbf{Y}\wedge\boldsymbol{\nabla}\wedge\mathbf{X} + \mathbf{X}\cdot\boldsymbol{\nabla}\mathbf{Y} + \mathbf{Y}\cdot\boldsymbol{\nabla}\mathbf{X} \tag{236}$$

Vector derivatives of waves (note: $f'(x) \equiv df/dx$)

$(\partial/\partial t)[t - \mathbf{n}\cdot\mathbf{r}/c] = 1$ so $\partial f/\partial t = f'$
$(\partial/\partial x)[-\mathbf{n}\cdot\mathbf{r}/c] = -n_x/c,$ so $\partial f/\partial x = (-n_x/c)\, f'$ and also for y, z.

Hence

$$\begin{aligned}
\boldsymbol{\nabla} f(t - \mathbf{n}\cdot\mathbf{r}/c) &= \mathbf{e}_x\partial f/\partial x + \mathbf{e}_y\partial f/\partial y + \mathbf{e}_z\partial f/\partial z \\
&= \mathbf{e}_x\,(-n_x/c)\, f' + \mathbf{e}_y\,(-n_y/c)\, f' + \mathbf{e}_z\,(-n_z/c)\, f' \\
&= -(\mathbf{n}/c)\, f' \\
&= -(\mathbf{n}/c)\,(\partial f/\partial t)
\end{aligned}$$

By similar arguments

$$\boldsymbol{\nabla}\cdot\mathbf{F}(t - \mathbf{n}\cdot\mathbf{r}/c) = \sum_{i=x,y,z} (\partial/\partial x_i)F_i = -(\mathbf{n}/c)\cdot(\partial\mathbf{F}\,/\partial t) \tag{237}$$

and

$$\begin{aligned}
\left[\boldsymbol{\nabla}\wedge\mathbf{F}(t - \mathbf{n}\cdot\mathbf{r}/c)\right]_x &= (\partial/\partial y)F_z - (\partial/\partial z)F_y \\
&= (-n_y/c)\, F_z' - (-n_z/c)\, F_y' \\
&= \left[(-\mathbf{n}/c)\wedge\mathbf{F}'\right]_x = \left[(-\mathbf{n}/c)\wedge(\partial\mathbf{F}\,/\partial t)\right]_x
\end{aligned}$$

and also for y, z, *i.e.*

$$\boldsymbol{\nabla}\wedge\mathbf{F}(t - \mathbf{n}\cdot\mathbf{r}/c) = -(\mathbf{n}/c)\wedge(\partial\mathbf{F}\,/\partial t) \tag{238}$$

A2 Poynting vector

Manipulate Maxwell's equations to get

$$\mathbf{E}\cdot\boldsymbol{\nabla}\wedge\mathbf{B} - \mathbf{B}\cdot\boldsymbol{\nabla}\wedge\mathbf{E} = \mu_0\left(\mathbf{j}\cdot\mathbf{E} + \frac{\partial}{\partial t}\left(\frac{\epsilon_0}{2}\mathbf{E}^2 + \frac{1}{2\mu_0}\mathbf{B}^2\right)\right) \tag{239}$$

The LHS is $-\boldsymbol{\nabla}\cdot(\mathbf{E}\wedge\mathbf{B})$. Integrate over a given volume:

$$-\frac{1}{\mu_0}\int \boldsymbol{\nabla}\cdot(\mathbf{E}\wedge\mathbf{B})\,dV = \left(\int \mathbf{j}\cdot\mathbf{E}\,dV + \frac{\partial}{\partial t}\int\left(\frac{\epsilon_0}{2}\mathbf{E}^2+\frac{1}{2\mu_0}\mathbf{B}^2\right)\,dV\right) \tag{240}$$

On the RHS the first integral is the rate at which the current does work, while the second integral is the rate of increase of energy stored in the fields. Therefore the LHS must be the energy entering the system. By the divergence (Gauss's) theorem, the integral on the LHS is equal to the outward-pointing surface integral of $-\mathbf{E}\wedge\mathbf{B}/\mu_0$, which implies that the energy flux density is $\mathbf{E}\wedge\mathbf{B}/\mu_0$. By thinking in terms of photons, each with momentum energy$/c$, we see that there is a corresponding momentum flux density of $\mathbf{E}\wedge\mathbf{B}/\mu_0 c$.

A3 Solution of $\Box\phi(\mathbf{r},t)=\rho(\mathbf{r},t)$

The wave equation for all components of the 4-potential is of this mathematical form, apart from a factor μ_0. For a time-independent source, this would just be Poisson's equation; we can make the equation time-independent by Fourier transforming with respect to time, and the resulting equation can be solved in much the same way as Poisson's equation. Fourier transforming simplifies time derivatives: $\partial^2/\partial t^2\to-\omega^2$, so that

$$\nabla^2\tilde{\phi}+k^2\tilde{\phi}=-\tilde{\rho}(\mathbf{r},\omega), \tag{241}$$

where $\tilde{\rho}$ is the Fourier Transform of ρ and $k=\omega/c$.

The neatest way to solve such an equation is to find the Green's function – the response for the case where the source $\tilde{\rho}$ is a point source:

$$\nabla^2 G(\mathbf{r},\mathbf{r}')+k^2G(\mathbf{r},\mathbf{r}')=-\delta^{(3)}(\mathbf{r}-\mathbf{r}')\,. \tag{242}$$

Because of the spherical symmetry, the solution must be a function of r only (if we shift the origin to the location of the spike, $\mathbf{r}'$). In this case, we can use

$$\nabla^2 G=\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}G\right)=\frac{1}{r}\frac{\partial^2}{\partial r^2}\left(r\,G\right)\,. \tag{243}$$

Using the second form, we see that the equation for $r\neq0$ is just the 1D wave equation for rG, so that the solution is

$$G=\frac{1}{r}A\exp(\pm ikr). \tag{244}$$

What is the constant A? For small r, the effect of kr is negligible and the solution tends to $G\propto 1/r$. We already know that this gives a delta-function in Poisson's equation:

$$\int\nabla^2\left(\frac{1}{r}\right)\,dV=\int\boldsymbol{\nabla}\left(\frac{1}{r}\right)\cdot d\mathbf{A}=\int\left(-\frac{1}{r^2}\right)\,r^2\,d\Omega=-4\pi. \tag{245}$$

Thus there must be a spike at the origin: $\nabla^2(1/r)=-4\pi\delta^{(3)}(\mathbf{r})$. The k^2G term is negligible compared to a delta-function, so we conclude that

$$G=\frac{1}{4\pi r}\exp(\pm ikr)=\frac{1}{4\pi|\mathbf{r}-\mathbf{r}'|}\exp(\pm ik|\mathbf{r}-\mathbf{r}'|) \tag{246}$$

(shifting the origin back to $\mathbf{r}'$).

Now we use the Green's function as usual: multiply by $\tilde{\rho}(\mathbf{r}')$ and integrate:

$$\tilde{\phi}(\mathbf{r},\omega)=\frac{1}{4\pi}\int\frac{\tilde{\rho}(\mathbf{r}',\omega)\,\exp(\pm i\omega|\mathbf{r}-\mathbf{r}'|/c)}{|\mathbf{r}-\mathbf{r}'|}\,dV'. \tag{247}$$

Finally, we take the inverse transform, noting that the transform of

$$\tilde{\rho}(\mathbf{r}', \omega) e^{\pm i\omega|\mathbf{r}-\mathbf{r}'|/c} \tag{248}$$

is $\rho(\mathbf{r}', t \pm |\mathbf{r} - \mathbf{r}'|/c)$. Thus we conclude that

$$\phi(\mathbf{r}) = \frac{1}{4\pi} \int \frac{\rho(\mathbf{r}', t \pm |\mathbf{r}' - \mathbf{r}|/c)}{|\mathbf{r} - \mathbf{r}'|} \, dV' \,. \tag{249}$$

The very last step is to resolve the sign ambiguity, and we see that there are two possibilities: *advanced* or *retarded*, where the field at $\mathbf{r}$ and time t is determined the source at times respectively after or before t. The former option seems to make no physical sense in terms of causality, so it is normal to reject the advanced solution on these grounds. But it makes sense that the advanced solution is present: Maxwell's equations are second order in time, so are unchanged under $t \to -t$. In other words, we can run a film of an electromagnetic process backwards and it must represent a valid physical process. This is familiar in Newtonian dynamics: toss a pebble in a pool and ripples spread out – but in principle ripples could be generated at the edge of the pool and spread inwards. So we see that the *mathematical* reason for rejecting advanced solutions is to do with boundary conditions: advanced solutions would require us to constrain the field at infinity. We therefore conclude that the solution to the wave equation can involve only the retarded source, written $[\rho]$, which is the source as it was at the time when the radiation we now receive was emitted.

A4 The Lagrangian and Hamiltonian of a particle in an electromagnetic field

We want to rewrite the Lorentz force, $\mathbf{F} = Q(\mathbf{E} + \mathbf{v} \wedge \mathbf{B})$ in the language of generalised dynamics. The requirement of stationary action, $\delta L(q, \dot{q}) \, dt = 0$ with respect to the trajectory $\mathbf{q}(t)$, gives *Euler's equation*:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) = \frac{\partial L}{\partial q_i}. \tag{250}$$

Note that we follow convention here and use q for coordinates, so temporarily we use Q for charge. For a fixed potential, the Lagrangian is $L = K - V$, so for electrostatics we would have

$$L = \frac{1}{2}mv^2 - Q\phi. \tag{251}$$

This is the non-relativistic case, which will suffice here. But even so, relativistic considerations suggest that the ϕ term will arise from $A^\mu U_\mu = \phi - \mathbf{v} \cdot \mathbf{A}$. This suggests that, for non-static fields, the non-relativistic Lagrangian might be

$$L = \frac{1}{2}mv^2 - Q\phi + Q\mathbf{v} \cdot \mathbf{A}. \tag{252}$$

It's not hard to show that this does indeed yield the Lorentz force. You need to remember the definition of the total, or 'convective' time derivative: $d/dt = \partial/\partial t + \dot{\mathbf{q}} \cdot \boldsymbol{\nabla}$, plus the triple vector product to deal with $\mathbf{v} \wedge \mathbf{B} = \mathbf{v} \wedge (\boldsymbol{\nabla} \wedge \mathbf{A})$. The only two subtleties are that (1) the required acceleration is $(d/dt)\mathbf{v}$; (2) spatial derivatives commute with $\mathbf{v} = \dot{\mathbf{q}}$, because Lagrangian dynamics treats q and $\dot{q}$ as independent.

Having the Lagrangian, we immediately get the generalized momentum, $p_i \equiv \partial L/\partial \dot{q}_i$:

$$\mathbf{p} = m\mathbf{v} + Q\mathbf{A}. \tag{253}$$

The Hamiltonian is

$$H \equiv \mathbf{p} \cdot \mathbf{v} - L = mv^2/2 + Q\phi = (\mathbf{p} - Q\mathbf{A})^2/2m + Q\phi. \tag{254}$$

Note the subtlety that $\mathbf{p} \neq m\mathbf{v}$. When making the transition to quantum mechanics, it is the generalized $\mathbf{p}$ that gets replaced by $-i\hbar\boldsymbol{\nabla}$.

For more detail on all of this, including the relativistic case, see *e.g.* section 12.1 of Jackson.

A5 Maxwell averaged cross section

The cross section for a beam of particles of type 1, travelling with velocity v, colliding with a static particle of type 2 ('colliding with' can mean reacting with or being deflected by) is $\sigma = \sigma(v)$. In a gas at temperature T consisting of n_1 m^{-3} particles of type 1 and n_2 m^{-3} particles of type 2, the rate of collision is

$$n_1 n_2 \langle \sigma v \rangle \quad \text{m}^{-3}\ \text{s}^{-1} \tag{255}$$

This is proven in two steps. First, choose a coordinate frame moving with velocity $\mathbf{v}_2$. Consider the flux of particles 1 on a particle 2 which is at rest in this frame. This will be given by σ times the flux of particles 1, *i.e.* n_1 times the relative velocity $v = |\mathbf{v}_1 - \mathbf{v}_2|$, integrated over the Maxwell velocity distribution of particles 1. This distribution is

$$f(\mathbf{v}_1; m_1)\ d^3\mathbf{v}_1 = \left(\frac{m_1}{2\pi kT}\right)^{3/2} e^{-m_1 v_1^2/2kT}\ d^3\mathbf{v}_1. \tag{256}$$

The result is

$$n_1 \int_{\mathbf{v}_1} \sigma v f(\mathbf{v}_1; m_1)\ d^3\mathbf{v}_1. \tag{257}$$

Now, in unit volume and in a small velocity range (nearly at rest in the moving frame) there are $n_2 f(\mathbf{v}_2; m_2)\ d^3\mathbf{v}_2$ particles 2. Thus the number of collisions m^{-3} s^{-1}of particles 1 with particles 2 is

$$n_1 n_2 \int_{\mathbf{v}_1} \int_{\mathbf{v}_2} \sigma v f(\mathbf{v}_1; m_1)\ f(\mathbf{v}_2; m_2)\ d^3\mathbf{v}_1\ d^3\mathbf{v}_2. \tag{258}$$

Now change variables from $\mathbf{v}_1, \mathbf{v}_2$ to $\mathbf{v} = \mathbf{v}_1 - \mathbf{v}_2$, $\mathbf{V} = (m_1\mathbf{v}_1 + m_2\mathbf{v}_2)/(m_1 + m_2)$ (*i.e.* to relative and centre-of-mass velocities). Note that $m_1 m_2 = \mu M$ where $M = m_1 + m_2$ and *reduced mass* $\mu = m_1 m_2/M$, and $m_1\mathbf{v}_1^2 + m_2\mathbf{v}_2^2 = \mu\mathbf{v}^2 + M\mathbf{V}^2$. Also (the Jacobian has magnitude unity) $d^3\mathbf{v}_1\ d^3\mathbf{v}_2 = d^3\mathbf{v}\ d^3\mathbf{V}$. Substituting in the integrals, they become

$$n_1 n_2 \int_{\mathbf{v}} \int_{\mathbf{V}} \sigma v f(\mathbf{v}; \mu)\ f(\mathbf{V}; M)\ d^3\mathbf{v}\ d^3\mathbf{V}. \tag{259}$$

The integrals separate, that over $\mathbf{V}$ integrating to unity (normalized velocity distribution), while that over $\mathbf{v}$ is the Maxwell average of σv for a particle of mass μ. this average is the one-dimensional integral (express $d^3\mathbf{v}$ in polar coordinates $v^2 dv\ d\Omega$):

$$\langle \sigma v \rangle = 4\pi \left(\frac{m_1}{2\pi kT}\right)^{3/2} \int_0^\infty \sigma v^3 e^{-\mu v^2/2kT}\ dv. \tag{260}$$

[Note:
1. the collision rate for a binary process – 2 particles – is proportional to density *squared* ($n_1 n_2 = x_1 x_2 n^2$ where x_1, x_2 are abundances). You can guess that the triple collision rate will be proportional to density *cubed*.

2. The coefficient $\langle \sigma v \rangle$ is a function of temperature. If $\sigma v \propto E^\alpha \propto v^{2\alpha}$ then inspection of the integral shows that $\langle \sigma v \rangle \propto T^\alpha$.]

A6 Spin

Rules of angular momentum:

$$[j_x, j_y] = i\hbar j_z, \quad \text{(cyclic)} \tag{261}$$

hence

$$[\mathbf{j}^2, j_z] = 0 \quad \text{(for any one cpt.)} \tag{262}$$

and the complete set of states are specified by the eigenvalues of these two operators. Invent ladder operators $j_\pm = j_x \pm i\, j_y$, and note $(j_+)^\dagger = j_-$ and $\mathbf{j}^2 = j_\pm j_\mp + j_z^2 \mp j_z$. The logic is similar to that of the harmonic oscillator. First we prove the raising and lowering properties: e.g. consider $j_z j_+|j,\, m\rangle$. Using the known commutators,

$$j_z j_+ = j_z j_x + i j_z j_y = (i\hbar j_y + j_x j_z) + i(j_y j_z - i\hbar j_x) = j_+(\hbar + j_z), \tag{263}$$

so that $j_+|j,\, m\rangle$ is an eigenstate of j_z with eigenvalue $(m+1)\hbar$. As before, consider the normalization $\langle j_+\psi|j_+\psi\rangle = \langle\psi|j_-j_+|\psi\rangle$. Since $j_-j_+ = j^2 - j_z^2 - \hbar j_z$, the eigenvalue of j_z cannot be too large, so we must have a top state where j_+ gives zero and a bottom state where j_- gives zero – and we must be able to get from one to the other by n steps. If the top j_z eigenvalue is t, and the bottom one is b, and the eigenvalue of j^2 is α, then setting $j_-j_+ = 0$ for the top state gives $\alpha = t(t+\hbar)$. Similarly at the bottom, $\alpha = (t-n\hbar)^2 - (t-n\hbar)$, which together gives $t = -b = n\hbar/2$: the algebra of angular momentum requires the possible existence of spin.

Hence we have $2j+1$ states, with z-component eigenvalues $m = -j, -(j-1), \cdots, (j-1), j$ in units of $\hbar$; the eigenvalue of squared AM is $j(j+1)$ in these units; and

$$j_\pm|j,\, m\rangle = \sqrt{(j \mp m)(j \pm m + 1)}\,\hbar\,|j,\, m \pm 1\rangle. \tag{264}$$

The wave functions corresponding to the states with $2j$ even are $\langle x|j, m\rangle = Y_{jm}(\theta, \phi)$, the spherical harmonics. for $2j$ odd there are no such wave functions (we use 2-component *spinors*).

Adding two AM operators (a system with two angular momenta), $\mathbf{J} = \mathbf{j}(1) + \mathbf{j}(2)$, gives complete sets of operators $\mathbf{j}(1)^2, \mathbf{j}(2)^2, \mathbf{J}^2, J_z$, or alternatively $\mathbf{j}(1)^2, \mathbf{j}(2)^2, j(1)_z, j(2)_z$ (check that in each group the operators commute). Since each is complete, the eigenstates of one group must be linear combinations of the eigenstates in the other. The coefficients are called *Clebsch-Gordan coefficients*. They can be found using the ladder operators $j_\pm$.

Adding two spins

We apply this to two spin $\frac{1}{2}$ particles. The spin operator has two eigenvalues, $\pm 1/2$, and the two eigenstates can be written $|\pm\rangle$. The commuting operators (neglect the $\mathbf{s}(1)^2$, $\mathbf{s}(2)^2$ operators, since their eigenvalues are always the same, 3/4) are $s_z(1), s_z(2)$ or $\mathbf{s}^2$, s_z, where $\mathbf{s} = \mathbf{s}(1) + \mathbf{s}(2)$. The corresponding eigenstates are then either $|+\rangle|+\rangle, |+\rangle|-\rangle, |-\rangle|+\rangle, |-\rangle|-\rangle$ or alternatively $|S, M_S\rangle$. We relate these alternative descriptions by noting that the $S = 1$, $M_S = 1$ state must be both spins 'up', *i.e.* $|1, 1\rangle = |+\rangle|+\rangle$.

To get the $|1,\, 0\rangle$ state, apply $S_- = s_-(1) + s_-(2)$ to the above equation and apply the formula for the action of $j_\pm$ (using S_- on the left and $s_-(1) + s_-(2)$ on the right):

$$|1,\, 0\rangle = \frac{1}{\sqrt{2}}\left(|+-\rangle + |-+\rangle\right), \tag{265}$$

and again to get $|1,\, -1\rangle$:

$$|1,\, -1\rangle = |--\rangle. \tag{266}$$

The remaining state, $|0, 0\rangle$, must be orthogonal to all of these, and in particular to $|1, 0\rangle$, from which we infer

$$|0,\, 0\rangle = \frac{1}{\sqrt{2}}\left(|+-\rangle - |-+\rangle\right) \tag{267}$$

(we could have derived this also by demanding $S_\pm|0,\, 0\rangle = 0$).

Adding two $\ell = 1$ states

Similarly, two particles with $\ell = 1$ gives $j = 2, 1, 0$. We start with the manifestly symmetric $m = 2$ state and lower progressively:

$$\begin{aligned}
|2,2\rangle &= |1\rangle|1\rangle \\
|2,1\rangle &= (1/\sqrt{2})\ (|1\rangle|0\rangle + |0\rangle|1\rangle) \\
|2,1\rangle &= (1/\sqrt{6})\ (|1\rangle|-1\rangle + 2|0\rangle|0\rangle + |-1\rangle|1\rangle) \\
|2,-1\rangle &= (1/\sqrt{2})\ (|-1\rangle|0\rangle + |0\rangle|-1\rangle) \\
|2,-2\rangle &= |-1\rangle|-1\rangle\ .
\end{aligned} \tag{268}$$

Now we need to find the maximal state of $j = 1$, with $m = 1$. This must be made from the individual $|1\rangle$ and $|0\rangle$ states, but we already have the symmetric combination of these for $|2,1\rangle$, so the only remaining possibility that is orthogonal to this is the antisymmetric

$$|1,1\rangle = (1/\sqrt{2})\ (|1\rangle|0\rangle - |0\rangle|1\rangle). \tag{269}$$

Lowering from this gives

$$\begin{aligned}
|1,1\rangle &= (1/\sqrt{2})\ (|1\rangle|0\rangle - |0\rangle|1\rangle) \\
|1,0\rangle &= (1/\sqrt{2})\ (|1\rangle|-1\rangle - |-1\rangle|1\rangle) \\
|1,-1\rangle &= (1/\sqrt{2})\ (|0\rangle|-1\rangle - |-1\rangle|0\rangle).
\end{aligned} \tag{270}$$

Finally, for $|0,0\rangle$, we need something orthogonal to $|2,0\rangle$ and $|1,0\rangle$. The only option is

$$|0,0\rangle = (1/\sqrt{3})\ (|1\rangle|-1\rangle - |0\rangle|0\rangle + |-1\rangle|1\rangle). \tag{271}$$

Spin and symmetry

In the absence of spin-dependent interactions, we factorize the single-particle wavefunctions into *product states*, $\psi(\mathbf{x}, \mathbf{S}) = \psi_{\text{space}}(\mathbf{x}) \times \psi_{\text{spin}}(\mathbf{S})$. If the particles do not interact amongst themselves, we can produce multi-particle states by multiplying product states for each particle, and then taking combinations of the appropriate symmetry. Consider how this works for two particles, where there are two single-particle spatial states A and B, and the two spin states α and β. A typical product state might be written in the following shorthand:

$$\psi(\mathbf{x}_1, \mathbf{x}_2, \mathbf{S}_1, \mathbf{S}_2) = A(\mathbf{x}_1)\alpha(\mathbf{S}_1)B(\mathbf{x}_2)\beta(\mathbf{S}_2) \equiv A(1)\alpha(1)B(2)\beta(2), \tag{272}$$

i.e. particle 1 is in state A with spin up; particle 2 is in state B with spin down.

There are 4 single-particle states ($A\alpha$, $A\beta$, $B\alpha$, $B\beta$), so there are ${}^4C_2 = 6$ distinct pairs to combine into 2-particle states (the exclusion principle says that the two states must be different). Written explicitly, the antisymmetric combinations are

$$\begin{aligned}
\psi_a &= [A(1)\alpha(1)A(2)\beta(2) - A(2)\alpha(2)A(1)\beta(1)]/\sqrt{2} \\
\psi_b &= [B(1)\alpha(1)B(2)\beta(2) - B(2)\alpha(2)B(1)\beta(1)]/\sqrt{2} \\
\psi_c &= [A(1)\alpha(1)B(2)\alpha(2) - A(2)\alpha(2)B(1)\alpha(1)]/\sqrt{2} \\
\psi_d &= [A(1)\beta(1)B(2)\beta(2) - A(2)\beta(2)B(1)\beta(1)]/\sqrt{2} \\
\psi_e &= [A(1)\alpha(1)B(2)\beta(2) - A(2)\alpha(2)B(1)\beta(1)]/\sqrt{2} \\
\psi_f &= [A(1)\beta(1)B(2)\alpha(2) - A(2)\beta(2)B(1)\alpha(1)]/\sqrt{2}.
\end{aligned} \tag{273}$$

The interesting thing to note is that linear combinations of these states can be rewritten as

a product of a space wavefunction and a spin wavefunction, each with a definite symmetry:

$$\begin{aligned}
\psi_a &= [A(1)A(2)] \times [\alpha(1)\beta(2) - \alpha(2)\beta(1)]/\sqrt{2} \\
\psi_b &= [B(1)B(2)] \times [\alpha(1)\beta(2) - \alpha(2)\beta(1)]/\sqrt{2} \\
\psi_c &= [A(1)B(2) - A(2)B(1)]/\sqrt{2} \times [\alpha(1)\alpha(2)] \\
\psi_d &= [A(1)B(2) - A(2)B(1)]/\sqrt{2} \times [\beta(1)\beta(2)] \\
(\psi_e + \psi_f)/\sqrt{2} &= [A(1)B(2) - A(2)B(1)]/\sqrt{2} \times [\alpha(1)\beta(2) + \alpha(2)\beta(1)]/\sqrt{2} \\
(\psi_f - \psi_e)/\sqrt{2} &= [A(1)B(2) + A(2)B(1)]/\sqrt{2} \times [\alpha(1)\beta(2) - \alpha(2)\beta(1)]/\sqrt{2}
\end{aligned} \tag{274}$$

In short,

$$\psi(\mathbf{x}_1, \mathbf{x}_2, \mathbf{S}_1, \mathbf{S}_2) = \psi_{\text{space}}(\mathbf{x}_1, \mathbf{x}_2) \times \psi_{\text{spin}}(\mathbf{S}_1, \mathbf{S}_2), \tag{275}$$

where each of the space and spin wavefunctions has a distinct symmetry. In order that the overall wavefunction be antisymmetric (for fermions) under exchange of space and spin labels, the symmetries of these two parts must be *opposite*, i.e. $(-) = (+)(-)$ or $(-) = (-)(+)$. For a boson, the wavefunction can be decomposed in the same way, but now the symmetries of ψ_{space} and ψ_{spin} must the same.

This factorization is very convenient for making symmetry arguments in a transparent way – but unfortunately it only works for two-particle systems; for 3 or more, we have to consider just the combination of product states. To prove this, consider the spin wavefunction for 3 particles: at least 2 must be either both spin-up or spin-down, so we cannot construct an antisymmetric spin wavefunction for three or more fermions. If factorization always applied, this says that the spatial wavefunction would have to be antisymmetric, so that all three particles would need to be in different states. However, we can put two particles in state A with spins up and down, and a third particle in state B with e.g spin up. The wavefunction is the antisymmetrized version of $A(1)\alpha(1)A(2)\beta(2)B(3)\alpha(3)$, which does not factorize, by the above argument.

A7 Review of atomic structure

LS coupling

The hydrogen atom has an exact solution. Helium and more complex atoms have at least three bodies, and therefore do not have a closed solution, but must be solved by approximation methods. However, H is a guide. The periodic system occurs as H-like angular momentum and energy states are filled by successive electrons, the exclusion principle for fermions forbidding more than one electron per energy, angular momentum and spin ('up' or 'down') state. Each energy shell is successively less bound, and as each shell is completed (i.e. all the electrons permitted are there) it behaves like a spherically symmetric system (each plus angular momentum is balanced by the corresponding minus). In a non-Coulomb radial force such as now applies, different angular momenta have different energies.

The method of Hartree assumes that each of the outer electrons moves in a spherically symmetric field due to the closed shells plus the angularly smoothed distribution of the other outer electrons. This process is iterated until the wave function of the outer electron is consistent. Then the Hamiltonian is

$$H = H_0 + H_1 + H_2 \tag{276}$$

where
$H_0 = \sum_i p_i^2/2m + U(r_i)$
$H_1 = \sum_i(-Ze/4\pi\epsilon_0 r_i - U(r_i)) + \sum_{i,j} e^2/4\pi\epsilon_0|\mathbf{r}_i - \mathbf{r}_j|$
$H_2 =$ relativistic effects.
Here $U(r)$ is the potential derived from the smoothed out distribution. The terms in the

Hamiltonian are successively smaller. The neglect of H_2 is the LS coupling approximation. The H_1 term splits the ground state into a few terms with well-determined (L, S), having energies differing by the residual e^-e^- interactions (*e.g.* in O^{++} with an ionization potential of 55 eV, the splitting is up to 5 eV).

Now, the arrangements. The individual electrons interact *via* their non-spherically symmetric wave functions so the individual **l**'s are not conserved; but $\mathbf{L} = \sum \mathbf{l}_i$ is. No spin terms occur in the Hamiltonian in LS coupling ($H_2 = 0$) so $\mathbf{S} = \sum \mathbf{s}_i$ is also conserved although, because of the coupling of spin states to angular momentum states to preserve overall antisymmetry of the electron states, the individual **s**'s are not.

What values of **L** are possible? The individual **l**'s can add to give a maximum (component in the 'z' direction) $M_L = \sum l_i$ corresponding to an **L** with that maximum, and the angular momenta can also combine to give **L**'s with all the lower values.

Example (1) The p^2 configuration: two $\ell = 1$ electrons with the same principal quantum number n also – electrons with the same n, ℓ are called *equivalent*.

There are six possible states for each electron (*cf. Landau & Lifshitz QM* §67), with $m_\ell,\ m_s$ given by

$(a)\ \ 1,\frac{1}{2}$ $\quad (a')\ \ 1,-\frac{1}{2}$

$(b)\ \ 0,\frac{1}{2}$ $\quad (b')\ \ 0,-\frac{1}{2}$

$(c)\ -1,\frac{1}{2}$ $\quad (c')\ -1,-\frac{1}{2}$

The exclusion principle forbids certain combinations. Noting only those seven combinations with positive M_L and M_S, $a + a'$ gives $M_L = 2,\ M_S = 0$, which is the only $M_L = 2$ component, and is therefore part of a ^1D term.[2] There are other combinations that will represent components of the ^1D term (*e.g.* $a+b'$ is the (1,0) component and one of the (0,0) components belongs to the D term). This leaves the only other $M_L = 1$ component $a + b$, which must be part of a triplet ^3P term. There is a triplet $M_L = 0$, part of the ^3P term, and two $M_L = 0$ singlets left. One belongs to the ^3P term and is the $M_S = 0$ component of that spin triplet. The other single one must represent a ^1S term. This accounts for all allowed combinations.

This counting scheme is legitimate, but in fact the eigenstates are symmetrised combinations of states with the same $M_L,\ M_S$.

Example (2) The p^3 configuration: three equivalent $\ell = 1$ electrons. There cannot be any F terms, as this would require the same l for each electron, and there are only two different spin states. Combining allowed states gives the following possible $(M_L,\ M_S)$:

$$a + a' + b \to (2, \frac{1}{2}), \quad a + b + b' \to (1, \frac{1}{2}), \quad a' + b + c \to (0, \frac{1}{2}) \tag{277}$$

$$a + a' + c \to (1, \frac{1}{2}), \quad a + b' + c \to (0, \frac{1}{2}) \tag{278}$$

$$a + b + c \to (0, \frac{3}{2}), \quad a + b + c' \to (0, \frac{1}{2}) \tag{279}$$

[2]The superscript is $(2S + 1)$, the spin *multiplicity*. In this case $S = 0$, a singlet term. $S = \frac{1}{2}$ gives a doublet, multiplicity 2, etc. If there is a subscript to the right it gives the total angular momentum quantum number $J = L + S$. The configuration eigenstate with a given $L,\ S$ is called a *term*.

The first row is the list of states of the ^2D term, the second is a ^2P term, and the third a ^4S term.

Hund's rule (1926) states that the lowest energy term is that with the highest spin (the most spin symmetric, therefore the most orbit-antisymmetric, therefore the term where the electrons are furthest from each other and the interaction energy – positive – is least); and thereafter the highest angular momentum (keeping furthest from the nucleus) term has lowest energy.

Thus the p^2 terms are (in increasing energy): ^3P, ^1S, ^1D.

The p^3 terms are: ^4S, ^2D, ^2P.

Spin-orbit coupling. The so-far neglected part of the Hamiltonian, H_2, spoils the LS scheme, but only as a perturbation that splits the terms into states with different J. This splitting is due to a relativistic term in H_2 proportional to $\mathbf{L}\cdot\mathbf{S}$ Since (angular momentum operators)

$$\mathbf{J} = \mathbf{L} + \mathbf{S} \tag{280}$$

and the eigenvalue of $\mathbf{J^2}$ is $J(J+1)$, etc,

$$J(J+1) = L(L+1) + S(S+1) + 2\mathbf{L}\cdot\mathbf{S} \tag{281}$$

Thus the term splitting, proportional to $\mathbf{L}\cdot\mathbf{S}$, differs between levels of different J, even though L and S are the same, lifting the degeneracy and converting the single term into closely-spaced levels.

For example, the $2p^2$ term becomes $^3P_0$, $^3P_1$, $^3P_2$.

This completes the quick review of atomic structure.

A8 Matrix elements of $\mathbf{s}_{\rm electron}$ between $S=1, M_S$ and $S=0, M_S=0$

The method applies to the spins of the electron and nucleus in a hydrogen atom. The spins combine as above, but transitions are caused by an operator on the electron state only (because the Bohr magneton for baryons is so small). The squared matrix element required is $|\langle Y|\mathbf{s}|X\rangle|^2 = |\mathbf{s}_{\rm YX}|^2$ given by

$$|\mathbf{s}_{\rm YX}|^2 = |(s_x)_{\rm YX}|^2 + |(s_y)_{\rm YX}|^2 + |(s_z)_{\rm YX}|^2 = \left(|(s_+)_{\rm YX}|^2 + |(s_-)_{\rm YX}|^2\right)/2 \; + |(s_z)_{\rm YX}|^2. \tag{282}$$

Take $|XM\rangle$ to be one of the $S=1$ states, and $|Y\rangle$ to be the $S=0$, $M=0$ state, also expressible as

$$|Y\rangle = |0,\,0\rangle = \frac{1}{\sqrt{2}}\left(|+\rangle_e|-\rangle_p - |-\rangle_e|+\rangle_p\right) \tag{283}$$

The $\mathbf{s}$ operator is the electron spin operator, and doesn't affect the proton spin. Check that $s_+|-\rangle_e = |+\rangle_e$, $s_-|+\rangle_e = |-\rangle_e$, $s_z|\pm\rangle_e = \pm(1/2)|\pm\rangle_e$ (and remember in what follows that the first state vector in the product is the electron one).

For $|XM\rangle$ representing the states with $S=1$ and $M_S = 1, 0, -1$ we get

M	$\lvert XM\rangle$	$(s_+)\lvert XM\rangle$	$(s_-)\lvert XM\rangle$	$(s_z)\lvert XM\rangle$	$(s_+)_{\rm YX}$	$(s_-)_{\rm YX}$	$(s_z)_{\rm YX}$
1	$\lvert+\rangle\lvert+\rangle$	0	$\lvert-\rangle\lvert+\rangle$	$\frac{1}{2}\lvert+\rangle\lvert+\rangle$	0	$-\frac{1}{\sqrt{2}}$	0
0	$\frac{1}{\sqrt{2}}\left(\lvert+\rangle\lvert-\rangle + \lvert-\rangle\lvert+\rangle\right)$	$\frac{1}{\sqrt{2}}\lvert+\rangle\lvert+\rangle$	$\frac{1}{\sqrt{2}}\lvert-\rangle\lvert-\rangle$	$\frac{1}{2\sqrt{2}}\left(\lvert+\rangle\lvert-\rangle - \lvert-\rangle\lvert+\rangle\right)$	0	0	$\frac{1}{2}$
-1	$\lvert-\rangle\lvert-\rangle$	$\lvert+\rangle\lvert-\rangle$	0	$-\frac{1}{2}\lvert-\rangle\lvert-\rangle$	$\frac{1}{\sqrt{2}}$	0	0

The result $|\mathbf{s}_{\mathrm{YX}}|^2$ is then $1/4$, the same for each value of M, as it should be for an angle-averaged quantity.

This value, in physical units then gives the term for magnetic dipole radiation

$$|2\mathbf{s}_{\mathrm{YX}}|^2 = \hbar^2. \tag{284}$$